DISSERTATION

LESSONS FROM THE MICROBIOME IN HEALTH AND IN DEATH

Submitted by

Victoria Nieciecki

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2025

Doctoral Committee:

Advisor: Jessica Metcalf

Kelly Wrighton
Zaid Abdo
Susan Lana

ABSTRACT

LESSONS FROM THE MICROBIOME IN HEALTH AND IN DEATH

The work presented in this dissertation explores two very different microbial ecosystems found within the broad field of microbiome science–tumors and cadavers. In health, recently described low-biomass tumor-associated microbial communities are implicated in disease, and therefore microbial intervention may represent a therapeutic target. In death, the microbial decomposers of human remains show potential as a novel forensic tool that could help solve homicide cases and provide relief to the families of victims. In Chapter 1, I give a brief review of each of my research areas and summarize current knowledge gaps that exist within these fields to help provide additional context for my work in Chapters 2–4. In Chapter 2, I use 16S rRNA gene amplicon sequencing to characterize the microbiome of mucin-secreting Pseudomyxoma peritonei human tumors, focusing on microbial contamination and data reproducibility. Moving away from human health and the tumor microbiome, Chapter 3 investigates the spatial and temporal responses of microbial communities found in soil near decomposing cadavers. Finally, in Chapter 4, I explore the effects of enclosed shelter on the cadaver microbiome during human decomposition and develop microbiome-based models to estimate the postmortem interval, or time since death. In summary, my dissertation presents new and valuable insights into microbial community structure and assembly in health and in death and provides new tools for assessing

environmental contamination that afflict low-biomass samples and for estimating time since death using microbial-based machine learning models.

ACKNOWLEDGEMENTS

It takes a village to raise a child, but also to finish a PhD program. There are numerous people in my life that have made this journey possible, and I would like to express my gratitude and appreciation for all their support and encouragement along the way.

I would first like to thank my mentors and colleagues from the “before times” that strongly encouraged me to pursue this PhD and believed that I could turn this dream into a reality. Thank you, Dr. Heather Callahan, Dr. Eddie Adams, Dr. Liz Dennett, and Dr. Joanne Liu, for inspiring and supporting me throughout my career and helping me discover my passion for microbiome science. I would not be here on this PhD journey without any of you.

I would like to thank Dr. Jessica Metcalf for taking a chance on me and admitting me into your awesome lab. You have been an incredible mentor and advisor. These last five years have been an amazing journey, and I have learned and grown so much both personally and professionally. I would not be the scientist I am today without your guidance and direction.

I am grateful to my committee members, Dr. Kelly Wrighton, Dr. Zaid Abdo, and Dr. Susan Lana for the continued support and challenging questions that have pushed me to grow as a critical researcher.

I would like to extend my thanks to the wonderful collaborators I’ve worked with during my PhD. Thank you, Dr. Scott Merrell, Dr. Traci Testerman, and Dr. Tom McAvoy, for your patience and guidance as I took years to work through analyzing and

writing the complicated and challenging Pseudomyxoma peritonei manuscript. And thank you to Dr. Sibyl Bucheli for leading decomposition field experiments and working with students to collect the thousands of samples this dissertation is based on.

Thank you to all the friends I've made throughout this adventure starting with Dr. Heather Deel. When I first got to Colorado and joined the Metcalf lab, you were kind and welcoming. I have learned so much about data analysis and decomposition from you and I'm grateful that we became friends. Lexi Keene-Snickers, thank you for your support and friendship over the last five years as we battled through prelims and now defense. Finally, thank you to Dr. Valerie Seitz. You made the CMB program feel welcoming and inclusive as CMBSA president, and I am appreciative of all your guidance and direction whenever I needed help writing or working through science problems.

Finally, I would like to thank my longtime friends and family, especially Jamie Sherlock. You have been my best friend and partner in crime long before I even began this program. Thank you for your never-ending support, cheerleading, and motivation throughout this long journey. And finally, I would like to thank my husband, Aaron Walker. You believed in me from day one even when I doubted myself, you selflessly moved halfway across the country with me during a global pandemic, you have been my support system throughout this entire process, and I would not be here at the end of this program without you.

TABLE OF CONTENTS

CHAPTER 1: AN INTRODUCTION TO THE TUMOR MICROBIOME AND MICROBIAL DECOMPOSERS

Microbes have been found inhabiting almost every environment on Earth, from hydrothermal vents on the ocean floor to the clouds floating above us. These microscopic organisms contribute to various biological processes around the globe from the generation of oxygen through photosynthesis to the decomposition of organic matter and are crucial to our survival. The invention of modern sequencing technologies and microbial culture-independent approaches [1] has enabled researchers to catalog microbes colonizing all of Earth's biomes revealing extraordinary biodiversity [2]. We have since made countless discoveries linking the microbes that live on and inside of us to health and disease [3,4]. Plummeting sequencing costs and increased availability of low-cost computational processing power has enabled the field to move towards reconstructing metagenomes and probing microbial metabolisms and interactions [5].

My dissertation focuses on two distinct areas within the broad field of microbiome research–the tumor microbiome and the microbiome of human decomposition. Although relatively unrelated, I found these topics immensely fascinating. I've always been captivated by the ability of microbes to survive and make a living in strange places like the soils of the Antarctic Dry Valleys [6] or 600 meters below in the subsurface [7]. Although I could have centered my thesis on human decomposition, the primary focus of the Metcalf lab, I also felt strongly about contributing to cancer research as I have personally seen close family members deal with this disease. I am grateful for the

unique opportunity I was given to address important scientific knowledge gaps within both microbial systems.

1.1 The tumor microbiome

Random genetic mutations are often considered to be the main drivers of tumor development but with recent technological advancements, it is becoming increasingly clear that microbes can play a role in cancer development and progression. Specific bacterial species such as *Fusobacterium nucleatum, Helicobacter pylori,* and *pks+ Escherichia coli* have been strongly implicated in the development of colon and stomach cancers through mechanisms involved in chronic inflammation, immune suppression, and genotoxin secretion [8–10]. Recently, numerous studies have revealed the colonization of tumor tissues by low-biomass, cancer-type specific microbial communities is strongly associated with the development of some cancers and treatment outcomes [11–13]. For instance, differences in tumor-associated microbial community diversity have been linked to survival outcomes in pancreatic cancers [14] and intracellular bacteria may promote survival of metastatic breast cancer cells [15]. Whether these recently described tumor-associated microbial communities are benign opportunists or contribute to disease is unclear and represents a significant knowledge gap in oncology.

1.2 Contamination in low-biomass microbiome systems

Characterizing the tumor microbiome and other low-biomass microbial systems using a next generation sequencing (NGS) approach is technically challenging, resulting in ambiguity between true biological signals and contamination. For example, tumors are estimated to contain 34 bacterial cells per mm^2 of tissue on average [16] which can

easily be overwhelmed by exogenous bacterial contamination (e.g. DNA extraction kit background) and co-purified genomic host DNA [17,18]. Because of low quantities of bacterial DNA in comparison to host DNA in tumors, most shotgun metagenomic sequencing reads originate from the host's genome [19]. Although not considered contamination in the traditional sense, host DNA is difficult to differentiate from bacterial DNA and can lead to the misclassification of host genomic content. In fact, the incomplete identification and computational removal of host DNA was revealed to contribute to false bacterial signals in a high-profile tumor microbiome paper and partially contributed to its retraction [20–22].

To contend with the immense background of host material, PCR and amplicon NGS have been used as an alternative approach to target and amplify microbial DNA. However, environmental contamination introduced during sample handling, DNA extraction, and PCR can be difficult to differentiate from true biological signals and has resulted in erroneous conclusions. For instance, a recent re-analysis of 16S amplicon sequencing data revealed that the placental microbiome, previously reported by others, is likely the result of extraction kit contamination and no evidence was found to support the existence of resident placental microbial communities [23]. Although progress continues to be made with the incorporation of technical controls [12] and the development of computational decontamination tools [24], new methods that improve detection of contamination and reproducibility are critically needed.

1.3 PMP tumor microbiome study goals

Often originating in the appendix, Pseudomyxoma peritonei (PMP) is a disease characterized by the dissemination of mucin-secreted tumors throughout the peritoneal

cavity (Figure 1.1). Devastating in nature, PMP is a rare disease with few treatment options [25]. Early studies suggested a possible link may exist between the development and progression of PMP and carcinogenic bacteria like *Helicobacter pylori* [26]. In a subsequent study, Gilbreath et al. 2013 characterized microbial communities associated with PMP tumors and PMP mucin. This study discovered a group of highly conserved taxa, which included *H. pylori,* colonized these tumors and showed antibiotic treatment may improve treatment outcomes [27]. However, negative controls that assess environmental contamination were not included, making it difficult to interpret the study's findings. Thus, the composition and extent of the PMP tumor microbiome remains unclear. In Chapter 2, we aimed to 1) characterize microbial communities associated with a new group of 70 PMP tumors. We included technical controls in addition to normal peritoneal tissue samples to evaluate the endogenous peritoneal microbiome. Additionally, we set out to 2) develop a cross-laboratory workflow to identify samples with reproducible biological signals while simultaneously uncovering laboratory-specific contamination.

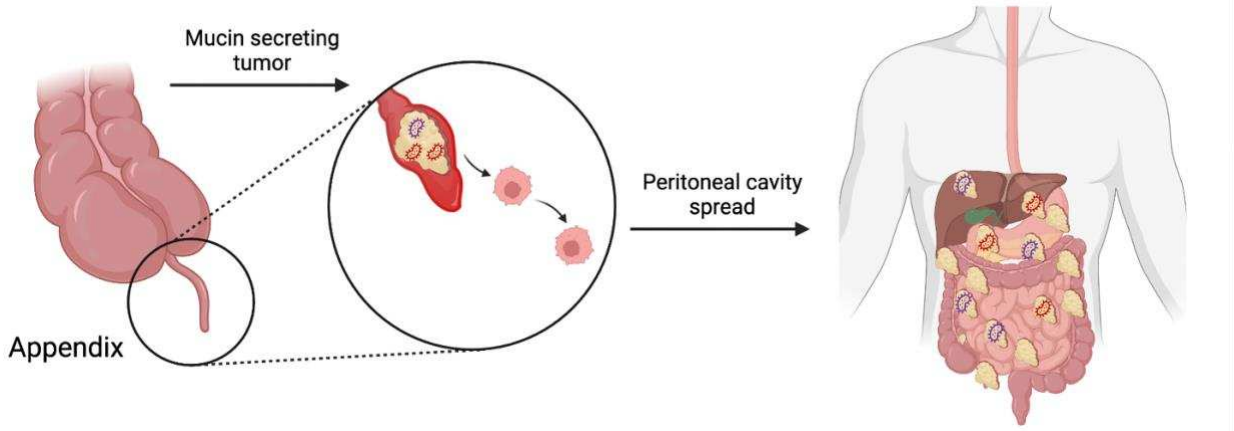


Figure 1.1. Conceptual model of Pseudomyxoma peritonei tumor microbiome origin and spread throughout the peritoneal cavity. A mucin-secreting tumor forms in the appendix (left). At this time, gut-associated microbes may invade the irregular growth. The neoplasm grows and ruptures through the appendix wall. Peritoneal fluids disseminate mucin-secreting tumor cells and tumor-associated bacteria throughout the peritoneal cavity (middle). During advanced stages, PMP metastases attach to the peritoneal wall and secrete jelly-like mucus that builds up in the abdomen and can press against organs (right). PMP tumors and associated microbes from this stage were collected for our study. Figure generated using BioRender.

1.4 Microbial decomposition of human remains

Human decomposition progresses in four major, and oftentimes co-occurring stages: fresh, early decomposition, advanced decomposition, and skeletonization [28]. By-products of decomposition become inputs into a complex nutrient influx that can affect multiple trophic levels (e.g. scavengers, insects, and microbes) (Figure 1.2). For instance, blow flies, attracted by the release of volatile organic carbons (VOCs) during fresh decomposition, begin to visit human remains almost immediately after death. As blow flies lay their eggs and feed on the human remains, they spread their associated microbes around the decomposing flesh [29]. Once hatched, maggots secrete digestive enzymes while feeding to aid in the breakdown of cadaver tissue. During feeding and burrowing, maggots further inoculate the decomposing flesh with maggot-associated

microbes that likely aid in breaking down the nutrient pool. Concurrently, the buildup of gases (methane, ammonia, hydrogen sulfide) within the abdominal cavity from anaerobic microbial metabolism results in the release of nutrient rich fluids and decomposition-associated microbes entering the surrounding soil environment [30]. This sudden influx of ammonium (NH_4^+), dissolved organic nitrogen (DON), and dissolved organic carbon (DOC), which marks the beginning of early decomposition, leads to shifts in microbial communities found in the surrounding soils [31].

Medical examiners are able to use physical changes associated with each stage (e.g. *rigor mortis*) [32] along with forensic entomology [33] to estimate the postmortem interval (PMI) or time since death. However, the rate of decomposition is largely driven by many environmental (e.g. humidity, temperature, body mass) and biotic factors (e.g. microbial decomposer communities, scavenger activity) and can make it challenging to accurately predict PMI. Present since the time of death, the microbial communities that help decompose human remains have been investigated as a powerful PMI estimation tool. As decomposition progresses and nutrient pools change, microbial community composition changes in response [34]. These clock-like microbial successional patterns are markedly similar across individual cadavers [35] and form the foundation of microbiome-based PMI machine learning models. Recent work has shown that PMI estimates based on microbial data are accurate within a +/- 3 day window from time of death regardless of climate region or season [36]. This study also revealed that, independent of geographical location, a universal group of microbes assemble during decomposition which enables these accurate predictions and a generalizable PMI model.

Despite their prevalence across decomposition sites, the origin of microbial decomposers that colonize cadavers is largely debated. Researchers have proposed that these microbial communities are likely human-associated and originate from the cadaver itself or possibly from the surrounding soils [34]. Burcham et al. 2024 demonstrated that the assembly of the decomposer network was location independent, suggesting soil may not be a primary reservoir. To help identify possible sources, I searched for the presence of these decomposer microbes in large publicly available datasets (i.e. Human Microbiome Project, Earth Microbiome Project). Surprisingly, most were scarce in these datasets except for some *Acinetobacter* and *Peptoniphilus* species. However, many were prevalent in blowfly and carrion beetle genomic datasets strongly suggesting these decomposer microbes are likely insect-associated or at least spread through insect-cadaver interactions.

1.5 Goals of decomposition studies

Time since death estimates are possible using microbial communities that assemble on cadaver skin and in the soil near the decomposing cadaver. While skin communities provide the best PMI estimates likely due to their similarity across individual cadavers, soils are still a suitable alternative and are important for understanding the breakdown and recycling of cadaver-derived nutrients [37] (Figure 1.2). Most studies investigating the ecology of microbial decomposers in the soil collect samples from the soil surface [35,36] or collect aggregate cores from 5-16 cm in depth [31,34,38]. Although these studies have demonstrated microbial communities respond to aboveground decomposition across most of these depths, it is currently unclear how deep these responses permeate into the subsurface. Therefore, the goal of Chapter 3

was to investigate microbial responses across multiple soil depths in order to improve soil collection strategies.

Microbial-based PMI estimation models are generalizable across individuals, climate zones, and seasons. However, most deaths occur indoors and may be left undiscovered for long periods of time [39]. Indoor decomposition studies indicate that insect arrival tends to be delayed due to obstructed access to the cadavers. Importantly, this delayed arrival was linked to slower indoor decomposition rates [39,40]. While only a few indoor studies exist, none have examined the impact of an indoor setting on the successional patterns of microbial communities that decompose cadaver skin or how these changes may differ over time. The goals of Chapter 4 were to 1) determine the impact of a sheltered decomposition environment on the cadaver microbiome and 2) to evaluate whether shelter affects microbial-based PMI model accuracy.

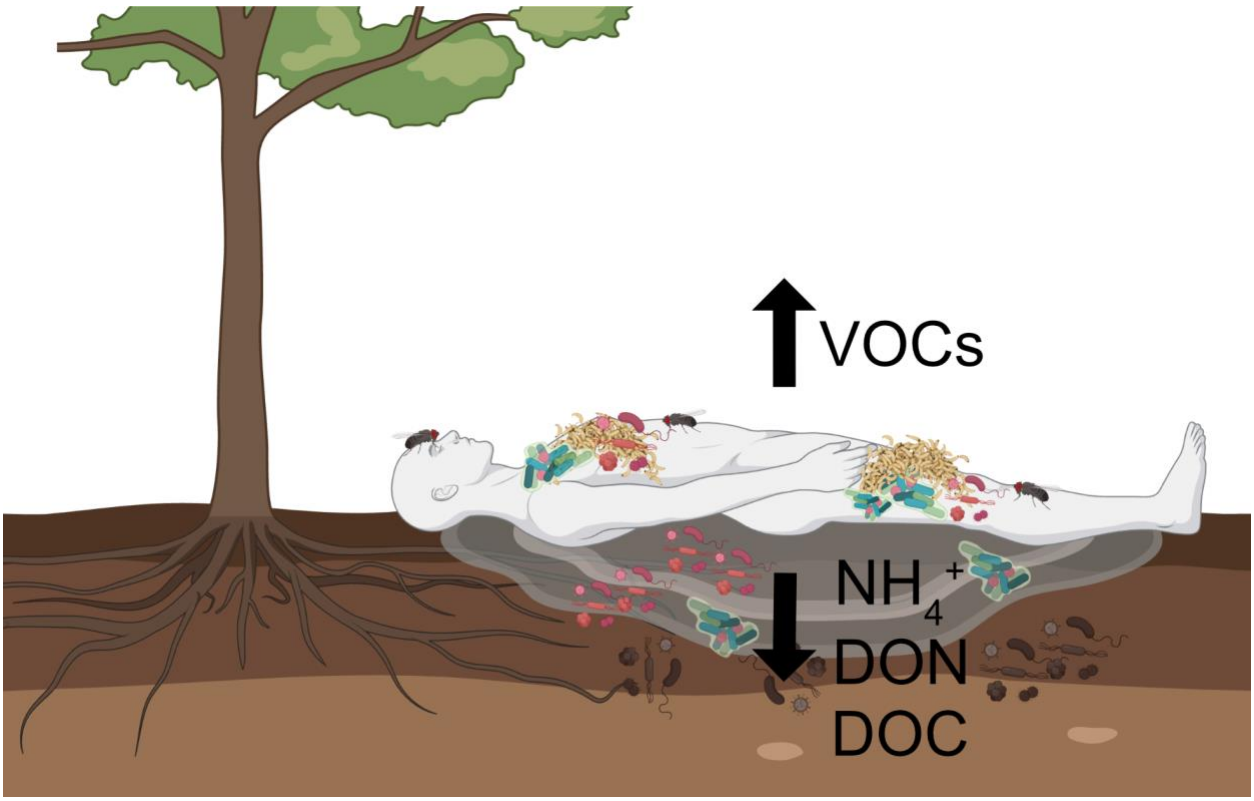


Figure 1.2. Conceptual figure of the major processes that occur during fresh and early outdoor decomposition. Figure created using BioRender.

CHAPTER 1: REFERENCES

1. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6: 1621–1624. doi:10.1038/ismej.2012.8

2. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551: 457–463. doi:10.1038/nature24621

3. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486: 207–214. doi:10.1038/nature11234

4. Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, et al. Gut Microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. Cell. 2016;167: 1469-1480.e12. doi:10.1016/j.cell.2016.11.018

5. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39: 499–509. doi:10.1038/s41587-020-0718-6

6. Cary SC, McDonald IR, Barrett JE, Cowan DA. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat Rev Microbiol. 2010;8: 129–138. doi:10.1038/nrmicro2281

7. Puente-Sánchez F, Arce-Rodríguez A, Oggerin M, García-Villadangos M, Moreno-Paz M, Blanco Y, et al. Viable cyanobacteria in the deep continental subsurface. Proc Natl Acad Sci U S A. 2018;115: 10702–10707. doi:10.1073/pnas.1808176115

8. Zepeda-Rivera M, Minot SS, Bouzek H, Wu H, Blanco-Míguez A, Manghi P, et al. A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. Nature. 2024;628: 424–432. doi:10.1038/s41586-024-07182-w

9. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med. 2001;345: 784–789. doi:10.1056/NEJMoa001999

10. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature. 2020;580: 269–273. doi:10.1038/s41586-020-2080-8

11. Geller LT, Barzily-Rokni M, Danino T, Jonas OH, Shental N, Nejman D, et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science. 2017;357: 1156–1160. doi:10.1126/science.aah5043

12. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science. 2020;368: 973–980. doi:10.1126/science.aay9189

13. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, et al. Commensal Microbiota Promote Lung Cancer Development via γδ T Cells. Cell. 2019;176: 998-1013.e16. doi:10.1016/j.cell.2018.12.040

14. Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. Cell. 2019;178: 795-806.e12. doi:10.1016/j.cell.2019.07.008

15. Fu A, Yao B, Dong T, Chen Y, Yao J, Liu Y, et al. Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. Cell. 2022;185: 1356-1372.e26. doi:10.1016/j.cell.2022.02.027

16. Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA, Knight R. The microbiome and human cancer. Science. 2021;371. doi:10.1126/science.abc4552

17. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27: 105–117. doi:10.1016/j.tim.2018.11.003

18. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, et al. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Front Microbiol. 2019;10: 1277. doi:10.3389/fmicb.2019.01277

19. Walker SP, Tangney M, Claesson MJ. Sequence-Based Characterization of Intratumoral Bacteria-A Guide to Best Practice. Front Oncol. 2020;10: 179. doi:10.3389/fonc.2020.00179

20. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020;579: 567–574. doi:10.1038/s41586-020-2095-1

21. Gihawi A, Ge Y, Lu J, Puiu D, Xu A, Cooper CS, et al. Major data analysis errors invalidate cancer microbiome findings. MBio. 2023;14: e0160723. doi:10.1128/mbio.01607-23

22. Sepich-Poore GD, McDonald D, Kopylova E, Guccione C, Zhu Q, Austin G, et al. Robustness of cancer microbiome signals over a broad range of methodological variation. Oncogene. 2024. doi:10.1038/s41388-024-02974-w

23. Panzer JJ, Romero R, Greenberg JM, Winters AD, Galaz J, Gomez-Lopez N, et al. Is there a placental microbiota? A critical review and re-analysis of published placental microbiota datasets. BMC Microbiol. 2023;23: 76. doi:10.1186/s12866-023-02764-6

24. Austin GI, Park H, Meydan Y, Seeram D, Sezin T, Lou YC, et al. Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data. Nat Biotechnol. 2023;41: 1820–1828. doi:10.1038/s41587-023-01696-w

25. O'Connell JT, Tomlinson JS, Roberts AA, McGonigle KF, Barsky SH. Pseudomyxoma peritonei is a disease of MUC2-expressing goblet cells. Am J Pathol. 2002;161: 551–564. doi:10.1016/S0002-9440(10)64211-3

26. Semino-Mora C, Liu H, McAvoy T, Nieroda C, Studeman K, Sardi A, et al. Pseudomyxoma peritonei: is disease progression related to microbial agents? A study of bacteria, MUC2 AND MUC5AC expression in disseminated peritoneal adenomucinosis and peritoneal mucinous carcinomatosis. Ann Surg Oncol. 2008;15: 1414–1423. doi:10.1245/s10434-007-9778-9

27. Gilbreath JJ, Semino-Mora C, Friedline CJ, Liu H, Bodi KL, McAvoy TJ, et al. A core microbiome associated with the peritoneal tumors of pseudomyxoma peritonei. Orphanet J Rare Dis. 2013;8: 105. doi:10.1186/1750-1172-8-105

28. Megyesi MS, Nawrocki SP, Haskell NH. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. J Forensic Sci. 2005;50: 618–626. Available: https://www.ncbi.nlm.nih.gov/pubmed/15932096

29. Deel HL, Montoya S, King K, Emmons AL, Huhn C, Lynne AM, et al. The microbiome of fly organs and fly-human microbial transfer during decomposition. Forensic Sci Int. 2022;340: 111425. doi:10.1016/j.forsciint.2022.111425

30. Carter DO, Yellowlees D, Tibbett M. Cadaver decomposition in terrestrial ecosystems. Naturwissenschaften. 2007;94: 12–24. doi:10.1007/s00114-006-0159-1

31. Taylor LS, Mason AR, Noel HL, Essington ME, Davis MC, Brown VA, et al. Transient hypoxia drives soil microbial community dynamics and biogeochemistry during human decomposition. FEMS Microbiol Ecol. 2024;100. doi:10.1093/femsec/fiae119

32. Lee Goff M. Early post-mortem changes and stages of decomposition in exposed cadavers. Exp Appl Acarol. 2009;49: 21–36. doi:10.1007/s10493-009-9284-9

33. Catts EP, Goff ML. Forensic entomology in criminal investigations. Annu Rev Entomol. 1992;37: 253–272. doi:10.1146/annurev.en.37.010192.001345

34. Cobaugh KL, Schaeffer SM, DeBruyn JM. Functional and Structural Succession of Soil Microbial Communities below Decomposing Human Cadavers. PLoS One. 2015;10: e0130201. doi:10.1371/journal.pone.0130201

35. Metcalf JL, Xu ZZ, Weiss S, Lax S, Van Treuren W, Hyde ER, et al. Microbial community assembly and metabolic function during mammalian corpse

decomposition. Science. 2016;351: 158–162. doi:10.1126/science.aad2646

36. Burcham ZM, Belk AD, McGivern BB, Bouslimani A, Ghadermazi P, Martino C, et al. A conserved interdomain microbial network underpins cadaver decomposition despite environmental variables. Nat Microbiol. 2024;9: 595–613. doi:10.1038/s41564-023-01580-y

37. DeBruyn JM, Keenan SW, Taylor LS. From carrion to soil: microbial recycling of animal carcasses. Trends Microbiol. 2024. doi:10.1016/j.tim.2024.09.003

38. Keenan SW, Schaeffer SM, Jin VL, DeBruyn JM. Mortality hotspots: Nitrogen cycling in forest soils during vertebrate decomposition. Soil Biol Biochem. 2018;121: 165–176. doi:10.1016/j.soilbio.2018.03.005

39. Ceciliason A-S, Andersson MG, Lindström A, Sandler H. Quantifying human decomposition in an indoor setting and implications for postmortem interval estimation. Forensic Sci Int. 2018;283: 180–189. doi:10.1016/j.forsciint.2017.12.026

40. Thümmel L, Lutz L, Geissenberger J, Pittner S, Heimer J, Amendt J. Decomposition and insect succession of pig cadavers in tents versus outdoors - A preliminary study. Forensic Sci Int. 2023;346: 111640. doi:10.1016/j.forsciint.2023.111640

CHAPTER 2: CROSS-LABORATORY REPLICATION OF PSEUDOMYXOMA PERITONEI TUMOR MICROBIOME REVEALS REPRODUCIBLE MICROBIAL SIGNATURES[1]

2.1 Summary

Recent work has demonstrated that cancer-specific microbial communities often colonize tumor tissues. However, untangling low biomass signals from environmental contamination makes this research technically challenging. We utilize Pseudomyxoma Peritonei (PMP), a cancer characterized by the spread of mucus-secreting cells throughout the peritoneal cavity, to develop a robust workflow for identifying reproducible tumor microbiomes. Typically originating from the rupture of an appendiceal tumor into the peritoneal cavity, metastasized tumors have been previously shown to harbor a core set of microbes. However, that work did not control for the potential contamination of these low microbial biomass samples. We expand upon these prior findings by characterizing the microbiome of 70 additional PMP tumors and 6 normal peritoneal control tissues along with appropriate laboratory controls. Additionally, DNA from a subset of 25 tissues was extracted and sequenced at an independent laboratory. We found evidence of reproducible microbial signatures between the replicates of 6 different PMP tumors that include a set of core taxa that may be introduced from surgical contamination, as well as patient-specific taxa that are also commonly implicated in colorectal cancer. In addition, preoperative chemotherapy treatment was found to reduce tumor microbiome diversity. Our findings demonstrate

[1] This work has been submitted for publication: Nieciecki, V., Blum, F., Johnson, R., Testerman, T., McAvoy, T., et al. (2025). Cross-laboratory replication of Pseudomyxoma Peritonei tumor microbiome reveals reproducible microbial signatures. Submitted to *mSphere*.

how independent sample replication can be a powerful approach to investigate low-biomass microbial communities associated with tumor tissues that will improve low microbial biomass research.

2.2 Importance

Recent work has demonstrated that microbial communities colonize over 30 different types of tumor tissues. The origin of these communities and their possible involvement in carcinogenesis or cancer treatment outcomes remains an unclear, yet important area of research. A current major challenge in characterizing low biomass, tumor-associated microbiomes is the introduction of environmental contamination during collection, handling, DNA extraction, PCR, and sequencing. Here, we provide a framework for replicating low biomass tumor microbiome samples to help identify tumors with robust microbial signals and low background contamination. Using this replication approach, we show that Pseudomyxoma Peritonei tumors host reproducible microbial communities, including organisms that have previously been associated with colorectal cancer. Incorporating sample replication into future tumor microbiome studies is a promising approach that will help identify robust signals and increase reproducibility in the field.

2.3 Introduction

Bacteria are known to contribute to cancers in human tissues such as the colon [1,2], stomach [3,4], and lung [5], all of which host commensal bacteria in their healthy states. Several recent, large-scale studies [6–9] revealed that bacteria may be prevalent in tumors throughout the body, even in tissues generally considered to be sterile. It has been hypothesized that leaky vasculature may allow microbes to enter the tumor, where

they are able to persist in the immunosuppressed microenvironment [10,11]. Using a combination of methods, these studies have demonstrated over 30 different tumor types host distinct microbial communities, which are diagnostically relevant. Moreover, a growing body of evidence suggests these bacterial communities are correlated with disease outcomes [12,13] and response to treatment [14,15]. Whether these bacteria are generally benign opportunists or contribute to disease outcomes is largely unknown and represents a critical knowledge gap in the field of cancer biology.

Although recent progress has been made in defining various tumor microbiomes, this research is technically challenging. Total DNA isolation from human tumor tissue generates a complex milieu of DNA molecules from various origins which researchers then need to disentangle using a combination of sample controls and bioinformatic approaches. Low microbial biomass makes it difficult to distinguish true biological signals from human DNA, which dominates the samples, or from surgical and laboratory-derived contamination. For example, in the first large-scale tumor microbiome paper, researchers leveraged The Cancer Genome Atlas (TCGA), a database of human whole-genome sequencing and whole-transcriptome studies, to extract microbial metagenomic reads [16]. After controlling for bacterial contamination using various computational methods, the authors reported cancer-specific microbial signatures. However, a fraction of human reads were misclassified as bacteria which generated artificial signatures [17]. These concerns were later corrected using the most recent human reference genome along with new filtering methods which confirmed that cancer-specific tumor microbiomes exist within TCGA [7], although the debate over microbial signals in these data continues. In another large-scale contemporaneous

publication, Nejman et al. 2020 utilized a 16S rRNA amplicon approach that avoids sequencing co-purified human genomic DNA. Although human DNA is avoided, background contamination becomes problematic as this method enriches bacterial reads. Hence, Nejman et al. included and evaluated extensive negative controls to identify and exclude bacterial contamination, and similar to Sepich-Poore et al. (2024) discovered cancer-specific tumor microbiome signatures, finding evidence for intracellular bacteria in tumor cells. Both studies utilized bioinformatic approaches for removing bacteria that were likely introduced during surgery or lab processing. Computational subtraction methods continue to improve with more sophisticated underlying models of contamination (e.g. SCRuB utilizes a plate map to account for well-to-well leakage).

Another less employed strategy that is compatible with these computational approaches is to identify and focus on microbial signals that are reproducible across extraction kit batches, suppliers, extraction facilities, etc. [18]. Moreover, this method can simultaneously be used to uncover environmental contamination. Lab environments kit reagents frequently exhibit unique contamination signatures that can be identified and computationally removed by comparing microbial profiles of samples replicated across different facilities. Microorganisms repeatably found across facilities are likely biological in origin while those unique to each facility or reagent are consistent with contamination. Ultimately, the ideal approach is likely a combination of data generation control measures and bioinformatic approaches. In a similar vein, the field of ancient DNA has utilized cross-lab replication for decades (e.g. [19] as part of the critical efforts to disentangle contamination from signal, particularly with hominid samples.

In a relatively early tumor microbiome publication that lacked controls and would benefit from cross-laboratory replication, Gilbreath et al. 2013 hypothesized that tumor-associated bacteria (e.g. *Helicobacter pylori*) may be associated with Pseudomyxoma Peritonei (PMP) disease progression. PMP typically originates in the appendix and usually spreads via a perforated appendiceal neoplasm into the peritoneal cavity, where it is characterized by implants of mucin-secreting, cancerous epithelial cells surrounded by gelatinous ascites [20]. Furthermore, there is a potential for tumor-associated appendiceal and intestinal bacteria to infiltrate the peritoneum as disease disseminates, and impact treatment outcomes. Using 16S rRNA gene sequencing and culturing approaches, they provided evidence that PMP tumors contained a conserved set of core bacteria. However, contamination tracking and mitigation were minimal in this study, and thus, it is unclear what proportion of the reported results are biological versus surgical and/or laboratory contamination.

To address limitations of appropriate controls in prior research, we analyzed a new set of 70 PMP tumor samples and established key quality metrics to better identify samples with a robust tumor microbiome signature, allowing us to determine whether a core microbiome exists for this cancer type. Following guidelines provided by (Eisenhofer et al. 2019) and (de Goffau et al. 2018), and the field of ancient DNA (Cooper and Poinar 2000), we utilized both within- and between-lab replication to identify taxa that are most robustly associated with PMP tumor tissue and highlight the utility of our methods. Furthermore, we explored our results in the context of their ecology, hypothesizing that PMP tumor-associated bacteria are likely common gut commensals that translocate via the appendix. Ultimately, these data allow us to better

understand the nature and patterns of contamination and biological signals in low-biomass tumor samples.

2.4 Results

2.4.1 Developing a cross-laboratory replication workflow

We first sought to determine whether PMP tumors harbor reproducible bacterial signatures by replicating DNA extraction and sequencing efforts across two independent laboratories. We hypothesized that a robust tumor-associated microbial signal should be reproducible across labs [18]. Furthermore, a comparison of bacterial taxa isolated from independently extracted PMP tumors would allow us to distinguish between tumor-associated bacteria and lab-specific background contamination more confidently, with the exception of contamination events that occurred during surgery. We also included several normal peritoneal (NP) control tissues that were collected during hernia repair or hysterectomy surgery from patients without a cancer diagnosis, as normal peritoneal tissue is difficult to obtain from PMP patients (Tables S1-2). NP control samples were utilized as a comparative set of amplicon sequence variants (ASVs) that may represent contamination from surgical procedures and environment or may alternatively represent microbes common to the peritoneal cavity that are not specific to tumors.

First, in lab 1, we isolated total DNA from PMP tumors and NP tissues and characterized tumor-associated microbes via 16S rRNA gene amplicon sequencing. A subset of PMP tumors and NP samples that appeared to sequence robustly (i.e. > 1000 sequencing reads & > 100 mg remaining tissue) were then selected from lab 1 for independent DNA extraction and sequencing replication in lab 2 (Figure 1A). To track

and report reagent and laboratory contamination, we included negative extraction controls (NECs; no tissue sample was added to the extraction) and no template controls (NTCs; in which no DNA template was added during the PCR amplification step). Additionally, mock microbial communities were also included as positive controls to assess DNA extraction efficiency and to further track background contamination (see methods).

With both lab 1 and 2 data sets, we then focused on high-quality tumor and control tissue samples by developing a stepwise filtering approach, which aimed to identify and remove samples with poor-quality and perceived gross bacterial contamination from our downstream analysis (Figure 1B). First, samples with fewer than 2000 sequencing reads were removed. We reasoned that samples with a robust microbial signature would have more reads than the median number of negative control reads (lab 1 = 586; lab 2 = 1404) (Table S3). Samples that largely deviate below this cutoff are likely to contain very low to undetectable amounts of tumor-associated bacteria, which would be indistinguishable from background contamination.

Next, the contribution of laboratory contaminants to the tumor microbiome communities was estimated via a Source-tracking Bayesian approach utilizing NECs and NTCs as potential microbial “sources” [21]. Samples that were estimated to contain more than 20% contribution from negative controls were removed. This conservative cut-off was chosen to eliminate samples with gross contamination that may be misinterpreted as genuine biological signals while retaining samples with low levels of environmental contamination which is expected from this sample type. Sample extractions that passed both filters were deemed high-quality and were used to

investigate the PMP tumor microbiome. Extractions that failed to pass either filter were deemed low-quality. Finally, to mitigate lab-specific contamination, the most prevalent taxa that were found in more than 50% of each lab's negative controls (e.g. *Turicella* and *Geobacillus*) were removed from high-quality samples (Figure 1B).

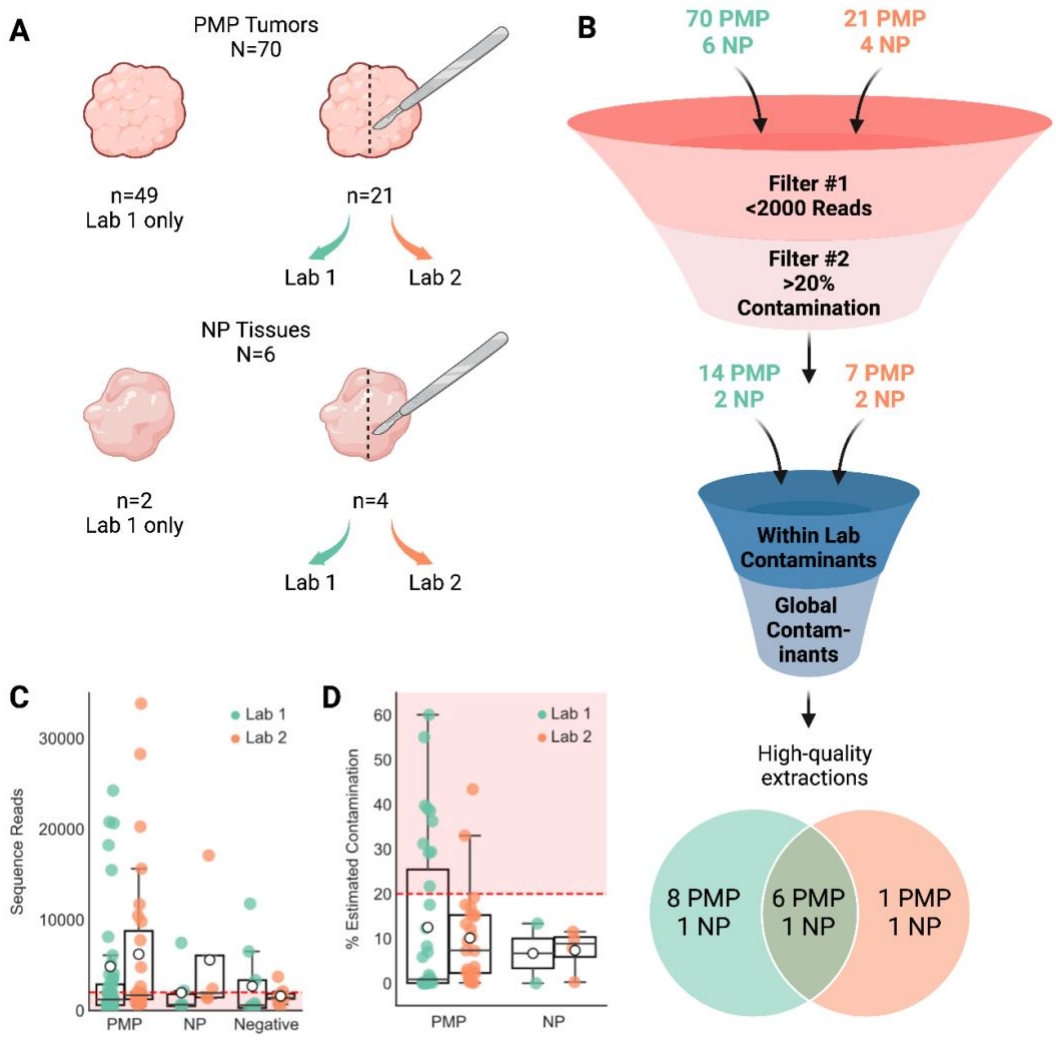


Figure 1. Experimental workflow and filtering pipeline. A) 70 PMP tumors and 6 NP samples were extracted and sequenced at lab 1 (teal). A subset of each tissue type was sent to lab 2 (orange) for replication. B). PMP and NP extractions underwent 2 sets of filtering steps. Sample extractions with <2000 reads were removed (filter #1). Extractions with >20% estimated contamination based on Source-tracking were removed (filter #2). Lab-specific and global contaminants were removed from high-quality extractions. Venn diagram shows 6 PMP tumors and 1 NP sample had 2 high-quality replicates. C) Number of sequence reads per extraction following DADA2 filtering and removal of non-microbial sequences. Samples with no sequencing reads and outliers were removed. D) SourceTracker2 estimated percent negative control contribution per sample. All samples were rarefied to 800 reads in order to retain as many negative controls for source-tracking as possible. Samples with <800 reads after filtering common contaminants were removed. Dashed red lines represent filter thresholds. White circles represent mean value.

2.4.2 Lab 1 Data Quality Assessment

In total, 70 PMP tumors, 8 NECs, 2 NTCs, 6 mock community dilutions (BEI Resources), and 6 NP tissues, were extracted and sequenced at lab 1. Of these 92 samples, 17 PMP tumors, 1 NEC, and 1 mock community had zero sequencing reads following DADA2 quality filtering and removal of non-bacterial reads and were subsequently removed. The 53 remaining PMP tumors retained an average of 5129 sequence reads (range: 46-94766) (Figure 1C). 7 NECs and 2 NTCs retained an average of 2681 reads (range: 55-11771) (Table S3). Of note, 2 NECs and 1 NTC had more than 3000 reads, suggesting possible random spot contamination. The 5 remaining mock community dilutions had an average of 41521 sequence reads (range: 1-160717). The 6 NP control tissue samples retained an average of 1964 reads per sample which ranged from 413 - 7460 reads.

To assess contamination while retaining enough negative extraction controls, samples were rarefied to 800. Source-tracking was performed on 35 of the 53 PMP tumors with >800 reads. On average, 12.5% of the PMP reads could be attributed to the NEC and NTC samples (range: 0 - 60.05%) (Figure 1D). SourceTracker2 identified 64 bacterial ASVs as potential contaminants. *Turicella* and an unidentified ASV in the Prevotellaceae family were the most abundant putative contamination contributors to lab 1 PMP samples (Table S4) and were found in 18 and 6 PMP samples, respectively. Although *Turicella* and Prevotellaceae are not commonly reported contaminants (Eisenhofer et al. 2019), several other taxa also identified via source-tracking, such as *Corynebacterium* and *Enterococcus,* have been previously reported as common laboratory or kit contaminants.

We also investigated whether NP samples were primarily dominated by laboratory contamination or hosted a distinct microbial community that might represent other sources (e.g. surgical environment or endogenous peritoneal microbes). We found that 2 of 6 NP samples contained sufficient reads for source-tracking. An average of 6.69% of reads were attributed to negative control sources.

Next, we examined the mock community dilutions with sufficient reads to determine whether background contamination could be identified in samples with a well-defined community. The BEI microbial mock community is composed of 22 bacterial species mixed in equal proportions (Figure S1A). We recovered 20 of the 22 known species but were not able to detect *P. acnes* and *S. epidermidis*. 19 bacterial contaminants were also identified and accounted for an average of 1.8% of reads. Only 6 of these 19 contaminants were also found in the lab 1 negative controls suggesting an alternative source of contamination (e.g. manufacturer facility). We also discovered that 5 mock community members were present in lab 1 negative controls – *E. coli*, *S. aureus*, *E. faecalis, A. odontolyticus,* and *H. pylori.* Whether these ASVs represent real, lab-associated background contamination or positive-to-negative control cross-contamination is unclear. While low-level cross-contamination is likely, these ASVs are common host-associated taxa and may represent true background contamination.

To distinguish between low- and high-quality sample extractions, we applied our two quality filters. We found 32% of PMP tumors (17/53) and 33% of NP samples (2/6) had more than 2000 sequencing reads (Figure 1C) and passed the first filtering step. To account for contamination, >20% estimated contamination was used as a cutoff; 14

PMP tumors and both NP samples passed and were retained in the dataset (Figure 1B & 1D).

2.4.3 Lab 2 Data Quality Assessment

Of the original 70 PMP tumors, a subset of 21 tumors and 4 of the 6 NP samples were chosen for replication at lab 2 (Figure 1A). In lab 2, 2 additional replicates of 3 PMP tumor samples, PMP-423, PMP-433, and PMP-485, were generated as a part of DNA optimization experiments (see methods). These 6 replicates have been included in our analysis to investigate within-lab reproducibility. In addition to the PMP and NP samples, 8 NECs, 4 NTCs, and 4 mock community controls (Zymo) were also extracted and sequenced. Lab 2 samples underwent DADA2 quality filtering and removal of non-bacterial reads in the same manner as lab 1 samples.

Following quality filtering, the 21 PMP tumors retained 5670 sequences per sample on average (range: 749 - 33825) (Figure 1C; Table S3). On average, the optimization samples were left with 8073 reads per sample. The 8 NECs and 4 NTCs retained 1592 sequence reads on average, which ranged from 364 - 3716 reads; 1 NEC contained more than 3000 reads. The 4 mock community samples on average retained 23169 sequence reads per sample (range: 18033 - 29209) and the 4 normal control tissue samples retained an average of 5578 reads per sample (range: 1364 - 17089). We observed that the range of sequencing reads derived from PMP tumors was highly variable in both labs and may be due to several factors including interference from human genomic DNA and variable bacterial load. However, we note that amplicon sequencing studies are generally characterized by unequal sequencing depth [22].

Source-tracking with lab-specific NECs and NTCs was again used to estimate the contribution of laboratory contamination to 20 PMP tumors with sufficient reads following rarefying (>800). We found that an average of 11.75% sequencing reads could be attributed to lab 2 negative control sources (range: 0.23 - 43.38%). The two most contaminating taxa identified by SourceTracker2, *Geobacillus* and *Acinetobacter* (Table S4), have previously been associated with negative extraction controls [23]. Notably, SourceTracker2 contamination signatures were lab-specific. Only 3 SourceTracker2 contaminants were found in both lab 1 and 2 PMP samples, and only 1 of these (*Enterococcus*) contributed more than 500 reads.

Negative control contribution to the normal peritoneal samples was 7.38% which was similar to results produced in lab 1.

Lab 2 mock communities were composed of 8 distinct bacterial species mixed in unequal proportions. We found 31 bacterial ASVs that accounted for an average of 2.4% reads were identified as lab 2 positive control contaminants and 15 of these 31 contaminants were also found in the lab 2 NECs and NTCs (Figures S1C). We again found several mock community members contaminating the lab 2 negative controls (*E. coli*, *S. aureus*, and *E. faecalis*). These findings taken together with the lab 1 positive control data demonstrate the risk of removing ASVs from biological samples solely based on their presence in negative control samples and highlights the frequency of cross-contamination.

Both filtering steps were applied to the lab 2 PMP and NP samples resulting in 9 of 21 PMP tumors and 2 of 4 NP samples containing > 2000 sequence reads and

passing the first filtering step. After the second contamination filter, 7 of these 9 PMP tumors and both NP control tissue samples passed (Figure 1B).

In total, we identified 14 high-quality PMP tumor extractions that had been processed at lab 1 and 7 extractions processed at lab 2. These 21 extractions represented PMP tumors from 15 different patients that may host robust tumor-associated microbial signatures. Additionally, 2 normal peritoneal extractions processed at lab 1, and 2 from lab 2 represent 3 distinct control tissues that may contain surgical contamination signatures or a microbial community endogenous to the peritoneal cavity (Figure 1B). The microbial communities of these 25 high-quality extractions were then used to distinguish tumor-associated microbes from potential contamination and to characterize the PMP tumor microbiome.

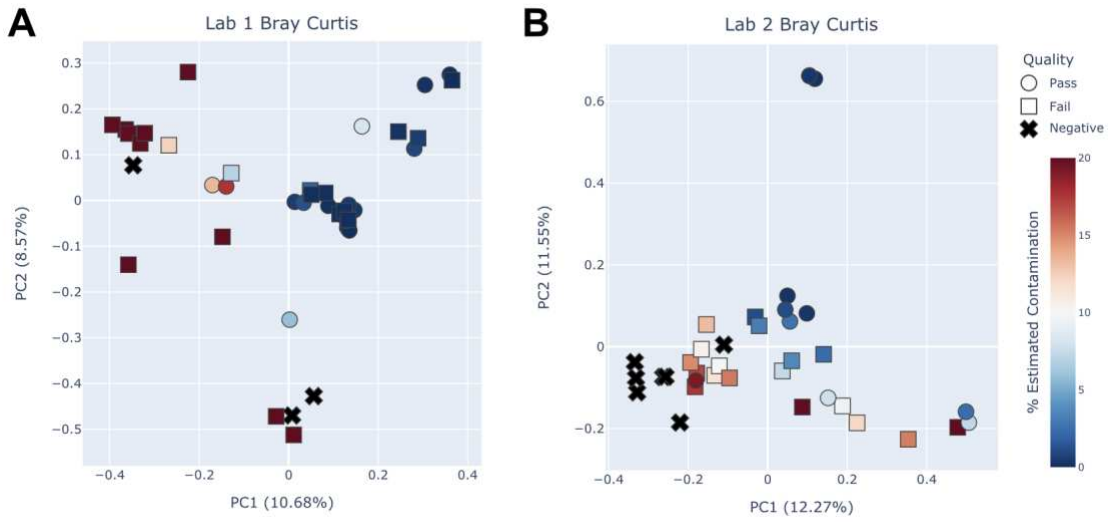


Figure 2. Low quality samples are similar to negative controls. A) Bray Curtis PCoA plot showing PMP tumors, normal peritoneal control tissues, and negative controls extracted at lab 1 and B) lab 2. All samples were rarefied to 800 reads per sample. Those with <800 reads were removed. Samples are colored according to % contamination as estimated by SourceTracker2. Max color was set to 20% which represents contamination cutoff. Negative control samples are colored in black. Shapes represent whether samples passed or failed both quality filters.

2.4.4 Cross-contamination influences sample quality

To evaluate whether our filtering criteria identified low-quality samples with gross background contamination, we visualized the PMP and NP tissue samples along with their respective negative controls using a Bray-Curtis principal coordinate analysis (Figure 2A-B). We expected that poor-quality samples would primarily be composed of background contamination and consequently cluster near negative extraction and negative template controls. In general, we found that the distance between tissue samples and negative controls tended to decrease as percent contamination increased (Figure S1F). Lab 1 low-quality PMP and NP samples were not statistically distinct from negative controls (PERMANOVA *q*-value=0.293) suggesting the groups were similar. However, several of the low-quality lab 2 samples clustered away from negative controls along PC1 suggesting their microbial communities were dissimilar even though they were estimated to contain gross contamination (PERMANOVA *q*-value=0.0015). We suspected this was likely driven by sample-to-negative control cross contamination. The relative abundance of *Acinetobacter* in these low-quality PMP samples was found to closely match the relative abundance found in two negative controls (Figure S1D) causing SourceTracker2 to overestimate contamination levels. Regardless, our goal was to identify and investigate unambiguously high-quality sample extractions. Therefore, these samples were excluded from further analysis. These findings further demonstrate how difficult complex background contamination can be to assess and mitigate.

2.4.5 Identifying and removing global contaminants

Although background contamination was largely lab-specific (Figure S1B & S1D; Table S5), we hypothesized that some contaminants may be reproducible across labs and their presence in tumor replicates could be falsely interpreted as true biological signals. A total of 9 global contaminants were present in the negative controls of both labs. We searched for and found 3 of these 9 ASVs were also reproducible between tumor replicates. *Escherichia-Shigella* and *Staphylococcus* ASVs, which were known members of the both positive controls (Zymo & BEI) were found in 1 (PMP-187) and 4 tissues (PMP-150, PMP-211, PMP-254, NP-24) respectively. *Acinetobacter* was found in 2 tumors (PMP-187, PMP-211).

To better understand whether these 3 ASVs were the result of contamination, or were truly tumor-associated but also present in the negative controls, we utilized Source tracking for Contamination Removal in microBiomes (SCRuB) [24] which accounts for and decontaminates well-to-well contamination. We also included the other 6 global contaminants in our analysis. *Escherichia-Shigella* reads were almost entirely identified as contamination by SCRuB. *Acinetobacter* and *Staphylococcus* retained 100% and 85% of their total reads respectively following decontamination (Figure 3A) indicating that these two taxa may truly be sample-associated. After removing poor quality samples and lab-specific contamination, we removed *Escherichia-Shigella* and 2 other global contaminants (*Enterococcus*, and *Brevibacterium*) identified by SCRuB from our dataset (Figure 1B, 3A, & S2A).

2.4.6 Tumor-associated microbial signals are repeatable across high-quality replicates

Following quality assessment and ASV decontamination, we next sought to determine whether tumor-associated microbial signals were reproducible between sample replicates extracted at different labs. We hypothesized that high-quality sample extractions are likely to contain reproducible tumor-associated microbial features while low-quality sample replicates with low reads and/or gross contamination will have few to no features in common. Although bacterial cell distribution may be heterogeneous within a subset of tumors, we expect ASVs should be more consistently reproducible between high-quality sample replicates. The 21 PMP and 4 NP samples that were extracted at both labs were categorized into three groups based on quality. The “High/High” group contained 6 tumors and 1 NP sample for which both replicates were high-quality. The “High/Low" group consisted of 4 tumors and 2 NP samples where one of the two replicates met high-quality criteria. The final group, “Low/Low”, was composed of 12 tumors and 1 NP sample where neither replicate passed the quality criteria (Table S6). We then considered the number of ASVs reproduced between sample replicates in each quality group and compared their relative abundances.

The 6 PMP tumor samples in the “High/High” quality group ranged in reproducibility. PMP-254 had the highest number of reproducible ASVs (n = 24) and PMP-187 had the next highest (n=8). PMP-150 and PMP-127 only had 2 ASVs in common between replicates, which may suggest that some microbes are heterogeneously distributed across the tissue in a subset of tumors. NP-24 was the only control tissue sample in the “High/High” quality group and had many reproducible ASVs

(n = 20). On average, 8.71 ASVs were reproduced across high-quality sample replicates (Figure 3B).

We found little to no reproducibility between replicates in the “High/Low” and “Low/Low" quality groups (Figure 3B & S2B-C; Table S7). On average, the “High/Low" group had fewer shared ASVs than the sample pairs in the “High/High” group (mean = 2.33) but this difference was not statistically significant (*q*-value = 0.3815). As expected, we found that the “Low/Low” quality group was the least reproducible. 5 of the 12 sample replicate pairs had 0 ASVs in common and the average number of reproducible ASVs was 1.17 which was significantly fewer than the “High/High” group (Figure 3B; *q*-value = 0.01568).

To further assess tumor microbiome reproducibility, we examined whether ASVs were recovered from each replicate at similar relative abundances. A positive Pearson’s correlation between relative abundances would indicate that microbial community composition is reproducible and would further suggest that at least some bacterial taxa are evenly distributed throughout the tumor tissue. We found a significant positive correlation between the relative abundance of ASVs isolated from tumor replicates PMP-254, PMP-433, PMP-187, PMP-211, and NP-24 in the “High/High" quality group (Figure 3C). Conversely, replicates between PMP tumors 127 and 150 had low reproducibility. Both tumors only shared 2 very low-abundance ASVs between replicates that were not reproducible (*r* = -0.095 and -0.27 respectively). This low reproducibility of high-quality samples may be due to technical variation introduced during DNA extraction and PCR, etc. or may be the result of heterogeneously distributed microbial communities as has been previously observed [8,9,12,25].

When we examined sample pairs in the “High/Low” and “Low/Low” quality groups, we found only one tumor with a significantly positive correlation between ASV relative abundances (Figure S2B). The ASVs recovered from the remaining tumors and control tissues within these quality groups either showed no significant correlation between replicates or exhibited a negative correlation. This result, along with the observation that low-quality samples are associated with fewer reproducible ASVs suggest that these lower quality samples contain little to no tumor-associated microbial signal. Alternatively, these findings may be the result of under sampling PMP tumor tissues that contain heterogeneously distributed bacterial communities as was seen with PMP-127 and PMP-150.

2.4.7 “Low/Low” quality group tumors lack detectable tumor-associated bacteria

To determine whether “Low/Low” quality tumors are the result of under sampling tissues with heterogeneously distributed bacteria, we examined a subset of PMP tumors (423, 433, and 485) for which we had generated an additional biological replicate during optimization experiments at lab 2 (two technical replicates per biological replicate; see methods). We hypothesized that if microbial cells reside within niches heterogeneously distributed throughout the tumor, we would expect to generate some high-quality extractions as more of these niches are sampled. In line with the lab 2 quality results, the additional PMP-423 and PMP-485 replicates were considered low-quality due to insufficient sequencing reads (<2000) while the additional PMP-433 replicates were found to be high-quality (>2000 reads and <20% contamination) (Table S3). These findings suggest that PMP tumor samples assigned to the “Low/Low” quality group (e.g. two failed extractions) were not under sampled but rather, may not be colonized by

microbial communities, at least not at a detectable level. However, for PMP-433 which was considered “High/High”, (e.g. two high-quality extractions) we were able to detect the most abundant taxa across all additional replicates demonstrating that at least in this case, the microbial community within a small (~200 mg) sample of tumor seems homogenous (Figure S4). We suspect samples classified as “High/Low” may represent tumors with a medium bacterial load that is marginally detectable, but more research is needed.

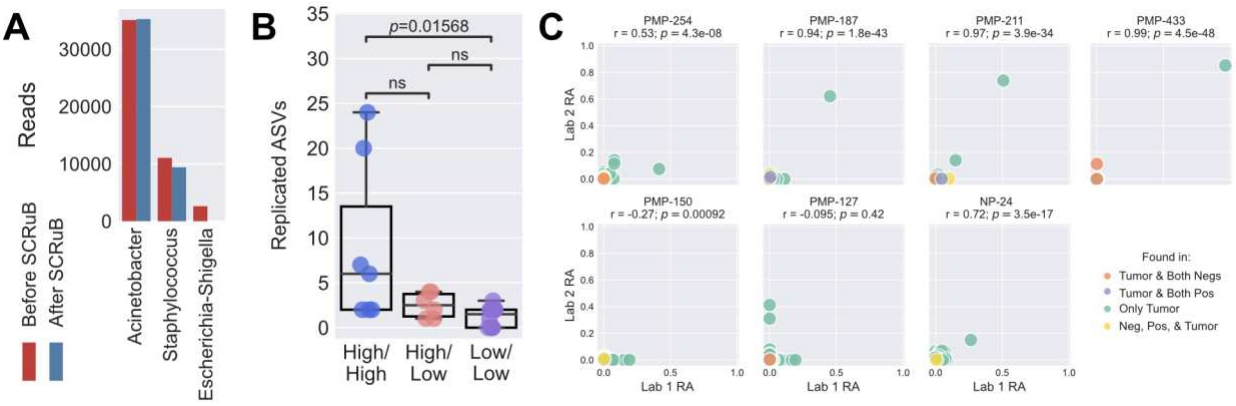


Figure 3. Bacterial signals are repeatable across high-quality DNA extraction replicates. A) Number of total reads for each global contaminant ASV before (red) and after (blue) running the SCRuB pipeline. Reads were summed across all PMP and NP samples. Only ASVs that were reproducible across replicates in the “High/High” quality group are shown. (B) Number of ASVs found in replicated PMP and NP samples grouped by quality category. “High/High” indicates both replicates met quality filtering criteria; “High/Low” indicates 1 passing and 1 failing replicate; “Low/Low” indicates both replicates did not meet quality criteria. Kruskal-Wallis test with Bonferroni correction. (C) Relative abundance (RA) of ASVs found in high-quality replicates from labs 1 and 2. Each point represents an ASV. ASVs that were also found in negative (orange), positive controls (blue), or both (yellow) are highlighted. Pearson correlation (r) was used to assess reproducibility of microbial communities.

2.4.8 Assessing surgical contamination

We next evaluated high-quality NP control extractions for possible surgical contamination. Of the 6 control tissues included in this study, only NP-24 had 2 high-quality replicates with reproducible taxa (n = 20). This sample was taken from a patient undergoing laparoscopic removal of focal fibrous adhesions and an oophorectomy. NP-24 was dominated by ASVs from the Peptoniphilaceae family (Table 1), many of which have previously been isolated from clinical specimens and are known to reside on the skin (e.g. *Corynebacterium*) [26], within the gastrointestinal tract (e.g. *Prevotella*) [27], or the vaginal tract (e.g. *Peptoniphilus, Fenollaria*, and *Ezakiella*) [28,29].

We assessed high-quality extractions from 2 additional control tissues NP-18 and NP-29 to determine whether they shared microbial signatures with NP-24 that may represent prevalent surgical contamination. These controls had been extracted in both labs but only had one high-quality extraction each (i.e. “High/Low” quality group). NP-18 was dominated by *Clostridia-UCG-014* and *Nocardioides* (Figure 4A), neither of which were reproducible between replicates. NP-29 was dominated by *Acinetobacter, Staphylococcus*, and *Lactobacillus* but similarly, none were reproducible. We found that many of the reproducible genera from the Peptoniphilaceae family were unique to NP-24 which indicates they may represent an isolated surgical contamination event or an underlying infection. Furthermore, *Bradyrhizobium*, a common kit contaminant found in lab 1 negative controls, was the only taxon found across all 3 control tissues suggesting surgery may not constitute a source of widespread and reproducible contamination.

Normal peritoneal samples and PMP tumors shared several taxa. *Corynebacterium*, *Staphylococcus, Pseudomonas,* and *Acinetobacter* were present in

numerous tumors and NP-24 and NP-29 extractions (Figure 4A). Notably, *Corynebacterium, Staphylococcus,* and *Pseudomonas* were only reproducible across NP-24 replicates and amounted to a small number of reads (range: 12-302 reads/replicate). Although NP and PMP samples shared some bacteria, PERMANOVA revealed these sample types were significantly different from each other (Figure S5A-B) (Jaccard p=0.031, Bray-Curtis p=0.043). We note that the NP sample size was small compared to PMP tumors which may impact findings. Overall, we did not find clear surgical contamination signatures across the high-quality NP control group. However, some reproducible, low-abundance taxa that were found in NP and PMP samples (e.g. *Corynebacterium*, *Staphylococcus*, and *Pseudomonas*) are likely contaminants but this finding is difficult to interpret as these tissues were isolated from cancer-free patients undergoing surgical procedures unrelated to the PMP cohort.

2.4.9 Comparison with previous PMP tumor study that did not control for contamination

Gilbreath et al. (2013) reported PMP tumors and mucin are colonized by a core microbiome containing bacteria found within the phylum Proteobacteria, Actinobacteria, Firmicutes, and Bacteroides. To confirm whether PMP tumors host a core set of microbial taxa, we first considered only PMP tumors with two high-quality replicates which were the most robust and reproducible in our dataset. ASVs were first collapsed at the phylum level and then phyla were totaled across samples. At the phylum level, we found Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidota were present in 80% or more of PMP tumor extractions, a finding similar to Gilbreth et al (2013). At the genus level, Gilbreth identified to the 34 core microbes, of which we only can confirm as core-

like *Acinetobacter*, *Staphylococcus, Pseudomonas* (50% of our high-quality PMP extractions), *Streptococcus* (50%), *Bacillus* (40%) and *Corynebacterium* (40%) (Figure S5C).

We expanded our analysis beyond those tumors that were only replicated across labs to include ASVs found in all of the high-quality tumor extractions (n = 21 samples). The prevalence of Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidota decreased to 70%. *Pseudomonas* and *Staphylococcus*, were found across 70% of tumor DNA (Figure 4B) while *Acinetobacter* and *Corynebacterium* were present in 50%. We also found *Streptococcus* and an unidentified genera in the Comamonadaceae family in 40% of PMP extractions. Despite a smaller group of core genera compared by Gilbreath et al. 2013, *Pseudomonas, Acinetobacter*, and *Streptococcus* were consistently identified in both PMP studies (Table 2). However, when taking the NP findings into consideration, it is unclear whether these core taxa are truly tumor-associated or represent surgical contamination.

Overall, we found 60 ASVs belonging to 14 different genera that were previously isolated from PMP tumors and mucin by the Glibreath et al. (2013) study. These ranged in prevalence and relative abundance (Table 2) and include genera that were identified here as core PMP members (*Pseudomonas*, *Staphylococcus*, *Acinetobacter*, and *Corynebacterium*). To determine whether these taxa may have been the result of contamination, we searched for their presence in the negative controls of both labs. We found several of the most prevalent genera had a relative abundance approaching or exceeding 1% in the negative control potentially signifying contamination. However, *Sphingomonas*, *Roseomonas,* and *Delfita* were present in more than one patient with

mean relative abundances ranging from 0.1-5%. Along with other taxa reported here, *Sphingomonas* and *Delfita* have previously been isolated from gastrointestinal tumor tissues [30,31]. In contrast to Gilbreath et al. (2013) where it was the most abundant genera, *Methylobacterium* was only found in one tumor sample in low abundance (0.3%).

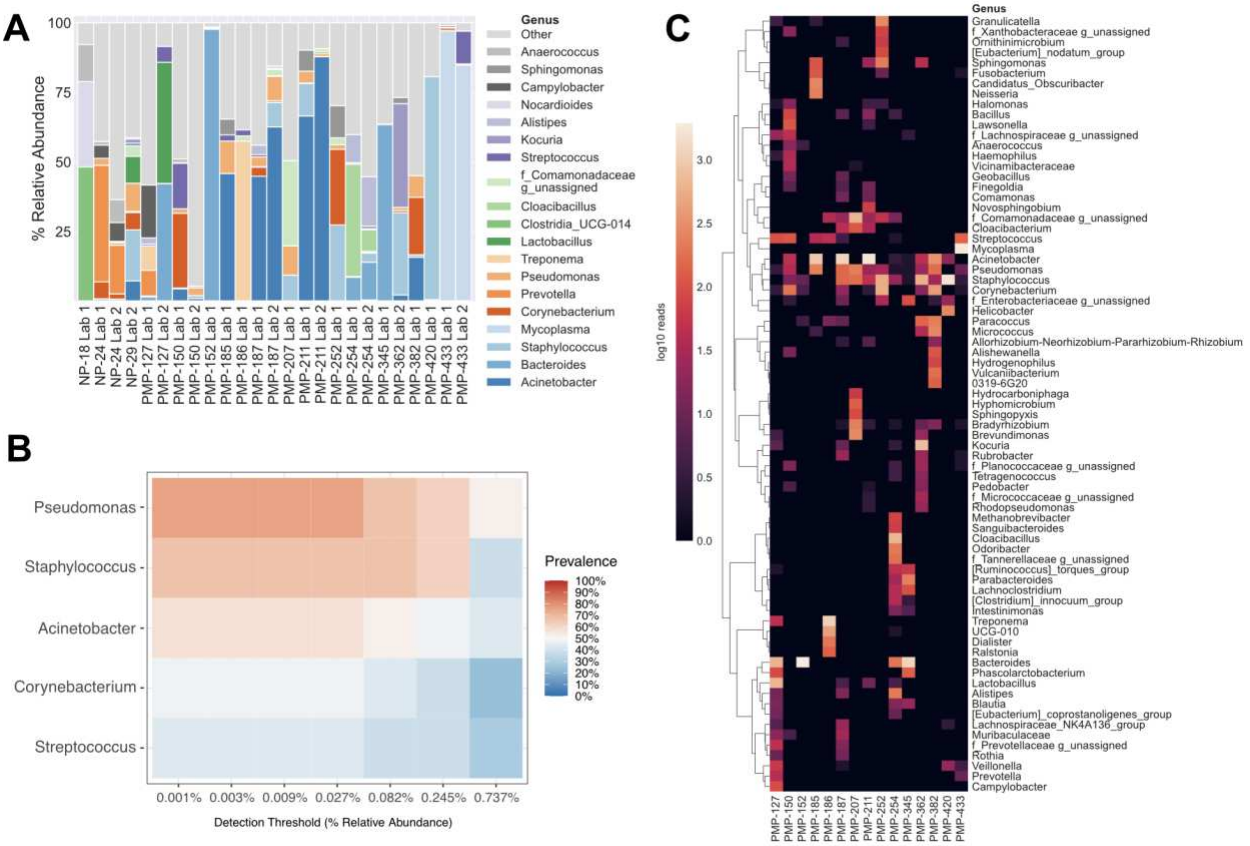


Figure 4. The Core PMP tumor microbiome. A) Taxa bar plot showing percent relative abundance of each high-quality normal peritoneal and tumor extractions grouped at genus level. Top 19 genera shown. B) Core microbiome at genus level. All high-quality PMP extractions were included in analysis. C) Heatmap showing $\log_{10}$ reads of the 75 most abundant genera found in high-quality PMP tumor extractions. Tumors with replicates were grouped and ASVs were summed together and then collapsed at genus level.

Several bacteria that were previously isolated and cultured by Glibreath et al. (2013) were also identified here. These include 17 ASVs within 3 genera and 1 family (*Chitinophagaceae*). Of note was an unclassified isolate within the *Chitinophagaceae* family that Gilbreath et al. (2013) found to interact with MUC2 secreting cells *in vivo,* suggesting that PMP-associated bacteria may potentially use mucin as a carbohydrate source. Although this unclassified *Chitinophagaceae* isolate was not detected in our data, we identified two different ASVs from this family (Table 2). *Chitinophagaceae* are typically associated with soil but were not found in controls associated with either lab, which indicates that they may be tumor-associated. Together, these findings suggest some of previously identified genera may be tumor-associated while others are likely the result of contamination.

Table 2. PMP tumor genera shared with Gilbreath et al. (2013). Column 1 shows genera found across both studies. Column 2 depicts the number of patients from this study each genera was found in (tumors with 2 high-quality extractions were grouped together). Columns 3-5 show the mean, minimum, and maximum percent relative abundance (RA) of each genus in this study. Column 6 shows % relative abundance of each genus in negative controls of either lab with > 100 reads (NECs + NTCs). * Denotes bacteria that Gilbreath et al. identified via culturing.

Genus	Number of Patients	Mean % RA	Min % RA	Max % RA	% RA in negative controls
Pseudomonas	12	3.953125	0.0625	12.8125	0.98
Acinetobacter	9	22.298611	0.0625	78.3125	0.91
Streptococcus	8	2.871094	0.25	8.75	0.77
Pedobacter	4	0.234375	0.03125	0.8125	0
Sphingomonas	4	5.546875	2.125	11.75	0
Delftia	2	0.15625	0.125	0.1875	0.11
Roseomonas	2	0.109375	0.03125	0.1875	0
Escherichia-Shigella	1	0.625	0.625	0.625	0
Flavobacterium	1	0.25	0.25	0.25	0
Helicobacter	1	15.875	15.875	15.875	0
Methylobacterium-Methylorubrum	1	0.34375	0.34375	0.34375	0
Moraxella	1	0.125	0.125	0.125	0
Ralstonia	1	7.0625	7.0625	7.0625	0.02
Stenotrophomonas	1	0.8125	0.8125	0.8125	0
*Corynebacterium**	9	7.336806	0.0625	28.5625	3.55
*Dermacoccus**	3	0.0625	0.03125	0.09375	0.14
*Chitinophaga**	1	0.03125	0.03125	0.03125	0
*f_Chitinophagaceae g_unassigned**	1	0.75	0.75	0.75	0

2.4.10 Reproducible PMP-associated taxa are associated with colorectal cancer

We next focused on the most robust bacterial signals that were not considered "core" taxa but were reproduced across high-quality tumor replicates which strongly indicates they represent true biological signals (Table 1). As appendiceal cancers are relatively rare and studies examining tumor-associated bacteria are limited, we sought to compare reproducible PMP-associated bacteria with those implicated in colorectal cancers (CRC). CRC and PMP tumors have distinct mutation profiles. A high percentage of CRC tumors contain *APC* and *TP53* gene mutations while *KRAS* and *GNAS* mutations are more common in PMP tumors [32]. Despite these differences, both tumor types are exposed to similar gut microbes that may play comparable roles in cancer progression and survival outcomes.

We found several reproducible PMP-associated taxa belonging to genera (see Table 1) that have been implicated in CRC. *Bacteroides fragilis*, *Fusobacterium nucleatum*, and *Helicobacter pylori* are most frequently associated with CRC development [1,33–35]. Although we were not able to achieve species-level resolution with our sequencing strategy, we identified 4 reproducible *Bacteroides* ASVs associated with PMP-254 totaling 367 reads. Other *Bacteroides* sequences were found in high-quality extractions of 3 additional tumors (PMP-127, PMP-152, and PMP-345) and together, accounted for more than 3000 reads. *Fusobacterium* was found in low abundance across 3 different tumors (PMP-185, PMP-252, and PMP-433) extracted at lab 1 but was not reproducible between tumor replicates. *Helicobacter* was only found in a single extraction from PMP-420 (n=236 reads).

Alistipes and *Streptococcus* have recently emerged as potential contributors to CRC and were found across several tumors [36,37]. PMP-254 contained 3 reproducible *Alistipes* ASVs totaling 450 reads. This genus was also found in individual PMP-127 and PMP-187 extractions in low abundance (n=87 total reads) (Figure 4C). Only a single *Streptococcus* sequence was reproducible between PMP-433 replicates (n=196 reads). However, we found *Streptococcus* ASVs in 7 other patients totaling 476 reads. The presence of several CRC-associated genera across a number of PMP tumors suggest these cancers may share a similar tumor microbiome.

Although typically associated with esophageal and other cancers [38], we found a robust *Mycoplasma* signal associated with one PMP tumor. *Mycoplasma* dominated PMP-433 (n=2911 reads) and was reproducible between and within labs (Figure 4A & Figure S4). Importantly, *Mycoplasma* was not detected in the negative controls of either lab and is not recognized as a common laboratory contaminant.

While many bacteria are associated with poor overall survival, we found several taxa within PMP tumors that are associated with improved CRC outcomes [33,39]. Two reproducible sequences within the *Odoribacter* genus were found in PMP-254 totalling 240 reads. We found reproducible *Blautia* (n=23 reads) and *Ruminococcus* (n=72 reads) ASVs also associated with PMP-254 along with 3 other sequences belonging to the *Lachnospiraceae* family. The current status of this patient is unknown, however, their disease-free survival was 60.8 months at the time of their last visit (median DPAM disease-free survival = 41.9 months) [40]. Both *Blautia* and *Ruminococcus* were isolated from several other tumors (n=48 and n=48 total reads respectively) (Figure 4C). These results highlight the complex relationships between different bacterial species, the host,

and disease, and further demonstrate the challenge of interpreting results due to our limited understanding of how these microbes function within the tumor microenvironment.

2.4.11 Chemotherapy treatment is correlated with lower tumor microbiome diversity

In addition to the standard treatment of cytoreductive surgery (CRS) and hyperthermic intraperitoneal chemotherapy (HIPEC), metastasized appendix cancers are commonly treated with colorectal-type systemic chemotherapy regimens [32,41]. Patients may receive systemic FOLFOX (folinic acid, fluorouracil, and oxaliplatin) or FOLFIRI (folinic acid, fluorouracil, and irinotecan) chemotherapy prior to surgery in an effort to shrink tumors and then again postoperatively to treat any residual tumor and delay recurrence in high-risk patients [42,43]. Several patients in this study received systemic FOLFOX, FOLFIRI, or Cisplatin before cytoreductive surgery where tumor samples were obtained or had previously been treated with CRS and HIPEC and were undergoing a subsequent procedure for PMP recurrence.

As chemotherapy agents are cytotoxic, can have antimicrobial properties [44,45], and are known to impact gut microbiome composition [46,47], we sought to better understand what effect previous treatment may have had on the PMP tumor microbiome. Unweighted Unifrac PCoA analysis of high-quality tumor extractions showed that patients who had previously received chemotherapy had a distinct tumor microbiome compared to those who had never received treatment (Figure 5A) (PERMANOVA p=0.007) which was associated with lower alpha diversity. We found that the number of unique bacterial ASVs and Faith's phylogenetic diversity was lower

in patients who had received chemotherapy before surgery (Figure 5B-C). ANCOM-BC was used to determine whether any taxa were differentially abundant between groups but found no significant difference. We additionally hypothesized that prior chemotherapy treatment may decrease sample quality but found no correlation (Chi-square p=0.2). The results presented here indicate that at least certain types of chemotherapy treatment can reduce bacterial diversity and potentially mask important biological signals.

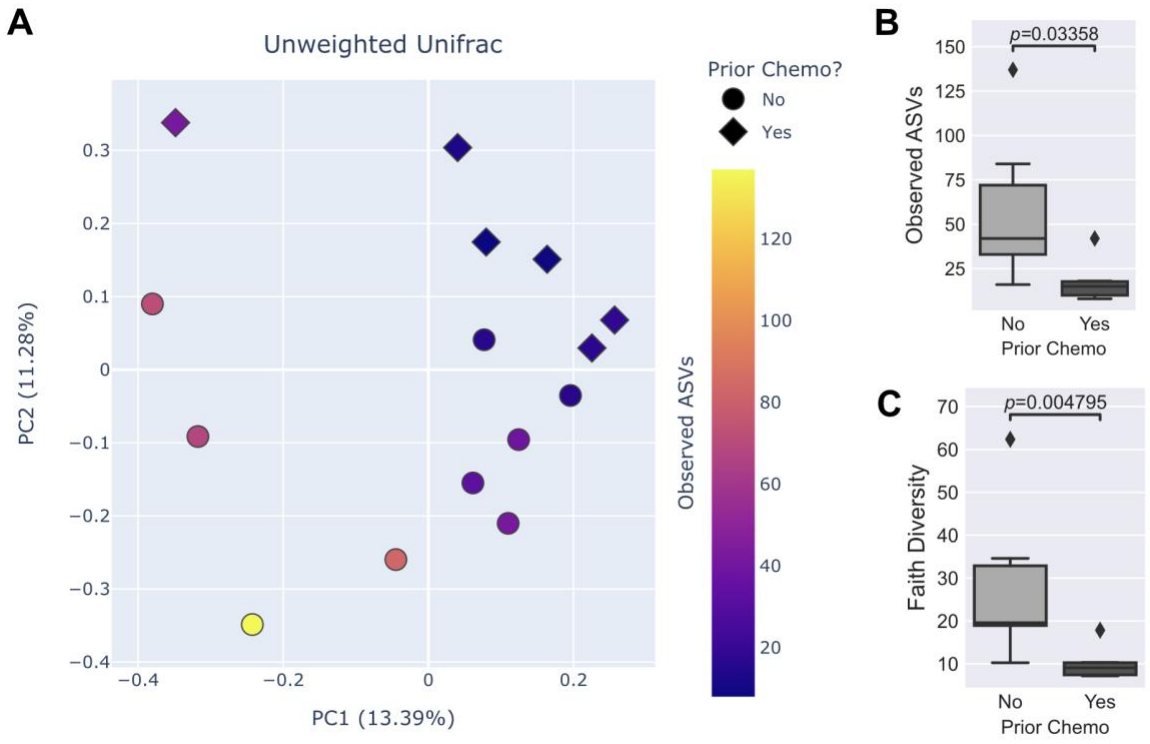


Figure 5. Prior Chemotherapy is associated with lower PMP tumor microbiome diversity. A) Unweighted Unifrac PCoA plot showing all high-quality tumor samples colored by the number of unique ASVs. Diamonds represent individuals who received prior chemotherapy. High-quality tumor extractions were rarefied to 2000 reads per sample and tumors with replicates were then grouped together. PERMANOVA p=0.007. B) Observed ASVs and C) Faith's Diversity metrics showing individuals who received systemic or HIPEC chemotherapy regimen prior to the cytoreductive surgery in which samples were obtained. Tumors with high-quality replicates were grouped together.

2.5 Discussion

Tumor-associated microbes are challenging to identify because of their low biomass, especially compared to human cell biomass. Typically, when low biomass samples are extracted and prepared for sequencing, even with great caution, their DNA becomes a fraction of the pool of DNA molecules that are ultimately sequenced. In this study, we examined the Pseudomyxoma Peritonei tumor microbiome from a new cohort of 70 patients using a cross-laboratory replication approach. This approach, alongside a quality control filtering workflow, allowed us to identify and focus on samples that likely contained tumor-associated microbes, and used normal peritoneal (NP) tissue to help assess surgical contamination. In contrast to the Gilbreath et al. study (2013), we did not find evidence that bacteria were associated with every tumor. This lower percentage of bacteria-positive tumors is consistent with [6] who reported that microbes colonize 14-63% of tumors. Furthermore, we found little evidence to support the finding that *Helicobacter* is a ubiquitous PMP tumor taxon (Glibreath et al. 2013). In addition, we detected a smaller core microbiome consisting of 4 taxa, 3 of which, *Pseudomonas*, *Acinetobacter*, and *Streptococcus,* were also reported by [48], but these were present in NP samples as well. We were surprised to find that several NP samples resulted in high-quality DNA extractions, including one sample with 20 reproducible ASVs. Of the core set of bacterial genera present in 40-70% of all high-quality tumor extractions– *Staphylococcus*, *Pseudomonas, Acinetobacter, Corynebacterium,* and *Streptococcus*, all but *Streptococcus* and *Acinetobacter* were present and reproducible across NP-24 (although in low relative abundance). Since these taxa are often associated with human skin [26], they may represent surgical or surgical suite contamination. Incorporating

additional samples to control for surgical contamination such as patient skin prior to surgical preparation, skin at the incision site during surgery, and samples of surgical equipment will be critical to untangle biological signals from contamination in the future.

We identified multiple bacterial taxa that are likely tumor-associated owing to their reproducibility across independently extracted and sequenced samples, their presence in other high-quality PMP tumors, and their absence from control tissues (Tables 1 & 2). Several of these taxa have been linked to colorectal cancer indicating they may also play a possible role in the development of progression of PMP, perhaps through increased inflammation and cell proliferation. In particular, *Fusobacterium* and *Bacteroides* have strongly been implicated in CRC development and were recovered from 3 and 4 PMP patients respectively. Several different *Alistipes* species, and *Streptococcus* have more recently been associated with CRC tumors and were recovered from 3 and 8 patients respectively. Finally, PMP-433 was dominated by *Mycobacterium* which has been linked with inflammation and cancer. Additionally, we provide preliminary evidence that chemotherapy treatment reduces tumor microbiome diversity. Interestingly, there was no correlation between chemotherapy treatment and sample quality suggesting treatment may not entirely eradicate detectable tumor-associated signals. Neoadjuvant chemotherapy is commonly used in the treatment of several cancer types including breast, colorectal, lung, and esophageal cancers which may potentially impact tumor microbiome studies. Recently, [49] reported anthracycline-based chemotherapies tended to decrease microbial richness in metastatic cancer suggesting that even chemotherapies with different modes of action may impact results.

Hence, future tumor microbiome studies should take care when interpreting results from patients who received chemotherapy treatments.

Identifying and mitigating background contamination remains a difficult obstacle with little consensus in the microbiome field about the best way forward. For example, we estimated that molecular lab background contamination can account for more than 50% of reads in low-biomass tumor samples. We demonstrate how cross-lab replication can be used to reveal and mitigate unique, lab-specific microbial contamination signatures (e.g. *Comamonadaceae, & Geobacillus)*. We found that some contaminants are not commonly reported, suggesting that low-biomass studies should not solely rely on previously reported contamination but should also utilize numerous controls that capture contamination along the sample processing pipeline (e.g. surgical controls, kit reagents, library prep., etc.). However, even when following gold-standard, low-biomass recommendations, we found several human-associated ASVs (e.g. *Staphylococcus*) that were present in lab 1 and 2 negative and positive controls, and tumor samples making it challenging to determine their source(s). Regardless, cross-laboratory replication remains a powerful tool for disentangling biological signals from contamination and is applicable to all low-biomass microbial systems. We anticipate integrating this approach into future studies as a validation check on a subset of samples for instance, will help improve reproducibility and strengthen confidence in recent tumor microbiome research enabling the field to focus on clinically relevant signals.

Here, we provide several additional recommendations based on our findings for future low-biomass studies to increase confidence and help untangle contamination

from biological signals. 1) Use positive controls composed of taxa from a non-related environment to improve identification of positive control to sample cross-contamination events. For example, consider using a soil community when studying human-associated microbiomes. 2) When possible, use sequencing methods that provide species-level resolution. An identical 300 base pair ASV present in controls and biological samples may be labeled as contamination but with higher resolution, may turn out to be different microbes. 3) Perform DNA extractions in a single tube format and away from positive controls to help mitigate cross contamination between samples and controls.

2.6 Limitations of the study

We identified several limitations in this study that will help guide future research. Most importantly, we lacked a robust, patient-matched, surgical sampling control. We intended for peritoneal samples taken from non-PMP patients to provide such a control, but our assumption that the peritoneal cavity, especially during acute trauma (i.e. surgery) would be sterile, may not be true. Thus, in future studies, collecting multiple types of controls that could collectively be used to assess surgical contamination may be warranted. Second, our sequencing approach did not provide species-level resolution and consequently, the specific classification of several taxa we found that may be associated with CRC (e.g. *Fusobacterium*, *Bacteroides,* etc.) was unknowable. Additional studies that comprehensively characterize the PMP tumor microbiome and test its effects on carcinogenesis, progression, and treatment outcomes are needed to move beyond correlation and establish causation.

2.7 Materials & Methods

2.7.1 PMP patients and sample collection

PMP specimens from 70 patients were collected peri-operatively at Mercy Medical Center in Baltimore, MD. All procedures were approved by the Institutional Review Boards of the Mercy Medical Center and the Uniformed Services University of the Health Sciences (MMC# 2010-31 or IRBnet# 1368159) in compliance with all applicable federal regulations governing the protection of human subjects. Written informed consent was obtained from all patients prior to enrollment in the study. Patients underwent Cytoreductive surgery and hyperthermic intraperitoneal chemotherapy (CRS-HIPEC) by one of two lead surgeons in one of two operating rooms (OR suit 5 or 6) both of which are on the same floor in a secure OR area. All personnel in the area, whether scrubbed in or not, are required to wear specific OR scrubs, shoe protection, hair cover, surgical mask, and eye protection. The day before surgery, patients are admitted to the hospital and undergo bowel prep, which includes Suprep (177 mL, x2 doxes), erythromycin (1000 mg, x3 doses), neomycin (1000 mg, x3 doses), and dulcolax (20 mg, x1 dose). The night before and the morning of surgery, patients receive a chlorhexidine bath. The day of surgery, once under anesthesia, patients also undergo a mechanical bowel prep using betadine and intravenous fluids (IV). The entire abdominal skin is prepped by being shaved, using a chlorhexidine wash, and finalized with an alcohol rinse to dry the area. Sterile drapes are used to isolate the prepped field. Once the incision is made, additional IV Cefotetan (2 g, x1 dose) is administered by the anesthesiologist. Once the abdominal cavity has been entered and an initial exploration is complete, research samples consisting of both tumor and mucin

are excised by the surgeon and handed to the scrub tech, who is also sterile/scrubbed into the procedure. The scrub tech uses a sterile instrument to place the sample directly into a DNA/RNA-free sterile collection tube. Once the sample is secured in the tube, research personnel (non-sterile, but using gloves to handle the tubes and wearing a surgical mask, eye protection, and hair/shoe protection) place the tube in liquid nitrogen to freeze it. Once sample collection is complete, all samples are removed from the liquid nitrogen, placed into a specimen bag, and stored in a -70C freezer.

Similarly, control peritoneal specimens were collected from 6 patients undergoing routine abdominal surgeries at Mercy Medical Center. Multiple surgeons who use various ORs on the same floor contributed samples. These included general surgeons, gynecologists, and other surgical oncologists from the same department as the two lead PMP surgeons. Patients arrive the day of surgery but were instructed to complete the chlorhexidine bath the night before and the morning of surgery. The same skin prep is used for the abdomen. However, depending on the surgery, the patient likely did not undergo bowel prep. All surgeries for normal cases had incisions in the abdomen that fully entered the abdominal cavity. Some had one midline incision (same as HIPEC) and others had multiple smaller incisions for robotic/laparoscopic surgeries. Once the abdominal cavity has been entered and an initial exploration is complete, research samples consisting of a biopsy of normal appearing abdominal wall peritoneum are excised by the surgeon and handled in the same manner as the PMP tumor samples above.

All procedures were approved by the Institutional Review Boards of the Mercy Medical Center and the Uniformed Services University of the Health Sciences. Written informed consent was obtained from all patients prior to enrollment in the study.

2.7.2 Lab 1 DNA extraction

Lab 1 DNA extractions,16S rRNA gene library construction, and sequencing were carried out by the Merrell Lab (Uniformed Services University of the Health Sciences, Bethesda, MD).

Total DNA from PMP and control tissue samples was extracted using the DNeasy PowerSoil HTP 96 Kit (Qiagen Cat. No. 12955-4), which combines chemical and mechanical lysis. Approximately 250 mg of each tissue sample was weighed out and added to the PowerSoil-HTP Bead Plate. As a positive DNA extraction control, a dilution series of Mixed Microbial Mock Community C (BEI Resources Cat No. HM-281) were extracted along with the tissue samples. The positive control series included dilutions from 10^7 cfu/mL down to 10^2 cfu/mL (6 total extractions, 50 ul per extraction). As a negative DNA extraction control, eight wells were left empty. Following the addition of the tissue samples or positive control samples, 750 ul of Bead Solution and 60 ul of C1 were added to the wells, and the plates were incubated at 65 °C for 15 minutes prior to mechanical lysis. Plates were homogenized using a Tissuelyser (Qiagen; Hilden, Germany) at speed 20 for 10 minutes, flipped, and then homogenized for an additional 10 minutes. Remaining steps of the PowerSoil HTP protocol were followed as written by the manufacturer. DNA was eluted from the spin filter plate with 100 ul of nuclease-free water (Integrated DNA Technologies; Coralville, IA). The tissue weighing, DNA extraction, and PCR setup steps were performed in a biosafety cabinet.

2.7.3 Lab 1 library preparation and sequencing

The V4 hypervariable region of the 16S rRNA gene was used to characterize tumor-associated microbial communities, using previously published primers: 515F (5' AATGATACGGCGACCACCGAGATCTACACGCT XXXXXXXXXXXX TATGGTAATT GT GTGYCAGCMGCCGCGGTAA 3') (Parada et al., 2016) and 806R (5' CAAGCAGAAGACGGCATACGAGAT AGTCAGCCAG CC GGACTACNVGGGTWTCTAAT 3') (Apprill et al., 2015). PCR reactions were prepared in a 35 ul volume as described below and then divided into 3 separate 10 ul reactions: 3.5 ul 10 uM barcoded 515F, 3.5 ul 10 uM 806R, 14 ul of 5Prime HotMasterMix (Quantabio Cat. No. 2200400), and 14 ul DNA template. PCR cycling parameters were as follows: initial heating step at 94 °C for 2 minutes, followed by 34 cycles of 94 °C for 20 seconds, 60 °C for 10 seconds, and 65 °C for 50 seconds, followed by a final elongation step at 72 °C for 10 minutes. Following PCR, the triplicate reactions were pooled back together. Approximate amplicon size and quantity were assessed using the Agilent BioAnalyzer (Agilent Technologies; Santa Clara, CA). Sequencing of the resulting pooled libraries was conducted using one run on the Illumina MiSeq platform, with 600 cycle, version 3 chemistry (Illumina; San Diego, CA) at the Naval Medical Research Command (Fort Detrick, MD).

2.7.4 Lab 2 Methods

Lab 2 experiments, DNA extraction, 16S rRNA gene library construction, and sequencing were carried out by the Metcalf Lab (Colorado State University in Fort Collins, CO). Tumor and control tissue samples were transferred, sequenced and analyzed under CSU IRB protocol #3467.

2.7.5 Lab 2 DNA Extraction Optimization Experiments

Due to the low biomass nature of tumor-associated bacterial communities, increasing the amount of tissue added to each extraction may help improve bacterial DNA concentrations. However, increasing tissue input may reduce yields by decreasing tissue lysis efficiency. We performed several experiments to optimize tissue input amount.

3 PMP tissue samples (PMP-423, PMP-433, PMP-485) were used to test 25 and 100 mg of tissue input using the legacy DNeasy PowerSoil Kit (Qiagen Cat. No. 12888). All tissue dissection and DNA extraction were carried out in a biosafety cabinet. Genomic DNA was measured using Nanodrop (Thermo Scientific). Yields increased from an average of 27 ng/ul to 97.5 ng/ul from 25 to 100 mg of tissue respectively (Table S8). PCR was performed in a laminar flow hood as described below. Of the three tumor samples, only PMP-433 produced visible 16S rRNA gene amplicons. Interestingly, a PCR gel showed that 16S rRNA gene amplicon concentrations appeared to be visibly higher when 25 mg of tissue was used (Figure S6A). Several off-target amplicons (approximately 300 and 500 bp) were produced in addition to the correct 16S rRNA gene amplicon. These amplicons were most likely produced by off-target primer interactions with host DNA as they were not readily seen in the host DNA-free positive control (data not shown).

Characterizing the tumor microbiome is technically challenging due to a combination of low microbial biomass and large amounts of co-purifying host DNA. We reasoned that mammalian cells easily lyse with chemical agents while bacterial cells require a harsher, mechanical lysis step. We hypothesized that reducing host DNA

would reduce off-target amplicon production and increase detection of rare abundance microbes. The UltraClean Tissue & Cells DNA Isolation Kit (MoBio Cat. No. 12334) was used in combination with the DNeasy PowerSoil Kit (Qiagen Cat. No. 12888) to first lyse host cells, remove the released host DNA, and then lyse bacterial cells with the PowerSoil Kit. 25 mg of tissue was added to a PowerSoil Bead Tube filled with 750 ul of Solution TD1. Proteinase K was omitted. Samples were homogenized for 15 minutes on the max setting (Scientific Industries, Vortex-Genie 2). Bead tubes were centrifuged at 12,000 x g for 1 minute to pellet cell debris and potential bacterial cells. Supernatant containing host DNA was removed and retained to check for bacterial DNA (host fraction). 750 ul of Bead Solution and 60 ul Solution SL were added to the bead tubes containing the cell pellet. The PowerSoil protocol was then followed to completion (bacterial fraction). PCR and 16S rRNA gene sequencing were performed as described below. On average, the host DNA fraction contained higher concentrations of DNA (34.1 ng/ul) than the bacterial fraction (14.9 ng/ul) (Table S9). Again, of the three tumor samples, only PMP-433 had visible 16S rRNA gene amplicons which were present in both fractions (Figure S6B). Off-target amplicons were present in both fractions of both tumor samples suggesting host DNA had not been sufficiently reduced.

Finally, we tested a commercially available host DNA depletion kit that utilizes a DNase digestion step to remove host DNA contamination prior to bacterial lysis. The HostZERO Microbial DNA Kit (Zymo Cat. No. D4310) was tested with the 3 PMP tumor samples listed above following the manufacturer's protocol which included an initial 15 minute tissue homogenization step using 2 mm ceramic beads (Zymo Cat. No. S6003-50). PCR and amplicon sequencing were performed as described below. The DNase

digestion step successfully removed host DNA which was too low to quantify via nanodrop. All samples produced visible 16S rRNA gene amplicons including the NECs. Off-target amplicons were not observed further suggesting that a high amount of background host DNA is the cause. It was discovered that the ProK solution was contaminated with an ASV from the *Rhizoibiaceae* family (Figure S6C) which was not detected when the Proteinase K solution was removed from the HostZero protocol.

2.7.6 Lab 2 Final DNA Extractions

All PMP tumor and normal peritoneal tissue dissection and DNA extraction were carried out in a biosafety cabinet. Samples were dissected using ethanol and flame sterilized forceps and scalpel. The DNeasy PowerSoil Pro Kit (Qiagen Cat. No. 47016) was chosen as a replacement for the recently discontinued DNeasy PowerSoil Kit. 25 - 50 mg of tissue sample was added directly to a 2 mm ceramic bead tube (Zymo Cat. No. S6003-50). 8 negative and 2 positive extraction controls were processed alongside tissue samples following the same protocol in order to capture environmental and reagent contamination and to assess bacterial DNA recovery. 1 mL of Solution CD1 was added to each tissue and control sample. 5 ul of ZymoBIOMICS Microbial Community Standard was added to each positive control. All samples and controls were homogenized on the highest setting for 15 minutes using a vortex adapter and Vortex-Genie 2. The entire sample volume and any remaining unhomogenized tissue was transferred to PowerBead Pro Bead Tubes. Samples were vortexed on the highest setting for an additional 10 minutes. DNA purification was carried out starting at step 3 of the manufacturer's protocol. Samples were eluted in 50 ul of C6 and stored at -20°C until further processing. DNA yields can be found in Table S10.

2.7.7 Lab 2 library construction and sequencing

The V4 hypervariable region of the 16S rRNA gene was amplified according to the Earth Microbiome Project [50]. 50 ul PCR reactions were made up as follows: 2 ul 10 uM barcoded 515F (5' AATGATACGGCGACCACCGAGATCTACACGCT XXXXXXXXXXXX TATGGTAATT GT GTGYCAGCMGCCGCGGTAA 3') [51], 2 ul 10 uM 806R (5' CAAGCAGAAGACGGCATACGAGAT AGTCAGCCAG CC GGACTACNVGGGTWTCTAAT 3') [52], 25 ul of Platinum™ Hot Start PCR Master Mix (2X) (ThermoFisher Cat. No. 13000012), 17 ul nuclease-free water, and 2 ul DNA template. Positive controls were diluted to 0.001 ng/ul prior to PCR. PCR cycling parameters were as follows: initial heating step at 94 °C for 3 minutes, followed by 35 cycles of 94 °C for 45 seconds, 50 °C for 60 seconds, and 72 °C for 90 seconds, followed by final elongation step at 72 °C for 10 minutes. PCR reactions were quantified using Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Cat. No. P11496). Samples were pooled in an equimolar ratio and the final pool was cleaned using MinElute PCR Purification Kit (Qiagen Cat. No. 28004). The 16S rRNA gene libraries were sequenced on an Illumina MiSeq instrument using a V2 500 cycle MiSeq Reagent Kit (Illumina Cat. No. MS-102-2003) at the Colorado State University Next Generation Sequencing Core (Fort Collins, CO).

2.7.8 Data Analysis

The QIIME 2 versions 2022.2 and 2022.8 (q2) bioinformatics platform were used for data processing and analysis [53]. Sample sequence data was generated across three separate sequencing runs. Lab 1 sample data and lab 2 optimization experiment sample data were provided as demultiplexed files which were separately imported into

q2. Lab 2 raw paired end sequence data was imported and demultiplexed using the q2-demux plugin. All three datasets were separately denoised using the q2-dada2 plugin [54]. Forward and reverse reads were trimmed to 233 and 211 respectively. Following denoising, feature tables and representative sequence files were combined. The merged dataset contained a total of 1,450,549 reads across 169 samples with a mean of 8,583 reads per sample.

ASVs were assigned taxonomy using the q2-feature-classifier [55,56] classify-sklearn naïve Bayes taxonomy classifier trained on the V4 (515f/806r) region of Silva 138 99% operational taxonomic units (OTUs) [57]. ASVs classified as mitochondria, mammalia, chloroplast, eukaryote, or were unclassified at the kingdom level were removed from the dataset. Following this additional filtering step, a total of 514,393 reads remained across 115 samples with a mean of 4,472 reads per sample.

SourceTracker2 [21] was used to estimate the proportion of negative control contribution to each PMP and NP samples. It is important to note that these negative controls only represent possible extraction kit, PCR reagent, and lab environmental contamination and therefore, contamination introduced during surgery cannot be estimated. Lab 1 and lab 2 datasets were processed separately using feature tables rarefied to 800 reads per sample. Negative controls were grouped by sample type (NEC vs NTC).

The SEPP method from the q2-fragment-insertion plugin was used to construct a rooted phylogenetic tree using the Silva 128 reference phylogeny [58,59]. Data were rarefied to 800 reads per sample and the core-metrics-phylogenetic pipeline from the q2-diversity plugin was used to calculate alpha and beta diversity metrics [60–62].

perMANOVA [63] was used to test for beta diversity differences between sample quality groups and negative controls using the beta group significance visualizer from the q2-diversity plugin. Following sample quality assessment, high-quality sample extractions were rarefied to 2,000 reads per sample and the core-metrics-phylogenetic pipeline was run again.

ASVs were considered global contaminants if they were found in any negative controls (NEC & NTCs) across both labs. SCRuB was used to estimate the percentage of well-to-well contamination and identify possible contaminants. SCRuB was run on lab 1 and lab 2 dataset separately. We observed the presence of positive control ASVs in negative controls and tumor samples in both lab dataset. For that reason, we set the positive controls as the first source of contamination followed by the negative extraction controls, and then negative template controls. Following decontamination, the lab 1 and lab 2 dataset were merged. Reads for each of the nine potentially global contaminant ASVs were summed and compared to pre-decontaminated reads. An ASV was removed from the dataset if SCRuB removed at least 50% of total reads (*Escherichia-Shigella, Enterococcus*, and *Brevibacterium*).

Core PMP tumor taxa were identified using the Microbiome R package [64]. Global contaminant ASVs (*Escherichia-Shigella, Enterococcus*, and *Brevibacterium*) that were found in the negative control samples across both labs were removed from the high-quality PMP tumor extractions prior to identifying core features. High-quality tumor extractions were collapsed at the phylum and genus levels. Features were considered “core” if they were found in 50% or more of samples. The Microbiome package in R was also used to generate core microbiome heatmaps.

Q2 artifacts were imported into Jupyter notebooks for plotting using the following packages: python (version 3.8.13), pandas (version 1.2.5), matplotlib (version 3.5.1), seaborn (version 0.11.2), and statannotations (0.4.4).

2.7.9 Data Availability

All raw amplicon sequencing data is available under European Nucleotide Archive project number PRJEB67895 (or study number ERP152899). Raw amplicon data is also available under QIITA study ID 14770 at https://qiita.ucsd.edu. Qiime2 commands, analysis code, files, and visualizations can be found at Open Science Framework project name PMP Microbiome (https://osf.io/e5d84/).

CHAPTER 2: REFERNECES

1. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science. 2017;358: 1443–1448. doi:10.1126/science.aal5240

2. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. Immunity. 2015;42: 344–355. doi:10.1016/j.immuni.2015.01.010

3. Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, et al. Helicobacter pylori infection and the risk of gastric carcinoma. N Engl J Med. 1991;325: 1127–1131. doi:10.1056/NEJM199110173251603

4. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med. 2001;345: 784–789. doi:10.1056/NEJMoa001999

5. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, et al. Commensal Microbiota Promote Lung Cancer Development via γδ T Cells. Cell. 2019;176: 998-1013.e16. doi:10.1016/j.cell.2018.12.040

6. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y. The human tumor microbiome is composed of tumor type–specific intracellular bacteria. 2020. Available: https://science.sciencemag.org/content/368/6494/973.abstract?casa_token=jY5Gg13B7RQAAAAA:PbUzBEfoynd1LqUi54cI-eX3fgoRjkWrrCW4PJ5T8WIwsdLbI4ID4YhL-CCC5IDTS-6r-8op-s7Ksg

7. Sepich-Poore GD, McDonald D, Kopylova E, Guccione C, Zhu Q, Austin G, et al. Robustness of cancer microbiome signals over a broad range of methodological variation. Oncogene. 2024;43: 1127–1148. doi:10.1038/s41388-024-02974-w

8. Fu A, Yao B, Dong T, Chen Y, Yao J, Liu Y, et al. Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. Cell. 2022;185: 1356-1372.e26. doi:10.1016/j.cell.2022.02.027

9. Niño JLG, Wu H, LaCourse KD, Kempchinsky AG, Baryiames A, Barber B, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. Nature. 2022; 1–8. doi:10.1038/s41586-022-05435-0

10. Cummins J, Tangney M. Bacteria and tumours: causative agents or opportunistic inhabitants? Infect Agent Cancer. 2013;8: 11. doi:10.1186/1750-9378-8-11

11. Baban CK, Cronin M, O'Hanlon D, O'Sullivan GC, Tangney M. Bacteria as vectors for gene therapy of cancer. Bioeng Bugs. 2010;1: 385–394. doi:10.4161/bbug.1.6.13146

12. Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. Cell. 2019;178: 795-806.e12. doi:10.1016/j.cell.2019.07.008

13. Zhou B, Shi L, Jin M, Cheng M, Yu D, Zhao L, et al. Caulobacter and Novosphingobium in tumor tissues are associated with colorectal cancer outcomes. Front Oncol. 2022;12: 1078296. doi:10.3389/fonc.2022.1078296

14. Geller LT, Barzily-Rokni M, Danino T, Jonas OH, Shental N, Nejman D, et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science. 2017;357: 1156–1160. doi:10.1126/science.aah5043

15. Scott Merrell D, McAvoy TJ, King MC, Sittig M, Millar EV, Nieroda C, et al. Pre- and post-operative antibiotics in conjunction with cytoreductive surgery and heated intraperitoneal chemotherapy (HIPEC) should be considered for pseudomyxoma peritonei (PMP) treatment. Eur J Surg Oncol. 2019;45: 1723–1726. doi:10.1016/j.ejso.2019.01.223

16. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020. pp. 567–574. doi:10.1038/s41586-020-2095-1

17. Gihawi A, Ge Y, Lu J, Puiu D, Xu A, Cooper CS, et al. Major data analysis errors invalidate cancer microbiome findings. MBio. 2023;14: e0160723. doi:10.1128/mbio.01607-23

18. de Goffau MC, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnock-Jones DS, et al. Recognizing the reagent microbiome. Nat Microbiol. 2018;3: 851–853. doi:10.1038/s41564-018-0202-y

19. Cooper A, Poinar HN. Ancient DNA: do it right or not at all. Science (New York, N.Y.). American Association for the Advancement of Science (AAAS); 2000. p. 1139. doi:10.1126/science.289.5482.1139b

20. Ronnett BM, Zahn CM, Kurman RJ, Kass ME, Sugarbaker PH, Shmookler BM. Disseminated peritoneal adenomucinosis and peritoneal mucinous carcinomatosis. A clinicopathologic analysis of 109 cases with emphasis on distinguishing pathologic features, site of origin, prognosis, and relationship to “pseudomyxoma peritonei.” Am J Surg Pathol. 1995;19: 1390–1408. doi:10.1097/00000478-199512000-00006

21. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8: 761–763. doi:10.1038/nmeth.1650

22. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics.

Microbiome. 2017;5: 27. doi:10.1186/s40168-017-0237-y

23. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27: 105–117. doi:10.1016/j.tim.2018.11.003

24. Austin GI, Park H, Meydan Y, Seeram D, Sezin T, Lou YC, et al. Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data. Nat Biotechnol. 2023;41: 1820–1828. doi:10.1038/s41587-023-01696-w

25. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science. 2020;368: 973–980. doi:10.1126/science.aay9189

26. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nat Rev Microbiol. 2018;16: 143–155. doi:10.1038/nrmicro.2017.157

27. Tett A, Pasolli E, Masetti G, Ercolini D, Segata N. Prevotella diversity, niches and interactions with the human host. Nat Rev Microbiol. 2021;19: 585–599. doi:10.1038/s41579-021-00559-y

28. Brown K, Church D, Lynch T, Gregson D. Bloodstream infections due to Peptoniphilus spp.: report of 15 cases. Clin Microbiol Infect. 2014;20: O857-60. doi:10.1111/1469-0691.12657

29. France MT, Clifford J, Narina S, Rutt L, Ravel J. Complete Genome Sequences of Ezakiella coagulans C0061C1 and Fenollaria massiliensis C0061C2. Microbiol Resour Announc. 2022;11: e0044422. doi:10.1128/mra.00444-22

30. An HJ, Partha MA, Lee H, Lau BT, Pavlichin DS, Almeda A, et al. Tumor-associated microbiome features of metastatic colorectal cancer and clinical implications. Front Oncol. 2023;13: 1310054. doi:10.3389/fonc.2023.1310054

31. Khan S, Banerjee G, Setua S, Jones DH, Chauhan BV, Dhasmana A, et al. Metagenomic analysis unveils the microbial landscape of pancreatic tumors. Front Microbiol. 2023;14: 1275374. doi:10.3389/fmicb.2023.1275374

32. Ang CS-P, Shen JP, Hardy-Abeloos CJ, Huang JK, Ross JS, Miller VA, et al. Genomic Landscape of Appendiceal Neoplasms. JCO Precis Oncol. 2018;2. doi:10.1200/PO.17.00302

33. Zhang X, Yu D, Wu D, Gao X, Shao F, Zhao M, et al. Tissue-resident Lachnospiraceae family bacteria protect against colorectal carcinogenesis by promoting tumor immune surveillance. Cell Host Microbe. 2023;31: 418-432.e8. doi:10.1016/j.chom.2023.01.013

34. Hatakeyama M. Oncogenic mechanisms of the Helicobacter pylori CagA protein. Nat Rev Cancer. 2004;4: 688–694. doi:10.1038/nrc1433

35. Boleij A, Hechenbleikner EM, Goodwin AC, Badani R, Stein EM, Lazarev MG, et al. The Bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. Clin Infect Dis. 2015;60: 208–215. doi:10.1093/cid/ciu787

36. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Front Immunol. 2020;11: 906. doi:10.3389/fimmu.2020.00906

37. Taylor JC, Kumar R, Xu J, Xu Y. A pathogenicity locus of Streptococcus gallolyticus subspecies gallolyticus. Sci Rep. 2023;13: 6291. doi:10.1038/s41598-023-33178-z

38. Yacoub E, Saed Abdul-Wahab OM, Al-Shyarba MH, Ben Abdelmoumen Mardassi B. The Relationship between Mycoplasmas and Cancer: Is It Fact or Fiction? Narrative Review and Update on the Situation. J Oncol. 2021;2021: 9986550. doi:10.1155/2021/9986550

39. Oh BS, Choi WJ, Kim J-S, Ryu SW, Yu SY, Lee J-S, et al. Cell-Free Supernatant of Odoribacter splanchnicus Isolated From Human Feces Exhibits Anti-colorectal Cancer Activity. Front Microbiol. 2021;12: 736343. doi:10.3389/fmicb.2021.736343

40. van Eden WJ, Kok NFM, Snaebjornsson P, Jóźwiak K, Woensdregt K, Bottenberg PD, et al. Factors influencing long-term survival after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for pseudomyxoma peritonei originating from appendiceal neoplasms. BJS Open. 2019;3: 376–386. doi:10.1002/bjs5.50134

41. Sugarbaker PH. Cytoreductive surgery and hyperthermic intraperitoneal chemotherapy in the management of gastrointestinal cancers with peritoneal metastases: Progress toward a new standard of care. Cancer Treat Rev. 2016;48: 42–49. doi:10.1016/j.ctrv.2016.06.007

42. Tejani MA, ter Veer A, Milne D, Ottesen R, Bekaii-Saab T, Benson AB 3rd, et al. Systemic therapy for advanced appendiceal adenocarcinoma: an analysis from the NCCN Oncology Outcomes Database for colorectal cancer. J Natl Compr Canc Netw. 2014;12: 1123–1130. doi:10.6004/jnccn.2014.0109

43. Hoehn RS, Rieser CJ, Choudry MH, Melnitchouk N, Hechtman J, Bahary N. Current Management of Appendiceal Neoplasms. Am Soc Clin Oncol Educ Book. 2021;41: 1–15. doi:10.1200/EDBK_321009

44. Rangel-Vega A, Bernstein LR, Mandujano-Tinoco EA, García-Contreras SJ, García-Contreras R. Drug repurposing as an alternative for the treatment of recalcitrant bacterial infections. Front Microbiol. 2015;6: 282. doi:10.3389/fmicb.2015.00282

45. Joyce K, Saxena S, Williams A, Damurjian C, Auricchio N, Aluotto S, et al. Antimicrobial spectrum of the antitumor agent, cisplatin. J Antibiot . 2010;63: 530–532. doi:10.1038/ja.2010.64

46. Montassier E, Gastinne T, Vangay P, Al-Ghalith GA, Bruley des Varannes S,

Massart S, et al. Chemotherapy-driven dysbiosis in the intestinal microbiome. Aliment Pharmacol Ther. 2015;42: 515–528. doi:10.1111/apt.13302

47. Oh B, Boyle F, Pavlakis N, Clarke S, Guminski A, Eade T, et al. Emerging Evidence of the Gut Microbiome in Chemotherapy: A Clinical Review. Front Oncol. 2021;11: 706331. doi:10.3389/fonc.2021.706331

48. Gilbreath JJ, Semino-Mora C, Friedline CJ, Liu H, Bodi KL, McAvoy TJ, et al. A core microbiome associated with the peritoneal tumors of pseudomyxoma peritonei. Orphanet J Rare Dis. 2013;8: 105. doi:10.1186/1750-1172-8-105

49. Battaglia TW, Mimpen IL, Traets JJH, van Hoeck A, Zeverijn LJ, Geurts BS, et al. A pan-cancer analysis of the microbiome in metastatic cancer. Cell. 2024;187: 2324-2335.e19. doi:10.1016/j.cell.2024.03.021

50. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551: 457–463. doi:10.1038/nature24621

51. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2016;18: 1403–1414. doi:10.1111/1462-2920.13023

52. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75: 129–137. doi:10.3354/ame01753

53. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37: 852–857. doi:10.1038/s41587-019-0209-9

54. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13: 581–583. doi:10.1038/nmeth.3869

55. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6: 90. doi:10.1186/s40168-018-0470-z

56. Robeson MS, O'Rourke DR, Kaehler BD, Ziemski M, Dillon MR, Foster JT, et al. RESCRIPt: Reproducible sequence taxonomy reference database management for the masses. bioRxiv. 2020. p. 2020.10.05.326504. doi:10.1101/2020.10.05.326504

57. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41: D590-6. doi:10.1093/nar/gks1219

58. Mirarab S, Nguyen N, Warnow T. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput. 2012; 247–258. doi:10.1142/9789814366496_0024

59. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. mSystems. 2018;3. doi:10.1128/mSystems.00021-18

60. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61: 1–10. doi:10.1016/0006-3207(92)91201-3

61. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71: 8228–8235. doi:10.1128/AEM.71.12.8228-8235.2005

62. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol. 2007;73: 1576–1585. doi:10.1128/AEM.01996-06

63. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online. Chichester, UK: John Wiley & Sons, Ltd; 2017. pp. 1–15. doi:10.1002/9781118445112.stat07841

64. Lahti L, Shetty S, Blake T, Salojarvi J, Others. Tools for microbiome analysis in R. Version. 2017;1: 28.

CHAPTER 3: MICROBIAL COMMUNITY RESPONSES TO ABOVEGROUND CADAVER DECOMPOSITION ARE DETECTABLE IN SUBSURFACE SOILS

3.1 Summary

During vertebrate decomposition, a purge of nutrient-rich fluids enters the surrounding soil producing a localized nutrient hotspot. The composition of microbial communities at the soil surface has been shown to predictably change in response to this influx of decomposition fluids over time. However, the depth to which these compositional changes permeate the soil near a decomposing human cadaver and which microbial members exhibit responses to aboveground decomposition both represent significant knowledge gaps. In this study, we longitudinally sampled forest soils adjacent to decomposing human cadavers and from nearby control sites in 2 cm increments to a depth of 10 cm. Soil samples were collected during purge, post purge, and the dry remains stage of decomposition over three seasons at Sam Houston State University's Southeast Texas Applied Forensic Science Facility. We detected microbial community responses in the upper 0–4 cm of soils adjacent to decomposition fluids, including several bacterial genera that significantly increased in response to decomposition. Microbial community diversity decreased, and community composition became increasingly different in comparison to the nearby control soils as decomposition progressed. We observed a more limited, yet still detectable response in the lower 6 – 10 cm of soil. In addition to improving our understanding of decomposition ecology, we anticipate that incorporating these findings into forensic investigations could help strengthen postmortem interval model estimation and better inform investigative

sampling practices.

3.2 Introduction

Human decomposition occurs in four major stages that can co-occur: fresh, early, advanced, and skeletonization [1]. Accumulation of gasses generated by microbial fermentation during early decomposition often results in abdominal rupture and the release of nutrient rich fluids into the surrounding soil creating a localized decomposition island [2]. This sudden influx of nutrients at the soil-cadaver interface primarily consisting of carbon, nitrogen, phosphorus, and salts results in rapid changes in soil pH and dissolved oxygen and nitrogen levels leads to a large shift in microbial community composition that predictably changes over the course of decomposition [3–5]. Following purge and during active decay, microbial respiration increases and members of *Pseudomonadota* (previously *Proteobacteria*), *Bacillota* (previously *Firmicutes*), and *Actinomycetota* (previously *Actinobacteria*) increase in relative abundance while members of *Acidobacteriota* and *Verrucomicrobiota* tend to decrease [6–8].

Computational models that rely on successional microbial patterns of the skin and adjacent soil are currently being developed to estimate the postmortem interval (PMI) of human remains [4,5,9–11]. Many of these PMI models integrate microbial community data from samples collected by swabbing the soil surface. While this sampling strategy is convenient and time saving, swabs can be expensive and are frequently in short supply. Additionally, backup swabs are often collected and archived in the event of mishaps. Others have utilized soil cores to describe the effects of human decomposition on microbial communities in soils directly below human cadavers suggesting cores may provide an alternative, cost-effective backup sampling solution

[7,8,12,13]. However, it remains unclear how deep in the soil these community responses to aboveground decomposition permeate, raising questions about best soil sampling practices. For example, composite soil cores that are primarily composed of deeper unimpacted soil layers may obscure decomposition responses occurring in more shallow layers. Collecting several, shallow core soils may be a more effective alternative sampling strategy.

In this study, we determine to what depth subsurface microbial community responses to aboveground decomposition are detectable, whether season impacts detection depth, and if these subsurface responses are detectable over the course of decomposition. To address these questions, we placed six cadavers at an outdoor forensics research facility in Huntsville, TX. Cadavers were placed in pairs during the winter, spring, and summer to investigate seasonality. 10 cm soil cores, which were subsampled into five, 2-centimeter sections, were taken from decomposition-associated soil near the human cadaver groin at the time of placement (fresh), purge, post-purge, and dry stages of decomposition. We hypothesized that microbial responses to aboveground decomposition would be detectable throughout the core samples, but the impacts from decomposition would decrease with increasing depth. Further, we hypothesized that decomposition-associated community composition will become increasingly different from control community composition over time. The findings from our study advance our understanding of the spatiotemporal responses of microbial communities in the soil near decomposing cadavers and provide valuable insights that will improve soil sampling practices for future studies.

3.3 Materials and Methods

3.3.1 Study site and temperature calculations

This study was conducted at the Outdoor Research Laboratory division of the Southeast Texas Applied Forensic Science (STAFS) facility. The facility is located roughly 5 km north of Huntsville, Texas in the Sam Houston State University at the Center for Biological Field Studies. STAFS is a willed-body donation center that conducts research in anthropology, forensics, and other fields. The site is located in the Pineywoods region of southeast Texas and has acidic, well-draining sandy loam that covers heavy clay soil. The climate is subtropical and experiences high humidity with abundant rainfall during the summer months and mild winters. This study took place in 2015 where the average temperature was 20.5°C. The site received 66.76 inches of total precipitation or an average of 0.18 inches per day. Weather data was accessed through the National Centers of Environmental Information public database and collected from Huntsville Municipal Airport weather station (USW00053903; www.ncei.noaa.gov).

Accumulated degree days (ADDs) were used to normalize cadaver exposure to heat over the duration of the study as described in [1]. Briefly, the average daily temperature (in Celsius) above 0°C was added to the sum of average temperatures of the preceding days.

3.3.2 Cadaver Placement

A total of six human cadavers were placed outdoors to decompose at the STAFS Facility. Cadavers were placed in pairs 5 meters apart during consecutive seasons (winter, spring, and summer) in 2015. The bodies did not undergo autopsy and were not

embalmed. All bodies were stored at 4°C until placement. The cadavers were placed on undisturbed ground without clothing in a supine position. Following placement, cadavers were allowed to decompose under natural conditions with the exception that wire cages were placed over the cadavers to protect against large-animal scavenging.

3.3.3 Sample collection

Soil was collected with a ten inch “T” core sample device near the groin but not directly in the purge fluid. The initial, purge, post-purge and dry samples were taken at the respective stages of decomposition. Additionally, control soil cores were taken at the time of placement and during each collection time point from a cleared area outside the effective decomposition range. This sampling scheme was followed for winter, spring, and summer cadaver pairs and generated a total of 36 core samples (n=18 impacted by decomposition and n=18 controls). The soil samples were wrapped in sterile aluminum foil and placed in a -80°C freezer until analysis.

3.3.4 Sample processing and sequencing

Samples were processed in collaboration with the Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston Texas following protocols benchmarked as part of the Human Microbiome Project. Prior to DNA extraction, each soil core was divided into 5 separate samples according to the following depths: 0 – 2 cm, 2 – 4 cm, 4 – 6 cm, 6 – 8 cm, and 8 – 10 cm. Samples were vortexed to homogenize prior to extraction. 250 mg of soil was transferred to a 2 ml PowerBead tube for extraction. Microbial DNA was extracted using the PowerSoil DNA kit (MoBio Laboratories, Carlsbad, CA) following the manufacturer's instructions. Negative

extraction controls were processed in addition to the 180 soil samples to assess extraction kit contamination. The purified DNA was sent to the University of California, San Diego Microbiome Core for 16S rRNA library preparation and sequencing. 30 cycles of PCR were performed using barcoded primers that target the V4 16S rRNA region (515F: 5' GTGCCAGCMGCCGCGGTAA; 806R: 5' GGACTACHVGGGTWTCTAAT). PCR products were assessed using gel electrophoresis. Samples that showed no distinguishable amplicon band were not selected for sequencing. Negative controls showed no signs of amplification and were not sequenced either. 16S amplicons were sequenced using an Illumina HiSeq 2500 instrument (2 x 125 bp reads).

3.3.5 16S rRNA amplicon data processing and statistical analysis

16S rRNA gene sequencing data processing and analysis was performed using QIIME2 (q2) version 2023.5 [14]. Forward reads were demultiplexed using the q2 demux-single plugin. Reads were trimmed to 106 bp and quality filtered using the DADA2 denoise-single plugin with the default parameters [15]. The DADA2-generated feature table was filtered to only include amplicon sequence variants (ASVs) that had at least 20 total reads. Taxonomy was assigned using a pre-trained Naive Bayes classifier and the q2 classify-sklearn plugin. The classifier was trained on the V4 region of 16S rRNA sequences (clustered at 99% identity) obtained from the SILVA v.138 database [16,17]. ASVs classified as mitochondria or chloroplast or had not been classified at the phylum level were removed. Following filtering, 20,921,891 total reads remained with a median of 126,799 reads per sample. The q2 SEPP fragment-insertion plugin [18] was used with the SILVA 128 reference database to construct a rooted phylogenetic tree.

The q2 core-metrics pipeline with a rarefaction depth of 10,000 sequences per sample was used to calculate alpha and beta diversity metrics. After rarefying, 148 of 180 samples were retained. Faith's phylogenetic diversity (PD) was used to assess alpha diversity [19]. Kruskal-Wallis test [20] from the alpha-group-significance plugin was used to evaluate alpha diversity differences. Bray-Curtis and Unweighted UniFrac [21,22] were used to measure beta diversity. Permutational multivariate analysis of variance (PERMANOVA) [23] from the beta-group-significance plugin was used to assess beta diversity differences between groups.

Analysis of composition of microbiomes with bias correction (ANCOM-BC) [24] was used to detect differentially abundant taxa. The feature table was filtered to only include samples that had been collected from cadavers placed during the spring and summer. ASVs were collapsed at the genus level (L6). The table was further filtered to only include genera that had a minimum frequency of 50 and were present in at least 4 samples prior to running the composition-ancombc plugin.

The q2 longitudinal linear-mixed-effects plugin was used to evaluate changes in alpha and beta diversity metrics over the course of decomposition [25] during the spring and summer. Shannon and Faith's Phylogenetic Diversity were used to assess changes in alpha diversity across decomposition stage. Bray-Curtis and unweighted UniFrac distance from the initial time point (first-distances plugin) was used to assess changes in beta diversity.

The MaAsLin 2 (Microbiome Multivariable Associations with Linear Models) package (v1.20.0) was used to identify microbial taxa that responded at each stage of decomposition [26]. The feature table was filtered to cadaver soils from the spring and

summer placements. This table was split into an upper and lower table containing samples from the upper 0-4 cm and lower 6-10 cm of soil. The features in each table were collapsed at the genus level (level 6) and then feature counts were converted to relative abundances. MaAsLin 2 was run separately on each log-transformed table using the decomposition stage as the fixed effect while donor and placement season were used as random effects. Seaborn was used to create heatmaps of significant results (upper *q-value* <0.01; lower *q-value* <0.05).

3.3.6 Data availability

Raw sequencing data and sample metadata are publicly available and can be accessed through QIITA (https://qiita.ucsd.edu/) study ID 10406. QIIME2 intermediate files, commands, and code used to generate visualization are available through Open Science Framework (https://osf.io/) under project title “Decomposition Soil Cores” (https://osf.io/gpsdf/).

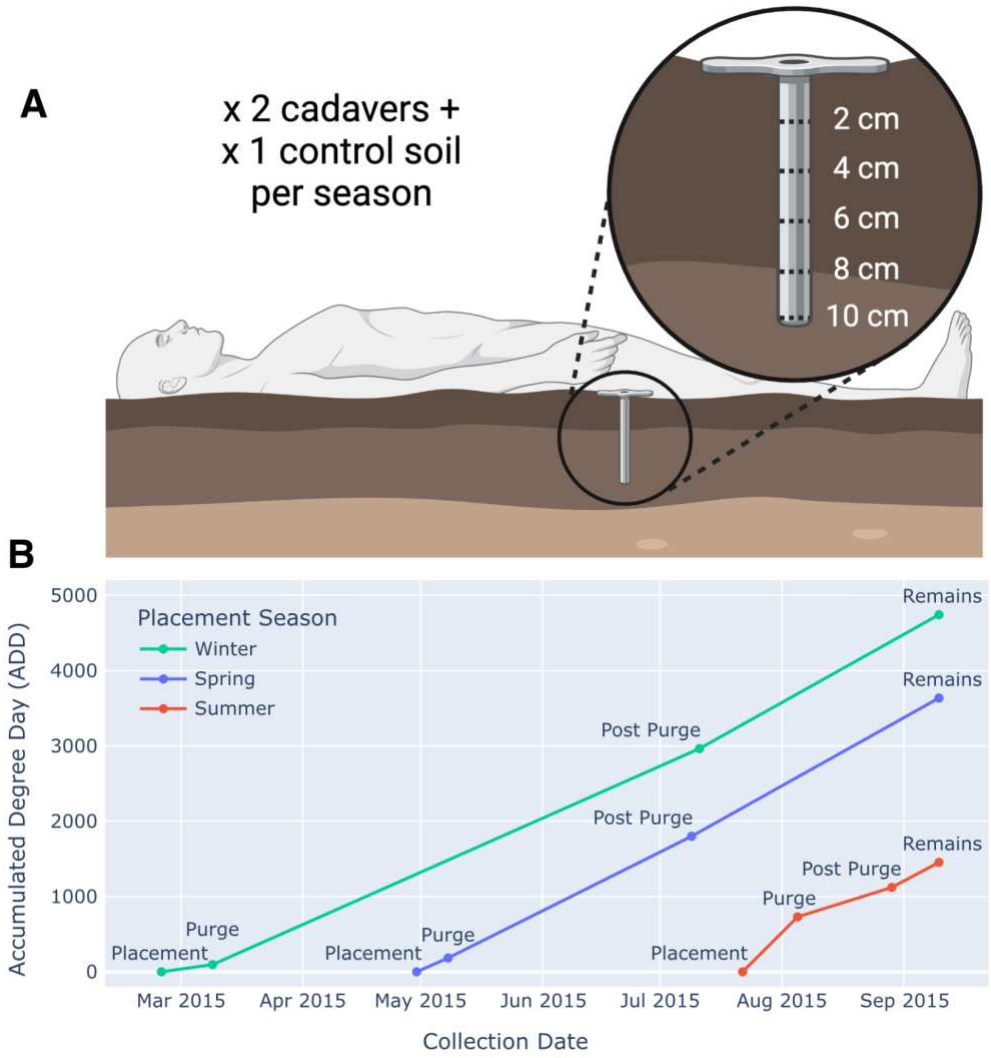


Figure 1. Experimental design. A) Soil cores were taken near the cadaver groin and from a soil control site not impacted by decomposition. Cores were divided into five sections every 2 cm prior to DNA extraction. Each depth was processed and sequenced separately. In total, 6 cadavers were placed in pairs for 3 consecutive seasons in 2015. One control soil site was sampled per cadaver pair per season. **B)** Soil sampling timepoints. Soil cores were collected at the time of placement ("Placement"), during purge ("Purge"), following purge ("Post Purge"), and once remains had dried ("Remains") during the respective season placement, indicated by green (winter), blue (spring), and red (summer). Accumulated degree days (ADDs) at each collection point are shown for each season.

3.4 Results

3.4.1 Microbial communities in the top 4 cm of soil respond to aboveground decomposition

To better understand the extent of spatial microbial responses to the ephemeral influx of decomposition fluids, we sought to characterize microbial community changes to decomposition across five soil depths (**Figure 1A**). Shannon and Faith's phylogenetic diversity both decreased in the top two cadaver soil depths (0-2 cm and 2-4 cm) compared to control soils at the same depths (Kruskal-Wallis *p*-values < 0.05; **Figure 2A** and **Supplementary Figure 1A**). Alpha diversity did not not significantly decrease in cadaver soils below 4 cm. Similarly, Bray-Curtis and Unweighted Unifrac distances were significantly larger between cadaver and control soils at 0-2 cm and 2-4 cm compared to cadaver and control soils at depths below 4 cm (PERMANOVA *q*-values < 0.05; Supplementary **Figure 1B-E**). Because we found the top two soil depths were significantly different from control soils, we designated these compositionally responsive soils (0-2 cm and 2-4 cm) as the "upper" layer while the less responsive soil depths (4-6 cm, 6-8 cm, and 8-10 cm) were considered the "lower" layer, indicating that decomposition may only affect the upper soil layer.

The microbial composition between control and cadaver soils varied significantly in the upper and lower layers, with a more pronounced shift in the upper layer (**Figure 2B-C**). We found that soil treatment (cadaver vs. control) explained 5.9% of community composition variability in the upper layer compared to 2.8% in the lower layer. However, in both layers, cadaver placement season explained more variability than soil treatment (R^2 upper = 8.8%; lower = 10%) suggesting soil community responses to decomposition

may be strongly impacted by environmental factors such as temperature at time of cadaver placement.

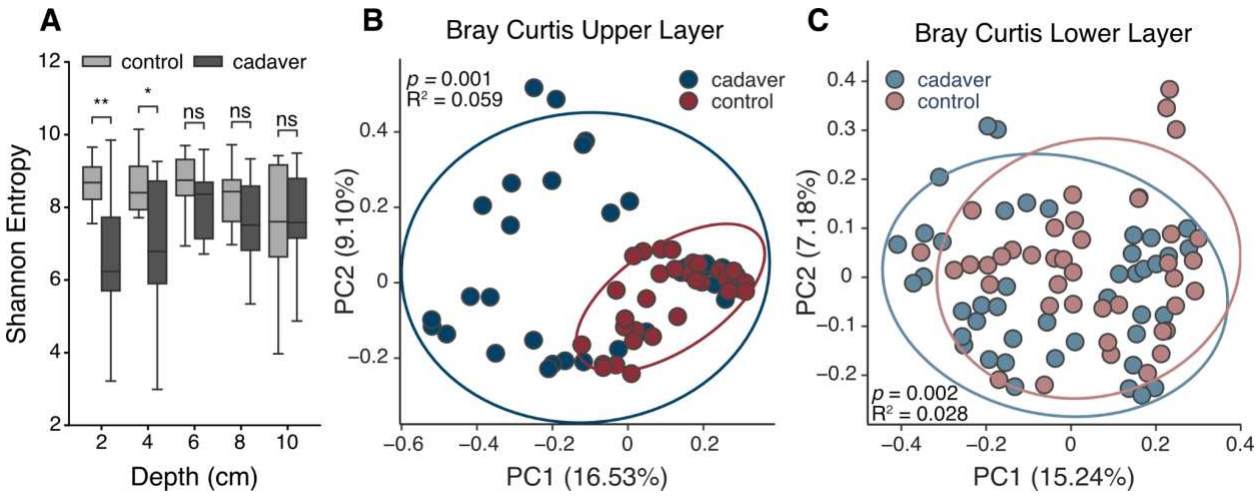


Figure 2. Spatial microbial responses to aboveground decomposition. A) Box plot depicting Shannon diversity of control soils (light gray) compared to cadaver soil (dark gray) from the same depth. A Kruskal-Wallis test was used to test for significance between soil depths (** *p*-value < 0.01; * *p*-value < 0.05; ns: not significant). **B**) PCoA plot based on Bray-Curtis distances between cadaver (blues) and control (reds) soils in the upper soil layer (PERMANOVA *p*-value = 0.001) and **C**) lower soil layer (PERMANOVA *p*-value = 0.002). Ellipses represent 95% confidence intervals for cadaver and control groups.

3.4.2 Placement season impacts belowground microbial communities

We next investigated how alpha and beta diversity trends in each layer were influenced by cadaver placement season. Alpha diversity significantly decreased in both layers during the spring and summer compared to the control soils (Kruskal-Wallis *p*-values < 0.05; **Figure 3A-B**), while winter alpha diversity was unaffected (**Figure 3C**). However, Faith's Phylogenetic Diversity did not decrease in the lower layer in the spring and summer suggesting that while decomposition depletes microbial taxa, these taxa are likely closely related (**Supplemental Figure 2A-B**).

Cadaver placement season also influenced below ground soil community

composition. Unweighted unifrac and Bray-Curtis metrics were both significantly altered in response to decomposition in both soil layers during the spring and summer compared to control soils (PERMANOVA *q*-values < 0.05; **Figure 3D-F** and **Supplementary Figure 2D-F**). Unexpectedly, unweighted unifrac showed that the lower soil layer adjacent to cadavers placed during the winter significantly clustered away from winter control soils (PERMANOVA *q*-value < 0.05; **Figure 3F**). However, Bray-Curtis distances were not significantly different (PERMANOVA *q*-value > 0.05), perhaps suggesting that decomposition drives changes in low-abundance, phylogenetically distinct taxa in the lower layer during the winter. We also found that soils in the upper layer collected during purge in the winter clustered away from a majority of samples (**Supplemental Figure 2F**) demonstrating that diffusion of nutrient-rich decomposition fluids may temporarily impact adjacent soil communities even during colder months. As winter cadaver placement produced a limited response in belowground soils, we excluded this group from further analyses.

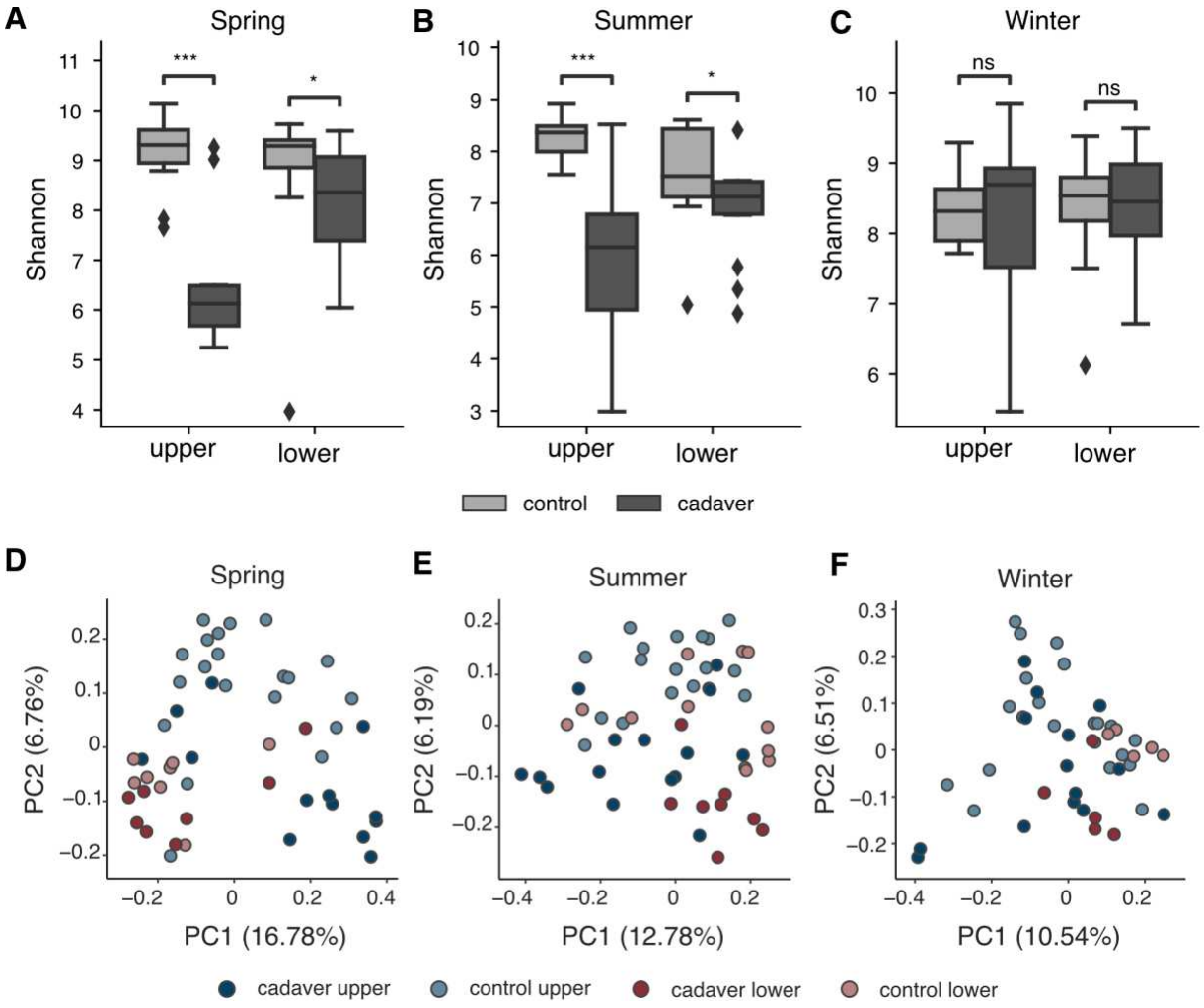


Figure 3. Cadaver placement season impacts belowground microbial communities' responses to decomposition. A-C) Box plots showing Shannon diversity of control soils (light gray) compared to cadaver soil (dark gray) from the same layer across cadaver placement seasons. A Kruskal-Wallis test was used to test for significance between soil treatment (*** p-value < 0.001; * p-value < 0.05; ns: not significant). **D-F**) PCoA plot based on unweighted unifrac distances between cadaver (blues) and control (reds) soils in the upper (dark) and lower (light) soil layers (PERMANOVA q-values < 0.05).

3.4.3 Decomposition enriches rare taxa in both layers

We next sought to evaluate whether decomposition selects for different microbes in each layer that could potentially be forensically important indicator taxa and may be impacted by sample collection practices. Several members of the bacterial genera

Terriglobus, *Acidisoma*, *Ignatzschineria*, and *Pandoraea* were significantly enriched in the upper layer of cadaver soils compared to the upper layer of control soils throughout the spring and summer. Conversely, we found that over 100 genera became depleted in response to decomposition in the upper soil layer, further demonstrating that select members are responsive to decomposition, while a larger proportion are not stimulated (**Supplemental Table 1**). Of these 100 depleted genera, *Pedosphaeraceae*, *Anaeromyxobacter*, *Haliangium*, and *Ellin6067* (*Nitrosomonadaceae* family) experienced the largest compositional shift (quantified by negative log fold change) in the upper layer. Decomposition had a more muted effect on the lower soil layer where in total, only 9 genera were differentially abundant (**Supplemental Table 2**). Interestingly, *Ignatzschineria,* a genera frequently associated with blow flies and may be a potential biological marker of decomposition [27], was also enriched in the lower soil layer.

We also examined control soils to determine whether the upper or lower layers were enriched for genera that could potentially utilize decomposition products (e.g. ammonium), making them useful in forensic investigations and therefore an important consideration when sampling soils. No genera were enriched in the upper layer. *Nitrospira* and *Candidatus Nitrosotalea* along with 4 other genera were enriched in the lower soil layer (**Supplemental Table 3**), suggesting soils may act as a reservoir for select microbes capable of N cycling. As these 2 genera are known to oxidize ammonia to nitrite [28,29], an important decomposition byproduct, sampling the lower layer may be important for studies investigating soil nitrogen cycling.

To better understand the distribution of these taxa, we examined the relative abundances of the 4 most enriched and 4 most depleted genera in the upper and lower

soil layers. We discovered these enriched genera tended to be rare in control soils. For instance, *Terriglobus* and *Acidisoma* only amounted to 0.16% relative abundance in the upper layer of control soils while *Ignatzschineria* and *Pandoraea* were not detected (**Figure 4A**). In contrast, these 4 genera made up 12.42% relative abundance in cadaver soils. However, depleted genera were readily detectable in both cadaver and control soils. A similar trend occurred in the lower soil layer where enriched genera were rare or not detected in the control soils (**Figure 4B**). These findings suggest enriched taxa such as *Ignatzschineria*, which are not readily detected in control soils, may prove to be useful microbial markers for detecting decomposition in soils in both layers.

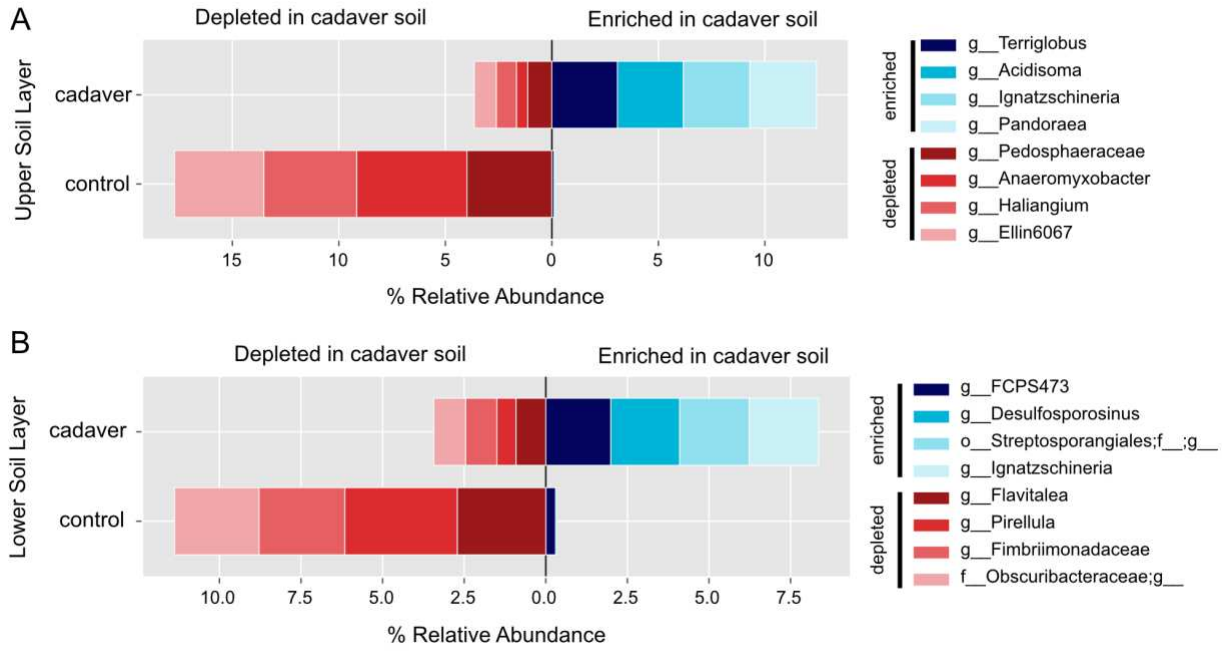


Figure 4. Enriched cadaver taxa are rare in control soils. A) Taxa barplot of the 4 most enriched and 4 most depleted taxa in the upper and B**)** lower soil layers during the spring and summer seasons. Average percent relative abundance of taxa enriched in cadaver soils (blue) are plotted to the right of zero while taxa depleted in cadaver soils (reds) are plotted to the left of zero.

3.4.4 Microbial communities change spatiotemporally in response to decomposition

Decomposition stage is known to play a role in microbial succession, as resources change over the time [4,6,9]. Therefore, we evaluated whether the observed changes in alpha and beta diversity to belowground microbial communities occurred temporally across 3 stages of decomposition (**Figure 1B**). Shannon and Faith's phylogenetic diversity significantly decreased in the upper layer of soils near decomposing cadavers across all decomposition stages whereas control soils saw an increase in alpha diversity over time (**Figure 5A** and **Supplemental Figure 3A**). We found no evidence that Shannon or Faith's phylogenetic diversity decreased in the lower soil layer over time demonstrating that the depletion response seen in the upper layer likely does not permeate soil to a detectable level during decomposition, at least in this soil type (**Figure 5B** and **Supplemental Figure 3B**).

Soil communities became more different from those at the time of cadaver placement as decomposition progressed. Interestingly, the Bray-Curtis distances from the time of placement to the time of purge were not significantly different between cadaver and control soils in the upper and lower layer (**Figure 5C-D**). However, by the final decomposition time point (remains), cadaver soils were significantly different from the initial timepoint while control soils became more similar which indicates that community composition responses in the subsurface are not instantaneous but develop over time.

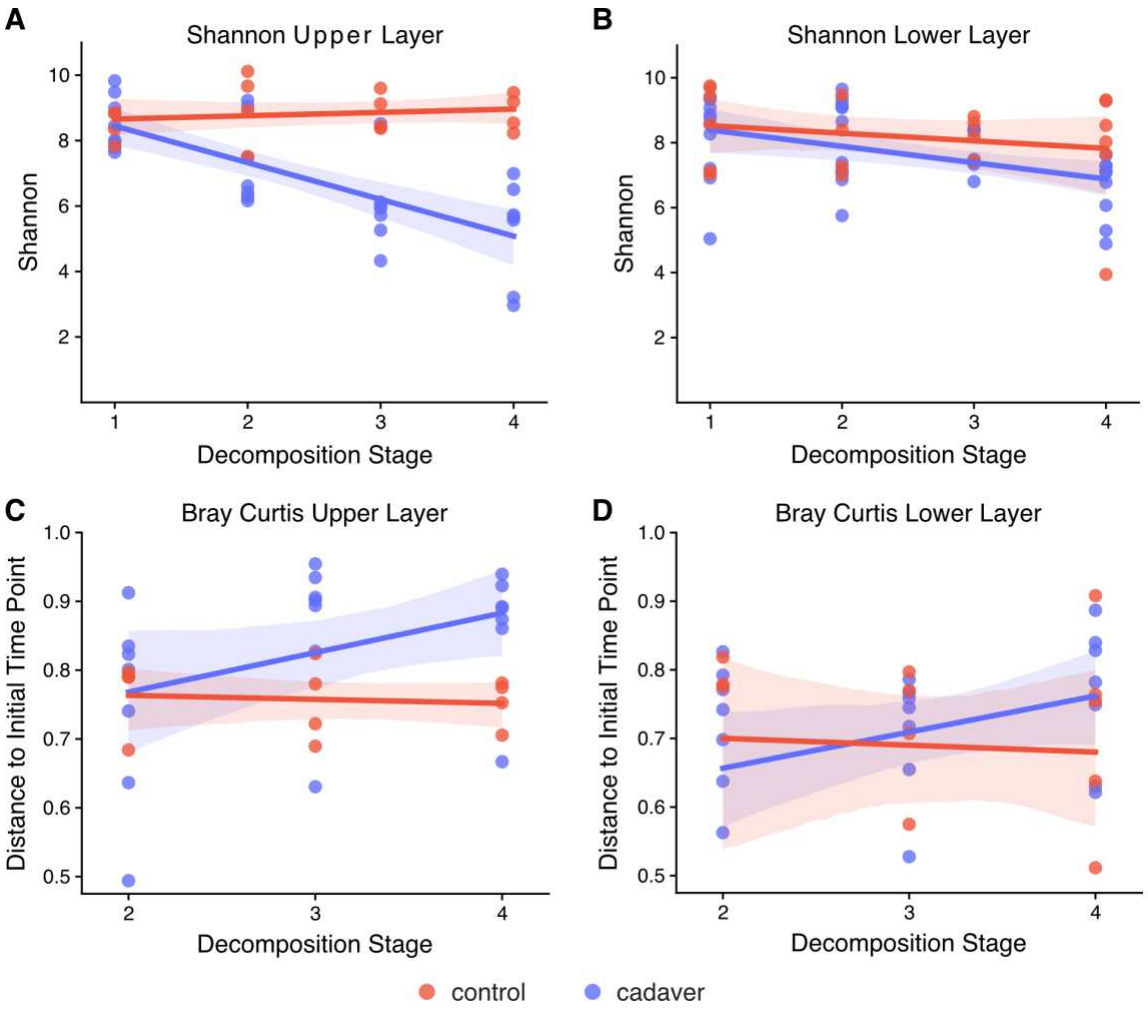


Figure 5. Below ground soil communities become increasingly different as decomposition progresses. Linear mixed effects (LME) models showing Shannon diversity for control (red) and cadaver (blue) soils across all decomposition stages (1= initial placement, 2 = purge, 3 = post purge, 4 = dry remains) in the upper (**A**) and lower (**B**) soil layers. LME models reveal Bray-Curtis distance to initial placement time point increases as decomposition progresses in upper (**C**) and (**D**) lower soil layers. Winter placement soils were excluded from all LME models.

Finally, we identified taxa with significant spatiotemporal responses to decomposition. In the upper layer of cadaver soils, the relative abundance of numerous bacterial genera decreased over the course of decomposition (**Supplemental Table 4**). However, for most taxa, this decrease was not significant until the stage following purge (**Figure 6**). A majority of these were members of the *Acidobacteriota* or

Verrucomicrobiota phylum. *Rhodanobacter*, *Subgroup 12* (*Acidobacteriota*), *Gemmataceae,* and *Fimbrimonadaceae* are all aerobes that were among the 6 genera to significantly decrease during purge suggesting they may be particularly sensitive to changes in oxygen levels that often decrease as decomposition fluids enter the surrounding soil.

Several genera, some of which may be human associated (i.e. *Fusobacterium*, *Enterococcus*, and *Proteus*), significantly peaked in percent relative abundance during purge. In contrast, *Escherichia-Shigella,* a well-known human gut-associated taxa, gradually increased throughout decomposition until it was significantly enriched by the final dry remains stage. *Ignatzschineria* and *Erysipelothrix*, which have both previously been associated with blowflies [30], also peaked during purge (**Figure 6**) but then returned to non-detectable levels by the final decomposition stage. We also found several acidophilic genera gradually increased throughout decomposition– *Acidisphaera*, *Acidisoma*, *Terriglobus*, and *Dyella*. These results demonstrate that bacteria from several potential sources (humans, insects, and soil) assemble at decomposition hotspots and dynamically respond to fluctuating nutrients and environmental variables (e.g. pH) over time in the upper soil layer.

Comparatively, far fewer taxa significantly changed in the lower layer (**Supplemental Table 5**). Only 12 genera decreased from their initial relative abundance and except for *Mucilaginibacter*, significant changes were not observed until the final decomposition stage (**Supplemental Figure 4**). Human-associated taxa *Peptostreptococcus*, and *Helcoccus* also increased in the lower soil layer. Only *Escherichia-Shigella* and *Ignatzschineria* followed consistent patterns in the upper and

lower layers. Collectively, these findings suggest decomposition has a significant, yet limited effect on microbial communities in the lower 6 cm of the soil.

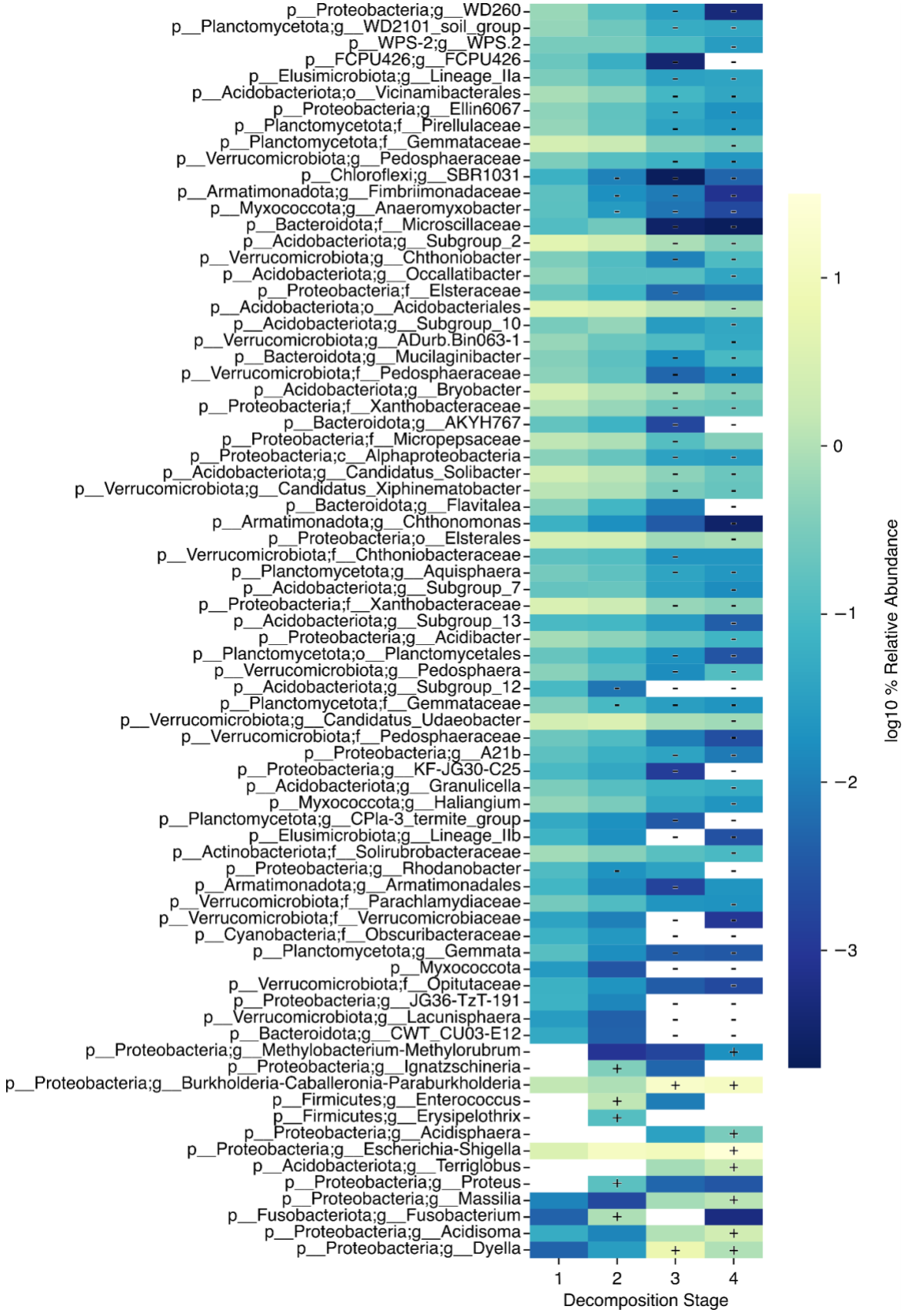

p__Proteobacteria;g__WD260
p__Planctomycetota;g__WD2101_soil_group
p__WPS-2;g__WPS.2
p__FCPU426;g__FCPU426
p__Elusimicrobiota;g__Lineage_IIa
p__Acidobacteriota;o__Vicinamibacterales
p__Proteobacteria;g__Ellin6067
p__Planctomycetota;f__Pirellulaceae
p__Planctomycetota;f__Gemmataceae
p__Verrucomicrobiota;g__Pedosphaeraceae
p__Chloroflexi;g__SBR1031
p__Armatimonadota;g__Fimbriimonadaceae
p__Myxococcota;g__Anaeromyxobacter
p__Bacteroidota;f__Microscillaceae
p__Acidobacteriota;g__Subgroup_2
p__Verrucomicrobiota;g__Chthoniobacter
p__Acidobacteriota;g__Occallatibacter
p__Proteobacteria;f__Elsteraceae
p__Acidobacteriota;o__Acidobacteriales
p__Acidobacteriota;g__Subgroup_10
p__Verrucomicrobiota;g__ADurb.Bin063-1
p__Bacteroidota;g__Mucilaginibacter
p__Verrucomicrobiota;f__Pedosphaeraceae
p__Acidobacteriota;g__Bryobacter
p__Proteobacteria;f__Xanthobacteraceae
p__Bacteroidota;g__AKYH767
p__Proteobacteria;f__Micropepsaceae
p__Proteobacteria;c__Alphaproteobacteria
p__Acidobacteriota;g__Candidatus_Solibacter
p__Verrucomicrobiota;g__Candidatus_Xiphinematobacter
p__Bacteroidota;g__Flavitalea
p__Armatimonadota;g__Chthonomonas
p__Proteobacteria;o__Elsterales
p__Verrucomicrobiota;f__Chthoniobacteraceae
p__Planctomycetota;g__Aquisphaera
p__Acidobacteriota;g__Subgroup_7
p__Proteobacteria;f__Xanthobacteraceae
p__Acidobacteriota;g__Subgroup_13
p__Proteobacteria;g__Acidibacter
p__Planctomycetota;o__Planctomycetales
p__Verrucomicrobiota;g__Pedosphaera
p__Acidobacteriota;g__Subgroup_12
p__Planctomycetota;f__Gemmataceae
p__Verrucomicrobiota;g__Candidatus_Udaeobacter
p__Verrucomicrobiota;f__Pedosphaeraceae
p__Proteobacteria;g__A21b
p__Proteobacteria;g__KF-JG30-C25
p__Acidobacteriota;g__Granulicella
p__Myxococcota;g__Haliangium
p__Planctomycetota;g__CPla-3_termite_group
p__Elusimicrobiota;g__Lineage_IIb
p__Actinobacteriota;f__Solirubrobacteraceae
p__Proteobacteria;g__Rhodanobacter
p__Armatimonadota;g__Armatimonadales
p__Verrucomicrobiota;f__Parachlamydiaceae
p__Verrucomicrobiota;f__Verrucomicrobiaceae
p__Cyanobacteria;f__Obscuribacteraceae
p__Planctomycetota;g__Gemmata
p__Myxococcota
p__Verrucomicrobiota;f__Opitutaceae
p__Proteobacteria;g__JG36-TzT-191
p__Verrucomicrobiota;g__Lacunisphaera
p__Bacteroidota;g__CWT_CU03-E12
p__Proteobacteria;g__Methylobacterium-Methylorubrum
p__Proteobacteria;g__Ignatzschineria
p__Proteobacteria;g__Burkholderia-Caballeronia-Paraburkholderia
p__Firmicutes;g__Enterococcus
p__Firmicutes;g__Erysipelothrix
p__Proteobacteria;g__Acidisphaera
p__Proteobacteria;g__Escherichia-Shigella
p__Acidobacteriota;g__Terriglobus
p__Proteobacteria;g__Proteus
p__Proteobacteria;g__Massilia
p__Fusobacteriota;g__Fusobacterium
p__Proteobacteria;g__Acidisoma
p__Proteobacteria;g__Dyella
1
2
3
4
Decomposition Stage
1
0
−1
−2
−3
log10 % Relative Abundance

Figure 6. Microbial responses to above ground decomposition over the course of decomposition. MaAsLin2 was used to identify microbial taxa that significantly changed over the course of decomposition in the upper soil layer. Heatmap colors correspond with average log10 transformed percent relative abundance. White boxes signify taxa was not detected at that decomposition stage. Only taxa with *q*-value < 0.01 are labeled with +/-. + annotation denotes taxa increased in percent relative abundance; - annotation denotes taxa decreased in percent relative abundance as compared to initial time point. ASVs were collapsed at genus level. Phylum along with lowest level taxonomic assignment are shown. Decomposition stage 1=initial, 2=purge, 3=post purge, 4=remains. Winter placement soils were excluded from analysis.

3.5 Discussion

3.5.1 Impact of placement season on depth of microbial responses

Similar to previous studies [7,8,31], we saw that microbial community richness in cadaver soils, as compared to controls, significantly decreased over time in the upper layer (0-4 cm) with the uppermost 0-2 cm experiencing the most significant reduction in taxa (**Figure 2A**). These responses were in large part only detectable in soils collected near cadavers that were placed in spring and summer which had an average temperature of 30.9°C (**Figure 3A-C**). This suggests responsive microbes in the top layer may have been within an optimal temperature range for utilizing newly added decomposition nutrients and sampling the upper soil layer in warmer months may be a suitable option for capturing microbial responses as purge fluid may not permeate deeply into the soil before it is utilized.

In contrast, we found little evidence that soil communities respond to aboveground decomposition when cadavers were placed during the colder winter months (average 15.5°C), a finding also reported by others [13,32]. Still, this was surprising given winter cadavers were left to decompose for the longest duration and still experienced warm temperatures (February - September). As such, we expected to

observe at least a delayed response that coincided with increased temperatures [33]. However, during purge (stage 2) we saw that soil samples in the upper 0-4 cm became very different from controls but then returned to baseline (**Supplemental Figure 2F**). Unweighted unifrac metrics, but not Bray-Curtis, revealed that on average, only the lower soil layer was distinct from controls during the winter. Combined, these observations suggest colder temperatures at the time of placement likely lowered the production of decomposition fluids and possibly arrested microbial activity in the upper soil layer following purge resulting in a temporary community shift. Once fluids entered the soil, the upper layer may have acted as insulation against low ambient temperatures allowing for a limited community response in the lower layer. These findings have implications on best sampling practices and indicate collecting soil from multiple depths (including deeper ones), may be more advisable for cold temperature studies.

3.5.2 Soil oxygenation may drive microbial community shifts in the upper layer

Changes in soil microbial community composition are known to largely be driven by changes in soil pH and dissolved oxygen content during vertebrate purge fluid release. As organic carbon enters the soil with purge fluids, microbial respiration increases which consequently produces CO_2, reducing the surrounding soil pH and oxygen levels creating anoxic conditions [12,13]. This soil biogeochemical process is associated with active and advanced cadaver decomposition aboveground [8] which suggests our stage 2 purge samples may have only captured the earliest stages of active decomposition where soil microbial communities had yet to completely respond to the nutrient influx. Because of our limited sampling approach, it is likely significant changes were not captured until stage 3. By our final collection time point, we found that

microbial community diversity and composition were still decreasing and changing. Long term decomposition studies have shown that alpha diversity does not begin to recover until dissolved soil oxygen levels exceed 75% during advanced decay and early skeletonization [8,12] and indicates that soil oxygen levels could have been low in our final time point, especially in the upper layer.

In support of this, the relative abundances of 4 obligate soil aerobes, *Rhodanobacter*, *Subgroup 12* (*Acidobacteriota*), *Gemmataceae,* and *Fimbrimonadaceae* were the first to decrease during purge (stage 2). This decrease was followed by a large-scale relative abundance depletion event of other aerobes following purge (stage 3). As has been reported by others, a majority of these taxa are found within the *Acidobacteriota* and *Verrucomicrobiota* phylum. In the lower layer, a limited number of aerobic taxa including *Fimbrimonadaceae, Bryobacter*, and *Pirellula* decreased in abundance by stage 4 suggesting oxygen levels may remain stable for a time in the lower layer as purge fluids and nutrients take time to spread throughout the soil. At the same time, several aerobic human and fly-associated microbial taxa (*Proteus, Ignatzschineria,* and *Erysipelothrix*) peaked in relative abundance during purge and became very rare or undetectable by stage 4. This trend indicates these taxa likely originated from an aerobic environment above the soil surface and likely could not survive after entering the anoxic soil. Together, these findings support the hypothesis that oxygen levels began to decrease during stage 2 and may have reached a minimum by stage 3 during spring and summer.

3.5.3 Slow recovery of aerobic taxa hints at future nitrification potential

During advanced and skeletonization stages of decomposition, soil oxygen levels

recover which permits the oxidation of accumulated ammonium to proceed through nitrification [3,34]. The relative abundance depletion of aerobic taxa following purge may hint at the possibility that aerobic nitrification may be initially inhibited due to low oxygen content and would explain the absence of nitrifying microbes during active decomposition. These results are in line with several decomposition studies that have shown soil nitrate concentrations only begin to increase following several months of decomposition once oxygen has recovered [8,12,13], a time point that is outside the window of our study.

The increasing relative abundance of some aerobes points towards recovery of dissolved oxygen towards advanced decomposition in the upper soil layer. For example, *Dyella, Massilia, Acidisoma and Acidisphaera* slowly increased during decomposition suggesting soil oxygen levels may have begun to recover by the final collected time point. Notably, we discovered several ammonia oxidizing taxa were enriched in the lower layer of control soils (i.e. *Candidatus Nitrosotalea* and *Nitrospira*) which may act as a sink for these key taxa [35] and will likely become important during later decomposition stages when oxygen levels recover allowing for nitrification to proceed.

Simultaneously, denitrification of nitrate to atmospheric N is known to occur within anoxic pockets of soil by anaerobic denitrifying bacteria [3,34]. The steady increase of *Burkholderia-Caballeronia-Paraburkholderia*, an observation previously reported by others [34], suggests denitrification may have been taking place in the upper layer and could potentially co-occur with nitrification during later time points.

3.5.4 Forensically informative taxa are present in upper layer

These predictable microbial responses to carbon and nitrogen nutrient pulses

from above ground human decomposition have become the foundation of microbial-based postmortem interval estimation models [4,5,11,36]. In our most recent work [4], we found a group of forensically important taxa emerge during decomposition across multiple geographical regions and seasons. While this study collected soils from the surface, we find some of these important indicator taxa permeate the soil surface. In addition to *Enterococcus* and *Ignatzschineria* (q-values < 0.01; **Figure 6**), we also found the relative abundances of *Savagea*, *Peptoniphilus*, *Vagococcus*, *Peptostreptococcus* (q-values < 0.05; **Supplemental Table 4)** increase during purge in the grave soils. In the lower layer, *Peptostreptococcus* and *Ignatzschineria* (q-values < 0.05; **Supplemental Table 5**) were also detected. Importantly, these taxa were only significantly enriched during the purge stage. Here, we included time points beyond the first 21 days of decomposition investigated by Burcham et al. 2024. and these findings suggest that taxa important for postmortem interval estimation models may not persist beyond active decomposition in subsurface grave soils. The possibility remains that these taxa may persist on the soil surface for longer periods of time, providing evidence of previous decomposition [31]. Even though some taxa only appear useful for predicting PMI during active decomposition, we found others such as *Terriglobus*, *Dyella*, *Massilia*, *Acidisoma* and *Acidisphaera* may be more informative indicator taxa during advanced decay and skeletonization, but further research is needed. Regardless of time frame, we recommend that soil sampling efforts should focus on the upper layer (0 - 4 cm) of grave soil in order to optimize detection of forensically important taxa. However, we acknowledge that our study only utilizes a single location with a single soil type (Ecological site F133BY007TX) and these findings may only be applicable to areas

with acidic, well-draining, sandy loam soils.

3.6 Conclusion

Our findings demonstrate that a majority of microbial responses to aboveground decomposition occur in the upper 4 cm of soil post-purge and into skeletonization in the spring and summer. As these responses were constrained to the upper 4 cm, mixing this layer with soils from deeper depths into a composite soil sample could obscure important biological signals. We also found small scale changes in the lower layer and taxa that may become important during more advanced stages of decomposition. Collectively, these findings suggest that future studies may benefit from collecting 10 cm soil cores and processing the upper and lower layers individually in order to avoid diluting signals that arise in different layers during different stages of decomposition.

CHAPTER 3: REFERENCES

1. Megyesi MS, Nawrocki SP, Haskell NH. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. J Forensic Sci. 2005;50: 618–626. Available: https://www.ncbi.nlm.nih.gov/pubmed/15932096

2. Carter DO, Yellowlees D, Tibbett M. Cadaver decomposition in terrestrial ecosystems. Naturwissenschaften. 2007;94: 12–24. doi:10.1007/s00114-006-0159-1

3. DeBruyn JM, Keenan SW, Taylor LS. From carrion to soil: microbial recycling of animal carcasses. Trends Microbiol. 2024. doi:10.1016/j.tim.2024.09.003

4. Burcham ZM, Belk AD, McGivern BB, Bouslimani A, Ghadermazi P, Martino C, et al. A conserved interdomain microbial network underpins cadaver decomposition despite environmental variables. Nat Microbiol. 2024;9: 595–613. doi:10.1038/s41564-023-01580-y

5. Metcalf JL, Xu ZZ, Weiss S, Lax S, Van Treuren W, Hyde ER, et al. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science. 2016;351: 158–162. doi:10.1126/science.aad2646

6. Cobaugh KL, Schaeffer SM, DeBruyn JM. Functional and Structural Succession of Soil Microbial Communities below Decomposing Human Cadavers. PLoS One. 2015;10: e0130201. doi:10.1371/journal.pone.0130201

7. Singh B, Minick KJ, Strickland MS, Wickings KG, Crippen TL, Tarone AM, et al. Temporal and Spatial Impact of Human Cadaver Decomposition on Soil Bacterial and Arthropod Community Structure and Function. Front Microbiol. 2018;8: 2616. doi:10.3389/fmicb.2017.02616

8. Taylor LS, Mason AR, Noel HL, Essington ME, Davis MC, Brown VA, et al. Transient hypoxia drives soil microbial community dynamics and biogeochemistry during human decomposition. FEMS Microbiol Ecol. 2024;100. doi:10.1093/femsec/fiae119

9. Pechal JL, Crippen TL, Benbow ME, Tarone AM, Dowd S, Tomberlin JK. The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. Int J Legal Med. 2014;128: 193–205. doi:10.1007/s00414-013-0872-1

10. Adserias-Garriga J, Hernández M, Quijada NM, Rodríguez Lázaro D, Steadman D, Garcia-Gil J. Daily thanatomicrobiome changes in soil as an approach of postmortem interval estimation: An ecological perspective. Forensic Sci Int. 2017;278: 388–395. doi:10.1016/j.forsciint.2017.07.017

11. Mason AR, McKee-Zech HS, Steadman DW, DeBruyn JM. Environmental predictors impact microbial-based postmortem interval (PMI) estimation models within human decomposition soils. PLoS One. 2024;19: e0311906.

doi:10.1371/journal.pone.0311906

12. Keenan SW, Schaeffer SM, Jin VL, DeBruyn JM. Mortality hotspots: Nitrogen cycling in forest soils during vertebrate decomposition. Soil Biol Biochem. 2018;121: 165–176. doi:10.1016/j.soilbio.2018.03.005

13. DeBruyn JM, Hoeland KM, Taylor LS, Stevens JD, Moats MA, Bandopadhyay S, et al. Comparative Decomposition of Humans and Pigs: Soil Biogeochemistry, Microbial Activity and Metabolomic Profiles. Front Microbiol. 2020;11: 608856. doi:10.3389/fmicb.2020.608856

14. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37: 852–857. doi:10.1038/s41587-019-0209-9

15. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13: 581–583. doi:10.1038/nmeth.3869

16. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6: 90. doi:10.1186/s40168-018-0470-z

17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41: D590-6. doi:10.1093/nar/gks1219

18. Mirarab S, Nguyen N, Warnow T. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput. 2012; 247–258. doi:10.1142/9789814366496_0024

19. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61: 1–10. doi:10.1016/0006-3207(92)91201-3

20. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47: 583–621. doi:10.1080/01621459.1952.10483441

21. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr. 1957;27: 325–349. doi:10.2307/1942268

22. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71: 8228–8235. doi:10.1128/AEM.71.12.8228-8235.2005

23. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online. Chichester, UK: John Wiley & Sons, Ltd; 2017. pp. 1–15. doi:10.1002/9781118445112.stat07841

24. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction.

Nat Commun. 2020;11: 3514. doi:10.1038/s41467-020-17041-7

25. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. Proceedings of the Python in Science Conference. SciPy; 2010. pp. 92–96. doi:10.25080/majora-92bf1922-011

26. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol. 2021;17: e1009442. doi:10.1371/journal.pcbi.1009442

27. Iancu L, Necula-Petrareanu G, Purcarea C. Potential bacterial biomarkers for insect colonization in forensic cases: preliminary quantitative data on Wohlfahrtiimonas chitiniclastica and Ignatzschineria indica dynamics. Sci Rep. 2020;10: 8497. doi:10.1038/s41598-020-65471-6

28. Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by Nitrospira bacteria. Nature. 2015;528: 504–509. doi:10.1038/nature16461

29. Lehtovirta-Morley LE, Sayavedra-Soto LA, Gallois N, Schouten S, Stein LY, Prosser JI, et al. Identifying potential mechanisms enabling acidophily in the ammonia-oxidizing archaeon “Candidatus Nitrosotalea devanaterra.” Appl Environ Microbiol. 2016;82: 2608–2619. doi:10.1128/AEM.04031-15

30. Xu W, Wang Y, Wang Y-H, Zhang Y-N, Wang J-F. Diversity and dynamics of bacteria at the Chrysomya megacephala pupal stage revealed by third-generation sequencing. Sci Rep. 2022;12: 2006. doi:10.1038/s41598-022-06311-7

31. Burcham ZM, Weitzel MA, Hodges LD, Deel HL, Metcalf JL. A pilot study characterizing gravesoil bacterial communities a decade after swine decomposition. Forensic Sci Int. 2021;323: 110782. doi:10.1016/j.forsciint.2021.110782

32. Carter DO, Metcalf JL, Bibat A, Knight R. Seasonal variation of postmortem microbial communities. Forensic Sci Med Pathol. 2015;11: 202–207. doi:10.1007/s12024-015-9667-7

33. Carter DO, Yellowlees D, Tibbett M. Temperature affects microbial decomposition of cadavers (Rattus rattus) in contrasting soils. Appl Soil Ecol. 2008;40: 129–137. doi:10.1016/j.apsoil.2008.03.010

34. Yang J, Su W, Yu Q, Shi Z, Huang X, Heděnec P, et al. The long-term decomposition of wild animal corpses leads to carbon and phosphorus accumulation and disturbs the ecological succession of the denitrification community encoded by narG. Appl Soil Ecol. 2022;175: 104455. doi:10.1016/j.apsoil.2022.104455

35. Keenan SW, Emmons AL, DeBruyn JM. Microbial community coalescence and nitrogen cycling in simulated mortality decomposition hotspots. Ecol Process. 2023;12. doi:10.1186/s13717-023-00451-y

36. Belk A, Xu ZZ, Carter DO, Lynne A, Bucheli S, Knight R, et al. Microbiome Data Accurately Predicts the Postmortem Interval Using Random Forest Regression Models. Genes . 2018;9. doi:10.3390/genes9020104

CHAPTER 4: EFFECTS OF INDOOR DECOMPOSITION ON THE CADAVER MICROBIOME

4.1 Summary

Estimating the postmortem interval (PMI) of human remains is a critical, but challenging step in death investigations. Traditional forensic methods that estimate PMI based on the physical changes that occur following death and forensic entomology may be less accurate when environmental variables such as temperature fluctuate rapidly. Recent work has shown that a microbial network assembles during human decomposition to break down organic matter regardless of climate and geography. These microbial decomposers act as valuable predictors in novel microbiome-based PMI estimation techniques. While these studies provide a framework for estimating PMI, they exclusively rely on data collected from outdoor forensic anthropological research facilities. Whether factors such as an indoor environment alter the trajectory of microbial communities decomposing human remains is unclear. In this study, we investigate the effects of enclosed shelter on microbial community assembly and successional patterns during human decomposition across two seasons and provide important considerations for PMI modeling. Compared to outdoor cadavers, we first show that indoor cadavers experienced delayed colonization of key decomposer microbes over the course of decomposition due to delayed maggot infestation. Consequently, machine learning models trained on outdoor cadavers frequently underestimated the PMI of cadavers decomposing indoors. Importantly, delayed maggot colonization was associated with higher PMI prediction errors suggesting that blow flies and maggots are important

sources of microbial decomposers that drive PMI model predictions. Ultimately, incorporating indoor cadaver data and insect activity into PMI models significantly improved prediction capabilities for both indoor and outdoor decomposition environments.

4.2 Introduction

The decomposition of human remains is a complex and dynamic process that is largely dependent on various environmental and biotic factors. In death investigations where no witnesses were present, medical examiners rely on well-established physical postmortem changes such as rigor mortis (stiffening of the muscles) and forensic entomology to estimate time since death [1–4]. However, the onset and rate at which these physical changes progress can be variable and is largely influenced by factors like temperature, humidity, and body mass [5–7] which can increase post mortem interval (PMI) estimates errors. Furthermore, traditional PMI estimations are most robust within particular postmortem timeframes [8], therefore developing novel forensic estimation techniques that can help span later time points is useful. Recently developed microbial-based machine learning models offer an additional forensic tool for predicting time since death, generally in the timeframe of days to weeks [9–15]. Microbial communities present in or on the cadaver, deposited by scavengers, and in the surrounding soil aid in the breakdown and nutrient cycling of human remains. As nutrient pools are consumed and availability shifts to other resources, community membership is known to predictably change in response [16–18]. These microbial successional patterns are the foundation of microbial-based forensic tools that enable robust PMI estimation.

Recent work has advanced our understanding of the ecology and assembly of microbial decomposer communities [10] by decomposing 36 cadavers across 3 outdoor forensic research facilities and characterizing microbial changes over time. Data revealed that an interdomain network of microbial decomposers universally assembles during human decomposition regardless of varied host communities, geography and season. Key members of this network included *Thiopseudomonas alkaliphila (*previously *Oblitimonas alkaliphila)*, *Ignatzschineria spp.*, *Wohlfahrtiimonas chitinclastica*, *Bacteroides*, *Vagococcus lutrae*, *Savagea*, *Acinetobacter rudis,* and *Peptoniphilaceae.* These central taxa exhibited high cross-feeding potential suggesting they likely co-metabolize the ephemeral nutrient pool that permeates the soil. The universal assembly and predictable successional patterns of the decomposer network underpinned a new microbiome-based machine learning model that accurately predicts PMI within +/- 3 consecutive days across multiple climate regions and seasons. Many of these key network taxa were rare in the environment and phylogenetically distinct from abundant human-associated taxa. Several are present in other vertebrate decomposition studies and are also frequently associated with blow flies and carrion beetles. While still unclear, it is hypothesized insects, and other scavengers, act as a play a large role in the dispersal of decomposer networks in the environment.

Although microbiome-based PMI estimation models are promising forensics tools, they are exclusively based on outdoor decomposition scenarios. A large number of human remains are found indoors, yet indoor decomposition is rarely studied and there has been no research to date that has investigated microbial community succession patterns under indoor conditions in humans [19]. Of the few studies that

have been conducted, several have observed that blow fly colonization of indoor cadavers is delayed compared to outdoor cadavers and this delay is associated with slower decomposition rates [19–21]. How insect colonization and enclosed shelter affects the cadaver microbiome and microbial succession is currently unclear and represents a large scientific knowledge gap. As insects are likely critical for the dispersal of the decomposer network, we hypothesized that microbial community succession will be significantly altered indoors if insect colonization is delayed. Furthermore, we suspect that these differences will likely negatively impact the accuracy of microbial-based PMI estimation models.

To better understand the impact of enclosed shelter on microbial community succession during human decomposition, we placed 12 cadavers inside unconditioned, enclosed shelters to naturally decompose along with 15 outdoor control cadavers at a forensic research facility in Huntsville, TX. Using 16S rRNA gene amplicon sequencing, we characterized the communities of the skin over the first 21 days of decomposition. We found that overall microbial community composition was strikingly similar between indoor and outdoor cadaver skin. However, indoor community succession lagged behind patterns observed in outdoor cadavers. Random forest models that incorporated microbiome data from indoor cadavers and environmental variables such as evidence of maggots and total body scores produced the most accurate PMI estimates, overcoming the different successional patterns observed during indoor decomposition. This novel research provides valuable insights about decomposer community assembly and demonstrates that machine learning model performance is closely tied to insect-cadaver

interactions, a critical discovery that will advance the development of PMI prediction models.

4.3 Materials & Methods

4.3.1 Study design, donors, and donor placement

We utilized the Sam Houston State University Southeast Texas Forensics Facility (STAFS) to place willed body donors in indoor and outdoor decomposition scenarios. STAFS is a morgue-like laboratory located approximately 5 km north of Huntsville, Texas with an outdoor area designed for decomposition studies and forensic science training. The facility is located in the southeast Texas Pineywoods ecoregion characterized by a humid and subtropical climate and a sparse forest covering of pine trees and a ground covering of herbaceous plants. The soil is a fine, friable sand that is moderately to strongly acidic and is moderately well-drained with medium available water content and slow permeability and runoff [22].

We placed a total of 27 human donors to decompose naturally beginning in the fall of 2020 and ending in the spring of 2022 (indoors n=12; outdoors n=15). During the fall of 2020, 3 donors were placed inside separate enclosed shelters along with 6 outdoor cadavers (3 = caged, 3 = uncaged). During the spring and fall of 2021 and the spring of 2022, 3 donors were placed indoors with 3 paired outdoor cadavers each season for a total of 9 indoor and 9 outdoor (all caged) cadavers. Donors were placed in indoor-outdoor pairs approximately one week apart for three consecutive weeks during each placement season.

Enclosed shelters were constructed from plastic and were not air conditioned. Cadavers were placed directly on the plastic floor and left to naturally decompose.

Doors remained closed unless sampling was taking place. Outdoor cadavers (*n*=15) were placed on undisturbed ground. All cadavers were placed unclothed in a supine position. A scavenger exclusion cage constructed from wood and wire was placed over twelve of the 15 outdoor bodies to prevent loss of limbs and other body parts to scavengers, as has been done in previous research at this site [9,10]. Scavenger exclusion cages were not utilized for three outdoor bodies, to compare potential effects of caging on outdoor donor decomposition. The three uncaged outdoor donors were placed with the indoor-outdoor pairs during the fall of 2020.

Cadavers did not undergo autopsy and were not embalmed. All bodies were stored at 4°C until placement. Individuals who died from SARS-CoV-2 or related complications were excluded from the study.

4.3.2 Temperature measurements, accumulated degree day calculations, and total body scores

Daily average air temperature was collected from the Easterwood Airport Station using Weather Underground (https://www.wunderground.com/). Temperature loggers were placed inside each of the three indoor structures (Onset HOBO MX2201 Pendant Wireless Temperature Data Logger). Indoor temperature was logged every 61 minutes and the average was taken for each day. Temperature loggers collected indoor temperatures from December 02, 2020 until May 13, 2021. The temperature of all three enclosed structures (sheds) was compared against the average outdoor temperatures during this period to determine whether outdoor temperatures could be used to calculate accumulated degree days (ADD). We found the indoor temperatures were highly correlated with outdoor temperatures (Pearson's $r = 0.96$, p-value < 0.001;

Supplemental Figure S1) and varied on average by +/- 1.4 °C. Thus, we used daily average outdoor temperature to calculate ADD for both indoor and outdoor donors.

ADD estimates were calculated following Megyesi et al. 2005 [1]. On day zero, the day of placement, ADD was set to zero for each donor. The ADD of the following day (day 1) was calculated as the average temperature of the preceding day (day 0). On day 2, the ADDs of all preceding days (day 0 + day 1) were summed. This was repeated for all succeeding days so that the final day's ADD (day 21) was the sum of all the preceding average daily temperatures. Values below 0 °C (baseline) were excluded. Total body scores (TBS) were calculated by summing visually estimated decomposition scores of head, trunk, and limbs following descriptions outlined in Megyesi et al. 2005 [1]**.**

4.3.3 Sample collection

Samples were collected using sterile dual-tipped BD SWUBE Applicator (REF 281130) swabs. Samples were taken daily for the first 21 days of decomposition beginning on the day of placement (day 0) from the skin of the face and the hip. Swabs of the indoor floor and outdoor soil near the hip were also collected daily. Skin samples were collected by gently rubbing the swab over the skin area for approximately 30 seconds with care not to break or damage the skin. For outdoor donors, the soil outside of the purge zone near the hip was sampled with a swab until visible soil material was adhered to the swab. Samples of the indoor floor and outdoor soil were also collected prior to donor placement. A control soil site that was not impacted by decomposition was sampled daily near each outdoor donor. Bulk soils (approximately 25 grams) were collected from the placement site and control site on the day of placement and every 7

days following placement. Swabs and bulk soils were immediately transferred to a -20°C freezer at Sam Houston State University until they were shipped on dry ice to the Metcalf lab at Colorado State University for DNA extraction and amplicon data generation.

4.3.4 DNA extraction

Each swab from the dual-tipped SWUBE was trimmed using sterile scissors and placed into a barcoded 96-well plate. One swab head was used for DNA extractions and the other was stored at -20°C as a backup in the respective 96-well plate. The MagAttract PowerSoil Pro DNA isolation kit (Qiagen, cat no. 47109) for KingFisher was used to extract microbial DNA following the manufacturer's instructions with the following modifications. DNA extraction plates 1- 2 were processed using the 96-well PowerBead Pro plate (Qiagen). All remaining DNA extraction plates were processed using a single bead tube format in place of a 96-well PowerBead Pro plate until lysate was transferred to a KingFisher Flex System (ThermoFisher). Following lysis and inhibitor removal, we utilized all lysate using a two-step binding approach outlined in the Earth Microbiome Project (EMP) DNA extraction protocol [23]. Briefly, all lysate was split between two deep-well KingFisher plates containing equal parts of binding solution (QSB1). MagAttract Suspension G beads were added to DNA binding plate 1 before transferring to a KingFisher for binding, purification, and elution. DNA was eluted in 60 µL of solution C6. Each DNA extraction plate included 6 blank extraction controls where no sample was added and 2 positive DNA controls where 25 µL of ZymoBIOMICS Microbial Community Standard was added prior to extraction. Samples were randomized across all plates to reduce batch effects.

4.3.5 Amplicon library preparation and sequencing

The V4 region of the 16S rRNA gene was amplified following the EMP protocol [24] using Golay barcoded primers 515f (5′GTGYCAGCMGCCGCGGTAA) [25] and 806rf (5′GGACTACNVGGGTWTCTAT) [26]. All samples were amplified to 30 cycles using the Platinum Taq DNA Polymerase Kit (Invitrogen) with the exception of DNA extraction plates 14-18 which were amplified to 27 cycles due to *Achromobacter* contamination originating from the MagAttract Suspension G beads (lot #172029496). Following amplification, amplicons were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen), each PCR plate was pooled together in equimolar amounts, and cleaned using UltraClean PCR clean-up kit (Qiagen). Libraries were shipped on ice to the CSU foothills campus sequencing core or to the Fierer lab at the University of Colorado, Boulder for sequencing. 16S rRNA amplicon libraries were sequenced using a 500-cycle kit on the Illumina MiSeq platform.

4.3.6 Data processing and analysis

Raw sequence data generated as a part of this study was processed and analyzed using QIIME2 bioinformatics platform versions 2023.5 and 2024.10 [27]. A total of 1,912 samples were sequenced which included 50 negative extraction controls and 44 positive extraction controls. Raw data from each sequencing run (n=8) was demultiplexed and quality filtered separately using the QIIME2 DADA2 plugin [28] with default parameters. All forward reads were truncated at 200 base pairs while all reverse reads were truncated at 100 base pairs before merging. All 8 frequency tables and amplicon sequence variant (ASV) tables were dereplicated and merged to create a one frequency table and one ASV table. After denoising and quality filtering, a total of

56,621,198 reads remained, with an average of 29,613 reads per sample. Taxonomy was assigned using the QIIME2 feature-classifier classify-sklearn method [29] with a naïve Bayes taxonomy classifier trained on the V4 (515f/806rb) region of Greengenes2 backbone (v.2024.09) [30]. ASVs classified as mitochondria, chloroplast, or *Achromobacter* (DNA extraction contamination) were removed from the dataset. The SEPP method from the QIIME2 fragment-insertion plugin was used to construct a rooted phylogenetic tree using the Greengenes2 reference phylogeny (v.2022.10) [31].

Alpha and beta diversity metrics were generated using the QIIME2 phylogenetic-diversity plugin using 5000 reads per sample as the rarefaction depth. A Kruskal–Wallis H-test was used to test for differences in alpha diversity metrics between groups. A Benjamini–Hochberg multiple-testing correction was used for comparisons with more than 2 groups. Permutational analysis of variance (PERMANOVA) was used to evaluate statistical differences in beta diversity metrics between groups. A Bonferroni correction was used when there were more than 2 groups per comparison. The Analysis of Compositions of Microbiomes with Bias Correction (ANCOMBC) package available as a QIIME2 plugin was used to identify differentially abundant taxa. Frequency tables were filtered to only include skin samples and then ASVs were collapsed at the species level (L7) for modeling. The ANCOM-BC plugin was run using a prevalence cutoff of 1% and minimum reads per sample set to 1000. The QIIME2 linear mixed-effects models package was used to test for statistical differences in beta diversity between decomposition environments (indoor vs outdoor) across ADDs using the formula “PCoA Axis 1 ~ ADD * decomposition_environment”. For time points with repeated measures, a value was chosen at random. Donor ID was used as a random intercept. Visualizations

were created with JupyterLab (v.4.1.5) using tidyverse (v.2.0.0), Plotly (v.5.20.0), Seaborn (v.0.11.2), or Matplotlib (v.3.6.0) in python (v.3.8.16).

4.3.7 Three outdoor facilities decomposition amplicon dataset

Decomposition data generated during a previous study conducted by our lab [10] was used here as a training dataset to create a machine learning model for estimating time since death. However, samples from Qiita prep ID 14929 were re-sequenced as described above and were included here (Qiita prep IDs 17684 and 17685). Raw forward and reverse 16S rRNA gene amplicon sequences were obtained from Qiita study ID 14989 and were processed as described above with the exception of the forward and reverse read truncation length which was set to 150 bp for all sequencing runs.

4.3.8 Random Forest regression modeling

A random forest regression (RFR) model was trained on the 3 outdoor facilities decomposition dataset to assess whether a model trained only on outdoor skin samples can accurately predict the PMI of cadavers decomposition indoors. This model was then used to predict the ADD of indoor and outdoor skin samples collected for this study. The dereplicated 3 outdoor facility frequency table was filtered to only contain skin samples (skin of face and hip), rarefied to 5000 reads per sample, and ASVs were collapsed on taxonomy level 7 (species) which has previously been shown to provide the best performance [10]. Skin samples from the face and hip of cadavers decomposing indoors and the paired outdoor control cadavers were used as the test set. The frequency table was also rarefied to a depth of 5000 and ASVs were collapsed to the species level to match the training set. The 3 outdoor facility and indoor-outdoor tables were then

filtered to only contain features that were present in both datasets (*n*=2153). Random forest models were built using the scikit-learn package (v.0.24.1) [32]. The best hyperparameters from Burcham et al. 2024 (bootstrap=False, max_depth=None, max_features=0.2, n_estimators = 1,000) were selected for training. This model was then used to predict the ADD of the indoor-outdoor skin dataset. Mean absolute error (MAE) was used to evaluate model accuracy by subtracting the predicted ADD from the actual ADD and taking the absolute value of all test set samples.

A cross-validation approach was used to evaluate the inclusion of indoor cadavers on model performance. Using LeaveOneGroupOut partitioning, we grouped the indoor-outdoor skin data by cadaver so that samples from the same donor were not split between the train and test datasets. For each random forest regression model, one cadaver was withheld as a test set and a model was built using the remaining donors (*n*=26). ADDs were then estimated for all skin samples in the test set using this 26-body model. The test cadaver was then moved back to the training set and a new cadaver was then assigned as the test set. This process was repeated until all donors (*n*=27) were designated as the test set. All models were run using the hyperparameters described above. Feature importances were averaged across all 27 models. Categorical variables such as environment were label encoded as numerical values prior to building models. General model accuracy was determined by averaging the mean absolute errors of all ADD predictions across all cadavers.

4.3.9 Data Availability

All 16S rRNA gene amplicon sequencing data is available through the open-source microbial study management platform Qiita (https://qiita.ucsd.edu/) under study

ID 13810. Raw data for the three-facility dataset can also be found on Qiita under study ID 14929 or through ENA project accession number PRJEB62460 (ERP147550). Greengenes2 data files can be found on the Greengenes 2 server (https://ftp.microbio.me/greengenes_release/). Code used for data processing and intermediate files are available through Open Science Framework under project "Indoor vs. Outdoor Human Decomposition" (https://osf.io/f9v8p/).

4.4 Results

4.4.1 Community differences between indoor and outdoor cadaver decomposition

Cadavers decomposing in an outdoor, aboveground environment are exposed to numerous environmental and biotic variables such as precipitation, insects, and soil, that may be limited in scenarios of indoor decomposition. Therefore, we first sought to understand how enclosed shelter impacts the microbial communities that assemble on decomposing cadaver skin. The microbial communities of indoor cadaver skin had lower alpha diversity compared to outdoor cadavers. Faith's phylogenetic diversity, Shannon diversity, and the number of unique amplicon sequence variants (ASVs) were all significantly impacted by indoor shelter (**Figure 1A-B** and **Supplemental Figure S2A**). We found that 7592 ASVs were unique to outdoor cadavers but this large number only accounted for a relative abundance of 2.5% on average (**Supplemental Figure S2B-C**), suggesting that although less diverse, the indoor skin communities share a large number of highly abundant ASVs with the outdoor cadavers. Bray-Curtis and Unweighted Unifrac beta diversity metrics both showed that community composition was significantly different between indoor and outdoor cadavers (**Figure 1C** and **Supplemental Figure S2D**). However, we found that accumulated degree day (ADD)

explained more variation in community composition (R^2 = 0.058, *p*-value < 0.01) than decomposition environment (R^2 = 0.03, *p*-value <0.001).

To better understand whether the decomposition environment alters the temporal patterns of the decomposer community, we evaluated changes in beta diversity across ADDs. Bray-Curtis distances between indoor and outdoor communities were more variable during early and active decomposition (50-200 ADD) along PCoA Axis 1(**Figure 1D**). Although statistically different (LME model *p*-value < 0.001), by advanced decomposition (>200 ADD), microbial communities appear to be converging on similar compositions suggesting that although the decomposition environment (indoor vs. outdoor) significantly influences the decomposer microbiome, especially at earlier time points, other factors may play a larger role in shaping community structure.

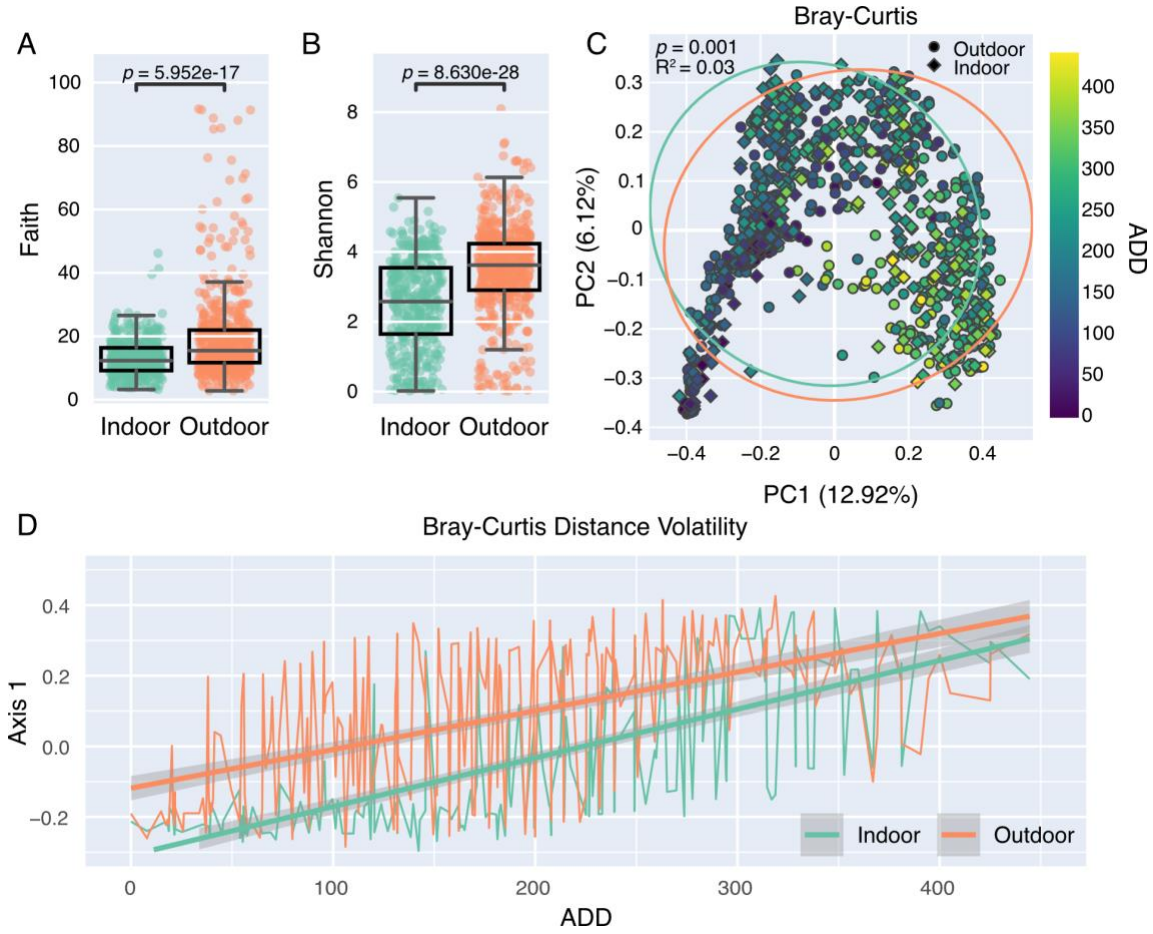


Figure 1. Difference between indoor and outdoor human decomposition. A) Box plots of Faith's phylogenetic diversity and **B**) Shannon's diversity compared between skin of indoor cadavers (teal) and outdoor cadavers (orange). The group median value is represented by the centerline of the box plots. Mann-Whitney test was used to test for significance between groups. **C**) Principal Coordinate Analysis (PCoA) plot based on Bray-Curtis distances showing cadaver skin samples collected from outdoor (circle) and indoor (diamond) cadavers. Samples are colored by ADD. 95% confidence intervals are shown for indoor (teal) and outdoor (orange) cadaver groups. PERMANOVA was used to test for significance between indoor and outdoor skin samples. **D**) Bray-Curtis PCoA Axis 1 coordinates are shown for indoor (teal) and outdoor (orange) cadaver skin samples across ADD with standard error (grey). A linear-mixed effects model was used to test for statistical significance with Donor ID as a random intercept. Timepoints with repeat measures were handled by randomly choosing one value. Indoor vs. outdoor across ADD p-value < 0.001.

We next evaluated which microbial taxa were differentially abundant between the skin of cadavers decomposing indoors and outdoors. Numerous microbes were found to

be enriched in the outdoor cadaver group while only a few were more abundant on indoor cadaver skin (**Figure 2**). Importantly, several taxa that were enriched outdoors, including *Ignatzschineria larvae, Savagea,* and *Wohlfahrtiimonas chitiniclastica* are key members of the universal decomposer network identified in previous research [10] and are strongly associated with blow flies, blow fly maggots, and beetles [33,34]. Enrichment of these taxa on outdoor cadavers, but not indoor, strongly suggest that insects may be important drivers of cadaver community assembly and composition.

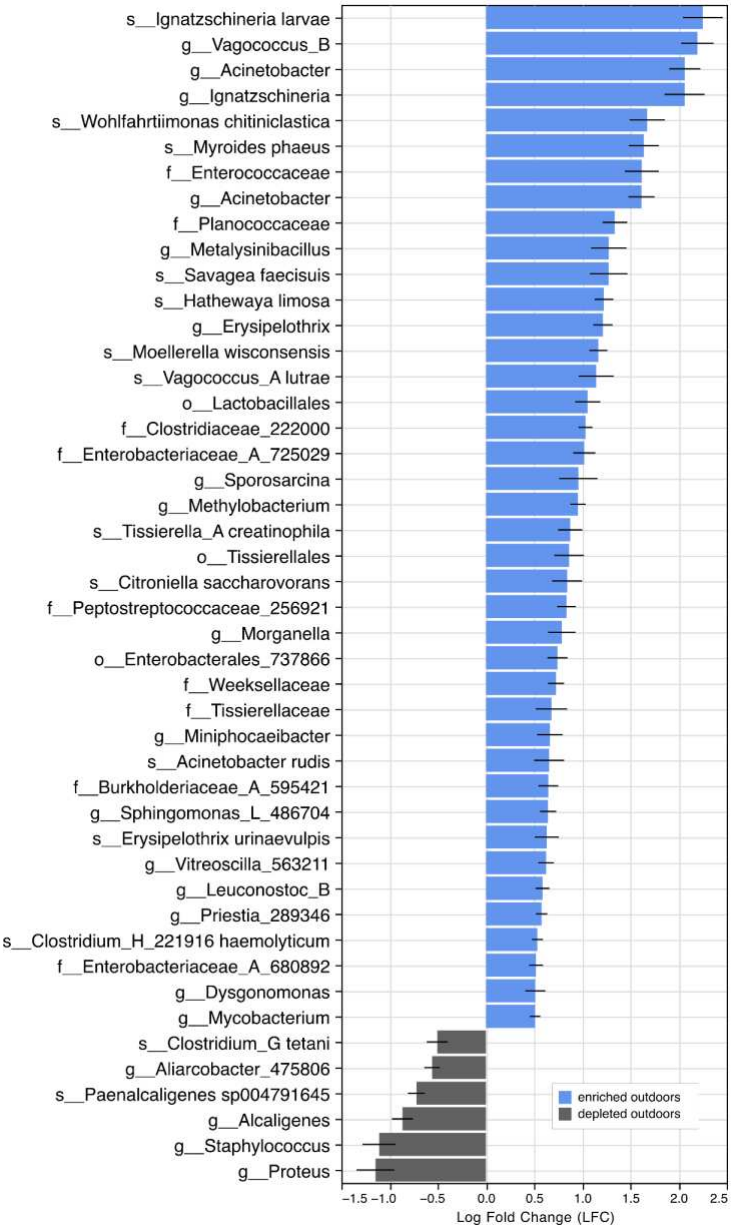


Figure 2. Taxa associated with indoor and outdoor decomposition. Bar plot with standard errors is shown. Taxa significantly enriched in outdoor cadaver skin samples are shown in blue. Taxa enriched in indoor samples are shown in grey (*q*-values < 0.05). Analysis was performed using ANCOM-BC. ASVs were collapsed on the species level. Lowest taxonomic classification is shown. Samples with fewer than 1000 reads and taxa found in less than 1% of samples were excluded. Only taxa with >= +/- 0.5 LFC are shown.

4.4.2 Influence of blow fly maggots on the decomposer microbiome

Blow fly access to and from the indoor structures was not controlled during this study and consequently, the first appearance of maggots on indoor cadavers lagged behind their appearance on the cadavers placed outdoors. This difference was most likely due to obstructed access into the structures that were only opened once per day for sampling. The average first day of a notable maggot presence was 11.1 days (ADD = 205.48) after the placement of indoor cadavers compared to 4.5 days (ADD = 74.9) for outdoor cadavers (**Figure 3A**). We observed that two indoor donors (D7 and D9) and two outdoor donors (D11 and D12) were not colonized by noticeable maggots during the timeframe of our study (21 days) which may have been due to the low outdoor temperatures during decomposition (mean outdoor temp = 11.7°C).

We found that cadaver skin without previous exposure to maggots was enriched for *Pseudomonas, Staphylococcus,* and *Corynebacterium* (**Supplemental Figure S3**). However, following the emergence of maggots, *Ignatzschineria, Savagea, Vagococcus, Thiopseudomonas alkaliphila* and many others became enriched. Taxa bar plots that were grouped by indoor and outdoor donors and ordered by days post placement showed the relative abundance of these maggot-associated taxa began to increase concurrent with maggot emergence (**Figure 3B, Supplemental Figures S4 & S5**). This was also reflected in Bray-Curtis diversity distances which showed indoor and outdoor communities became more similar to each other along Axis 1 following the appearance of indoor maggots at approximately 200 ADD (**Figure 1C-D**). Interestingly, cadavers in which maggots were never detected were almost entirely dominated by *Pseudomonas, Staphylococcus,* and *Corynebacterium* across the entire 21-day decomposition trial

(**Supplemental Figures S4 & S5**). Together, this evidence supports the idea that significant changes in microbial community composition are associated with notable maggot mass development and likely begins when blow flies first visit cadavers.

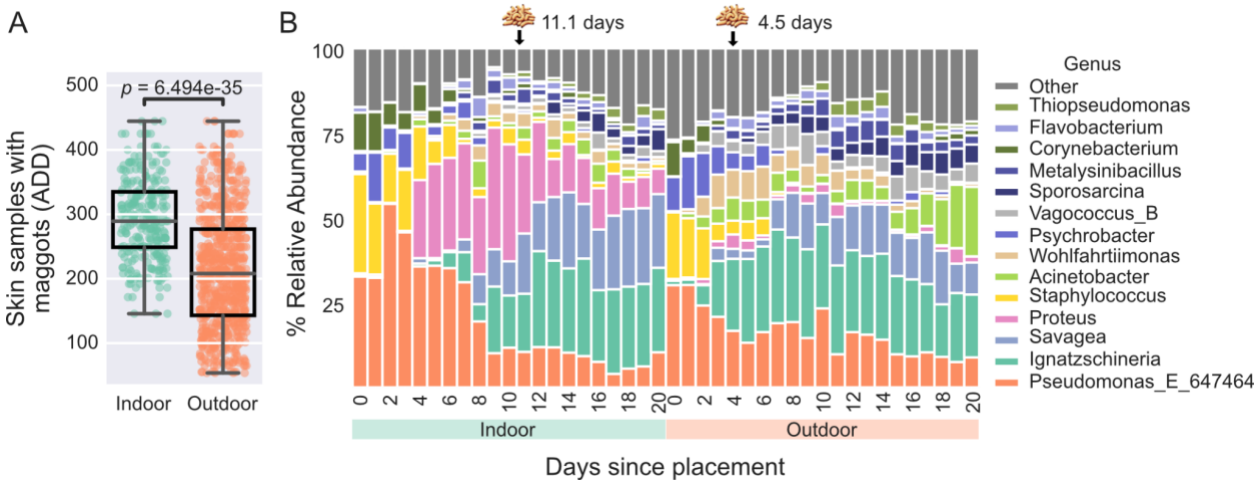


Figure 3. Maggot emergence is delayed indoors. A) Box plot shows ADD of skin samples that had evidence of maggot exposure are shown for indoor (teal) and outdoor (orange) cadavers. Mann-Whitney test was used to test for significance between groups. **B**) Taxa bar plots show average percent relative abundance of the 14 most abundant genera found in indoor (left) and outdoor (right) cadaver skin samples across all 21 days of decomposition. Maggot mass icons denote the average first day of notable maggot mass for each group.

4.4.3 Predicting indoor postmortem intervals with a model trained on outdoor cadavers

Although microbial succession patterns of indoor cadavers appeared to vary from outdoor cadavers, both cadaver groups shared similar taxa (**Supplemental Figure 2B-C**) which may enable the use of our large and robust 3 outdoor facilities model for predicting PMI regardless of decomposition environment (indoor vs. outdoor) [10]. We found that the 3 outdoor facilities and indoor/outdoor datasets shared 2153 model features which represent ASVs grouped at the species level (L7). Unsurprisingly, the 3

outdoor facilities random forest regression (RFR) model that was trained on data collected from cadavers decomposing outdoors performed better with the outdoor cadavers in this study. The mean absolute prediction error (MAE) was 50.26 ADD for outdoor cadavers (*n*=15) compared to 72.05 ADD for indoor cadavers (*n*=12; **Figure 4A**). We also found that this model generally underpredicted the ADD of indoor cadavers, especially at later time points (**Supplemental Figure S6B)**.

We next looked to identify potential differences between indoor and outdoor cadaver communities that could be responsible for the significant differences in prediction accuracy by examining the 20 most important taxa from the 3 outdoor facilities model. The change in relative abundance of each taxon across ADD was used to survey possible differences in arrival time and the magnitude of change which may reduce the prediction accuracy of indoor samples. Many of these important features exhibited similar trends (e.g. relative abundance increased across ADDs) but these changes were generally delayed in the indoor group. For example, the most important 3 outdoor facilities model feature, *Citroniella saccharovorans*, began to increase at approximately 150 ADD in outdoor cadavers and at approximately 275 ADD in indoor samples (**Figure 4B**). Similarly, Peptostreptococcus russelilli_A increased at approximately 100 ADD in outdoor samples, while in indoor samples this taxon was delayed until 200 ADD.

Of the 20 most important taxa from the 3-facility model,16 were also ones previously found to be significantly enriched in cadaver skin samples with notable maggots (**Supplemental Figures S3 & S6A**) suggesting a possible link between model performance and maggot-associated taxa. We also found that many of the important

features were generally absent from the donors that were never colonized by maggots (**Supplemental Figure S7**). Based on the likely association between these important model features and maggots, we evaluated whether PMI prediction MAEs and ADD at the time of first notable maggot appearance were correlated. For most indoor donors, PMI errors increased as the time between cadaver placement and first maggot appearance increased (**Figure 4C**). This trend strongly indicates that blow fly and maggot-associated microbes drive PMI modeling.

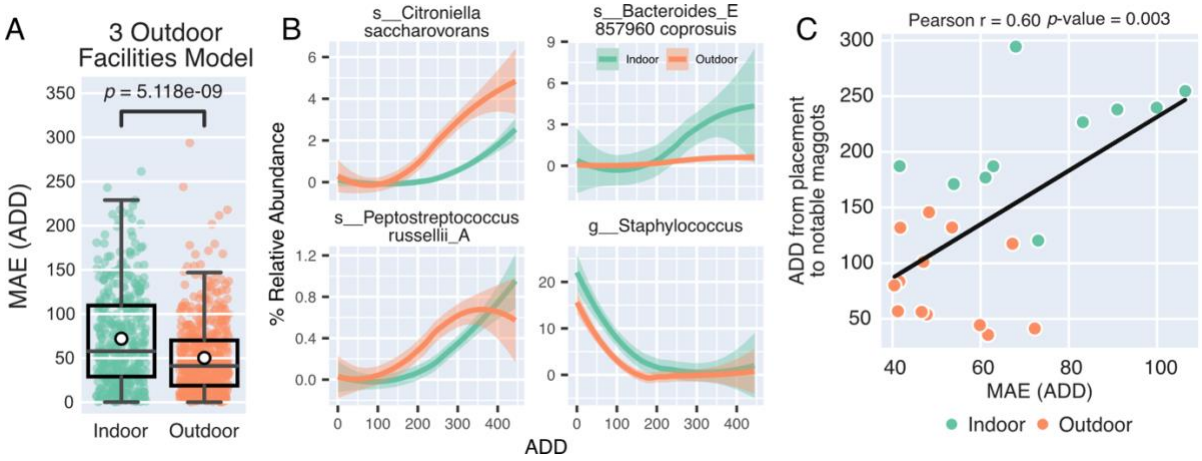


Figure 4. Maggots drive outdoor based PMI model predictions. A) RFR model trained on skin samples collected from cadavers decomposing at 3 different outdoor forensic research facilities was used to predict the PMI of a new indoor and outdoor cadaver dataset. Box plot of mean absolute error (MAE) of prediction is shown for both groups. White marker denotes group means. Mann-Whitney test was used to test for statistical significance. **B**) The four most important 3 outdoor facilities model features are shown from most (top left) to fourth most important (bottom right). The percent relative abundance of each feature is compared between indoor (teal) and outdoor (orange) cadaver skin samples across decomposition. Percent relative abundances were smoothed using loess regression. **C**) Scatter plot shows MAE of each donor plotted against the time (ADD) between cadaver placement and the first appearance of maggots. Donors that were never colonized by maggots were omitted. Colors represent decomposition environments. Pearson correlation (r) was used to test for significant relationships between variables. Linear regression of MAE and maggot ADD is shown in black.

4.4.4 Incorporating indoor cadavers and environmental variables into PMI models

Given that most outdoor bodies are almost immediately visited by blow flies, RFR models trained solely on outdoor data likely fail to capture microbial succession patterns that arise when maggot colonization is delayed. To account for this, we next trained several RFR models on the indoor/outdoor dataset from this study to determine if including indoor cadavers and environmental data like decomposition environment, evidence of maggots, and total body score (TBS) could help improve PMI prediction accuracy of indoor bodies. In total, 4 different models were trained on the indoor and outdoor cadaver data (**Table 1**). The first model, “indoor & outdoor”, was trained on microbial features collected from this new dataset of 27 donors (*n*=12 indoor, *n*=15 outdoor). In a second model, “indoor & outdoor & env”, we included additional data about the decomposition environment (whether the cadaver was inside or outside) and maggot exposure (whether there was evidence of maggots at time of sampling) in addition to microbial features. The third model (“indoor & outdoor & env & TBS”) expanded on the previous model by incorporating TBSs. In a fourth model, “env & TBS”, only environmental data and body scores were included (no microbial features).

Table 1. RFR model summary. Training dataset refers to data used to build the PMI prediction model. Test set refers to data used for testing model accuracy. A leave-one-donor-out cross-validation approach was used to train and test the indoor/outdoor dataset. 16S data was either included ("x") or excluded (" ") from modeling. Information about the decomposition environment (indoor/outdoor column), whether maggots had ever been observed (maggots column), and total body score (TBS column) were included ("x") in several models.

Model	Training data	Test data	16S	Decomposition Environment	Maggots	TBS
3 Outdoor Facilities	3 outdoor facility	Indoor & Outdoor	x			
Indoor & Outdoor (this study)	Indoor & Outdoor	Indoor & Outdoor	x			
Indoor & Outdoor & Env	Indoor & Outdoor	Indoor & Outdoor	x	x	x	
Indoor & Outdoor & Env & TBS	Indoor & Outdoor	Indoor & Outdoor	x	x	x	x
Env & TBS	Indoor & Outdoor	Indoor & Outdoor		x	x	x

Combining microbial data from indoor and outdoor cadavers into one model ("indoor & outdoor") significantly improved indoor prediction errors from 72 ADD to 52.2 ADD (**Figure 5A**). Importantly, the addition of indoor bodies had no significant effect on outdoor predictions (50.3 ADD vs. 49.35 ADD). Including decomposition environment and maggot exposure data (indoor & outdoor & env model) did not further improve predictions (**Supplemental Figure S8**). However, we found when total body scores were also included ("indoor & outdoor & env & TBS" model), PMI MAE significantly decreased to 42.6 ADD (approximately 2.42 days) for both indoor and outdoor bodies.

When only taking MAE into consideration, the “environment & TBS” model appeared to perform as well as the “indoor & outdoor” and “indoor & outdoor & env” models (**Supplemental Figure S8**). However, this model tended to predict the same or very similar PMI values for a large range of ADDs (**Supplemental Figure S6H**). This tendency was most likely due to TBSs plateauing or regressing for several days across multiple donors (**Figure 5D**) which demonstrates the utility and robustness of microbiome-based PMI estimation models. Overall, we found the “indoor & outdoor & env & TBS” model had the highest prediction accuracy across both decomposition environments (indoors vs. outdoors) despite the variation in maggot emergence times.

We found the most important microbial features of the best performing model (“indoor & outdoor & env & TBS”) were similar to those found in the 3 outdoor facilities model. However, the order of importance was altered and *Miniphocaeibacter* became the most important taxa followed by *Thiopseudomonas alkaliphila* and *Ignatzschineria* (**Figure 5C**). The relative abundance of these three microbial taxa appears to follow similar patterns in both decomposition environments (**Supplemental Figure S9**) suggesting some microbes may not rely on maggot colonization as a dispersal mechanism. To test this, we examined the relative abundance of these important features in maggot-free donors and found *Miniphocaeibacter* was absent while *Thiopseudomonas alkaliphila* and *Ignatzschineria* were found in low abundance at later time points (**Supplemental Figure S10A**) further suggesting that blow flies or other insects may be involved in inoculating these cadavers. Interestingly, we found that 9 of the top 18 microbial features were entirely absent or almost undetectable from maggot-free donors. However, the prediction errors of maggot-free donors were below the

average indoor and outdoor MAEs (**Supplemental Figure S10B**) which demonstrates the generalizability of this new RFR PMI prediction model.

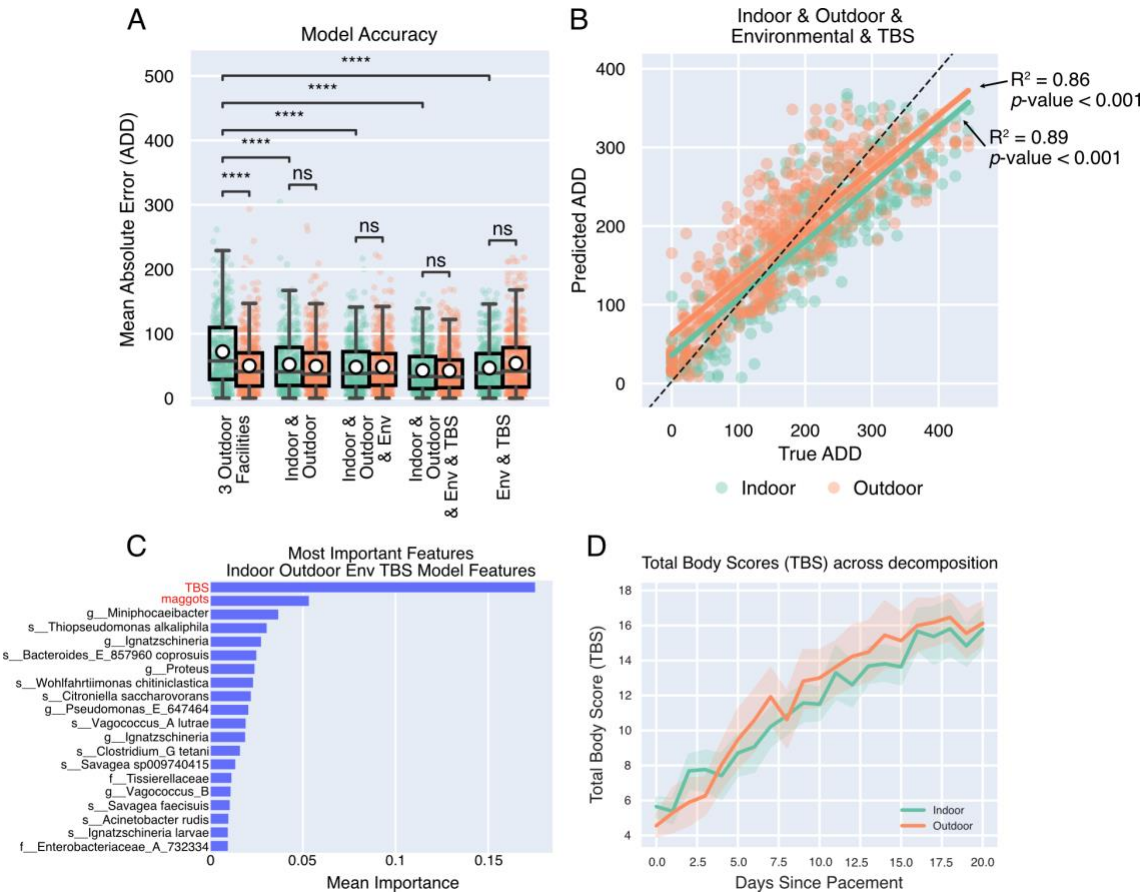


Figure 5. Indoor cadavers and environmental variables improve model accuracy. **A)** Box plot shows cross-validation mean absolute errors for each model that was constructed. Samples were grouped by indoor (teal) and outdoor (orange) cadavers. The model trained on 16S amplicon sequences, environmental variables (maggots, indoor/outdoor), and total body scores (TBS) was the best performing model and had the lowest errors. White markers represent group means. ANOVA with post-hoc Tukey HSD correction for multiple comparisons was used to test for significance. **** = *p*-value <0.0001; ns = not significant. **B)** Scatter plot displaying predicted ADDs compared to true ADDs for indoor (teal) and outdoor (orange) skin samples for the best performing model. Linear regressions for indoor and outdoor groups are shown as solid lines. The dashed black line represents perfect prediction accuracy. **C)** The 20 most important features for the best performing model's accuracy are shown. Environmental variables are colored in red. The average of each feature was taken across all cross-validation folds. **D)** Mean cumulative total body score across the first 21 days of decomposition for indoor (teal) and outdoor (orange) cadavers. Shaded areas represent 95% confidence intervals.

4.5 Discussion

Our results demonstrate that cadaver microbial succession during decomposition is altered by indoor structures due to delayed colonization by blow flies and their associated microbes. This finding is similar to other indoor decomposition studies that also reported delayed necrophagous insect colonization (but did not study their microbes) [19–21]. As has been previously shown by Burcham et al., necrophagous insects, particularly blowflies, appear to be the main dispersal mechanism of most key decomposer microbes involved in decomposing animal flesh, which in outdoor scenarios occurs almost immediately upon death. We confirm here that delays in blowfly maggot colonization on indoor cadavers has large implications for microbial-based PMI estimation models. However, we find that PMI prediction errors can be overcome by including indoor cadavers in PMI models and several additional environmental variables, including whether the cadaver was decomposing inside or outside, whether there was evidence of maggot colonization, as well as the total body score (**Figure 5**). Interestingly, the presence of maggots was a much more important variable for prediction than whether a body was decomposing indoors or outdoors. Therefore, a generalized PMI model for cadavers, regardless of whether it was discovered in an indoor structure, may be possible if insect data are also included.

Although several indoor and outdoor donors were never colonized by blow fly maggots, PMI was still accurately predicted for these individuals. Burcham et al. demonstrated that the most important decomposers for PMI estimation exhibited similar trends across decomposition locations and across multiple climates enabling a generalizable PMI model. Here, we show that even when some of the most important

model features are missing from a subset of samples (i.e. maggot-free donors; **Supplemental Figure S10**), RFR models are still able to accurately predict PMI by leveraging other important features. For example, these maggot-free donors exhibited a meat spoilage-like microbiome which is often characterized by a low-diversity microbial community dominated by *Pseudomonas* [35–37]. In a study investigating chicken carcass chilling methods, Belk et al. noted that water chilling decreased bacterial load and hypothesized that this reduced microbial competition which allowed for the overgrowth of *Pseudomonas*. We suspect a similar ecological process occurs during human decomposition when maggot-associated microbes fail to inoculate human remains and out-compete *Pseudomonas* species which are likely native to the corpse. This overgrowth was observed in most indoor and outdoor maggot-free cadavers and likely drives PMI estimates of these cadavers, enabling a general PMI model.

Consolidating multiple large decomposition datasets into a single model that is able to accurately estimate PMI across various geographical locations, seasons, decomposition environments, and insect activity levels within a forensically useful timeframe is a large challenge. Whether one large, mixed decomposition scenario model is feasible or whether multiple smaller, decomposition scenario-specific models are needed is currently unclear and additional research is needed to inform best practices. In this current study, our best performing model was composed of 44% indoor donors and we found this percentage did not negatively impact outdoor predictions while enabling accurate indoor predictions. These results imply a large model with a mixture of microbial and insect data from indoor and outdoor cadavers may be possible to construct. We suspect this type of model may even improve the prediction of

cadavers decomposing during colder months when insect activity is absent. However, it is currently unclear whether incorporating data from different outdoor geographical locations will decrease indoor performance or if additional indoor cadavers will be needed to maintain a near 1:1 ratio. Furthermore, the enclosed structures used in the study were unconditioned and closely followed daily temperature fluctuations. We hypothesize that indoor decomposition patterns may be further altered in conditioned spaces with stable temperatures. This outcome may necessitate the creation of multiple, decomposition-specific models to maximize model performance across disparate decomposition scenarios.

Our results also bring up several interesting questions about the relationship between necrophagous insects and their associated microbes. For example, do these microbes provide fitness advantages to the insects that host them? Are some key decomposers only found on particular species of insects such as blow flies, carrion beetles, or centipedes? Furthermore, do insect-associated microbial communities from different insect species compete for decomposition resources or do they cooperatively break down the nutrient pool? Future research is needed to unravel these relationships and answer these questions which could additionally help refine PMI models as our understanding of the interdependency between insects and decomposer microbes increases.

CHAPTER 4: REFERENCES

1. Megyesi MS, Nawrocki SP, Haskell NH. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. J Forensic Sci. 2005;50: 618–626. Available: https://www.ncbi.nlm.nih.gov/pubmed/15932096

2. Catts EP, Goff ML. Forensic entomology in criminal investigations. Annu Rev Entomol. 1992;37: 253–272. doi:10.1146/annurev.en.37.010192.001345

3. Amendt J, Campobasso CP, Gaudry E, Reiter C, LeBlanc HN, Hall MJR, et al. Best practice in forensic entomology--standards and guidelines. Int J Legal Med. 2007;121: 90–104. doi:10.1007/s00414-006-0086-x

4. Hewadikaram KA, Goff ML. Effect of carcass size on rate of decomposition and arthropod succession patterns. Am J Forensic Med Pathol. 1991;12: 235–240. doi:10.1097/00000433-199109000-00013

5. Byard RW. Estimation of the time since death in the early postmortem period (24–48 hours). Estimation of the Time since Death. Elsevier; 2020. pp. 11–27. doi:10.1016/b978-0-12-815731-2.00002-9

6. Carter DO, Yellowlees D, Tibbett M. Moisture can be the dominant environmental parameter governing cadaver decomposition in soil. Forensic Sci Int. 2010;200: 60–66. doi:10.1016/j.forsciint.2010.03.031

7. Mann RW, Bass WM, Meadows L. Time since death and decomposition of the human body: variables and observations in case and experimental field studies. J Forensic Sci. 1990;35: 103–111. doi:10.1520/jfs12806j

8. Lee Goff M. Early post-mortem changes and stages of decomposition in exposed cadavers. Exp Appl Acarol. 2009;49: 21–36. doi:10.1007/s10493-009-9284-9

9. Metcalf JL, Xu ZZ, Weiss S, Lax S, Van Treuren W, Hyde ER, et al. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science. 2016;351: 158–162. doi:10.1126/science.aad2646

10. Burcham ZM, Belk AD, McGivern BB, Bouslimani A, Ghadermazi P, Martino C, et al. A conserved interdomain microbial network underpins cadaver decomposition despite environmental variables. Nat Microbiol. 2024;9: 595–613. doi:10.1038/s41564-023-01580-y

11. Mason AR, McKee-Zech HS, Steadman DW, DeBruyn JM. Environmental predictors impact microbial-based postmortem interval (PMI) estimation models within human decomposition soils. PLoS One. 2024;19: e0311906. doi:10.1371/journal.pone.0311906

12. Belk A, Xu ZZ, Carter DO, Lynne A, Bucheli S, Knight R, et al. Microbiome Data Accurately Predicts the Postmortem Interval Using Random Forest Regression Models. Genes . 2018;9. doi:10.3390/genes9020104

13. Deel H, Emmons AL, Kiely J, Damann FE, Carter DO, Lynne A, et al. A Pilot Study of Microbial Succession in Human Rib Skeletal Remains during Terrestrial Decomposition. mSphere. 2021;6: e0045521. doi:10.1128/mSphere.00455-21

14. Pechal JL, Crippen TL, Benbow ME, Tarone AM, Dowd S, Tomberlin JK. The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. Int J Legal Med. 2014;128: 193–205. doi:10.1007/s00414-013-0872-1

15. Adserias-Garriga J, Hernández M, Quijada NM, Rodríguez Lázaro D, Steadman D, Garcia-Gil J. Daily thanatomicrobiome changes in soil as an approach of postmortem interval estimation: An ecological perspective. Forensic Sci Int. 2017;278: 388–395. doi:10.1016/j.forsciint.2017.07.017

16. Metcalf JL, Wegener Parfrey L, Gonzalez A, Lauber CL, Knights D, Ackermann G, et al. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. Elife. 2013;2: e01104. doi:10.7554/eLife.01104

17. Cobaugh KL, Schaeffer SM, DeBruyn JM. Functional and Structural Succession of Soil Microbial Communities below Decomposing Human Cadavers. PLoS One. 2015;10: e0130201. doi:10.1371/journal.pone.0130201

18. Mason AR, Taylor LS, DeBruyn JM. Microbial ecology of vertebrate decomposition in terrestrial ecosystems. FEMS Microbiol Ecol. 2023;99. doi:10.1093/femsec/fiad006

19. Anderson GS. Comparison of decomposition rates and faunal colonization of carrion in indoor and outdoor environments: Decomposition rates and faunal colonization of carrion. J Forensic Sci. 2011;56: 136–142. doi:10.1111/j.1556-4029.2010.01539.x

20. Thümmel L, Lutz L, Geissenberger J, Pittner S, Heimer J, Amendt J. Decomposition and insect succession of pig cadavers in tents versus outdoors - A preliminary study. Forensic Sci Int. 2023;346: 111640. doi:10.1016/j.forsciint.2023.111640

21. Ceciliason A-S, Andersson MG, Lindström A, Sandler H. Quantifying human decomposition in an indoor setting and implications for postmortem interval estimation. Forensic Sci Int. 2018;283: 180–189. doi:10.1016/j.forsciint.2017.12.026

22. McClintock WR Jr, Castille JJ, Stewart M, Andrew LE. Soil Survey of Walker County, Texas. 1979. Available: http://hdl.handle.net/2346/47819

23. Berg-Lyons D, Lauber CL, Humphrey G, Thompson L, Gilbert JA, Jansson JK, et al. EMP DNA Extraction Protocol. 2018 [cited 15 Jan 2025]. Available: https://www.protocols.io/view/emp-dna-extraction-protocol-nutdewn.pdf

24. Greg Caporaso J, Ackermann G, Apprill A, Bauer M, Berg-Lyons D, Betley J, et al. EMP 16S Illumina Amplicon Protocol. 2018 [cited 15 Jan 2025]. Available: https://www.protocols.io/view/emp-16s-illumina-amplicon-protocol-nuudeww.pdf

25. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2016;18: 1403–1414. doi:10.1111/1462-2920.13023

26. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75: 129–137. doi:10.3354/ame01753

27. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37: 852–857. doi:10.1038/s41587-019-0209-9

28. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13: 581–583. doi:10.1038/nmeth.3869

29. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6: 90. doi:10.1186/s40168-018-0470-z

30. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, et al. Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol. 2024;42: 715–718. doi:10.1038/s41587-023-01845-1

31. Mirarab S, Nguyen N, Warnow T. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput. 2012; 247–258. doi:10.1142/9789814366496_0024

32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12: 2825–2830. Available: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

33. Deel HL, Montoya S, King K, Emmons AL, Huhn C, Lynne AM, et al. The microbiome of fly organs and fly-human microbial transfer during decomposition. Forensic Sci Int. 2022;340: 111425. doi:10.1016/j.forsciint.2022.111425

34. Shukla SP, Plata C, Reichelt M, Steiger S, Heckel DG, Kaltenpoth M, et al. Microbiome-assisted carrion preservation aids larval development in a burying beetle. Proc Natl Acad Sci U S A. 2018;115: 11274–11279. doi:10.1073/pnas.1812808115

35. Dorn-In S, Mang S, Cosentino RO, Schwaiger K. Changes in the Microbiota from fresh to spoiled meat, determined by culture and 16S rRNA analysis. J Food Prot. 2024;87: 100212. doi:10.1016/j.jfp.2023.100212

36. Barcenilla C, Cobo-Díaz JF, Puente A, Valentino V, De Filippis F, Ercolini D, et al. In-depth characterization of food and environmental microbiomes across different

meat processing plants. Microbiome. 2024;12: 199. doi:10.1186/s40168-024-01856-3

37. Belk AD, Duarte T, Quinn C, Coil DA, Belk KE, Eisen JA, et al. Air versus water chilling of chicken: A pilot study of quality, shelf-life, microbial ecology, and economics. mSystems. 2021;6. doi:10.1128/mSystems.00912-20

CHAPTER 5: CONCLUSION AND FUTURE DIRECTIONS

My research spans multiple niches within the broader microbiome field where numerous knowledge gaps exist. While this dissertation aimed to address some of those gaps, many questions remain but I believe the path forward is bright and promising. Here, I highlight future research directions based on my findings that may help advance our understanding of PMP disease and develop better forensic tools.

Low-biomass human microbiome research is technically challenging. Environmental microbes can contaminate samples at every point along the processing pipeline. Furthermore, human-associated contamination is often indistinguishable from real biological signals. In Chapter 2, I developed a cross-laboratory replication pipeline that helped distinguish laboratory-specific contamination from tumor-associated microbial signatures. Using this approach, I showed that a subset of PMP tumors were colonized by microbes that likely originated from the human gut. However, it is unclear whether these microbes colonized the peritoneal cavity following PMP tumor metastasis or whether these microbes may have traveled with the primary appendiceal tumor. The latter would suggest these PMP-associated microbes may play a significant role in disease development. A study has recently shown there is a strong link between *Fusobacterium nucleatum* colonization and appendiceal cancer development [1]. Given this strong association, I believe future studies should focus on characterizing primary PMP appendiceal tumors and explore the potential role *Fusobacterium nucleatum* may have in PMP development and progression. While the discovery of intratumoral microbial communities is intriguing, I suspect that a significant portion of the PMP tumor

microbiome may be an artifact arising from surgical contamination [2,3]. As this possibility is broadly applicable, I believe a closer look should be taken at surgical contamination signatures before, or in conjunction with, further sequencing-based tumor microbiome studies are conducted.

In Chapter 3, I investigated how soil microbial communities respond to aboveground decomposition. Surprisingly, I discovered that only microbial communities in the uppermost 4 cm (approx. 1.5 inches) of soil adjacent to a cadaver showed a significant response to this influx of nutrients. I believe this finding will impact how grave soil samples will be collected in future forensic and decomposition ecology studies. However, it is unclear whether the depth of these responses is only applicable within the study location (Huntsville, TX) or whether these results are generally applicable to locations with southern sandy loam upland and similar soils [4]. A future study that investigates microbial community responses across different soil types and locations will be needed to fully decipher spatial microbial responses to cadaver decomposition inputs.

Finally, the work in Chapter 4 demonstrates the assembly of the decomposer community is closely tied to the arrival of blow flies and blow fly maggots, a discovery that has large implications for machine learning-based postmortem interval estimation techniques. While this finding represents a step forward for the field of forensic science, much work still needs to be done to determine the exact ecological role insects play in microbial decomposer dispersal. Our understanding of how different insects may contribute to the cadaver microbiome is still limited. For example, beetles are known to visit human remains during advanced decomposition and studies suggest they may

impart their own gut-associated microbes to the corpse [5]. I believe future studies that further investigate insect-cadaver-microbe exchanges and interactions will help advance our understanding of microbial decomposer network assembly. Incorporating this knowledge into PMI models may help improve accuracy across decomposition environments and along with bone and soil-based decomposition models [6], could help extend the timeframe in which actionable PMI estimates are possible.

Although still under development, I anticipate that law enforcement agencies will adopt this microbiome-based forensic tool as a supplementary method for determining the PMI in the future. Few methods are available to investigators for accurately estimating the PMI following the first 24-48 hours [7], in death investigations where insects are entirely absent, or where the interval between time of death and insect colonization time is unclear [8]. In these scenarios, I expect microbiome-based forensic tools will be crucial for providing actionable PMI estimates. Beyond these instances, this tool could be used to further substantiate PMI estimates that were calculated using traditional methods (e.g. rigor mortis). Ultimately, this powerful tool presented in my dissertation, may one day help the families of victims pursue justice.

As the microbiome field advances and new computational tools and approaches are developed, researchers will need to gain additional training to contend with these more advanced workflows. In my experience, short, intensive workshops that focus on the basics of a data processing pipeline (e.g. 16S rRNA sequencing using QIIME2) have been popular and productive ways of training novices. However, in the past, resources outside of these classes have been few and far between making it challenging to properly analyze data without the help of experienced mentors or

colleagues. I believe new AI-based models like ChatGPT which are trained on publicly available code will become exceptionally helpful computational resources for both beginners and experts. As these tools become more refined, more accurate, and more widely adopted, they will increase the rigor and reproducibility of microbiome science overall.

In this dissertation, I presented important contributions that will help advance tumor microbiome science and accelerate the development of powerful PMI estimation tools. I am grateful for the opportunities I have received over the last five years to contribute to these research fields, and I look forward to seeing how the future unfolds.

CHAPTER 5: REFERENCES

1. Park H, Knotts C, Blodgett R, Lewis CR, Bartlett DL, Dadgar N, et al. Fusobacterium nucleatum in appendiceal cancer: Prevalence and influence on the tumor immune microenvironment. Surgical Oncology Insight. 2025;2: 100122. doi:10.1016/j.soi.2025.100122

2. Seidelman JL, Mantyh CR, Anderson DJ. Surgical site infection prevention: A review. JAMA. 2023;329: 244–252. doi:10.1001/jama.2022.24075

3. Long DR, Bryson-Cahn C, Waalkes A, Holmes EA, Penewit K, Tavolaro C, et al. Contribution of the patient microbiome to surgical site infection and antibiotic prophylaxis failure in spine surgery. Sci Transl Med. 2024;16: eadk8222. doi:10.1126/scitranslmed.adk8222

4. McClintock WR Jr, Castille JJ, Stewart M, Andrew LE. Soil Survey of Walker County, Texas. 1979. Available: http://hdl.handle.net/2346/47819

5. Shukla SP, Plata C, Reichelt M, Steiger S, Heckel DG, Kaltenpoth M, et al. Microbiome-assisted carrion preservation aids larval development in a burying beetle. Proc Natl Acad Sci U S A. 2018;115: 11274–11279. doi:10.1073/pnas.1812808115

6. Deel H, Emmons AL, Kiely J, Damann FE, Carter DO, Lynne A, et al. A Pilot Study of Microbial Succession in Human Rib Skeletal Remains during Terrestrial Decomposition. mSphere. 2021;6: e0045521. doi:10.1128/mSphere.00455-21

7. Byard RW. Estimation of the time since death in the early postmortem period (24–48 hours). Estimation of the Time since Death. Elsevier; 2020. pp. 11–27. doi:10.1016/b978-0-12-815731-2.00002-9

8. Amendt J, Campobasso CP, Gaudry E, Reiter C, LeBlanc HN, Hall MJR, et al. Best practice in forensic entomology--standards and guidelines. Int J Legal Med. 2007;121: 90–104. doi:10.1007/s00414-006-0086-x

Appendix A: Chapter 2 Supplemental Figures and Tables

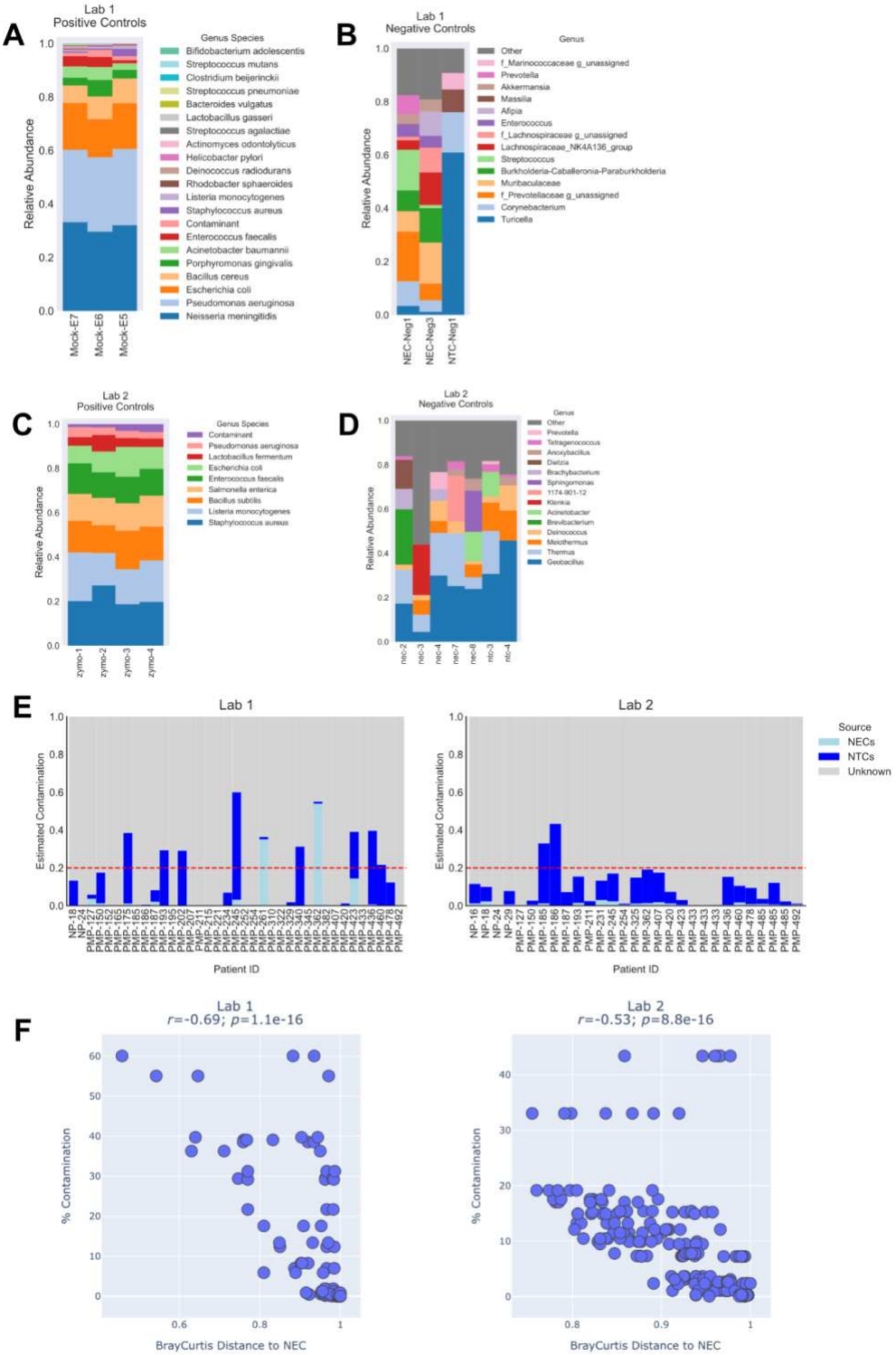


Figure S1. Sample quality assessment. A) Relative abundance of bacterial taxa found in lab 1 positive control dilutions (BEI). Taxa not known to be in the control community were labeled as contaminants. **B)** Taxa barplot of most abundant bacterial taxa found in lab 1 negative controls. ASVs were grouped at genus level. **C**) Relative abundance of bacterial taxa found in lab 2 positive controls (Zymo). Taxa not known to be in the control community were labeled as contaminants. **D)** Taxa barplot of most abundant bacterial taxa found in lab 2 negative controls. ASVs were grouped at genus level. **E**) % estimated negative control contribution from lab 1 (left) and lab 2 (right). Sources represent lab specific NEC and NTC. NEC and NTCs were grouped together to create one representative sample. Samples were rarefied to 800 reads. Unknown represents sequences that could not be attributed to any source sample. Red dashed line represents the quality control cut off of 20%. **F**) Bray-Curtis distances between negative extraction controls (NECs & NTCs) and samples (PMP & NP) was plotted against % contamination estimated by SourceTracker2 for lab 1 (left) and lab 2 (right). All negative controls, PMP, and NP sample extractions with more than 800 reads were included. Distances were calculated separately for each lab. Pearson's r was used to assess correlation between distance and contamination.

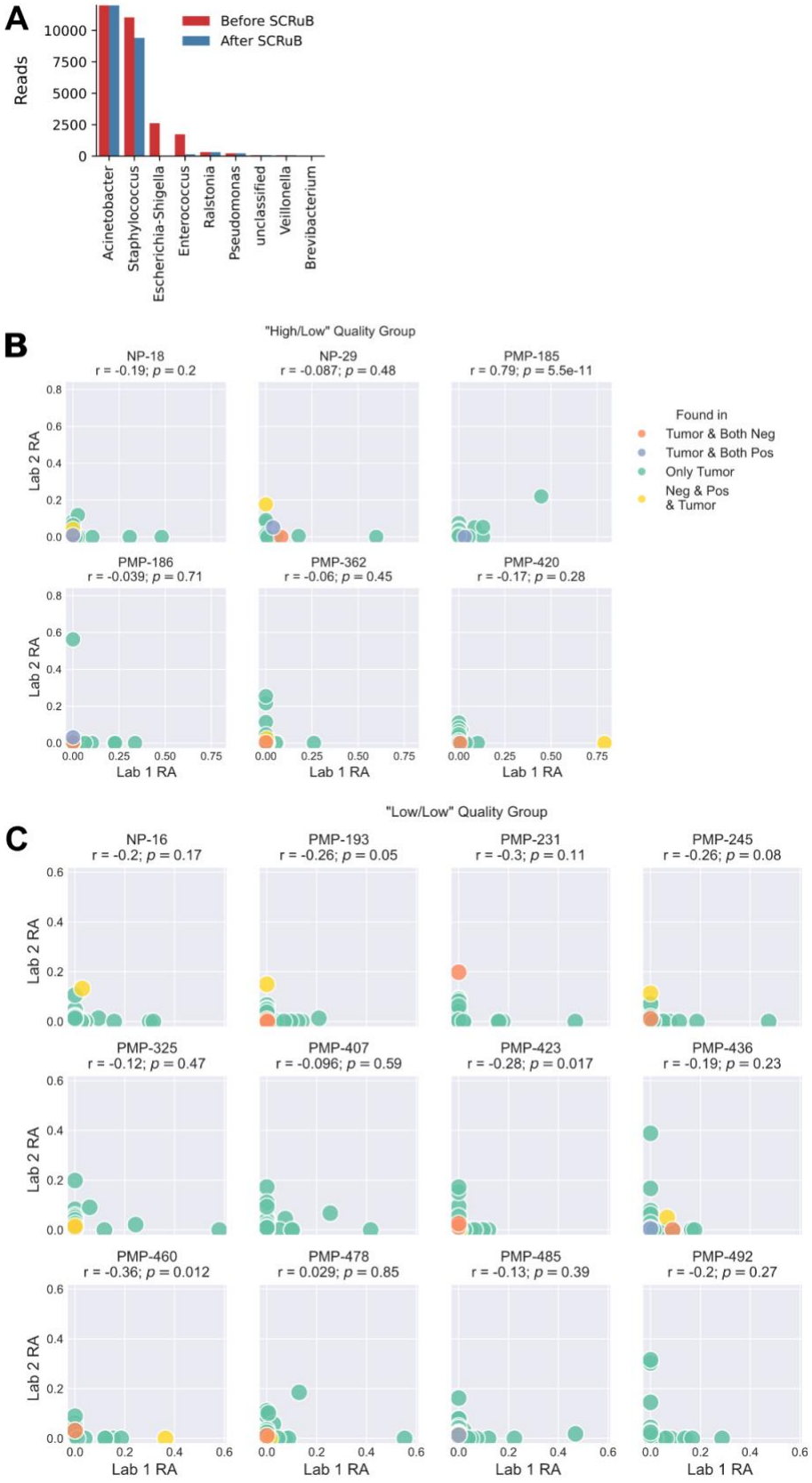

A
Before SCRuB
After SCRuB
Reads
10000
7500
5000
2500
0
Acinetobacter
Staphylococcus
Escherichia-Shigella
Enterococcus
Ralstonia
Pseudomonas
unclassified
Veillonella
Brevibacterium
B
"High/Low" Quality Group
NP-18
r = -0.19; p = 0.2
NP-29
r = -0.087; p = 0.48
PMP-185
r = 0.79; p = 5.5e-11
PMP-186
r = -0.039; p = 0.71
PMP-362
r = -0.06; p = 0.45
PMP-420
r = -0.17; p = 0.28
Lab 2 RA
Lab 1 RA
Found in
Tumor & Both Neg
Tumor & Both Pos
Only Tumor
Neg & Pos & Tumor
C
"Low/Low" Quality Group
NP-16
r = -0.2; p = 0.17
PMP-193
r = -0.26; p = 0.05
PMP-231
r = -0.3; p = 0.11
PMP-245
r = -0.26; p = 0.08
PMP-325
r = -0.12; p = 0.47
PMP-407
r = -0.096; p = 0.59
PMP-423
r = -0.28; p = 0.017
PMP-436
r = -0.19; p = 0.23
PMP-460
r = -0.36; p = 0.012
PMP-478
r = 0.029; p = 0.85
PMP-485
r = -0.13; p = 0.39
PMP-492
r = -0.2; p = 0.27
Lab 2 RA
Lab 1 RA

Figure S2. “High/Low” and “Low/Low” quality groups have low reproducibility. **A**) Number of total reads for each global contaminant ASV before (red) and after (blue) running the SCRuB pipeline. Reads were summed across all PMP and NP samples. All global contaminants are shown. Only *Escherichia-Shigella*, *Enterococcus*, and *Brevibacterium* had a >50% reduction in reads and were removed from dataset. **B**) Relative abundance (RA) of ASVs found in High/Low quality replicates from labs 1 and 2 and **C**) “Low/Low’ quality group. Each point represents an ASV. ASVs that were also found in negative (orange), positive controls (blue), or both (yellow) are highlighted. Pearson correlation (r) was used to assess reproducibility of microbial communities.

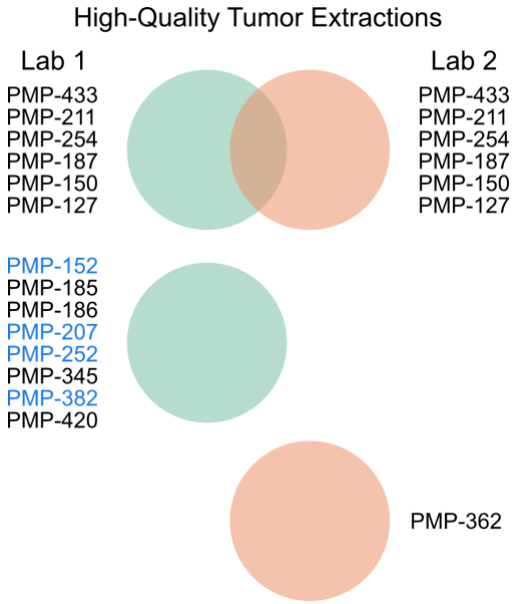


Figure S3. High-quality tumor replicates. Diagram shows all high-quality tumors and where they were extracted. Venn diagrams depict whether each tumor was paired (top) or unpaired (middle & bottom) with another high-quality replicate. Samples highlighted in blue were only extracted in lab 1.

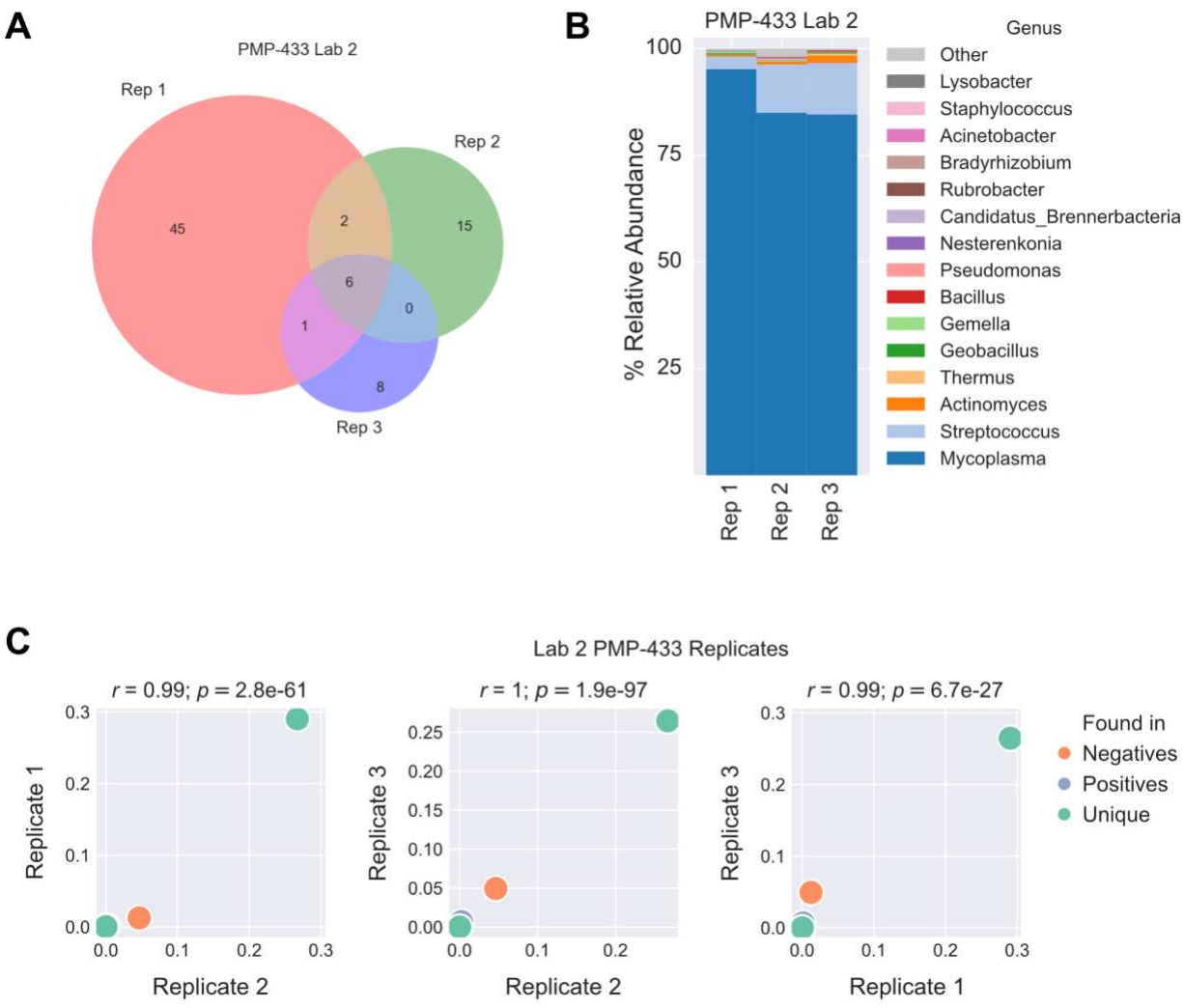


Figure S4. PMP-433 tumor microbiome is reproducible within lab 2. A) Venn diagram showing number of overlapping ASVs between all 3 PMP-433 replicates extracted and sequenced at lab 2. **B**) Taxabar plot showing each PMP-433 replicate (rep) extracted at lab 2. Lab-specific and global contaminants have not been removed (e.g. *Geobacillius* and *Escherichia-Shigella*). **C)** Scatter plots showing relative abundance of each ASV found in Replicate 1 vs Replicate 2 vs Replicate 3.

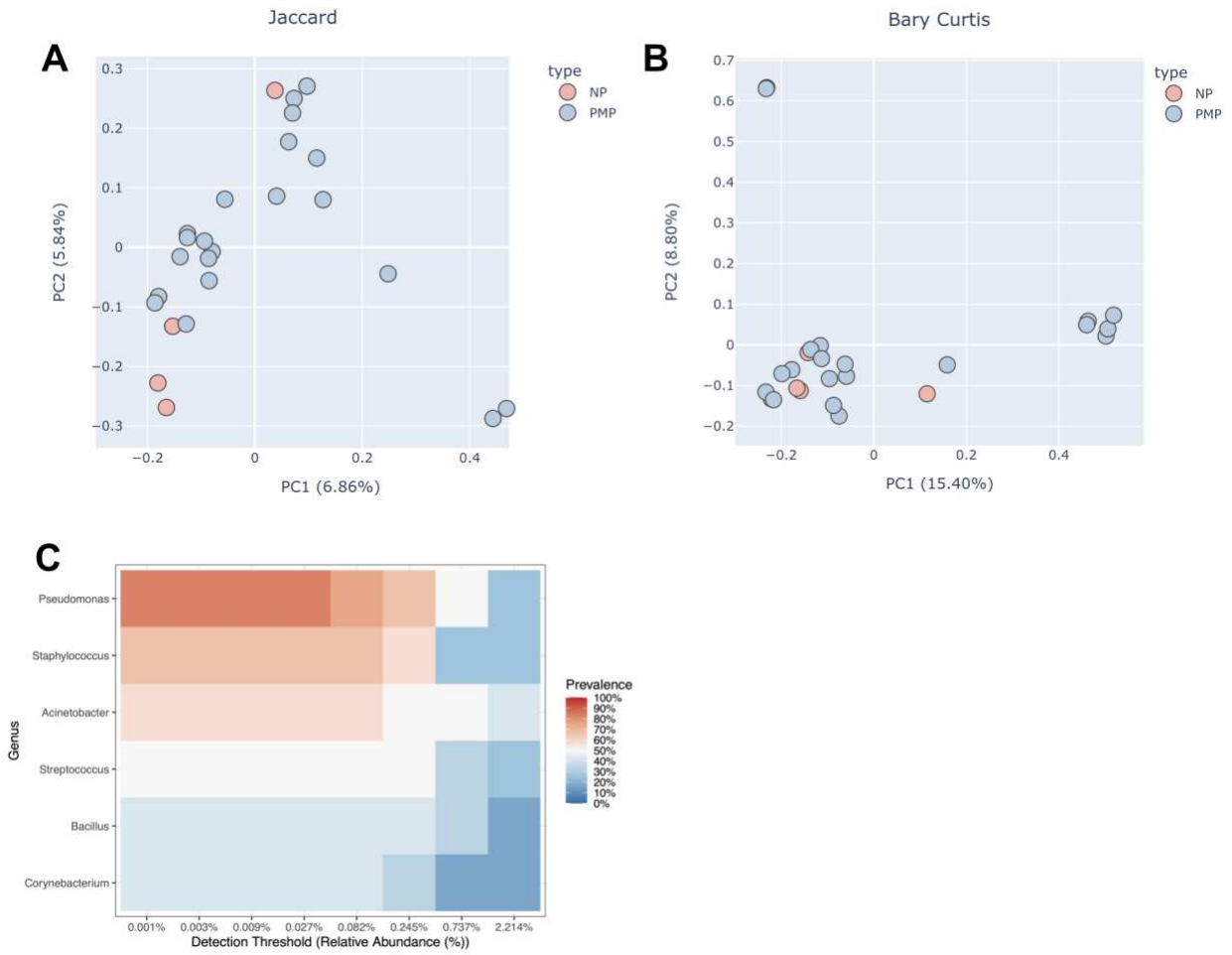
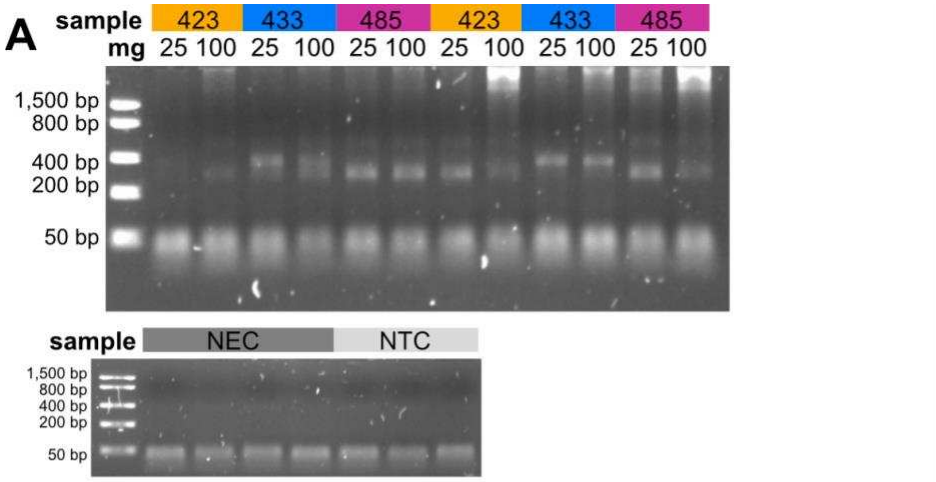
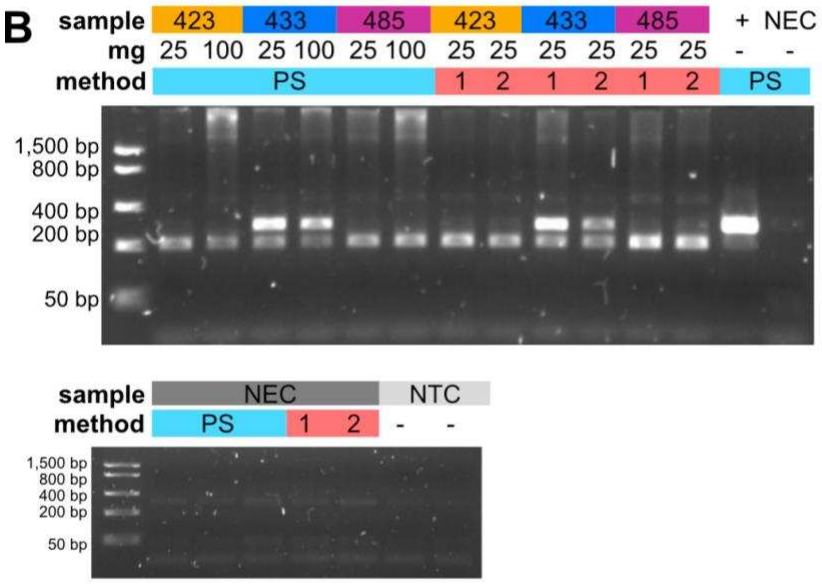


Figure S5. The PMP tumor microbiome differs from NP control tissues. A) Jaccard PCoA plot showing high-quality PMP tumors (blue) and NP control tissues (pink). PERMANOVA *p*=0.031. **B**) Bray-Curtis PCoA plot showing high-quality PMP tumors (blue) and NP control tissues (pink). PERMANOVA *p*=0.043. **C**) Heatmap showing core microbiome genera present across PMP tumor replicates in the “High/High” quality group (n=12 extractions). ASVs were grouped at genus level.


A
sample
423 433 485 423 433 485
mg
25 100 25 100 25 100 25 100 25 100 25 100
1,500 bp
800 bp
400 bp
200 bp
50 bp
sample
NEC
NTC
1,500 bp
800 bp
400 bp
200 bp
50 bp


B
sample
423 433 485 423 433 485 + NEC
mg
25 100 25 100 25 100 25 25 25 25 25 25 - -
method
PS 1 2 1 2 1 2 PS
1,500 bp
800 bp
400 bp
200 bp
50 bp
sample
NEC NTC
method
PS 1 2 - -
1,500 bp
800 bp
400 bp
200 bp
50 bp

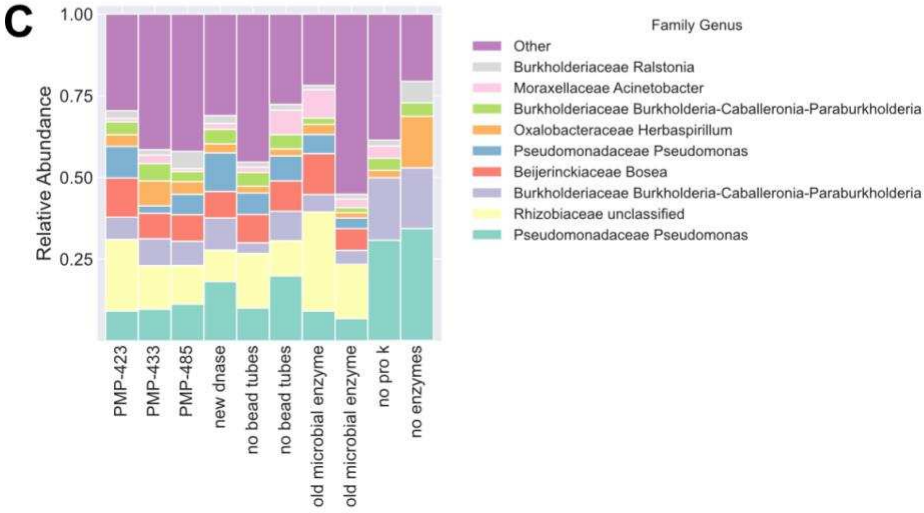

C
1.00
0.75
0.50
0.25
Relative Abundance
PMP-423
PMP-433
PMP-485
new dnase
no bead tubes
no bead tubes
old microbial enzyme
old microbial enzyme
no pro k
no enzymes
Family Genus
Other
Burkholderiaceae Ralstonia
Moraxellaceae Acinetobacter
Burkholderiaceae Burkholderia-Caballeronia-Paraburkholderia
Oxalobacteraceae Herbaspirillum
Pseudomonadaceae Pseudomonas
Beijerinckiaceae Bosea
Burkholderiaceae Burkholderia-Caballeronia-Paraburkholderia
Rhizobiaceae unclassified
Pseudomonadaceae Pseudomonas

Figure S6. Optimization Experiments. A) Agarose gel depicts PCR amplicons from DNA input experiment (see methods). Total DNA from 25 or 100 mg of PMP tumor 423, 433, and 485 was extracted. Gel shows generation of off-target amplicons in addition to desired product (~400 bp). **B**) Agarose gel depicts PCR amplicons from host DNA depletion experiment (see methods). Method PS refers to total DNA extraction following the PowerSoil Kit. Method 1 refers to the "host" fraction containing the depleted host DNA. Method 2 refers to the "bacterial" fraction containing host depleted bacterial DNA. Zymo Microbial Community was used as "+" control. **C**) Taxa plot depicting unclassified *Rhizobiaceae* ASV (light yellow) was found contaminating Zymo HostZERO ProK solution. Removing ProK resulted in lost detection of *Rhizobiaceae* ASV ("no pro k" and "no enzymes").

See AppendixA.xlsx for supplemental tables.

Table S1 PMP Patient Metadata. Metadata associated with each PMP patient that was extracted in this study.

Table S2 NP Patient Metadata. Metadata associated with normal peritoneal control tissue samples that passed quality control filtering metrics and were considered "high-quality".

Table S3 Sample Quality. Table includes information about each tumor, normal control tissue, negative, and positive control sample. Reads-filtered indicates how many reads remained after removal of non-bacterial reads. NaN indicates no sequencing reads were obtained the extraction or reads were too low to be included in contamination source tracking analysis.

Table S4 SourceTracker2 contamination. Background contamination was estimated using SourceTracker2 for each lab separately. Table S4 shows how many reads SourceTracker2 attributed to each ASV found in each lab's sink (NECs + NTCs). For each lab, reads were summed across all PMP and normal control tissue samples (counts columns). The number of PMP or NP extractions (>800 reads) each ASV was found in is shown in columns 6 & 7.

Table S5 Negative Control ASVs. All ASVs found negative controls (NECs & NTCs), their total reads, and prevalence in each lab are shown. Only negative controls with > 100 reads were included. ASVs that were found in both labs were considered to be global contaminants. ASVspresent >50% of negative control samples were removed from the dataset during the lab-specific contaminants filtering step.

Table S6 Quality Groups. PMP and normal peritoneal control tissues samples extracted at both labs were assigned quality groups based on individual replicate status. Tumors in "High/High" group had 2 high-quality replicates. Tumors in "High/Low" group had 1 high-quality and 1 low-quality replicate. Tumors in "Low/Low" group had no high-quality extractions.

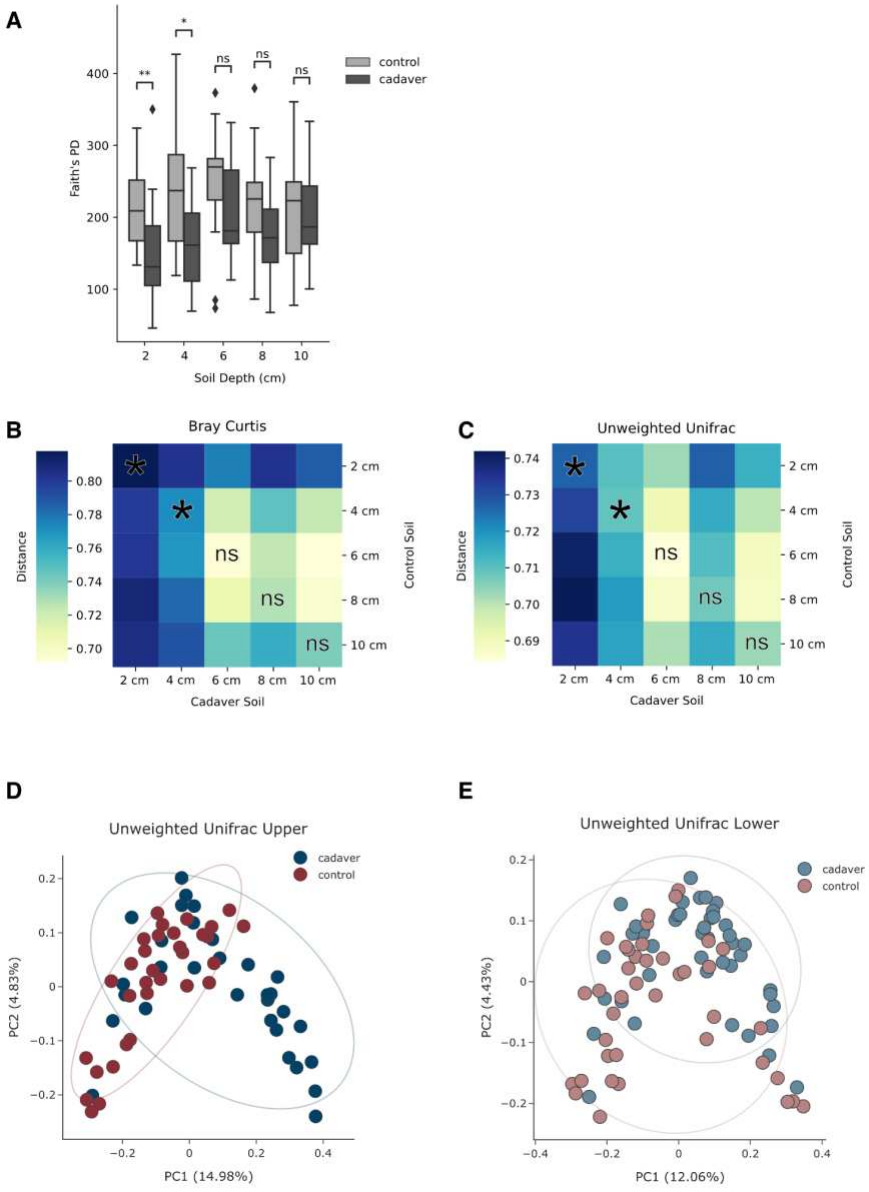
Table S7 Reproducible ASVs. ASVs found in both tumor replicates and normal control tissue replicates are listed. ASV ID and classifications are provided. ASVs that were also found across lab 1 and 2 NECs and NTCs are denoted with *. To retain as many low-quality replicates as possible, feature table was not rarefied.

Table S8 PMP Tissue Input Optimization Experiment. Nanodrop DNA yields and purity of PMP samples that were used for tissue input optimization experiment. Extractions were not used for analysis. Sample IDs beginning with N are negative extraction controls.

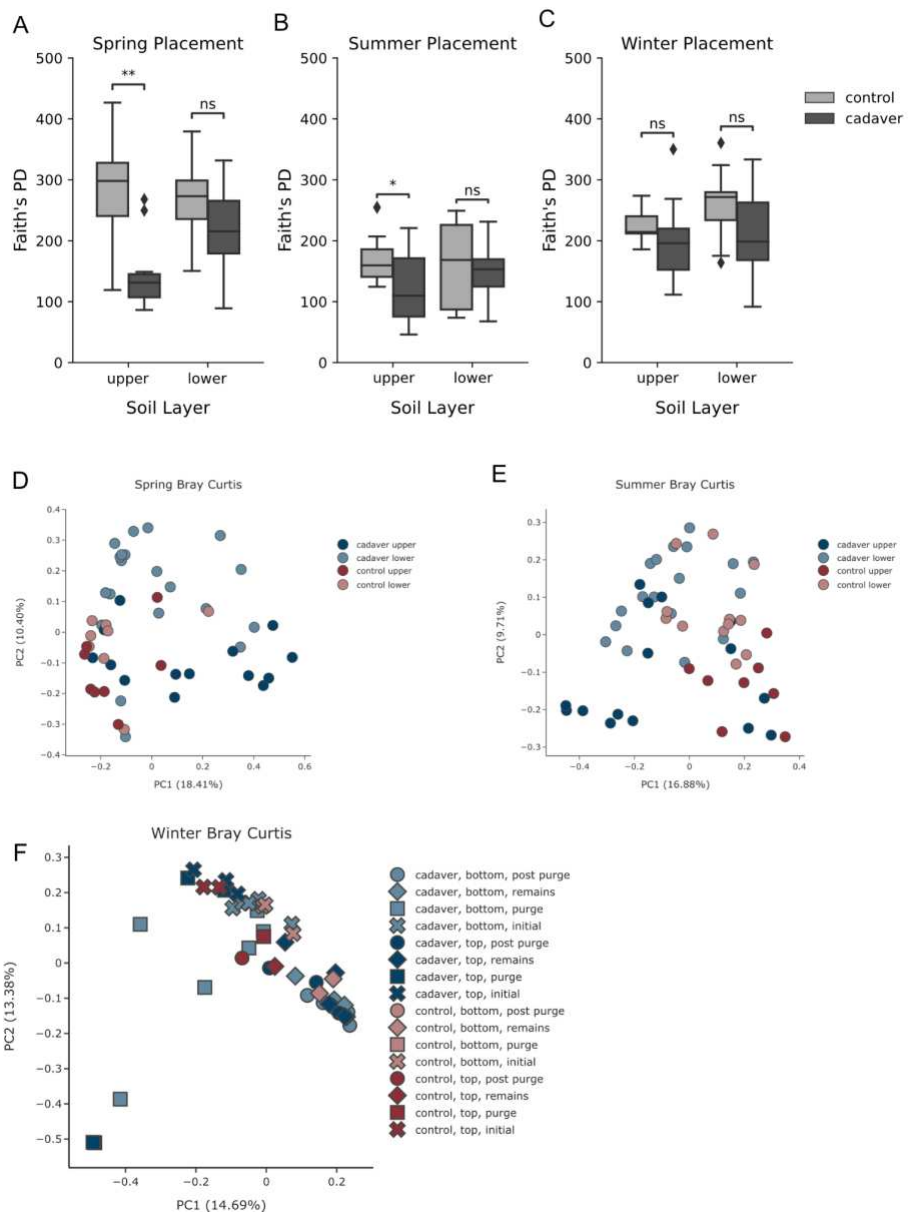
Table S9 Two-step Host DNA Depletion Experiment. Nanodrop DNA yields and purity of PMP tumors used for optimization experiments. Each PMP tumor yielded 2 extractions from same starting material: one host depletion fraction and one bacterial fraction. See methods for description of each fraction. Extractions were used for within lab analysis.

Table S10 Final Lab 2 Extractions. DNA yield and purity of all PMP and NP samples extracted at lab 2. Extractions were used as input for constructing final lab 2 16S rRNA libraries.

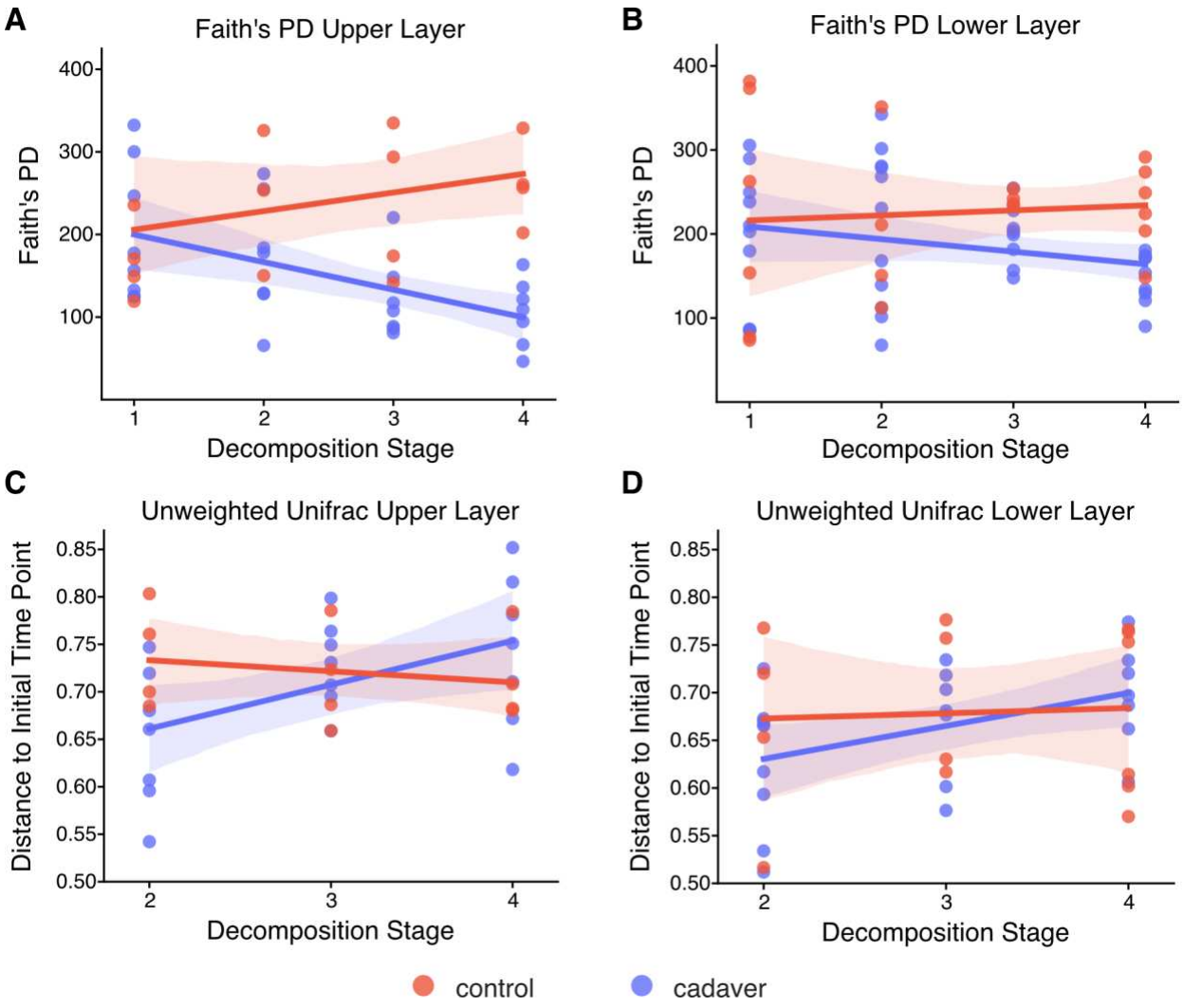
Appendix B: Chapter 3 Supplemental Figures and Tables



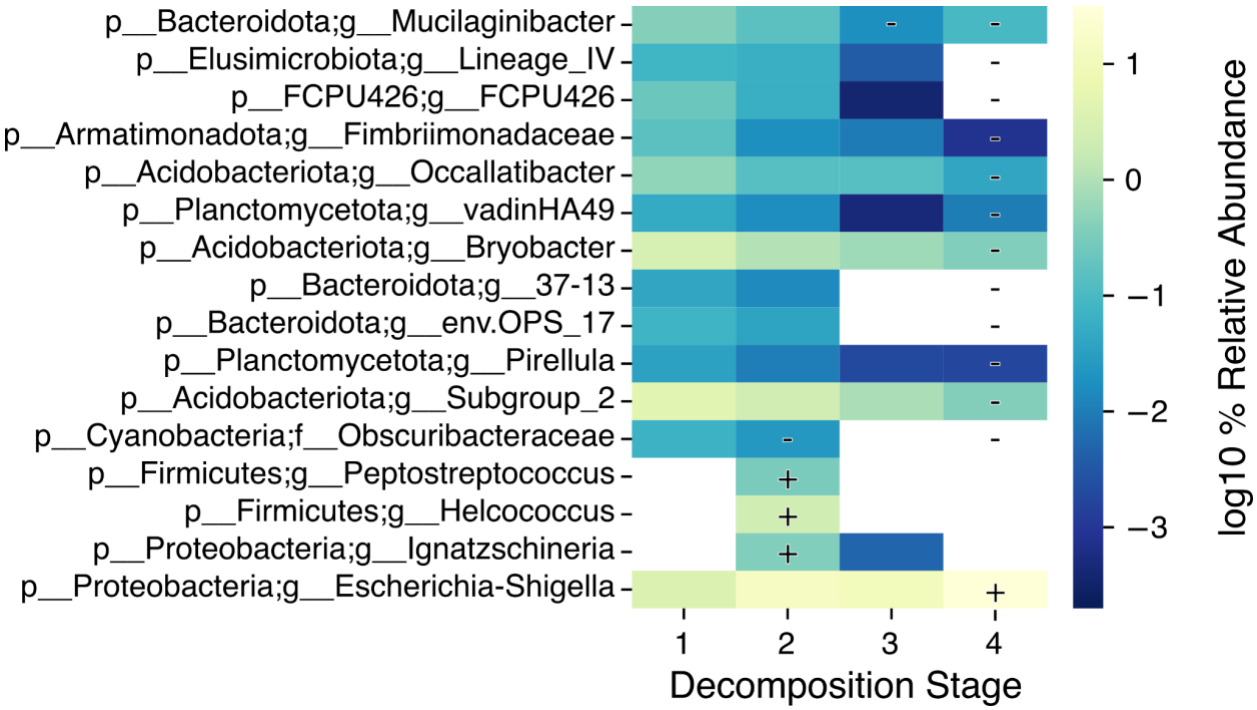
Supplementary Figure 1. A) Faith's phylogenetic diversity of control soils (light gray) compared to cadaver soil (dark gray) at each depth. A Kruskal-Wallis test was used to test for significance between soil depths (**: *p-value* < 0.01; *: *p-value* < 0.05; ns: not significant). **B**) Bray-Curtis distances between cadaver (x-axis) and control soils (y-axis) at each depth. PERMANOVA, *: *q-value* < 0.05; ns: not significant. **C)** Unweighted unifrac distances between cadaver (x-axis) and control soils (y-axis) at each depth. PERMANOVA, *: *q-value* < 0.05; ns: not significant. **D**) PCoA plot of unweighted unifrac distances between cadaver (blues) and control (reds) soils in the upper soil layer (PERMANOVA *p-value* = 0.001) and **E**) lower soil layer (PERMANOVA *p-value* = 0.005). Ellipses represent 95% confidence intervals for cadaver and control groups.



Supplementary Figure 2. Cadaver placement season impacts belowground microbial communities' responses to decomposition. A-C) Faith's phylogenetic diversity of control soils (light gray) compared to cadaver soil (dark gray) from the same layer across cadaver placement seasons. A Kruskal-Wallis test was used to test for significance between soil treatment (**: *p*-value < 0.01; *: *p*-value < 0.05; ns: not significant). **D-E**) PCoA plot of Bray-Curtis distances between cadaver (blues) and control (reds) soils in the upper (dark) and lower (light) soil layers (PERMANOVA *q*-values < 0.05) during spring and summer placement seasons. **F**) PCoA plot of Bray-Curtis distances between cadaver (blues) and control (reds) soils in the upper (dark) and lower (light) soil layers during winter placement (PERMANOVA *q*-values < 0.05). Shapes represent the decomposition stage (circle = post purge; diamond = remains; square = purge; cross = initial).



Supplemental Figure 3. Below ground soil communities become increasingly different as decomposition progresses. Linear mixed effects (LME) models showing Faith's phylogenetic diversity for control (red) and cadaver (blue) soils across all decomposition stages (1= initial placement, 2 = purge, 3 = post purge, 4 = dry remains) in the upper (**A**) and lower (**B**) soil layers. LME models reveal unweighted unifrac distance to initial placement time point increases as decomposition progresses in upper (**C**) and (**D**) lower soil layers. Winter placement soils were excluded from all LME models.



Supplemental Figure 4. MaAsLin2 was used to identify microbial taxa that significantly changed over the course of decomposition in the lower soil layer. Heatmap colors correspond with average log10 transformed percent relative abundance. White boxes signify taxa was not detected at that decomposition stage. Only boxes with *q-value* < 0.05 are labeled. + annotation denotes taxa increased in percent relative abundance; - annotation denotes taxa decreased in percent relative abundance as compared to initial time point. ASVs were collapsed at genus level. Phylum along with lowest level taxonomic assignment are shown. Decomposition stage 1=initial, 2=purge, 3=post, 4=dry. Winter placement soils were excluded from analysis.

See AppendixB.xlsx for supplemental tables.

Table S1. Differentially abundant taxa in upper layer of grave soils (ANCOM-BC). Only spring and summer seasons are included.

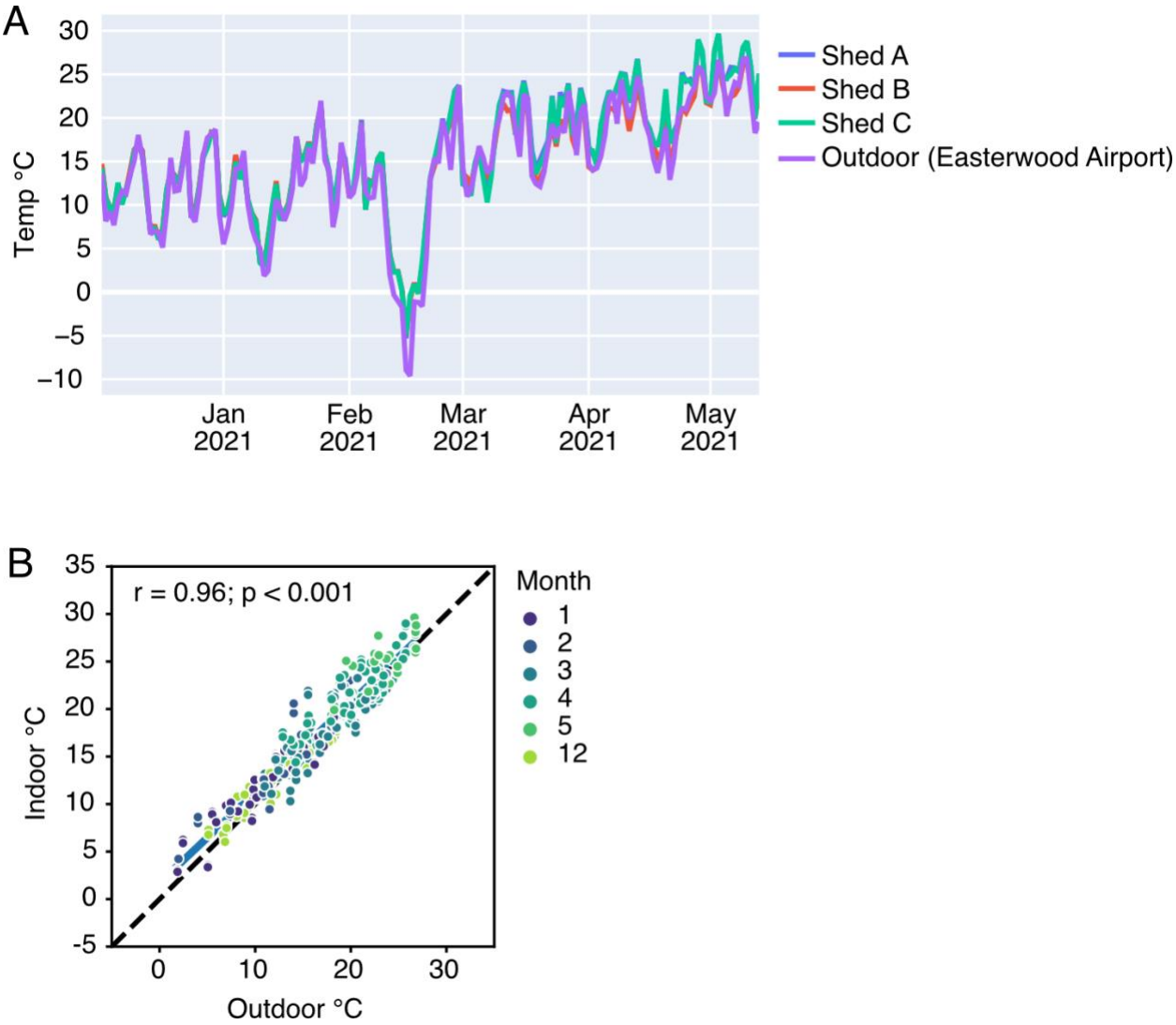
Table S2. Differentially abundant taxa in lower layer of grave soils (ANCOM-BC). Only spring and summer seasons are included.

Table S3. Differentially abundant taxa in upper layer of control soils (ANCOM-BC). Only spring and summer seasons are included.

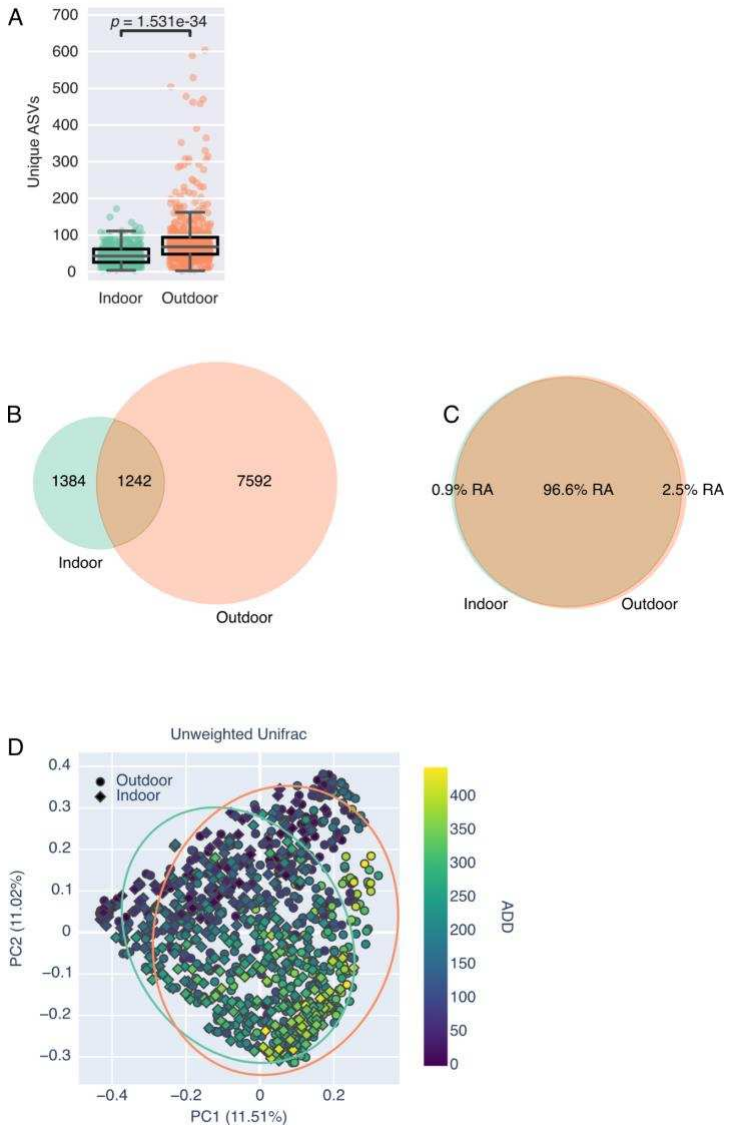
Table S4. Taxa associated with each decomposition stage in the upper grave soil layer (MaAsLin2). Only spring and summer seasons are included.

Table S5. Taxa associated with each decomposition stage in the lower grave soil layer (MaAsLin2). Only spring and summer seasons are included.

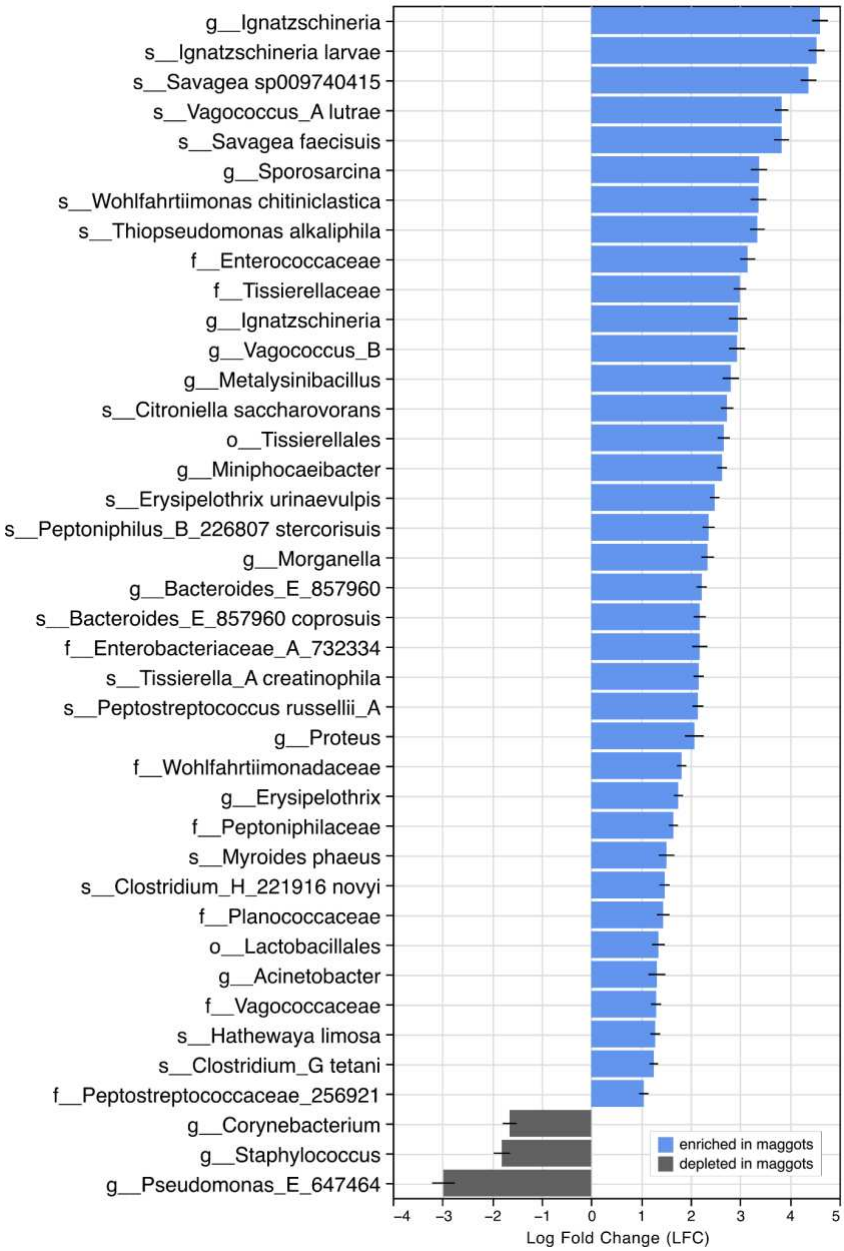
Appendix C: Chapter 4 Supplemental Figures



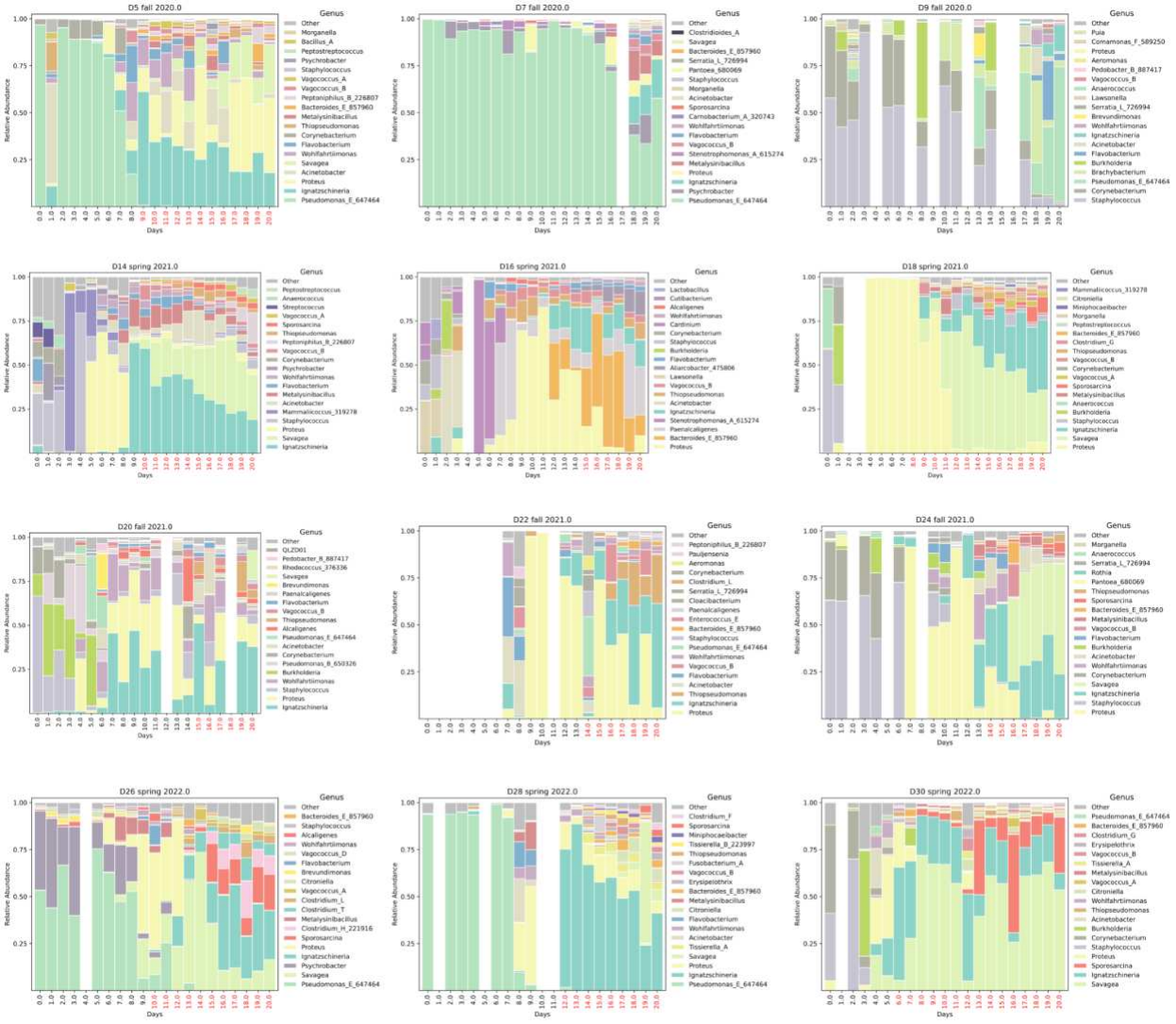
Supplemental Figure S1. A) Average daily temperatures between December 02, 2020 and May 13, 2021. Temperatures were collected from inside each individual structure (shed) and compared to outdoor temperatures. Outdoor temperatures were collected from Easterwood Airport located in College Station, Texas. **B)** Indoor temperatures were highly correlated with outdoor temperature. Pearson correlation was used to test for statistical significance. Dashed line represents perfect correlation.



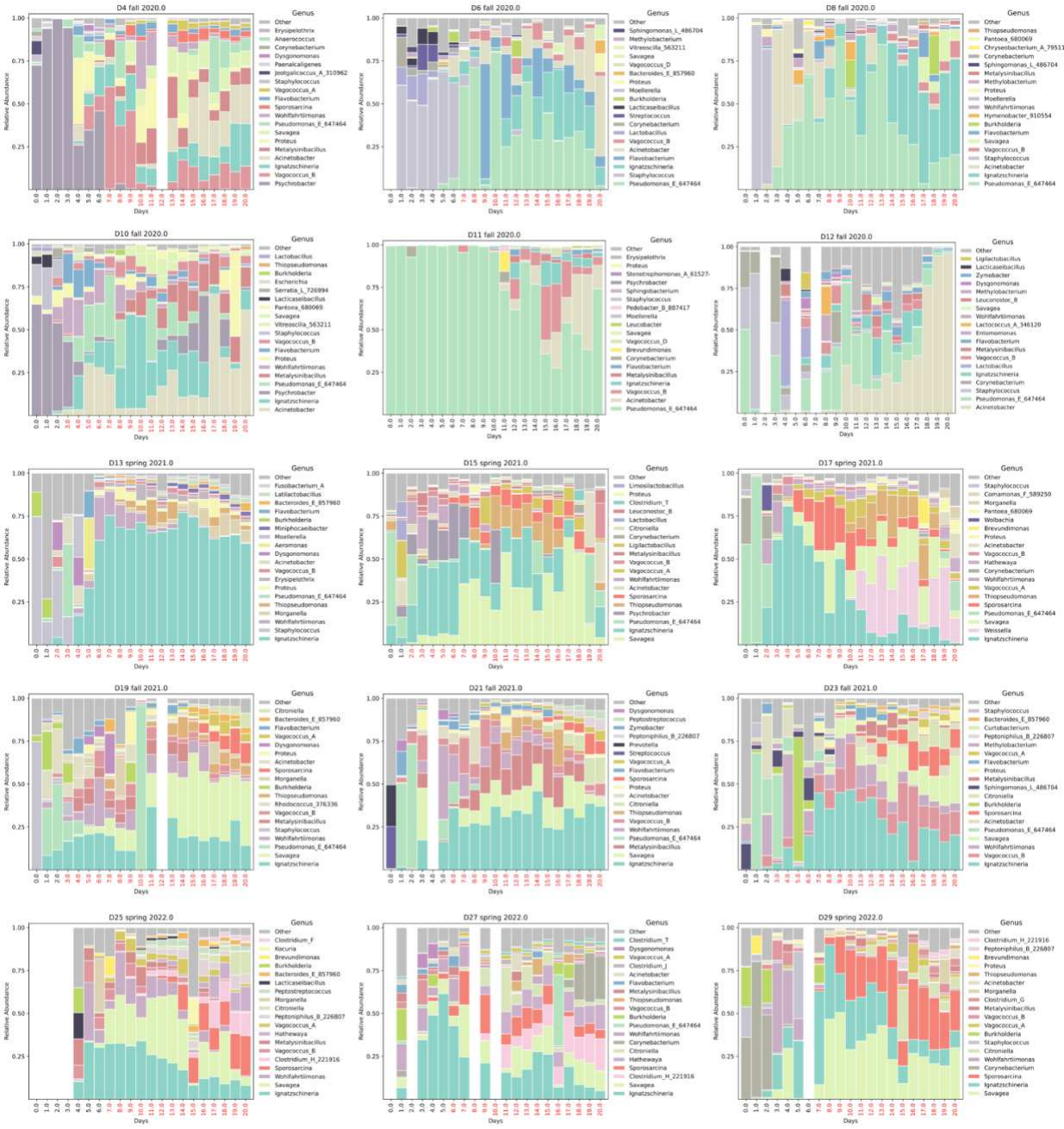
Supplemental Figure S2. A) Number of unique ASVs compared between skin of indoor cadavers (teal) and outdoor cadavers (orange). Mann-Whitney test was used to test for significance between groups. **B)** Venn diagram shows the number of unique ASVs only found in indoor cadaver skin samples (*n* = 1384) and outdoor samples (*n* = 7592). In total, 1242 ASVs were found in both decomposition environments. **C)** Venn diagram shows the unique indoor ASVs (*n* = 1384) only accounts for 0.9% relative abundance of indoor skin samples while the 7592 unique outdoor ASVs only accounted for 2.5% relative abundance of outdoor samples. **D)** Unweighted unifrac Principal Coordinate Analysis (PCoA) plot showing cadaver skin samples collected from outdoor (circle) and indoor (diamond) cadavers. Samples are colored by ADD. 95% confidence intervals show indoor (teal) and outdoor (orange) cadaver groups. PERMANOVA was used to test for significance between indoor and outdoor skin samples (*p*-value < 0.001; R^2 = 0.021679).



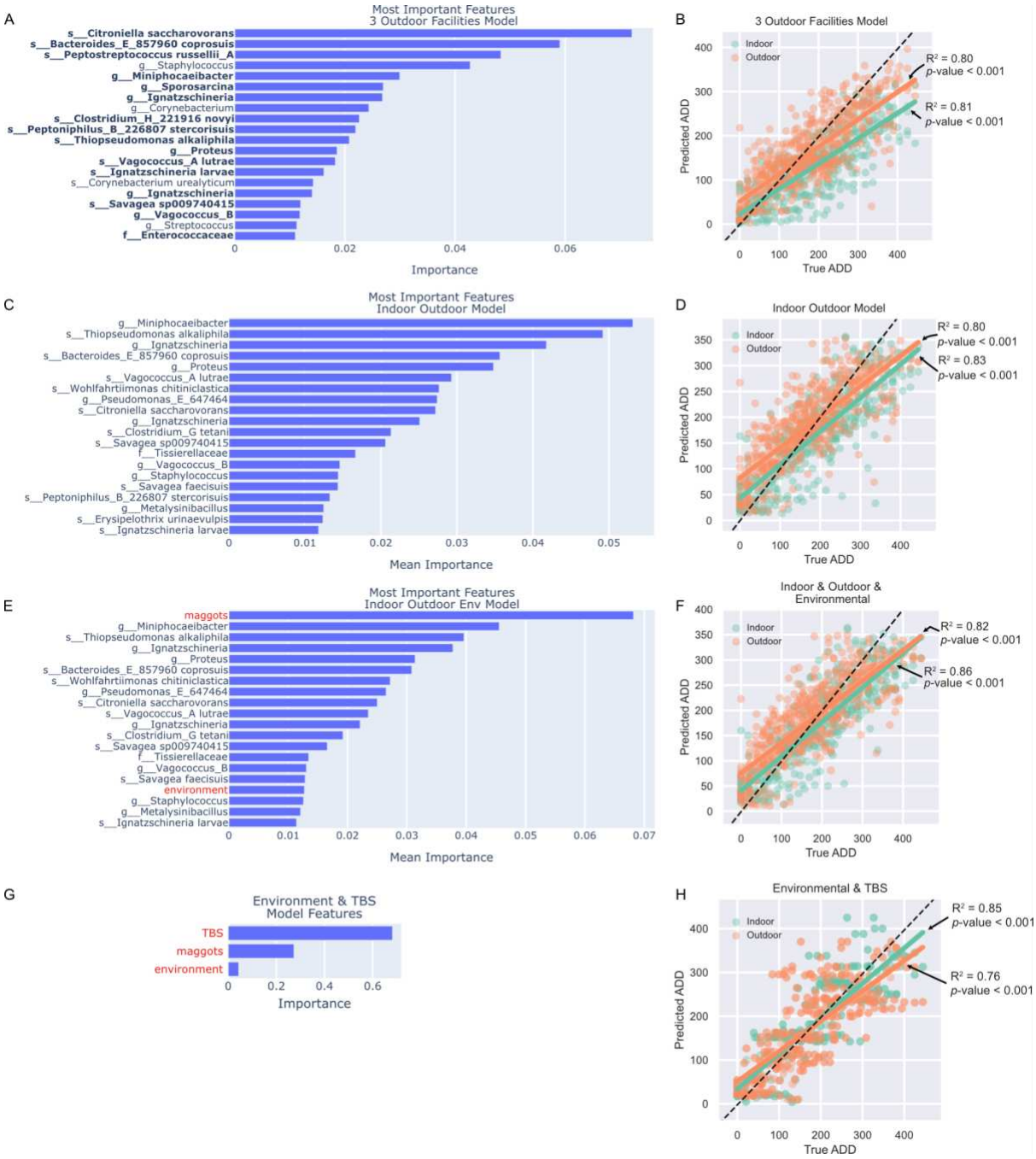
Supplemental Figure S3. Taxa associated with maggot exposure. Taxa that were significantly enriched in cadaver skin samples that had previously been exposed to notable maggots are shown in blue. Taxa enriched in skin samples with no previous maggot presence are shown in grey (*q*-values < 0.05). Analysis was performed using ANCOMBC. ASVs were collapsed on the species level. Lowest taxonomic classification is shown. Samples with fewer than 1000 reads and taxa found in less than 1% of samples were excluded. Only taxa with >= +/- 1 LFC are shown.



Supplemental Figure S4. Indoor taxa bar plots. Bar plots for each donor placed indoors are shown starting on the day of placement (day=0). Skin samples (face and hip) were averaged each day. ASVs were grouped at the genus level. Samples were rarefied to 5000 reads per sample. Blank bars indicate samples had fewer than 5000 reads per sample at that time point. Days colored in red indicate the first day of notable maggot presence.

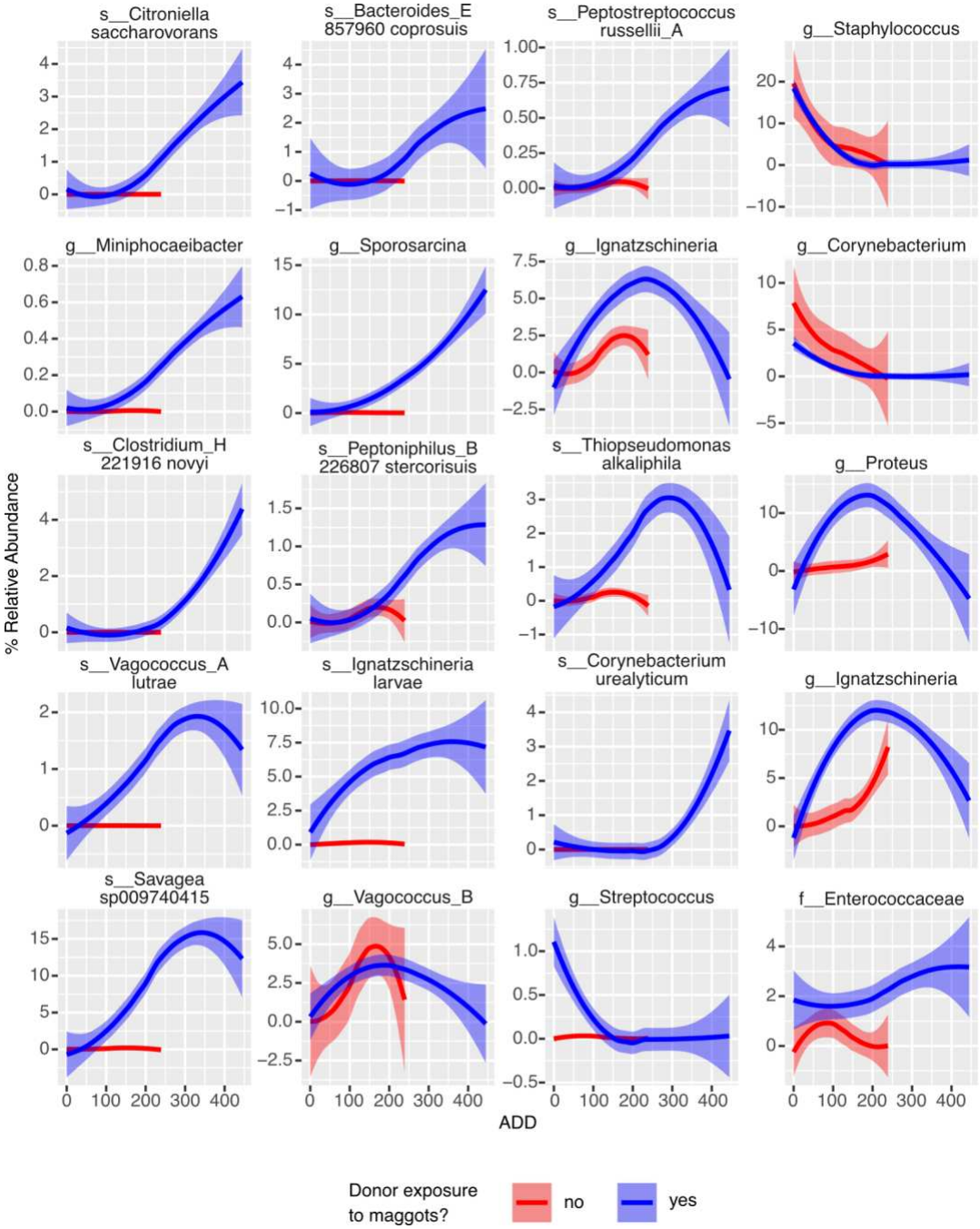


Supplemental Figure S5. Outdoor taxa bar plots. Bar plots for each donor placed outdoors are shown starting on the day of placement (day=0). Skin samples (face and hip) were averaged each day. ASVs were grouped at the genus level. Samples were rarefied to 5000 reads per sample. Blank bars indicate samples had fewer than 5000 reads per sample at that time point. Days colored in red indicate the first day of notable maggot presence.

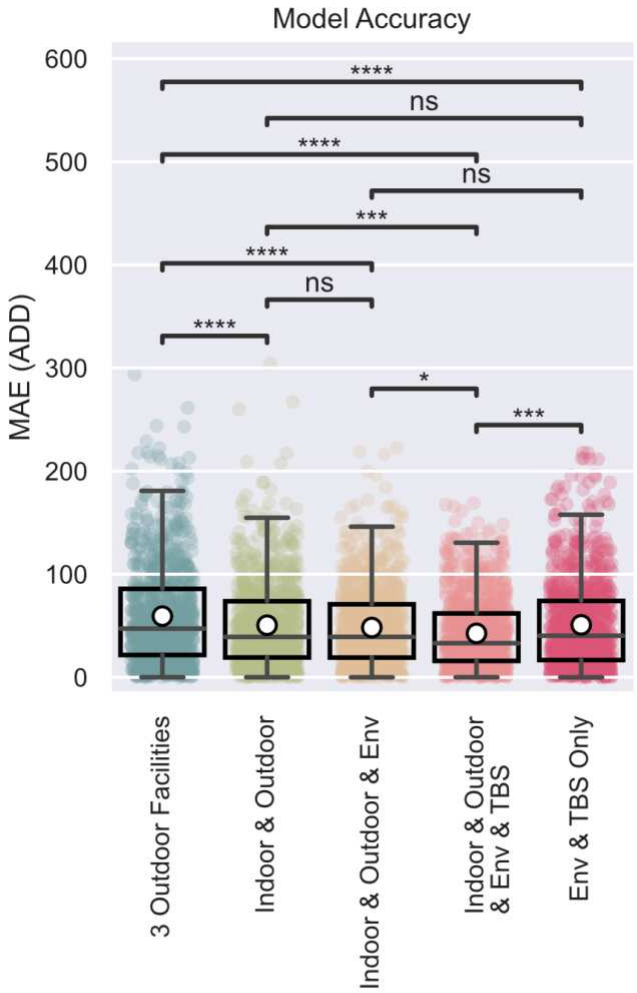


Supplemental Figure S6. Important model features and performance. The 20 most important features for each model are shown in **A**, **C**, **E**, and **G**. Cross-validation was performed for models C and E and feature importance was averaged across all cross-validation folds (mean importance). Features highlighted in red represent non-microbial, environmental features. Bolded taxa in **A** were also enriched in cadaver samples with notable maggots present (see Figure S3). **B, D, F, H**) Predicted ADDs are compared to true ADDs for indoor (teal) and outdoor (orange) skin samples for each model. Linear regressions for indoor and outdoor groups are shown. Dashed black line represents

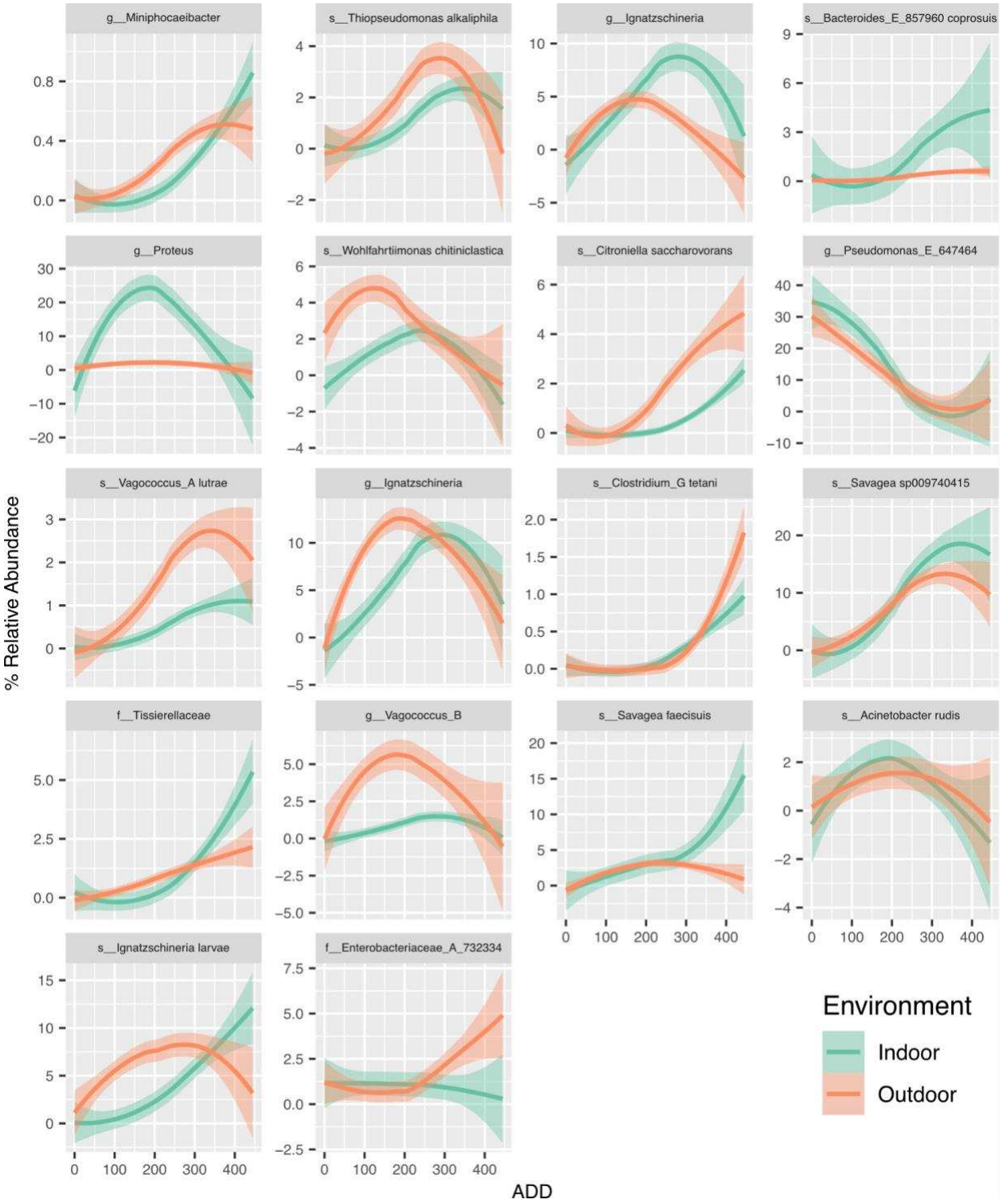
perfect prediction accuracy.



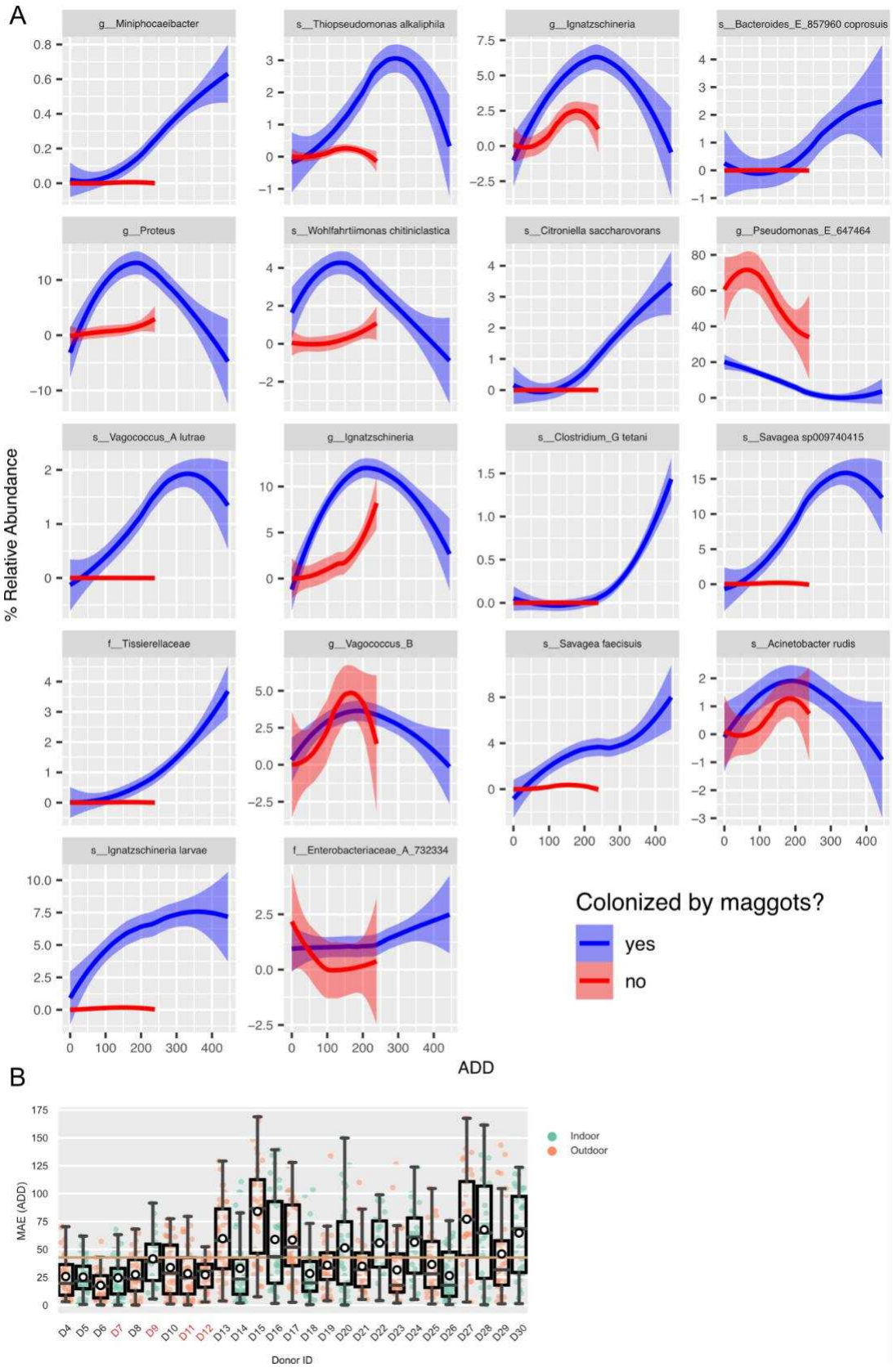
Supplemental Figure S7. Several important outdoor model features are not found in donors that were not colonized by maggots. The 20 most important "3 outdoor facilities model" features are shown from most (top left) to least important (bottom right). The percent relative abundance of each feature is compared between the four cadavers (D7, D9, D11, and D12) with no notable maggot colonization (red) and cadavers with maggot mass (blue) across decomposition (ADD). Percent relative abundances were smoothed using loess regression.



Supplemental Figure S8. Overall model performance. Boxplots show cross-validation mean absolute errors for each model that was constructed. The model trained on 16S amplicon sequences, environmental variables (maggots, indoor/outdoor), and total body scores (TBS) was the best performing model and had the lowest errors. White markers represent group means. ANOVA with post-hoc Tukey HSD correction for multiple comparisons was used to test for significance. **** = *p*-value < 0.0001, *** = *p*-value < 0.001, * = *p*-value < 0.05; ns = not significant.



Supplemental Figure S9. Important model features share similar patterns. The 18 most important microbial features for the “Indoor & Outdoor & Env & TBS” model are shown from most (top left) to least important (bottom right). The percent relative abundance of each feature is compared between the indoor (teal) and outdoor (orange) cadavers across decomposition (ADD). Percent relative abundances were smoothed using loess regression.


A
g__Miniphocaeibacter
s__Thiopseudomonas alkaliphila
g__Ignatzschineria
s__Bacteroides_E_857960 coprosuis
g__Proteus
s__Wohlfahrtiimonas chitiniclastica
s__Citroniella saccharovorans
g__Pseudomonas_E_647464
s__Vagococcus_A lutrae
g__Ignatzschineria
s__Clostridium_G tetani
s__Savagea sp009740415
f__Tissierellaceae
g__Vagococcus_B
s__Savagea faecisuis
s__Acinetobacter rudis
s__Ignatzschineria larvae
f__Enterobacteriaceae_A_732334
% Relative Abundance
ADD
Colonized by maggots?
yes
no
B
MAE (ADD)
Donor ID
Indoor
Outdoor

Supplemental Figure S10. Missing maggot-associated features do not appear to impact donor MAE. A) The 18 most important microbial features for the “Indoor & Outdoor & Env & TBS” model are shown from most (top left) to least important (bottom right). The percent relative abundance of each feature is compared between the four cadavers (D7, D9, D11, and D12) with no notable maggot colonization (red) and cadavers with maggot mass (blue) across decomposition (ADD). Percent relative abundances were smoothed using loess regression. **B)** Mean absolute error for each donor is shown. White marker reflects the mean value. Indoor cadavers are colored in teal and outdoor cadavers are colored in orange. Donors with no notable maggot colonization are labeled in red. Horizontal lines represent MAE of indoor (42.950 ADD) and outdoor cadavers (42.396 ADD).