

From Field to Bioreactor: Metagenomic and Physiological Insights into Microbial Nutrient
Cycling Across Oxygen Gradients

Zachary Flinkstrom

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Mari K. H. Winkler, Chair

David E. Butman

Wei Qin

Program Authorized to Offer Degree:

Department of Civil and Environmental Engineering

©Copyright 2025

Zachary Flinkstrom

University of Washington

Abstract

From Field to Bioreactor: Metagenomic and Physiological Insights into Microbial Nutrient
Cycling Across Oxygen Gradients

Zachary Flinkstrom

Chair of the Supervisory Committee:

Mari K.H. Winkler

Department of Civil and Environmental Engineering

Microbial communities drive essential nutrient fluxes in diverse ecosystems, and their activity is strongly influenced by oxygen availability, which governs the balance between aerobic respiration and anaerobic metabolisms such as denitrification, sulfate-reduction, fermentation and methanogenesis. This dissertation explores how oxygen gradients shape microbial community structure and function across three systems: a global metagenomic dataset, a freshwater wetland, and twin aerobic and anaerobic chemostat bioreactors. Clustering of thousands of metagenomic functional gene profiles revealed oxygen as a major axis of differentiation, alongside light, with distinct traits such as antibiotic production and archaeal abundance. In the wetland study, microbial communities were structured primarily by soil depth rather than seasonal variation, reflecting the persistent influence of vertical oxygen gradients despite shifts in porewater chemistry and greenhouse gas fluxes. In bioreactors fed with N-acetylglucosamine—a model compound for microbial necromass—both aerobic and anaerobic communities mineralized the substrate, but with divergent metabolic byproducts and community

trajectories. The aerobic system exhibited strong coupling of carbon and nitrogen cycles, including nitrification and denitrification, while the anaerobic system became increasingly methanogenic and resilient to oxygen perturbations. Metagenomic and transcriptomic analyses revealed distinct pathways of compound utilization between treatments. Overall, this work integrates field data, laboratory experiments, and bioinformatics to deepen our understanding of how oxygen availability governs microbial community dynamics and ecosystem-scale biogeochemical processes.

Acknowledgements

This dissertation would not have been possible without the unwavering support of incredible mentors, lab-mates, collaborators, friends, and family. I'm especially grateful to have joined Mari Winkler's research group five years ago, in the midst of the COVID pandemic, and to have had the opportunity to explore the fascinating world of wetlands and microbial ecology. Thank you, Mari, for your mentorship and for your genuine care for the wellbeing of our group.

I'm deeply appreciative of my colleagues for their endless knowledge and camaraderie—especially Britt Abrahamson, Pieter Candry, Sam Bryson, Bruce Godfrey, Kris Hunt, Tom Lie, Dave Stahl, Bao Nguyen Quoc, Stephany Wei, Brian Roman, Korena Mafune, Pei-Hsin Wang, and Renee Davis. I also thank our collaborators Dongyu Wang, Chongle Pan, Julie Johnston, and Xavier Mayali, as well as my committee members Wei Qin, David Butman, and Bob Morris, for their guidance and insight throughout this journey.

The biggest thanks of all go to my life-partner, Colby, who moved across the country with me to start this chapter and has always given me the confidence to keep going. Thank you to my family in Maine for their support from afar, and to my friends in Seattle for always being up for a beer, a trip to the mountains, or both.

Table of Contents

Chapter 1 – An Introduction to Environmental Microbiology1
 Background1
 Research objectives and outline4
 References6
Chapter 2 – Metagenomic clustering links specific metabolic functions to globally relevant ecosystems8
 Abstract9
 Introduction10
 Materials and Methods12
 Results15
 Discussion20
 References25
 Figures29
 Supplementary figures33
Chapter 3 – Soil depth shapes microbial community composition and function in a freshwater wetland44
 Abstract45
 Introduction46
 Materials and Methods48
 Results57
 Discussion65
 References72
 Figures78
 Supplementary tables and figures86
Chapter 4 – N-acetylglucosamine degradation by a wetland microbial community under varying oxygen concentrations91
 Abstract92
 Introduction93
 Materials and Methods95
 Results99
 Discussion105

References	109
Figures	112
Supplementary Material.....	118
Chapter 5 – Other scientific contributions	127
Comparative physiological and genomic characterization of a novel <i>Nitrobacter vulgaris</i> strain from a nitrate-contaminated subsurface.....	128
Resource availability governs polyhydroxyalkanoate (PHA) accumulation and diversity of methanotrophic enrichments from wetlands	132
Wetlands harbor lactic acid-driven chain elongators	134
Identifying microbial functional guilds performing cryptic organotrophic and lithotrophic redox cycles in anaerobic granular biofilms	138
Chapter 6 – Conclusions and perspective	143
Conclusions	143
Perspective	145

List of Tables

Table 3-1 – Bulk wetland soil characteristics.....	86
Table 3-2 – Surface greenhouse gas fluxes at each sampling timepoint.....	86
Table 4-1 – Averaged biomass-specific carbon and nitrogen production rates of the anaerobic (R1) and aerobic (R2) chemostats.....	118
Table 4-2 – Modified lacustrine media composition	120

List of Main Figures

Figure 2-1 – Global distribution of dataset and functional clustering.	29
Figure 2-2 – KEGG pathways associated with cluster-specific marker genes.....	30
Figure 2-3 – Top 15 cluster marker genes based on median fold-change.....	31
Figure 2-4 – GC content variation and correlated genes of interest.	32
Figure 3-1 – Study site geochemical characterization.	78
Figure 3-2 – Targeted metabolomics of wetland soil sections from May.....	79
Figure 3-3 – Gene-centric metagenomics of wetland soil cores.....	80
Figure 3-4 – Genome-centric metagenomics of wetland soil cores.....	82
Figure 3-5 – Abundant proteins and taxa based on metaproteomics.	84
Figure 4-1 – Biomass-specific production rates ($\mu\text{mol}\cdot\text{mg}_{\text{DCW}}^{-1}\cdot\text{h}^{-1}$) of GlcNAc degrading, wetland sediment chemostats over the entire operational period.	112
Figure 4-2 – Class-level taxonomic abundance across reactors and operational phase.....	114
Figure 4-3 – Trajectory of taxonomic and functional composition.	115
Figure 4-4 – Functional gene variation across reactors and phases.....	116
Figure 4-5 – N-acetylglucosamine utilization pathways.....	117
Figure 5-1 – Phylogeny and operon structure of nitrite oxidation and nitrous oxide reduction in <i>Nitrobacter vulgaris</i> MLSD-S22.	131
Figure 5-2 – Resource availability governs polyhydroxyalkanoate (PHA) accumulation and diversity of methanotrophic enrichments from wetlands.	134
Figure 5-3 – Enrichment of a lactic acid-driven chain elongation community from wetland soil.	137
Figure 5-4 – Functional and taxonomic characterization of granular biofilm MAGs.	141

List of Supplemental Figures

Figure S 2-1 – Characteristics of dataset metagenomes.....	33
Figure S 2-2 – Bootstrap testing of k-means and Ward’s hierarchical clustering methods over a range of cluster numbers.	34
Figure S 2-3 – Distribution of Silhouette coefficients across the dataset.	35
Figure S 2-4 – Characteristics of metagenomes separated by cluster assignment.....	36
Figure S 2-5 – Domain-level taxonomic composition of metagenome contigs in each cluster. ...	37
Figure S 2-6 – Assignment of cluster-specific marker genes.	38

Figure S 2-7 – Number of cluster-specific marker genes associated with KEGG Brite Categories.	39
Figure S 2-8 – KEGG carbohydrate and amino acid metabolism.	40
Figure S 2-9 – KEGG terpenoid, polyketide, and vitamin metabolism.	41
Figure S 2-10 – KEGG glycan and lipid metabolism.	42
Figure S 2-11 – Sub-clustering analysis.	43
Figure S 3-1 – Indicator of Reduction in Soils (IRIS) tubes from three seasons.	87
Figure S 3-2 – Additional measurements of site geochemistry.	88
Figure S 3-3 – MAG quality statistics.	89
Figure S 3-4 – Proteins detected in different taxa.	90
Figure S 4-1 – Biomass-specific carbon production rates.	121
Figure S 4-2 – Biomass-specific nitrogen production rates.	123
Figure S 4-3 – Taxonomic and functional diversity.	124
Figure S 4-4 – Order-level taxonomic composition across reactors and operational phases.	125
Figure S 4-5 – Family-level taxonomic composition across reactors and operational phases.	126

Chapter 1 – An Introduction to Environmental Microbiology

Background

Though small in size, bacteria, archaea, and fungal organisms comprise about 16% (or ~89 Gt C) of all biomass on earth, second only to plants in abundance¹. With diverse catalytic capabilities, microorganisms are the “engines” that drive earth’s geochemical cycles across aquatic, terrestrial, and engineered ecosystems². Consequently, microbes mediate globally significant fluxes of the greenhouse gases carbon dioxide, methane, and nitrous oxide in both natural and manmade settings. Furthermore, they control the breakdown of organic matter and transformation of nitrogen, phosphorus, sulfur, and metals into different forms with varying bioavailability, mobility, and toxicities. Understanding how microorganisms in the environment function and cooperate on a cellular level is essential for predicting larger-scale biogeochemistry. It is of great importance for predicting emissions under future climate scenarios, bioremediation, improving soil health and engineering solutions for climate change mitigation.

The study of environmental microbiology has traditionally relied on the enrichment and isolation of microorganisms using selective growth media. In this approach, a sample from a target environment is used as an inoculum in liquid or solid media containing macro- (e.g., sugars, amino acids, lipids) and micro-nutrients (e.g., trace metals, vitamins, cofactors) of interest. The media is incubated under controlled environmental conditions (e.g., temperature, pH, oxygen levels) to cultivate microbial communities or single organisms capable of specific metabolic functions. The resulting cultures can be characterized in various ways: microscopy to assess morphology, kinetics to evaluate growth rates and yields, versatility tests to explore different energy sources, stress response analysis, and production of secondary metabolites.

These traditional culture-based methods have significantly advanced our understanding of microbes essential to environmental nutrient cycles.

Notable examples of culture-based discoveries include Sergei Winogradsky's pioneering work on nitrifying and sulfur-oxidizing bacteria, which established the concept of chemolithoautotrophy (the ability to derive energy from inorganic compound oxidation and fix inorganic carbon for growth)³. Similarly, Martinus Beijerinck's studies on nitrogen-fixing and sulfate-reducing bacteria, and Robert E. Hungate's development of anaerobic culturing techniques for isolating strict anaerobes like cellulose degraders and methanogens⁴, have provided foundational insights. These groundbreaking studies share a common theme: the application of innovative culturing methods to uncover the critical roles of microorganisms in biogeochemical cycling.

The advent of DNA sequencing has further revolutionized microbiology. One early milestone was the comparative analysis of the conserved 16S rRNA gene, which revealed archaea as a domain of life distinct from bacteria⁵. This groundbreaking discovery gave rise to molecular-based taxonomy, significantly expanding our understanding of microbial diversity and evolution. Recent advances in sequencing technology have further propelled the field through metagenomics, which enables comprehensive and detailed analysis of environmental DNA. Bioinformatic techniques for assembling, annotating, and binning metagenomic sequences now allow for the reconstruction of genomes (metagenome-assembled genomes, or MAGs) and the prediction of metabolic pathways. Metagenomics has greatly expanded our understanding of global functional and phylogenetic diversity within microbiomes from diverse habitats, including the gut, soils, and oceans⁶⁻⁸.

Moreover, metagenomic analysis serves as a powerful hypothesis generator, driving the discovery of novel microbial processes. This approach has facilitated breakthroughs such as the identification and isolation of ammonia-oxidizing archaea^{9,10}, complete ammonia-oxidizing bacteria (comammox)^{11,12}, and previously unknown methanogenic organisms^{13,14}. By integrating metagenomic insights with traditional culture-based methods and environmental geochemical measurements, researchers now wield a robust toolbox for advancing the study of environmental microbiology.

Much of this dissertation will focus on wetlands as an environment of interest due to their outsized role in the global carbon cycle, as evidenced by the fact that they store approximately 20% of terrestrial organic carbon in just 5-8% of land area¹⁵. Wetlands have a number of unique properties because they exist at the transition between aquatic and terrestrial ecosystems: their saturated anaerobic soils store immense amounts of organic carbon, they harbor aquatic plants specially adapted to these conditions, and they also emit globally significant quantities of methane (~30% of global emissions)¹⁶. Understanding how wetland carbon cycling dynamics will change under future climate conditions (i.e. increased temperatures, prolonged droughts, more intense flooding, etc.) is of prime importance, especially given the potential of wetlands to drive a positive climate feedback loop that would accelerate climate change¹⁷.

The wetland microbiome largely controls the fate of carbon and other nutrients in this ecosystem. In the absence of oxygen, alternative electron acceptors like nitrate, sulfate, or iron may be used for respiration but once those are depleted, certain microbes must rely on thermodynamically challenging metabolisms like methanogenesis, fermentation, or acetogenesis. As a result, redox state is an important parameter dictating carbon metabolism within a wetland. The specific nature and quality of carbon compounds can also greatly affect their persistence

within a wetland soil. For instance, simple sugars are more bioavailable compared to polyphenols or cellulose which require special enzymes to be broken down. Finally, dead microbial cells (aka necromass) have been implicated as important constituents of soil carbon^{18,19}, further illustrating how microbial activity drives the carbon cycle. Taken together, studying the complex microbial food webs of wetland soils under climate-relevant perturbations is important for understanding how these ecosystems function at a larger scale. This backdrop motivates the following research objectives.

Research objectives and outline

The overarching objective of this dissertation is to understand how oxygen gradients influence microbial nutrient cycling, with a focus on wetland ecosystems. This work integrates microscale investigations of the wetland microbiome with macroscale measurements of biogeochemistry to link microbial processes with broader ecosystem patterns. With this in mind, more specific objectives (in italics) and how they are addressed in various chapters are outlined below:

(i) Assess the utility of metagenomic functional gene profiles for describing microbial communities across oxygen gradients

Chapter 2 – Metagenomic clustering links specific metabolic functions to globally relevant ecosystems addresses this objective by analyzing a large and ecosystem-diverse dataset from the Joint Genome Institute’s Integrated Microbial Genomes and Metagenomes database. Unsupervised clustering was used to identify broad-scale patterns in functional gene profiles and showed clear discrimination between terrestrial, aquatic, and anaerobic ecosystems. The approach also revealed the specific metabolic functions that were overrepresented in each cluster

and associated ecosystem. This chapter provides both a framework for gene-centric metagenome analysis and unique findings from this large dataset²⁰.

(ii) Characterize the biogeochemistry of a freshwater wetland field site over temporal and spatial scales to identify dominant drivers of microbial community assembly

Chapter 3 - Soil depth shapes microbial community composition and function in a freshwater wetland presents the results of a field study conducted in the Washington Park Arboretum wetland in Seattle. Measurements of surface greenhouse gas flux, depth-resolved pore water chemistry, and metagenomics were taken across four seasons to create a dataset integrating various elements of the site's biogeochemistry. In addition, depth-resolved metabolomics and metaproteomics were analyzed from single timepoints. We found soil depth to be the strongest driver of microbial community assembly based on metagenomics despite seasonal fluctuations in water level and temperature. Building upon this finding, we identified numerous microbial taxa, metabolic pathways, and specific metabolites that were significantly correlated with soil depth. The study raises interesting questions about how microbial communities respond and adapt to varying oxygen concentrations.

(iii) Characterize the pathways for necromass utilization by wetland microbial communities under aerobic and anaerobic conditions

Chapter 4 – N-Acetylglucosamine Degradation by Wetland Microbial Communities Under Varying Oxygen Concentrations investigates the degradation of a model necromass compound, N-acetylglucosamine, by microbial communities in low-dilution-rate chemostat bioreactors. Aerobic and anaerobic bioreactors were operated for 270 days, allowing microbial communities to stabilize and their metabolic end-products to be defined. DNA and RNA

sequencing revealed the dominant microbes and their active metabolic pathways under both steady-state and oxygen-perturbed conditions. The findings provide insight into how necromass is processed by microbial communities in contrasting oxygen environments. Constrained mass balances further enabled stoichiometric modeling of greenhouse gas emissions from this compound under aerobic and anaerobic conditions.

References

1. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc Natl Acad Sci U S A* **115**, 6506–6511 (2018).
2. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science (1979)* **320**, 1034–1039 (2008).
3. Dworkin, M. Sergei Winogradsky: A founder of modern microbiology and the first microbial ecologist. *FEMS Microbiology Reviews* vol. 36 364–379 Preprint at <https://doi.org/10.1111/j.1574-6976.2011.00299.x> (2012).
4. Chung, K.-T. & Bryant, M. P. Robert E. Hungate: Pioneer of Anaerobic Microbial Ecology. *Anaerobe* **3**, 213–217 (1997).
5. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**, 4576–4579 (1990).
6. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
7. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science (1979)* **348**, (2015).
8. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
9. Venter, J. C. *et al.* Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science (1979)* **304**, 66–74 (2004).
10. Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).

11. Van Kessel, M. A. H. J. *et al.* Complete nitrification by a single microorganism. *Nature* **528**, 555–559 (2015).
12. Daims, H. *et al.* Complete nitrification by *Nitrospira* bacteria. *Nature* **528**, 504–509 (2015).
13. Krukenberg, V., Kohtz, A. J., Jay, Z. J. & Hatzenpichler, R. Methyl-reducing methanogenesis by a thermophilic culture of Korarchaeia. *Nature* **632**, 1131–1136 (2024).
14. Kohtz, A. J. *et al.* Cultivation and visualization of a methanogen of the phylum Thermoproteota. *Nature* **632**, 1118–1123 (2024).
15. Nahlik, A. M. & Fennessy, M. S. Carbon storage in US wetlands. *Nat Commun* **7**, 1–9 (2016).
16. Saunio, M. *et al.* The global methane budget 2000-2017. *Earth Syst Sci Data* **12**, 1561–1623 (2020).
17. Zhang, Z. *et al.* Emerging role of wetland methane emissions in driving 21st century climate change. *Proc Natl Acad Sci U S A* **114**, 9647–9652 (2017).
18. Kästner, M., Miltner, A., Thiele-Bruhn, S. & Liang, C. Microbial Necromass in Soils—Linking Microbes to Soil Processes and Carbon Turnover. *Frontiers in Environmental Science* vol. 9 Preprint at <https://doi.org/10.3389/fenvs.2021.756378> (2021).
19. Kallenbach, C. M., Frey, S. D. & Grandy, A. S. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nat Commun* **7**, (2016).
20. Flinkstrom, Z., Bryson, S., Candry, P. & Winkler, M.-K. H. Metagenomic clustering links specific metabolic functions to globally relevant ecosystems. *mSystems* **9**, (2024).

Chapter 2 – Metagenomic clustering links specific metabolic functions to globally relevant ecosystems

Zachary Flinkstrom^{1*}, Samuel Bryson², Pieter Candry^{1,3}, Mari-Karoliina H. Winkler¹

¹ Department of Civil and Environmental Engineering, University of Washington, Seattle, WA

² Phase Genomics, Seattle, WA

³Current address: Laboratory of Systems and Synthetic Biology, Wageningen University & Research. 6708 WE, Wageningen, The Netherlands

*Correspondence to: Zachary Flinkstrom, Civil and Environmental Engineering, University of Washington, 201 More Hall, Box 352700, Seattle, WA 98195-2700, USA; E-mail: zflinky@gmail.com

Peer-reviewed and published in *mSystems* **9**, (2024) <https://doi.org/10.1128/msystems.00573-24>

Abstract

Metagenomic sequencing has advanced our understanding of biogeochemical processes by providing an unprecedented view into the microbial composition of different ecosystems. While the amount of metagenomic data has grown rapidly, simple-to use methods to analyze and compare across studies have lagged behind. Thus, tools expressing the metabolic traits of a community are needed to broaden the utility of existing data. Gene abundance profiles are a relatively low dimensional embedding of a metagenome's functional potential and are thus tractable for comparison across many samples. Here, we compare the abundance of KEGG Ortholog Groups (KOs) from 6,539 metagenomes from the Joint Genome Institute's Integrated Microbial Genomes and Metagenomes (JGI IMG/M) database. We find that samples cluster into terrestrial, aquatic, and anaerobic ecosystems with marker KOs reflecting adaptations to these environments. For instance, functional clusters were differentiated by the metabolism of antibiotics, photosynthesis, methanogenesis, and surprisingly GC content. Using this functional gene approach, we reveal the broad-scale patterns shaping microbial communities and demonstrate the utility of ortholog abundance profiles for representing a rapidly expanding body of metagenomic data.

Importance

Metagenomics, or the sequencing of DNA from complex microbiomes, provides a view into the microbial composition of different environments. Metagenome databases were created to compile sequencing data across studies, but it remains challenging to compare and gain insight from these large datasets. Consequently, there is a need to develop accessible approaches to extract knowledge across metagenomes. The abundance of different orthologs (i.e. genes that perform a similar function across species) provides a simplified representation of a

metagenome's metabolic potential that can easily be compared with others. In this study, we cluster the ortholog abundance profiles of thousands of metagenomes from diverse environments and uncover the traits that distinguish them. This work provides a simple to use framework for functional comparison and advances our understanding of how the environment shapes microbial communities.

Introduction

Environmental metagenomics has uncovered the diversity of microbial life driving nutrient cycling in different ecosystems. An overarching goal of the field has been to connect environmental conditions to genomic characteristics to identify “who’s there?” and “what are they doing?”. As DNA sequencing technology has become more accessible the amount of metagenomic data has ballooned. For example, the Joint Genome Institute’s Integrated Microbial Genomes and Metagenomes (JGI IMG/M) database contains more than 30,000 metagenomes¹ and the European Nucleotide Archive (ENA) contains upwards of 50,000 metagenomes². Large cross-study analyses of this data can increase our understanding of the patterns in microbiome function across environments. However, these efforts are resource intensive given the size and complexity of the data. To illustrate this, consider the hierarchical nature of a processed metagenome consisting of sequencing reads assembled into contigs which contain genes that are assigned a function, taxonomy, and relative abundance. Furthermore, contigs can be binned into metagenome assembled genomes (MAGs) based on characteristics such as tetranucleotide frequency and read coverage^{3,4}.

Genome-centric analysis of metagenomes can shed light on the metabolic potential of uncultivated microorganisms. For instance, MAGs have expanded our view of nitrogen fixation

in the surface ocean⁵, methane metabolism in archaea⁶⁻⁸, and carbon degradation in permafrost⁹. Multiple large collections of MAGs have been compiled using metagenomes from the human gut¹⁰, anaerobic digestors¹¹, and a mix of different environments¹². While these collections provide a valuable resource for the exploration of the vast functional and taxonomic diversity of microbiomes, they do not reveal the functional composition of the whole community or enable facile comparison across samples. To this end, a gene-centric approach may be preferred.

Early metagenomic studies, using limited sample sets, observed how ortholog abundance differed across environments according to the geochemical context^{13,14}. Studies of specific ecosystems have linked functional variation to parameters such as temperature, pH, and latitude within topsoil¹⁵, permafrost soil¹⁶, and ocean microbiomes^{17,18}. Together, these studies demonstrate how ortholog abundance is a valuable metric for representing a metagenome and linking it to its source environment. In addition, abundance profiles have a simple data structure that allows for large scale comparisons and the placement of new data in the context of existing samples.

In this study, we evaluate the characteristics of KEGG Ortholog Group (KO) abundance profiles across a large and diverse set of metagenomes from JGI IMG/M. We use unsupervised clustering to show how functional potential stratifies samples by environment and identify the marker KOs that differentiate sample groupings. Interestingly, we note differences in GC content between functional clusters and explore the KOs associated with this trait. The compiled dataset can be explored and expanded upon with new metagenomes. Overall, this study provides insight into how environment shapes the functional composition of microbial communities and supports the use of ortholog abundance profiles for understanding metagenomic data.

Materials and Methods

Data retrieval and preprocessing

Ortholog abundance data was retrieved from the JGI IMG/M database in January 2022. Metagenomes were selected if they had (i) no data use restrictions, (ii) were sequenced and processed through JGI standard workflows¹⁹, (iii) contained more than 100,000 genes assembled and more than 30% of the genes had a KEGG KO (Kyoto Encyclopedia of Genes and Genomes Orthology)²⁰ annotation assigned, (iv) had estimated gene copy data, and (v) were not a combined assembly of multiple read libraries. The JGI IMG/M functional profile tool was then used to retrieve the estimated gene copies, based on read mapping coverage, of all KOs for each sample. In other words, the KO abundance count represents the sum of the abundance of every instance of that KO in a metagenome. All genes without a KO assignment were excluded from the downstream analyses.

Additionally, the metadata associated with each metagenome was exported. The sample ecosystem label was defined by the “Ecosystem Type” metadata field if more than 100 samples shared the label, otherwise the broader “Ecosystem” field was used to describe the sample. This was done to reduce the number of ecosystem categories containing few samples. JGI taxonomic composition report tables describing contig-level taxonomic assignment were available for 4,988 of the 6,539 total metagenomes. This data provided a count of contigs assigned to different domains and phyla in each metagenome. Taxonomic assignment was determined based on the lineage that received a majority of hits in a database homology search of the contig’s protein coding genes¹⁹.

Subsequent data processing and analysis was performed in Python Jupyter Notebooks that are available in the study’s github repository

(https://github.com/zflink/Metagenome_functional_clustering). The SCANPY²¹ gene expression data analysis Python package and its ANNDATA data structure class were used to facilitate processing and exploration of the KO profiles and associated metadata. While SCANPY was developed for the analysis of single cell RNA sequencing data, its data structure, analysis, and plotting tools remain useful for any high-dimensional dataset that has lots of metadata associated with its constituent variables and observations. To construct the ANNDATA object, ortholog abundance tables, sample metadata, and KO names were first loaded into pandas dataframes before being passed to the “anndata.AnnData” constructor function with ortholog abundance as the data matrix, sample metadata as the observations table, and KO names as the variable table. Rare or nonexistent KOs were removed from the dataset if they were present in less than 10% of samples. The resulting KO abundance profiles were center-log₂ ratio (clr) normalized on a per sample basis based on the method from ALDEx2²², yielding normalized abundance values per KO in each sample.

Unsupervised clustering

We explored the unsupervised clustering of the dataset with hierarchical clustering using the Ward variance minimization algorithm, implemented in scipy²³, and k-means clustering, implemented in sci-kit learn²⁴. To evaluate the performance of the two methods and to find the optimal number of clusters, we used a bootstrapping approach where the dataset was randomly sampled and clustered 100 times. Bootstrap cluster performance was evaluated using the average silhouette score and adjusted rand index. To prevent overfitting, the final k-means model used for clustering was trained using a random sample of the dataset containing half of the samples. For the sub-clustering analysis, the data were subset based on cluster assignment, then k-means clustering was performed again to find two sub-clusters within each of the three main clusters.

Clusters were visualized on the uniform manifold approximation and projection (UMAP)²⁵ of the normalized dataset implemented in SCANPY²¹ using 50 neighbors and 5 principal components. UMAP is a nonlinear dimensionality reduction technique that creates a less overcrowded projection of the dataset compared to a linear method like principal component analysis (PCA)²⁵.

Marker KO analysis

Significant differences in normalized KO abundance across clusters were defined using an ANOVA-like Kruskal-Wallis H test (implemented in scipy²³) followed by Dunn's test for post-hoc multiple comparisons (implemented in scikit-posthocs²⁶). Cluster specific marker genes were defined if a KO had a Bonferroni-corrected p-value < 0.001 for all Dunn's tests and had a median $\log_2(\text{fold-change})$ of greater than 2 compared to every other cluster. The false discovery rate of the p-value and fold-change thresholds was estimated using permutation testing, which entailed randomly assigning cluster membership 100 times followed by calculating Dunn's test p-values and median fold-changes for each KO which produced no false positive marker KOs. Marker KOs were mapped to their associated pathways by searching for their ortholog number in the KEGG pathway definition file downloaded from the KEGG database.

GC content KO correlation

Differences in assembled metagenome GC content (derived from JGI dataset metadata) across clusters were evaluated like the marker KOs with an ANOVA-like Kruskal-Wallis H test followed by Dunn's test for pairwise comparisons between clusters. Correlated KOs were evaluated by fitting a linear regression for each set of KO abundances against GC content. The Pearson correlation coefficient was used as a measure of goodness of fit and significance was evaluated using the Wald Test as implemented in Scipy's linregress function.

Results

Ortholog abundance profiles cluster metagenomes by ecosystem

The final dataset consisted of 6,539 metagenomes with 9,871 KO gene orthologs, representing a total of 4.85 terabases of assembled sequence data with an average of 742 ± 10.8 megabases per sample (Fig. S 2-1A). The metagenomes contained an average of 1.50 million genes with 38.8% having an associated KO annotation (Fig. S 2-1B-C). The dataset contained samples from across the globe, although they predominantly originated from the USA (Fig. 2-1A). A variety of ecosystems are represented with soils ($n = 2,464$), freshwater ($n = 1,506$), and marine ($n = 1,184$) being the most prevalent.

The UMAP of normalized KO abundance, which aids in the visualization of this high dimensional dataset, showed separation of aquatic and terrestrial ecosystems with some overlap occurring in the center (Fig. 2-1B). On the aquatic side, there was a distinct grouping of freshwater and marine samples. To investigate the groupings of metagenomes in this functional gene space, independent of the ecosystem label, we performed unsupervised clustering on the normalized KO abundance profiles, testing both k-means and hierarchical clustering using Ward's method for their abilities to produce consistent and cohesive clusters as measured by the adjusted Rand Index and Silhouette Coefficient. Bootstrapping of the dataset over a range of cluster numbers showed that k-means performed consistently better than Ward's hierarchical clustering in terms of consistency and cohesiveness and showed the optimum number of clusters was three (Fig. S 2-2). Accordingly, we proceeded with analyzing three clusters generated by the k-means method.

Membership of the resulting clusters followed environmental origin with some exceptions. For example, Cluster 1 was dominated by samples from soil environments (70.7%)

though it also contained samples associated with plant roots (8.1%) and some freshwater samples (7.1%). Cluster 2 consisted almost entirely of freshwater (49.7%) and marine (42.0%) samples. Lastly, Cluster 3 contained a mixture of ecosystem categories. It is worth noting that all 142 samples associated with anaerobic digestors and 93 out of 106 samples associated with digestive systems were found in Cluster 3 (Fig. 2-1C). Plotting cluster membership on the UMAP projection illustrated how the high-dimensional functional gene space is split by the k-means algorithm and confirms the broad grouping of samples based on environmental origin (Fig. 2-1D). Still, based on the Silhouette Coefficient, the k-means approach produced more cohesive groups of samples compared to using the ecosystem label (Fig. S 2-3).

The characteristics of the underlying metagenomes within each cluster were compared to evaluate any differences in sequence assembly and annotation. Cluster 1 metagenomes had a lower proportion of sequencing reads assembled and lower average coverage compared to Cluster 2 and 3 (Fig. S 2-4A, C), which is not surprising given the exceptional diversity of soil microbial communities¹⁵. Total gene count and the proportion of genes with an assigned KO were more consistent across the clusters (Fig. S 2-4B, D).

Marker KO analysis reveals metabolic functions that differentiate clusters

Marker KOs were identified for each cluster to uncover the differentially abundant functional traits driving cluster formation. Marker KOs were assigned if they had a median $\log_2(\text{fold-change})$ greater than two and a Bonferroni corrected p-value < 0.001 , both metrics were robust to false-positives when compared to random cluster assignment (Fig. S 2-6). A total of 2,546 marker KOs were identified with 684, 945, and 917 belonging to Clusters 1, 2, and 3. Clusters 2 and 3 were enriched in energy metabolism related KOs (106 in Cluster 2 & 131 in Cluster 3 vs 29 in Cluster 1), while Cluster 1 had more marker KOs related to the metabolism of

terpenoids and polyketides (77 in Cluster 1 vs 17 in Cluster 2 & 6 in Cluster 3) as well as xenobiotics degradation and metabolism (112 in Cluster 1 vs 13 in Cluster 2 & 23 in Cluster 3) (Fig. 2-2A). Specifically, Cluster 1 had 76 markers related to polyketide biosynthesis and 34 related to antimicrobial resistance (Fig. S 2-5). Within the energy metabolism category, Cluster 2 was dominated by photosynthesis related marker KOs (n = 40) while Cluster 3 was dominated by marker KOs related to methane metabolism (n = 68), specifically methanogenesis (Fig. 2-2B). Cluster 3 also contained 26 marker KOs associated with carbon fixation pathways in prokaryotes and 18 with sulfur reduction metabolism. Within the xenobiotics degradation and metabolism category, Cluster 1 had 19 marker KOs associated with benzoate degradation, 18 with aminobenzoate degradation, and 12 with xylene degradation (Fig. 2-2C). Many of these KOs were annotated as dioxygenases associated with a variety of aromatic compounds that could be derived from lignin.

Beyond KOs related to metabolism, we found cluster-specific marker KOs associated with other core cellular processes such as transcription and translation. For example, Clusters 2 and 3 had marker KOs in the categories of ribosome, ribosome biogenesis, and transcription machinery (Fig. S 2-7). Cluster 2 also had 90 marker genes in the category of chromosome and associated proteins and Cluster 3 had 48 markers associated with the prokaryotic defense system which are mostly CRISPR-associated proteins (Fig. S 2-7). Closer inspection of marker KOs in these categories revealed many specific to eukaryotic cell functioning in the case of Cluster 2 and archaeal cell functioning in Cluster 3. To validate this finding, we analyzed the taxonomic composition of cluster metagenomes. Cluster 2 metagenomes had a significantly higher proportion of contigs assigned to eukaryotes (Kruskal-Wallis: $p < 1e-306$; Dunn's multiple comparisons: $p < 1e-216$) while Cluster 3 had significantly more assigned to archaea (Kruskal-

Wallis: $p < 1e-287$; Dunn's multiple comparisons: $p < 1e-217$) and Cluster 1 had more assigned to bacteria (Kruskal-Wallis: $p < 1e-308$; Dunn's multiple comparisons: $p < 1e-251$) (Fig. S 2-5). Interestingly, Cluster 2 was also enriched in viral contigs (Kruskal-Wallis: $p < 1e-308$; Dunn's multiple comparisons: $p < 1e-105$) (Fig. S 2-5D). Despite these differences, all clusters were still dominated by bacterial contigs with a median prevalence of 99.0% in Cluster 1, 95.2% in Cluster 2, and 95.4% in Cluster 3 (Fig. 2-S5A).

To look at specific examples of marker KOs, we selected the top 15 based on median fold-change for further investigation. The heatmap shows the abundance of top genes (rows) in every sample (columns) and reveals a clear separation of clusters and highlights the uniqueness of these KOs (Fig. 2-3). Surprisingly, four of the top Cluster 1 marker KOs were annotated as the propane monooxygenase (*prm*), with other KOs associated with antibiotic resistance and synthesis, transport of manganese, lactate, and citrate. Cluster 2 markers were dominated by genes related to photosynthesis and the biosynthesis of related pigments. One KO was also related to the metabolism of dimethylsulfoniopropionate, an important algal metabolite that sustains many marine bacteria. Cluster 3 top KOs were related to anaerobic energy metabolisms; for example, the carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) complex, several hydrogenases and more intriguingly, a damage-control phosphatase and a transcription factor related to nitrogen regulation.

Functional clustering separates metagenomes by underlying GC content

Guanine-cytosine (GC) content is a fundamental characteristic of DNA sequences that represents the proportion of GC vs AT base pairings. Exploration of dataset metadata revealed that the metagenomes from the functional clusters had significantly different GC contents (Kruskal-Wallis: $p < 1e-308$; Dunn's multiple comparisons: $p < 1e-25$) despite being formed

solely based on KO abundance and no other sequence characteristics (Fig. 2-4A). Cluster 1 metagenomes had consistently high GC contents, while Cluster 2 and Cluster 3 GC contents were lower and more varied. To gain insight into the potential functional drivers of this trait, we performed simple linear regression to identify highly correlated genes. Two genes related to non-homologous end-joining DNA repair, *LigD* and *Ku*, were positively correlated with GC content (Fig. 2-4B, D). In contrast, DNA polymerase V (*polV*), which is associated with the SOS response to UV damaged DNA²⁷, and a DNA cytosine methyltransferase KO were negatively correlated with GC content (Fig. 2-4C, E). Aside from genes related to DNA processes, four genes related to trehalose metabolism were positively correlated with GC content and a pseudaminic acid synthase and Na⁺-transporting NADH:ubiquinone oxidoreductase complex were among the most negatively correlated.

Sub-clustering reveals finer scale differences within the functional landscape

To investigate how our approach performed on a more similar set of metagenomes, we further divided each of the three main clusters into two sub-clusters. The sub-clusters appeared as cohesive groups within the UMAP of the dataset (Fig. S 2-11B, D, F). Cluster 2 was clearly separated into marine and freshwater dominated sub-clusters, Cluster 3 produced one sub-cluster enriched in digestive systems and engineered environment samples, while Cluster 1 was split relatively evenly with respect to ecosystem label (Fig. S 2-11A, C, E). Top marker KOs, based on median fold-change, for Cluster 2 were associated with the metabolism of compatible solutes ectoine and betaine in the marine sub-cluster, while KOs associated with magnesium, sulfate, potassium transport were enriched in the freshwater sub-cluster. The Cluster 3 digestive system associated sub-cluster had numerous marker KOs associated with sporulation. Finally, Cluster 1 markers were associated with either fungal metabolism (e.g. chitin and 1,3-beta-glucan synthesis,

yeast amino acid transporter, and other eukaryotic cell functions) in one sub-cluster while a multicomponent $K^+ : H^+$ antiporter involved in bacterial pH adaptation and the Rnf H^+ / Na^+ -translocating ferredoxin: NAD^+ oxidoreductase complex were top markers in the other sub-cluster.

Discussion

Metagenomics is an increasingly valuable tool for interrogating the structure and function of microbiomes, nonetheless analysis of this complex data across large sample sets remains challenging. Here, we focus on KO abundance profiles as a consistent and low-dimensional embedding of functional potential to enable comparison across a large set of metagenomes. This dimension reduction results in our entire dataset occupying a few hundred megabytes of storage while the associated raw sequencing data likely occupies hundreds of terabytes. Some of the earliest metagenomics studies made similar observations of ecosystem-specific clustering of function using only a few samples^{13,14}. We extend these findings to a much larger and more diverse dataset and analyze cluster-specific marker KOs. Splitting the markers into their associated KEGG pathway categories highlighted which metabolic functions defined each cluster, and by extension their respective environments. Broadly we found that the availability of oxygen for respiration and light for photosynthesis were strong drivers of differentiation (aquatic vs terrestrial vs sludge). Furthermore, KOs related to domain-specific cellular functioning were enriched in different clusters (i.e. archaea in Cluster 3, eukaryotes in Cluster 2). Finally, the top cluster-specific marker KOs pointed at several interesting adaptations to different environments.

Cluster 1 was dominated by soils and characterized by marker genes related to the degradation of a diverse set of aromatic compounds as well as antibiotic synthesis and resistance. There were marker genes associated with the breakdown of vanillate, syringate, catechol, and

anthranilate which are all key intermediates in the bacterial degradation of lignin²⁸. Lignin is an important component of plant cell walls that accounts for approximately 30% of global organic carbon stocks²⁹, as such it represents a key carbon source in soils^{28,30}. Interestingly, the *prm* propane monooxygenase was one the most differentially abundant genes in Cluster 1. In addition to propane oxidation³¹, studies have associated this enzyme with N-nitrosodimethylamine degradation³² and 1,4-Dioxane degradation³³, suggesting that it may participate in the breakdown of other recalcitrant compounds in the environment. The apparent importance of this enzyme complex in soils warrants a more detailed exploration of its diversity and function. Previous reports have identified soils as rich reservoirs of both antibiotics^{13,34} and antibiotic resistance^{35,36} and our results reinforce that this as a defining functional characteristic of soils compared to other ecosystems. Lastly, sub-clustering suggested soils could be grouped based on fungal versus bacterial abundance, possibly driven by pH, which has been shown to be one of the strongest drivers of microbial community assembly in soils^{15,37}.

Cluster 2 contained almost exclusively freshwater and marine samples and was clearly distinguished by genes related to photosynthesis and associated pigments. This cluster had the highest proportion of eukaryotic contigs and had marker KOs associated with eukaryotic cell biology presumably from an abundance of algae. Aside from photosynthesis, dimethylsulfoniopropionate (DMSP) demethylase was one of the top marker genes. DMSP is an important osmolyte for phytoplankton that can account for as much as 10% of their fixed carbon and also acts as a key carbon source for bacteria³⁸. The identification of the DMSP demethylase as one of the top marker genes illustrates the extent to which phototrophs shape the food web in aquatic systems. Sub-clustering divided freshwater and marine metagenomes based on the metabolism of compatible solutes and transport of inorganic ions, highlighting the unique

adaptations of microbial life to aquatic environments of differing ionic strength. Furthermore, these observations illustrate how the clustering method can reveal finer scale differences in functional potential when the broader environmental context is held constant.

Cluster 3 consisted of a variety of anaerobic environments from aquatic sediments to anaerobic digestors and gut microbiomes. This points to the conservation of anaerobic metabolisms (i.e. methanogenesis, sulfur reduction, and fermentation) in spite of the broader environmental context. Indeed, an evolutionary study of the CODH/ACS complex, a key enzyme in the Wood-Ljungdahl pathway of carbon fixation and the top marker KO for Cluster 3, revealed an astounding level of conservation across many anaerobic bacterial and archaeal lineages³⁹. Cluster 3 harbored numerous marker genes specific to archaeal cell functioning and had the highest proportion of archaeal contigs, highlighting the unique dominance and adaptation of this domain of life in anaerobic settings. We also found many CRISPR-associated genes as markers for Cluster 3, which is in alignment with a study of isolate genomes that found CRISPRs were more prevalent in anaerobes compared to aerobes⁴⁰. This suggests that defense against phage predation is especially important in anaerobic settings in contrast to Cluster 1 where genes for antibiotic production and resistance were uniquely abundant. Cluster 3 had one sub-cluster associated with digestive system samples and spore formation. These environments may experience fluctuating conditions that require sporulation for the survival of anaerobes compared to the anaerobic sediments, found in the other sub-cluster, that have more steady environmental conditions.

We observed significant differences in GC content across functional clusters, which corroborates previous reports of metagenome GC content partitioning across environments^{41,42}. This phenomenon has been discussed from a variety of viewpoints, such as differences in carbon

to nitrogen ratios, oxygen requirements, genome size, and DNA repair pathway⁴³. Using our functional gene approach, we found that NHEJ-related genes were correlated with high GC content while *polV* and a cytosine methyltransferase were correlated with low GC content. The role that NHEJ and *polV* play in shaping GC content has been explored in isolate genomes^{43,44}. In addition, methylated cytosine bases are mutational hotspots⁴⁵ that have been suggested to play a role in the evolution of GC content in vertebrate genomes⁴⁶. Our results suggest that differences in DNA repair and methylation play a role in shaping GC content even at the whole community level. This reinforces the notion that environmental conditions shape community functional potential, which in turn may contribute to the evolution of distinct genomic traits, such as GC content.

While this approach provides an insightful glimpse into the functional potential of microbial communities across ecosystems, it is still limited by our ability to comprehensively assemble and annotate short metagenomic sequencing reads. The sequence assembly difficulties are particularly acute in soil samples due to their immense diversity. Furthermore, we are only able to assign KOs to roughly 40% of genes in this dataset. A key question becomes, is the composition of microbial dark matter⁴⁷, what we cannot assemble or annotate, vastly different from what we can assemble and annotate? Given the functional redundancy of microbial communities, it's possible that the elucidation of the "dark matter" will not significantly change our overall view of community function. Lastly, it's worth noting that metagenomes only tell us about functional potential and should not be misconstrued as proxies of actual activity, for which proteomics, transcriptomics, or stable-isotope probing would be more appropriate.

Overall, this study demonstrates the value of ortholog abundance profiles for representing microbial communities and supports their use in modeling community function. Ortholog

abundance-based models have shown promise, for example, in explaining geochemical patterns of nutrient cycling in the ocean^{17,48,49} and in predicting metabolic preference for sugars versus amino and organic acids in a collection of isolates⁵⁰. This approach takes advantage of the functional redundancy of microbial communities, that is the fact that functional traits are more stable than taxonomy^{51,52}. Focusing on function simplifies the comparison across microbial communities as evidenced by this study, which includes thousands of metagenomes. The dataset provides a valuable resource for the exploration of the functional gene landscape across a range of ecosystems. It is amendable to expansion with additional samples and metadata describing key parameters like pH, temperature, and oxygen availability. Integrating ortholog abundance profiles with quality metadata will bring us closer to a predictive understanding of how microbial communities assemble and function in the face of varying environmental conditions.

Acknowledgements

We thank all of the researchers involved in the sample collection and processing of the metagenomes used in this study. We thank JGI for developing and maintaining the database infrastructure that made this analysis possible. This work was funded by the US Department of Energy under award number DE-SC0020356.

Data and code availability statement

Compiled ortholog abundance profiles and metadata with associated ascension numbers are available from the associated figshare project (https://figshare.com/projects/Metagenome_functional_clustering/187989). Analysis code can be found in the associated github repository

(https://github.com/zflink/Metagenome_functional_clustering). Raw metagenomic data and ortholog abundance profiles are available from the JGI IMG/M database.

References

1. Chen, I. M. A. *et al.* The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res* **51**, D723–D732 (2023).
2. Richardson, L. *et al.* MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res* **51**, D753–D759 (2023).
3. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, (2015).
4. Wu, Y. W., Tang, Y. H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, (2014).
5. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* **3**, 804–813 (2018).
6. Liu, Y.-F. *et al.* Genomic and Transcriptomic Evidence Supports Methane Metabolism in *Archaeoglobi*. *mSystems* **5**, (2020).
7. Borrel, G. *et al.* Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat Microbiol* **4**, 603–613 (2019).
8. Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science (1979)* **350**, 434–438 (2015).
9. Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
10. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* **39**, 105–114 (2021).
11. Campanaro, S. *et al.* New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnol Biofuels* **13**, (2020).
12. Nayfach, S. *et al.* A genomic catalog of Earth’s microbiomes. *Nat Biotechnol* **39**, 499–509 (2021).
13. Tringe, S. G. *et al.* Comparative Metagenomics of Microbial Communities. *Science (1979)* **308**, 554–557 (2005).

14. Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632 (2008).
15. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
16. Waldrop, M. P. *et al.* Permafrost microbial communities and functional genes are structured by latitudinal and soil geochemical gradients. *ISME Journal* (2023) doi:10.1038/s41396-023-01429-6.
17. Raes, J., Letunic, I., Yamada, T., Jensen, L. J. & Bork, P. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol Syst Biol* **7**, (2011).
18. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* (1979) **348**, (2015).
19. Clum, A. *et al.* DOE JGI Metagenome Workflow. *mSystems* **6**, D723–D733 (2021).
20. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, 457–462 (2016).
21. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol* **19**, (2018).
22. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
23. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).
24. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
25. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–47 (2019).
26. Terpilowski, M. scikit-posthocs: Pairwise multiple comparison tests in Python. *J Open Source Softw* **4**, 1169 (2019).
27. Pham, P., Rangarajan, S., Woodgate, R. & Goodman, M. F. Roles of DNA polymerases V and II in SOS-induced error-prone and error-free repair in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **98**, 8350–8354 (2001).

28. Buchan, A., Neidle, E. L. & Moran, M. A. Diversity of the Ring-Cleaving Dioxygenase Gene *pcaH* in a Salt Marsh Bacterial Community. *Appl Environ Microbiol* **67**, 5801–5809 (2001).
29. Boerjan, W., Ralph, J. & Baucher, M. Lignin Biosynthesis. *Annual Review of Plant Biology* vol. 54 519–546 Preprint at <https://doi.org/10.1146/annurev.arplant.54.031902.134938> (2003).
30. Black, J. E., Wagner, T. & Abbott, G. D. Assessing Lignin Decomposition and Soil Organic Carbon Contents Across a Tropical Savannah-Rainforest Boundary in Guyana. *Frontiers in Forests and Global Change* **4**, (2021).
31. Kotani, T., Yamamoto, T., Yurimoto, H., Sakai, Y. & Kato, N. Propane Monooxygenase and NAD⁺-Dependent Secondary Alcohol Dehydrogenase in Propane Metabolism by *Gordonia* sp. Strain TY-5. *J Bacteriol* **185**, 7120–7128 (2003).
32. Sharp, J. O. *et al.* An inducible propane monooxygenase is responsible for N-nitrosodimethylamine degradation by *Rhodococcus* sp. strain RHA1. *Appl Environ Microbiol* **73**, 6930–6938 (2007).
33. Deng, D., Li, F. & Li, M. A Novel Propane Monooxygenase Initiating Degradation of 1,4-Dioxane by *Mycobacterium dioxanotrophicus* PH-06. *Environ Sci Technol Lett* **5**, 86–91 (2018).
34. Singh, H. W. *et al.* Metagenomic data reveals type I polyketide synthase distributions across biomes. *mSystems* **8**, (2023).
35. Qian, X. *et al.* Metagenomic analysis reveals the shared and distinct features of the soil resistome across tundra, temperate prairie, and tropical ecosystems. *Microbiome* **9**, (2021).
36. Nesme, J. *et al.* Large-scale metagenomic-based study of antibiotic resistance in the environment. *Current Biology* **24**, 1096–1100 (2014).
37. Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME Journal* **4**, 1340–1351 (2010).
38. Moran, M. A., Reisch, C. R., Kiene, R. P. & Whitman, W. B. Genomic insights into bacterial DMSP transformations. *Ann Rev Mar Sci* **4**, 523–542 (2012).
39. Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci U S A* **115**, E1166–E1173 (2018).
40. Weissman, J. L., Laljani, R. M. R., Fagan, W. F. & Johnson, P. L. F. Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME Journal* **13**, 2589–2602 (2019).

41. Foerstner, K. U., von Mering, C., Hooper, S. D. & Bork, P. Environments shape the nucleotide composition of genomes. *EMBO Rep* **6**, 1208–1213 (2005).
42. Chuckran, P. F., Hungate, B. A., Schwartz, E. & Dijkstra, P. Variation in genomic traits of microbial communities among ecosystems. *FEMS Microbes* **2**, (2022).
43. Wu, H., Zhang, Z., Hu, S. & Yu, J. On the molecular mechanism of GC content variation among eubacterial genomes. *Biol Direct* **7**, (2012).
44. Weissman, J. L., Fagan, W. F. & Johnson, P. L. F. Linking high GC content to the repair of double strand breaks in prokaryotic genomes. *PLoS Genet* **15**, (2019).
45. Duncan, B. K. & Miller, J. H. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560–561 (1980).
46. Mugal, C. F., Arndt, P. F., Holm, L. & Ellegren, H. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. *G3: Genes, Genomes, Genetics* **5**, 441–447 (2015).
47. Pavlopoulos, G. A. *et al.* Unraveling the functional dark matter through global metagenomics. *Nature* **622**, 594–602 (2023).
48. Louca, S. *et al.* Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proc Natl Acad Sci U S A* **113**, E5925–E5933 (2016).
49. Reed, D. C., Algar, C. K., Huber, J. A. & Dick, G. J. Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proc Natl Acad Sci U S A* **111**, 1879–1884 (2014).
50. Gralka, M., Pollak, S. & Cordero, O. X. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nat Microbiol* (2023) doi:10.1038/s41564-023-01458-z.
51. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science (1979)* **353**, 1272–1277 (2016).
52. Louca, S. *et al.* Function and functional redundancy in microbial systems. *Nat Ecol Evol* **2**, 936–943 (2018).

Figures

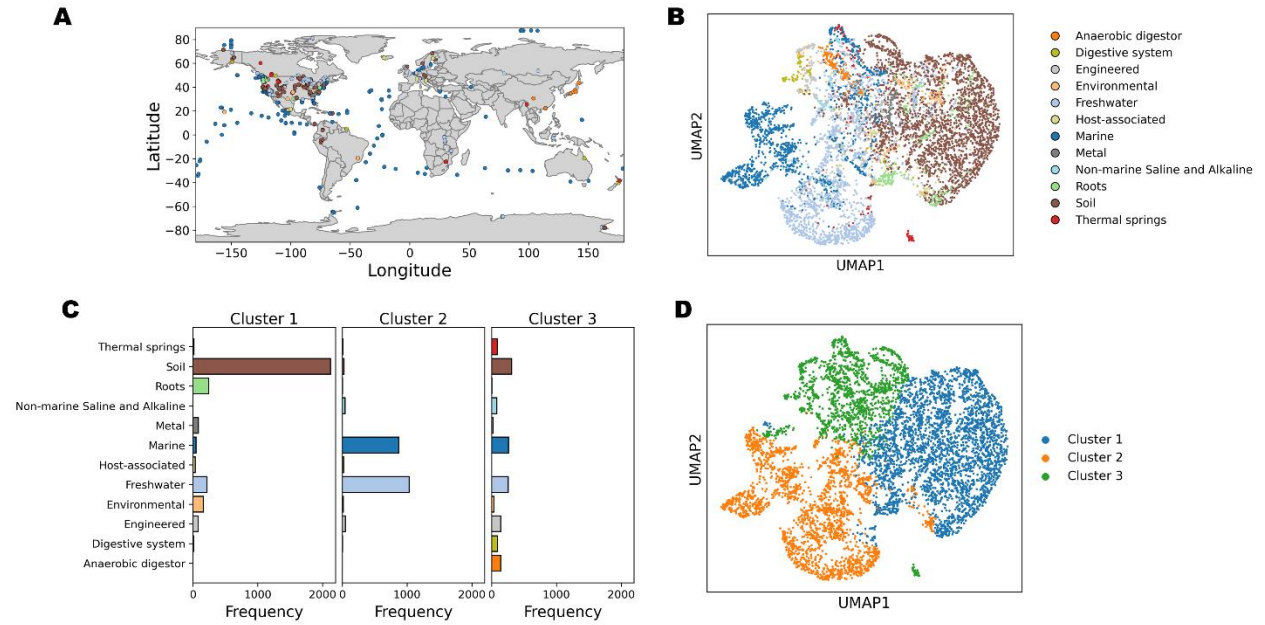


Figure 2-1 – Global distribution of dataset and functional clustering.

(A) Sampling location colored by ecosystem label plotted on a global map. (B) UMAP projection of the normalized KO profiles colored by ecosystem label. (C) Frequency of different ecosystem labels in the three k-means clusters. (D) The same UMAP projection as panel (B) but colored by cluster membership.

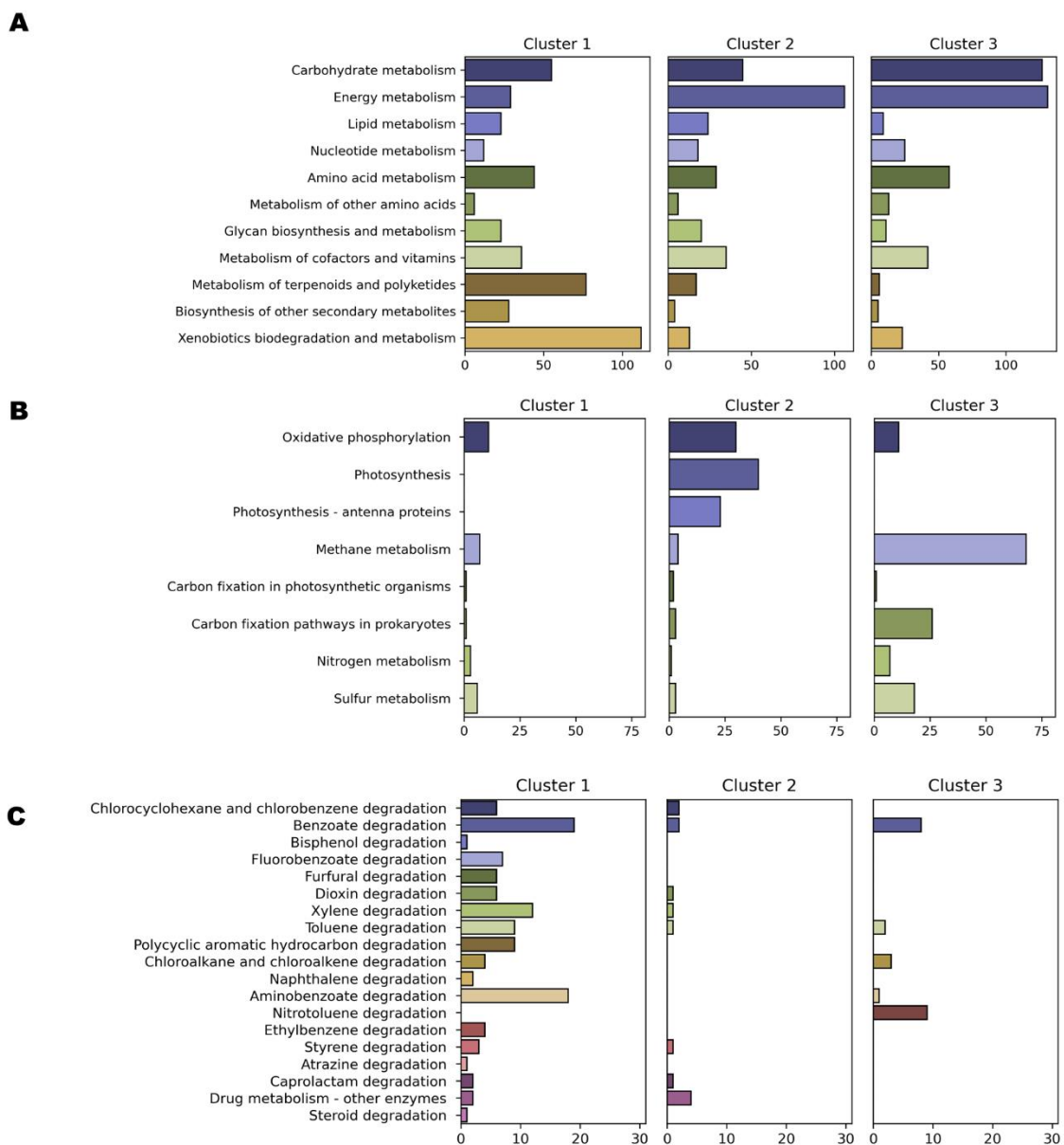


Figure 2-2 – KEGG pathways associated with cluster-specific marker genes.

Number of cluster-specific marker genes belonging to the different KEGG Metabolism pathways

(A), Energy metabolism pathways (B), and Xenobiotics degradation and metabolism pathways

(C).

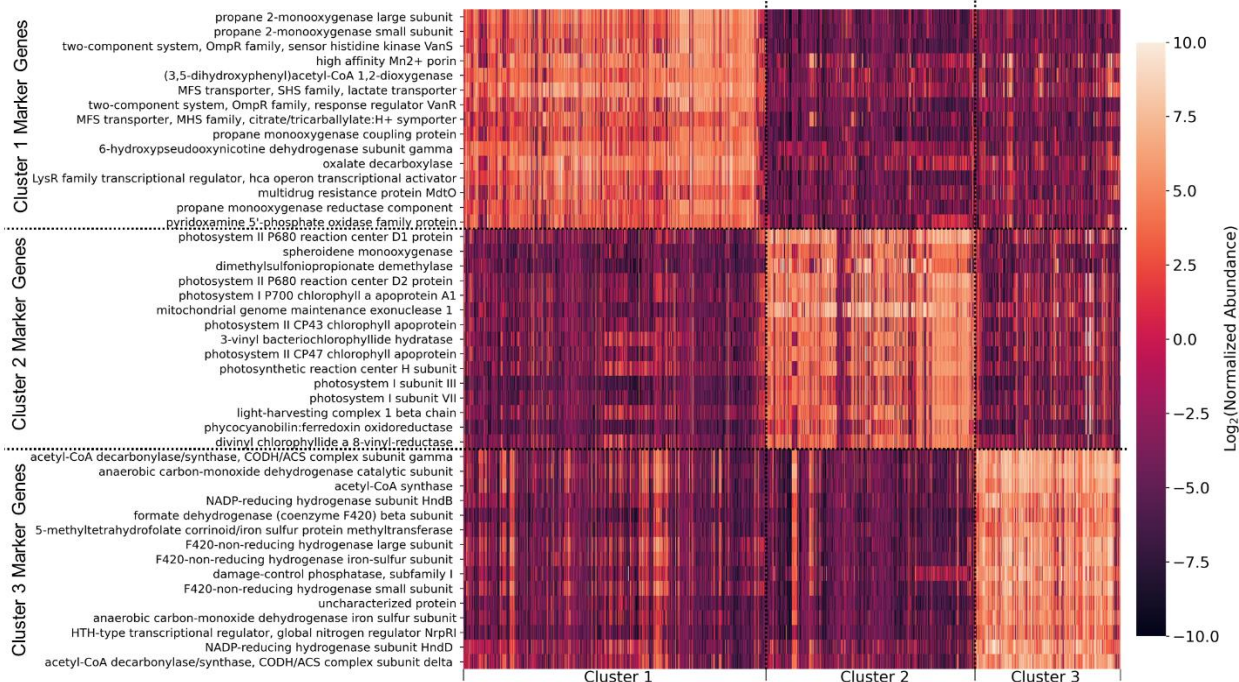


Figure 2-3 – Top 15 cluster marker genes based on median fold-change.

Each row represents a gene, and each column represents a metagenome, sorted by cluster. Cells are colored according to the clr normalized KO abundance in each sample, where zero represents a KO abundance equal to the geometric mean of the sample.

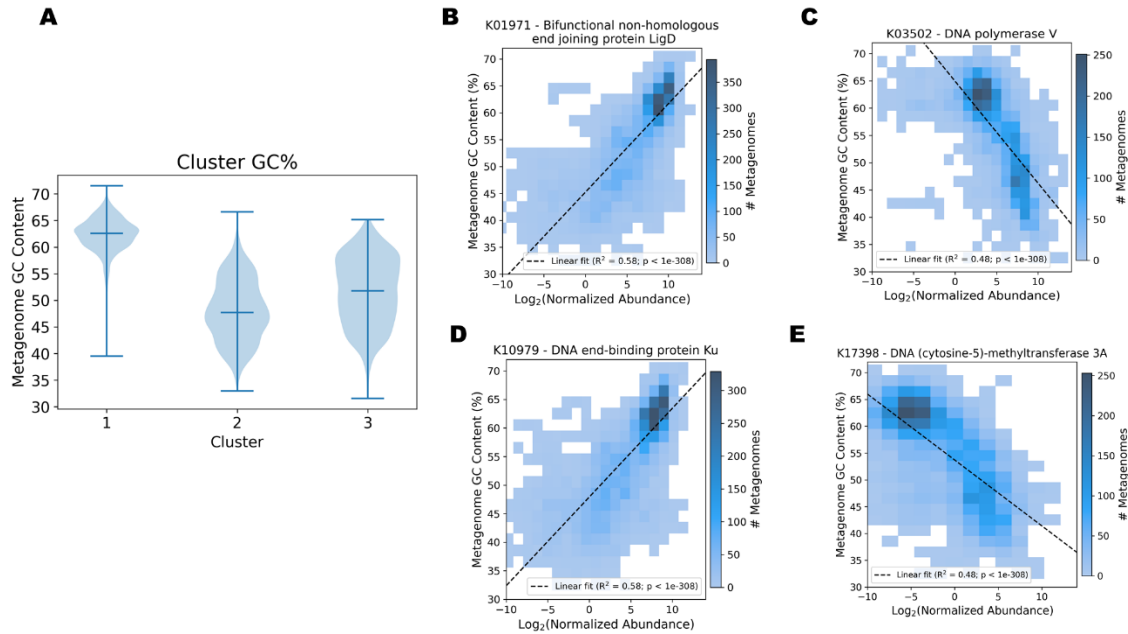


Figure 2-4 – GC content variation and correlated genes of interest.

(A) Distribution of assembled metagenome GC content in different clusters. Binned scatter plots showing GC content versus the abundance of the bifunctional non-homologous end joining protein LigD (B), DNA polymerase V (C), DNA end-binding protein Ku (D), and DNA (cytosine-5)-methyltransferase 3A (E).

Supplementary figures

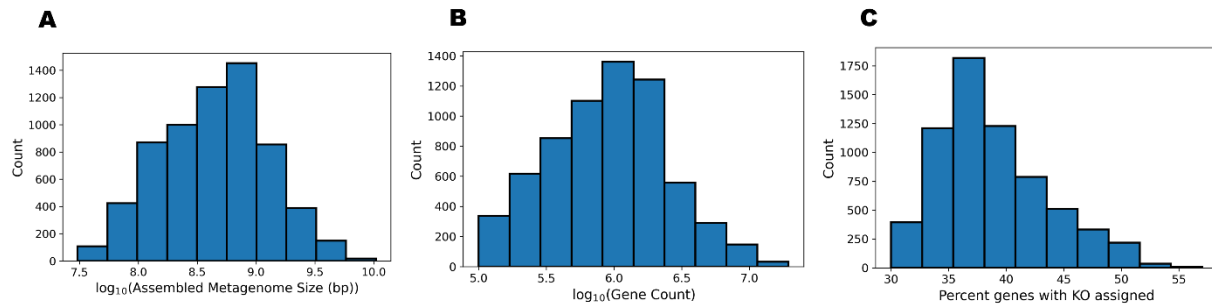


Figure S 2-1 – Characteristics of dataset metagenomes

(A) Histogram of assembled metagenome size in base pairs. (B) Histogram of assembled metagenome gene count. (C) Histogram of percent of genes with KO assigned.

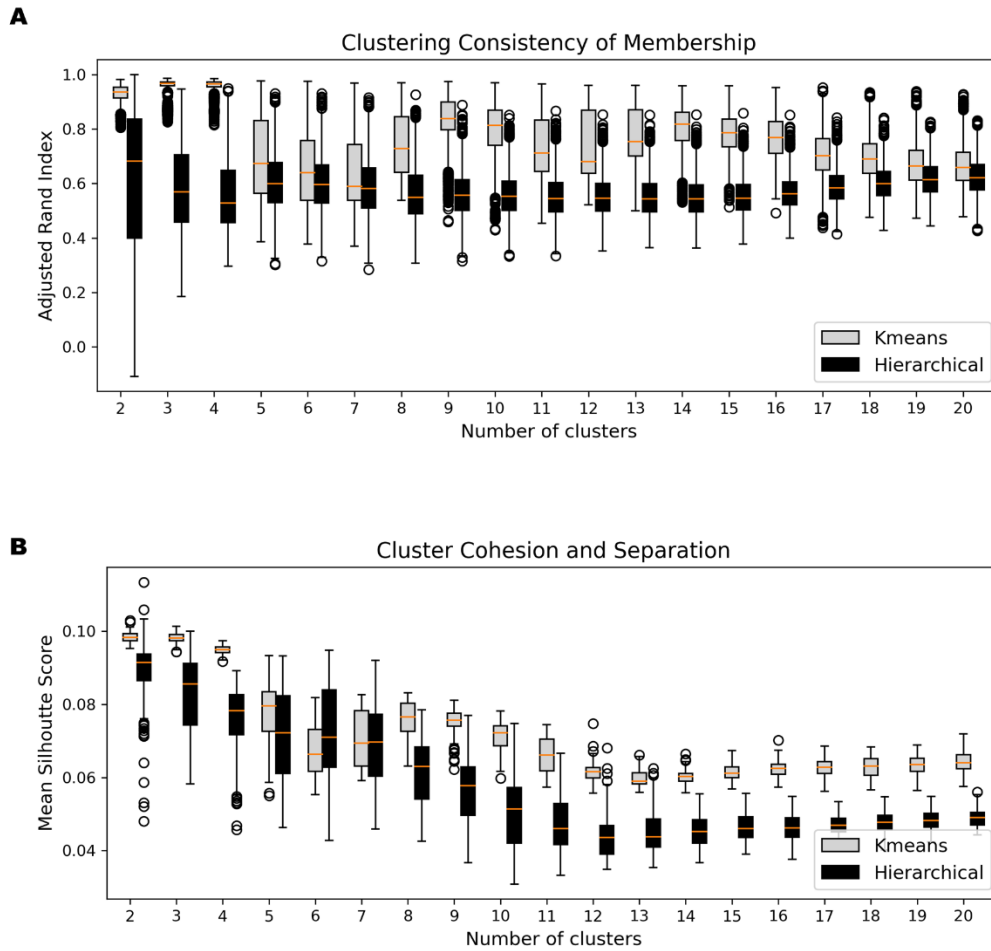


Figure S 2-2 – Bootstrap testing of k-means and Ward’s hierarchical clustering methods over a range of cluster numbers.

(A) Adjusted Rand Index results from clustering of 100 bootstrap samples of the dataset over a range of 2 to 20 clusters. (B) Mean Silhouette Coefficient results from the same clusters.

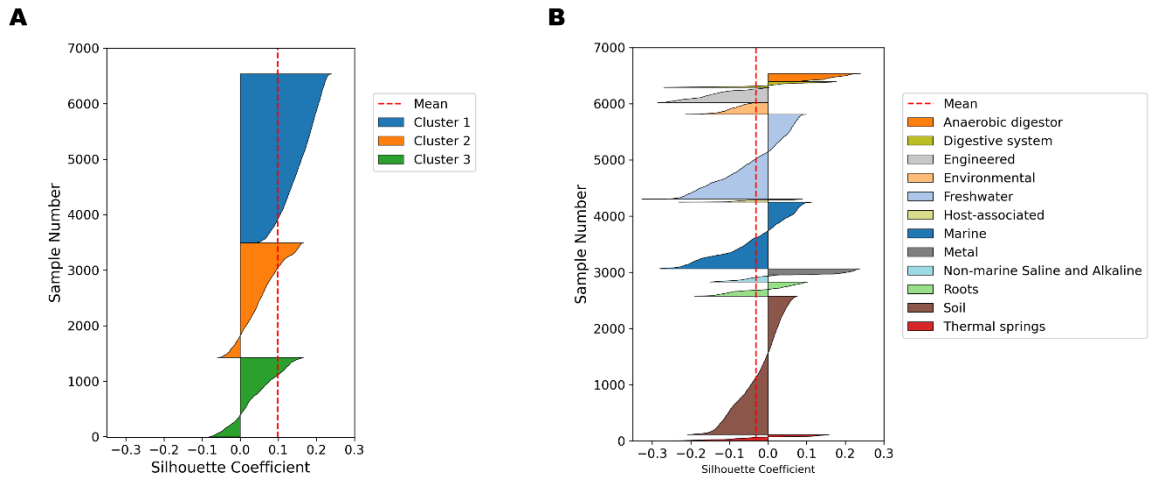


Figure S 2-3 – Distribution of Silhouette coefficients across the dataset.

Silhouette coefficient values for each sample based on k-means (A) versus ecosystem label grouping (B). Mean Silhouette Coefficient for all samples is shown with red dashed line.

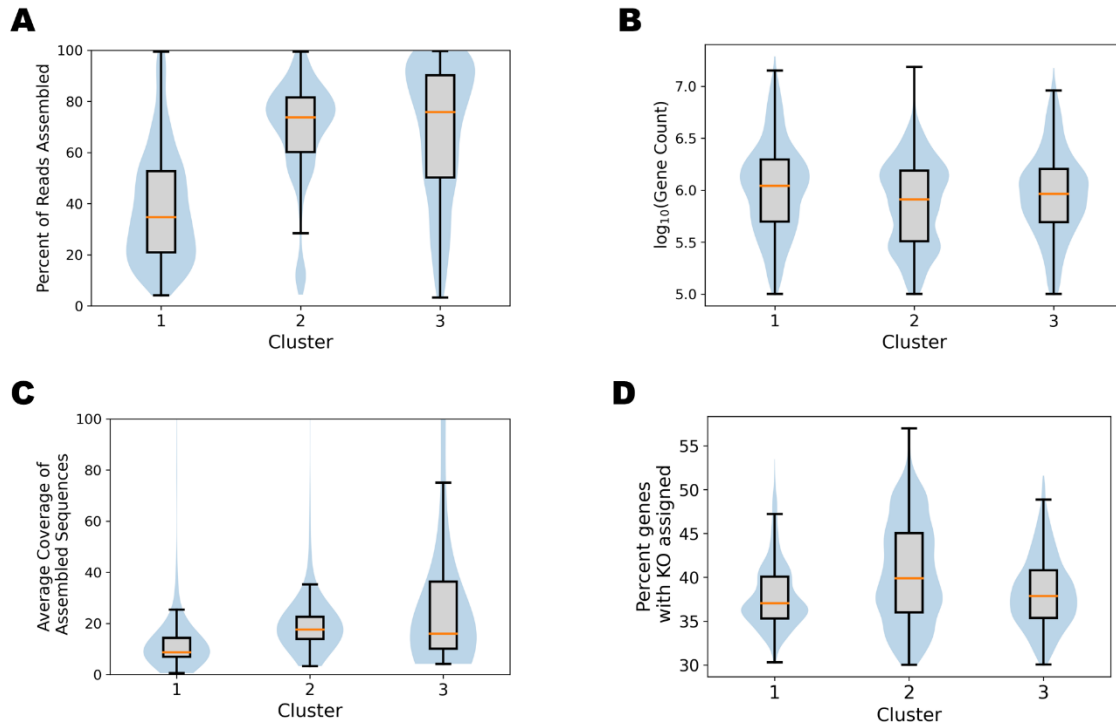


Figure S 2-4 – Characteristics of metagenomes separated by cluster assignment.

Boxplot and violin plots, broken down by cluster, showing (A) the percent of sequencing reads assembled, (B) \log_{10} of gene count, (C) average coverage of assembled sequences, and (D) percent of genes with a KO assigned.

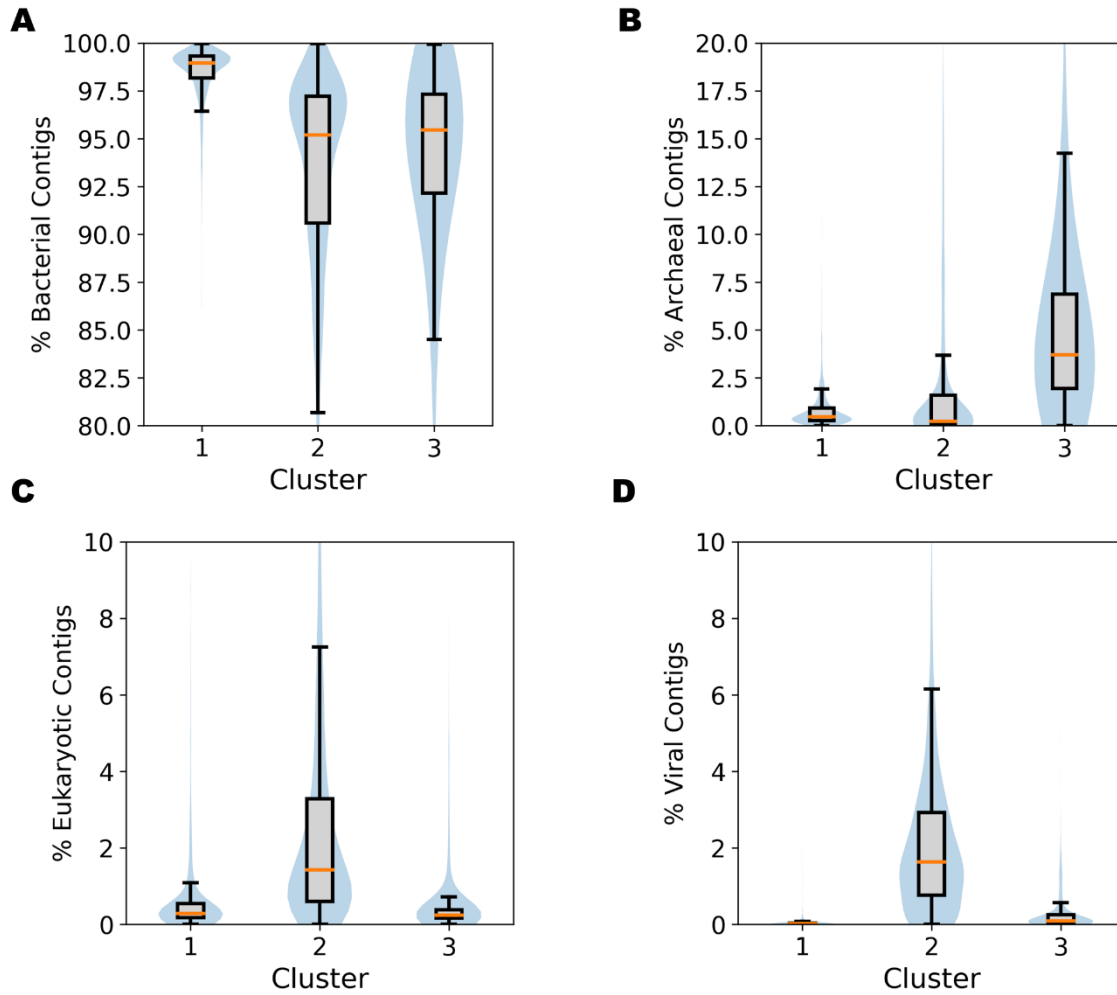
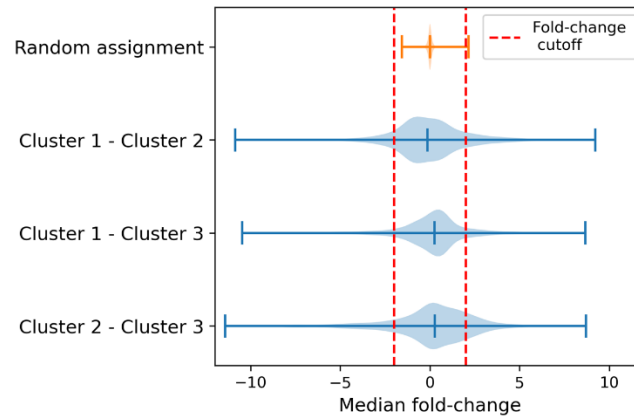
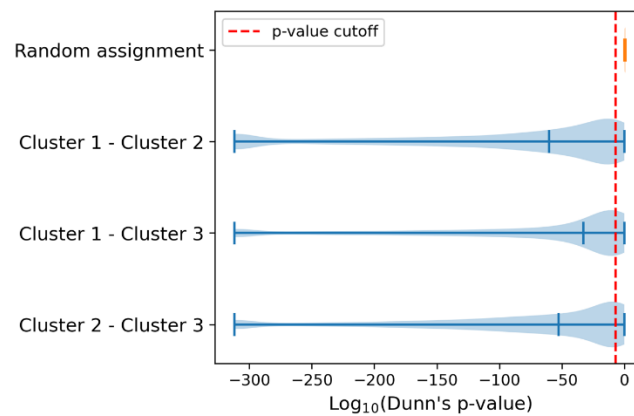


Figure S 2-5 – Domain-level taxonomic composition of metagenome contigs in each cluster.

Boxplot and violin plots showing (A) the percent of contigs assigned to bacteria, (B) archaea, (C) eukaryotes, and (D) viruses. Data derived from JGI IMG/M’s taxonomic composition tables which assign contig phylogeny based on database homology search of protein-coding genes.

A**B****Figure S 2-6 – Assignment of cluster-specific marker genes.**

(A) Distributions of median fold-change in gene abundance between clusters compared to if clusters were randomly assigned with fold-change cutoff shown in red. (B) Distributions of Dunn's test p-values compared to random assignment with p-value cutoff shown in red. Results based on 100 random cluster assignments.

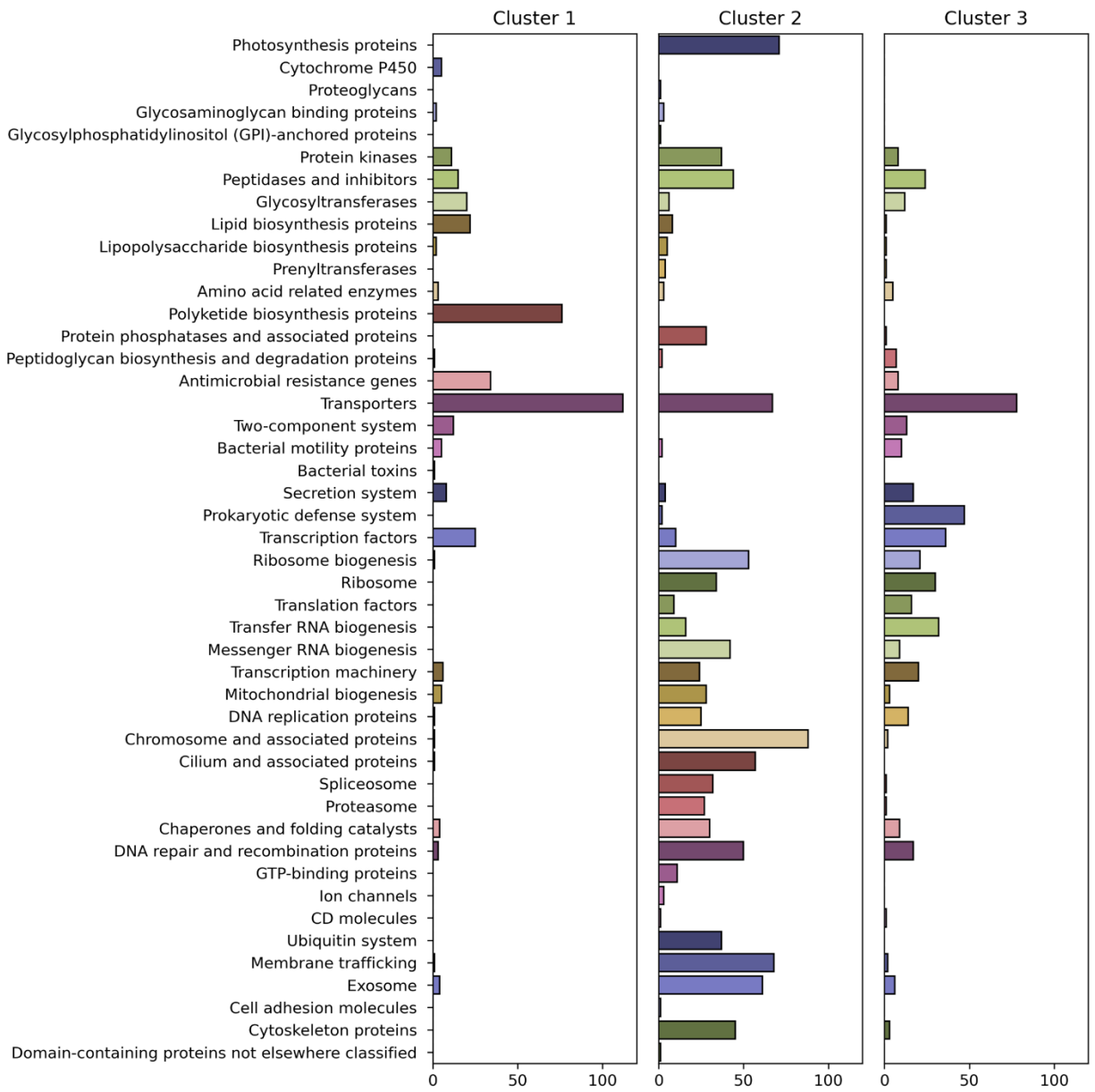


Figure S 2-7 – Number of cluster-specific marker genes associated with KEGG Brite Categories.

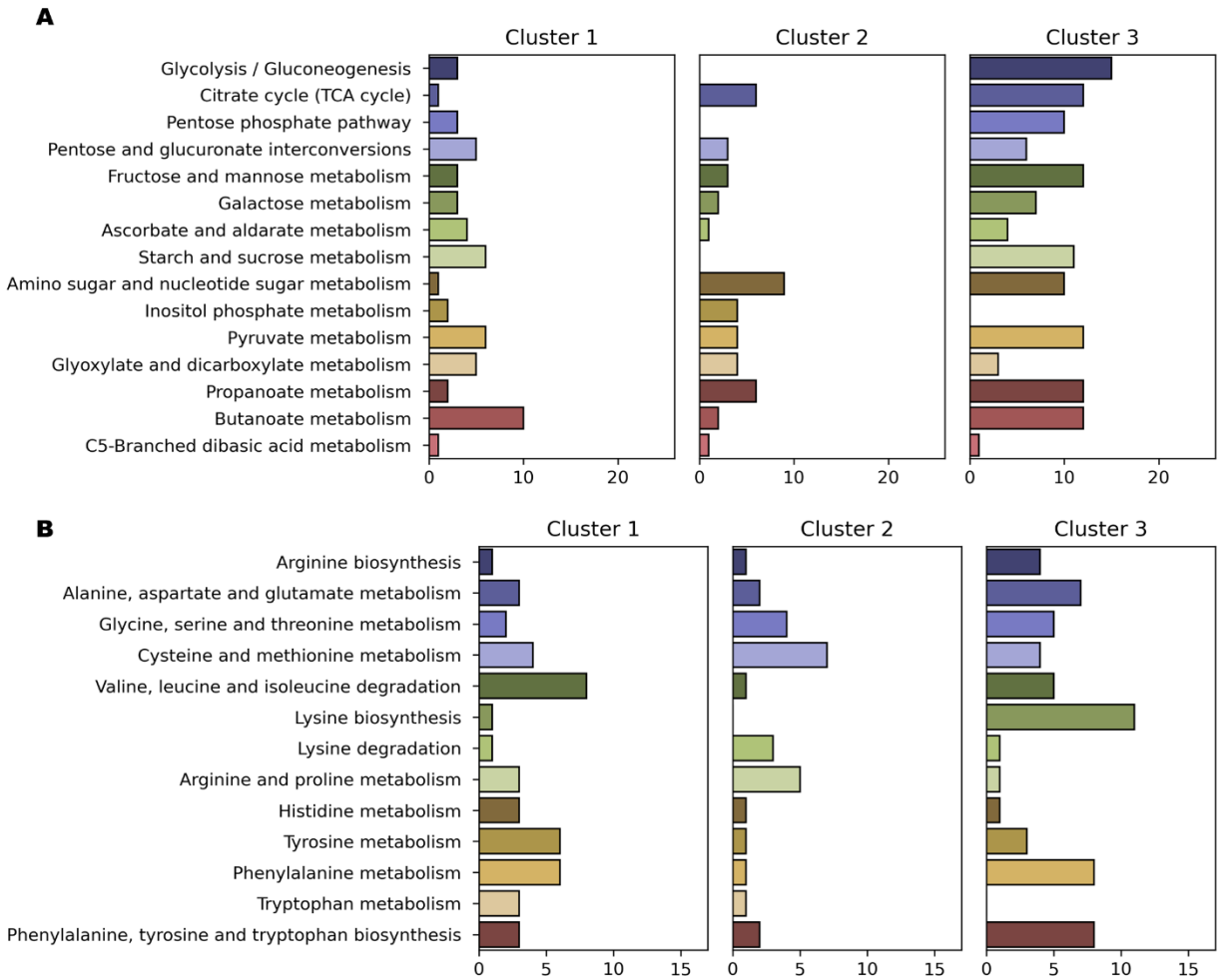


Figure S 2-8 – KEGG carbohydrate and amino acid metabolism.

Number of cluster-specific marker genes associated with KEGG carbohydrate metabolism (A) and amino acid metabolism (B).

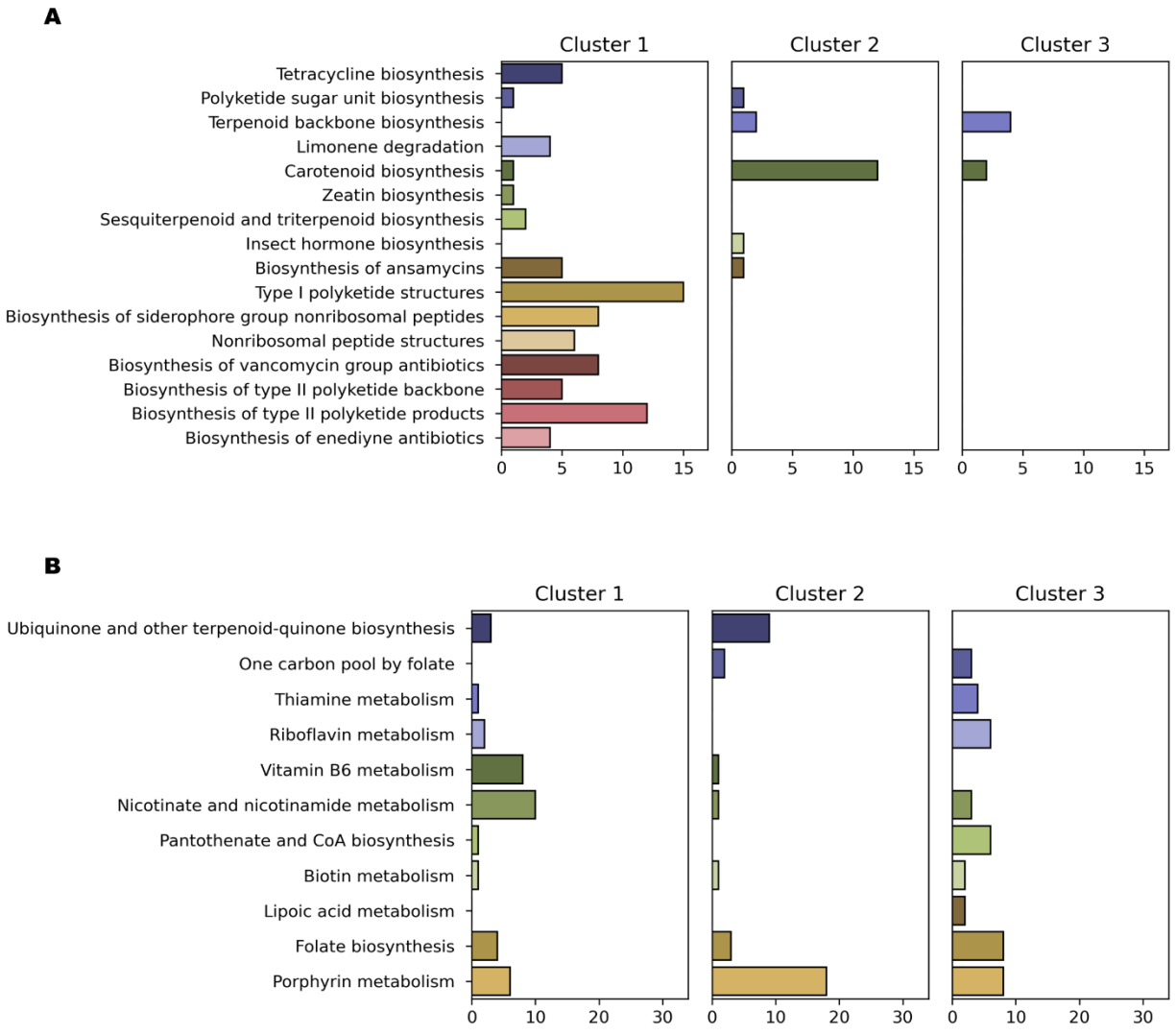


Figure S 2-9 – KEGG terpenoid, polyketide, and vitamin metabolism.

Number of cluster-specific marker genes associated with KEGG metabolism of terpenoids and polyketides (A) and vitamins and cofactors metabolism (B).

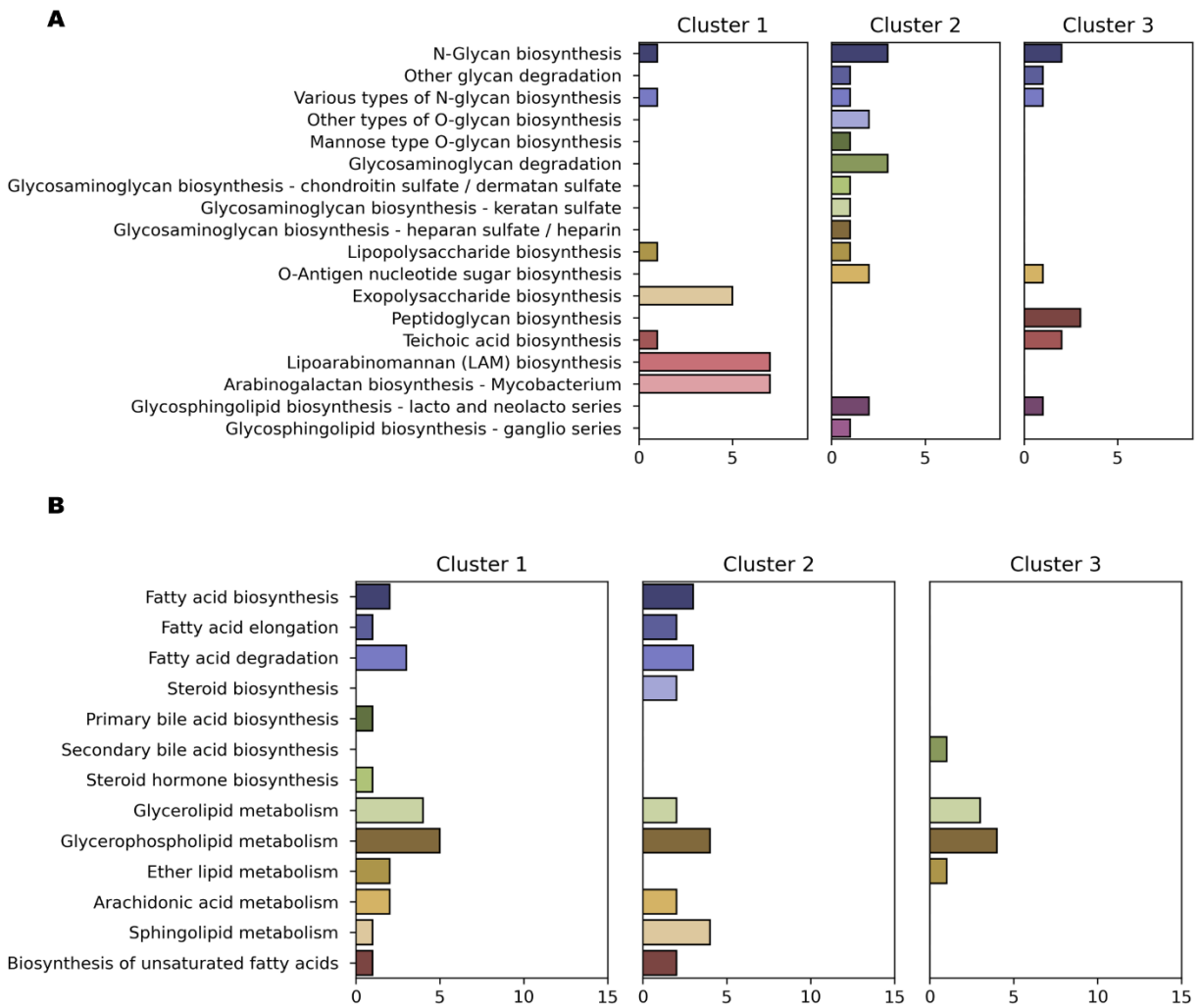


Figure S 2-10 – KEGG glycan and lipid metabolism.

Number of cluster-specific marker genes associated with KEGG glycan biosynthesis and metabolism (A) and lipid metabolism (B).

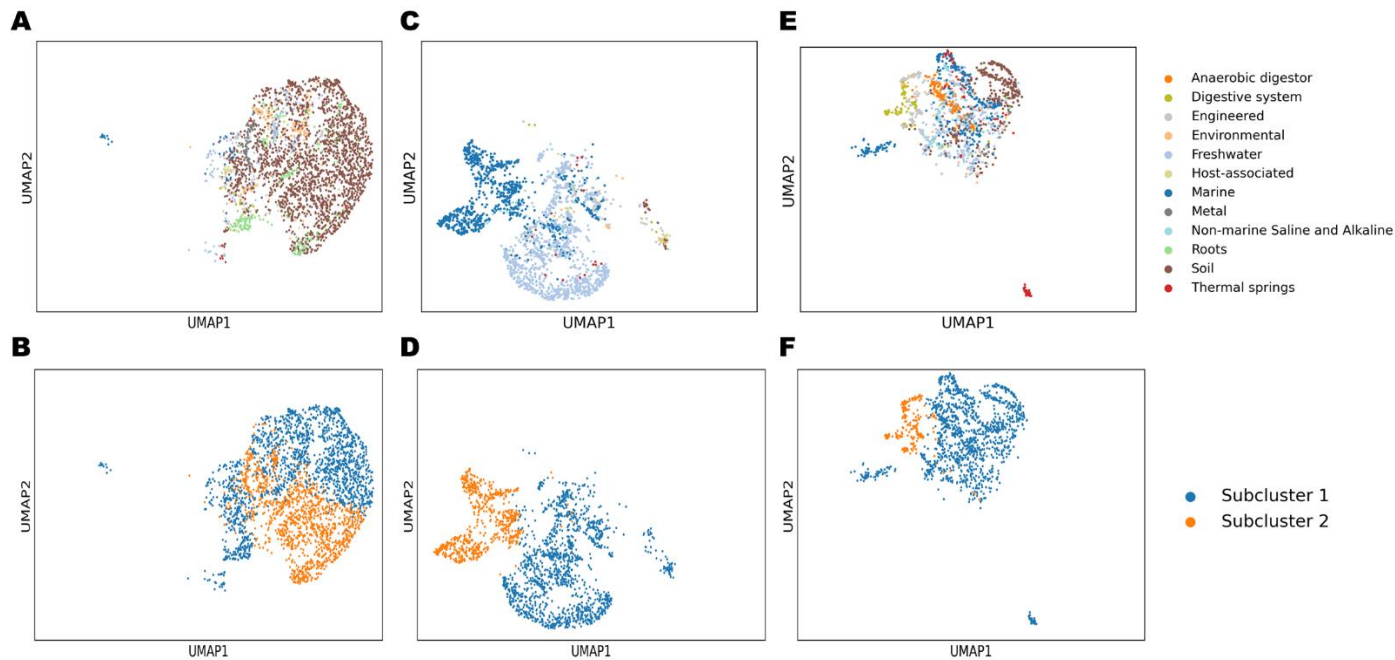


Figure S 2-11 – Sub-clustering analysis.

UMAP plots, using the same projection coordinates as Fig. 1B, showing just Cluster 1 (A) and the sub-clusters of Cluster 1 (B), Cluster 2 (C) and the sub-clusters of Cluster 2 (D), and Cluster 3 (E) and the sub-clusters of Cluster 3 (F).

Chapter 3 – Soil depth shapes microbial community composition and function in a freshwater wetland

Zachary Flinkstrom^{1*}, Pieter Candry^{1,2}, Bruce Godfrey¹, Britt Abrahamson¹, Dongyu Wang³, Katherine B. Louie^{4,5}, Benjamin P. Bowen^{4,5}, Trent R. Northen^{4,5}, Chongle Pan^{3,6}, Mari-Karoliina H. Winkler¹

¹ Department of Civil and Environmental Engineering, University of Washington, Seattle, WA

² Present address: Laboratory of Systems & Synthetic Biology, Wageningen University & Research, Wageningen, Netherlands

³ School of Biological Sciences, University of Oklahoma, Norman, OK

⁴ Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

⁵ Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

⁶ School of Computer Science, University of Oklahoma, Norman, OK

*Correspondence to: Zachary Flinkstrom, Civil and Environmental Engineering, University of Washington, 201 More Hall, Box 352700, Seattle, WA 98195-2700, USA; E-mail: zflinky@gmail.com

Abstract

Wetland microbiomes play a crucial role in the global carbon cycle by modulating soil organic carbon (SOC) and greenhouse gas (GHG) emissions. Understanding how microbial communities respond to environmental changes is essential for predicting wetland carbon fluxes under future climate scenarios. Here, we investigated the biogeochemistry of a temperate lacustrine wetland across four seasons and three soil depths, integrating greenhouse gas flux measurements, porewater metabolite profiles, metagenomics, metabolomics, and metaproteomics. While seasonal shifts in GHG fluxes and porewater chemistry were evident, microbial community composition and function were primarily structured by soil depth, suggesting resilience to short-term seasonal fluctuations and a high degree of metabolic plasticity. Depth-correlated microbial taxa and metabolic pathways revealed distinct stratification: surface soils were enriched in metabolically versatile Gammaproteobacteria capable of oxygen and nitrate respiration, as well as methane and sulfur oxidation, whereas deeper layers favored strict anaerobic metabolism, with increasing abundances of Anaerolinea and Methanomicrobia. Metabolomics showed an enrichment of purine nucleotides and amino acids at the surface, while deeper soils accumulated amino sugars and phenolic compounds, highlighting differences in carbon processing. Metaproteomics confirmed active metabolic pathways, linking functional potential to microbial activity. By integrating multi-omics with biogeochemical measurements, this study provides a system-level view of wetland microbial function and resilience, contributing to predictive models of wetland carbon cycling under future climate scenarios.

Introduction

Wetlands play a disproportionate role in the global carbon cycle, storing approximately 20% of terrestrial carbon despite covering only 5–8% of Earth’s land surface¹. They also contribute around 30% of global methane emissions^{2,3}. Positioned at the interface between aquatic and terrestrial environments, wetlands harbor unique ecological and biogeochemical dynamics. However, these ecosystems face growing threats from land use and climate change, which risk disrupting their carbon balance^{3–5}. Hydrological disturbances—such as wetland drainage, drought, and flooding—can also significantly alter sediment biogeochemistry, affecting carbon storage and greenhouse gas emissions¹. Understanding how wetlands respond to these stressors is critical for predicting their role in future climate scenarios.

Driving the wetland carbon cycle is a dynamic and rich microbial community, adapted to low-oxygen conditions where alternate electron acceptors must be used to sustain energy metabolisms. The classic redox ladder concept posits that along the depth-profile of a saturated sediment, the concentration of the most thermodynamically favorable electron acceptors will be depleted in succession: oxygen will decrease first, followed by nitrite, manganese, nitrate, iron, sulfate, and carbon dioxide^{6,7}. In reality, the picture is more complicated than a simple layered structure. In a variably saturated soil, the microbiome experiences fluctuating redox conditions, while soil structure and biofilms create small-scale redox gradients on different spatial axes^{8–10}. Still, depth is a strong driver of microbial community assembly, especially in anoxic sediments where deeper layers approach the thermodynamic limits of life¹¹.

Changes in water table level, temperature, and aquatic chemistry on multiple time and length scales also affect wetland carbon cycling dynamics. For example, a meta-analysis found that a critical water table level of 50 cm above the soil surface is most favorable for methane

emissions due to the interplay of microbial activities and gas transport¹². Furthermore, the strong temperature dependence of methanogenesis drives seasonal peaks in wetland methane emission in the summer months¹³. By studying the microbial response to these fluctuations, we gain mechanistic insight into the micro-scale processes that drive ecosystem-scale changes in the carbon cycle.

Studies combining metagenomics with field measurements have provided valuable insight into microbial-mediated nutrient cycling in wetlands. For example, this approach has been used in various wetland types to investigate the nature of alcohol cycling and sulfate reduction^{14,15}, to identify potentially oxygen-tolerant methanogens¹⁶, to disentangle the effects of salinity and sulfate on methane and sulfur metabolism^{17,18}, to study the functional genes associated with methane emissions in a restored wetland complex¹⁹, and to analyze carbon processing in thawing permafrost²⁰. Integrating metagenomics with metaproteomics and metabolomics to study the wetland microbiome at different soil depths across seasons can reveal the microbes and processes that control nutrient cycles under changing temperature and redox conditions. It could also indicate whether the community is "resistant, resilient, or redundant"^{21,22} in response to environmental change. For instance, does a dropping water table cause deeper soil layers to transition to a more surface-like state, or does the same community adapt by altering its metabolism?

To investigate this, we characterized the biogeochemistry of a seasonally flooded temperate freshwater wetland, incorporating measurements of porewater chemistry, surface greenhouse gas flux, metagenomics, metabolomics, and metaproteomics. By linking multi-omics data with site-level information across four seasons and three soil depths, we gained insights into the wetland's microbiome and its response to natural changes across temporal and spatial scales.

We hypothesized that significant compositional and functional shifts would be evident across both seasons and depths. Interestingly, we found that depth was a much stronger determinant of microbial community composition, whereas broader site geochemistry varied substantially with season and depth. This study sheds light on the composition and function of wetland microbial communities under varying environmental conditions and highlights the compounds and organisms showing the greatest depth dependence. Broadly, these findings contribute to our understanding of wetland carbon cycling in a changing climate.

Materials and Methods

Study site description and sampling

This study was performed at a seasonally flooded urban freshwater wetland connected to Union Bay of Lake Washington in Seattle, WA, USA (Coordinates: 47.642196° N, -122.296236° W). Lake levels are controlled at 6.1 – 6.7 m above sea level at the Hiram M. Chittenden Locks with draw down in the Fall and refill in the Spring. Porewater samples, soil cores, and surface gas measurements were collected from the same location in October 2021, February 2022, May 2022, and August 2022. All samples were taken within the same variably submerged portion of the wetland. The soil was a hydric silt that contained a high proportion of organic matter (details in Table S1). Site vegetation included *Hydrocotyle* (Floating Pennywort), Lemnoideae (Duckweed), and *Salix* (Willow) (Fig. 3-1a).

Porewater was collected using a porewater equilibrators (peeper) constructed using polyvinyl chloride (PVC) sheet, stainless-steel wire mesh, and an aluminum handle. Holes were drilled in a 3 by 19 matrix in the acrylic sheet to house 5 mL LDPE vials at 19 different depths down to 54 cm (Cat. No. 6250-0005, Thermo Scientific, Waltham, MA). A 0.2 µm pore size 47 mm diameter PES filter (P/N 66234, Pall Corporation, Ann Arbor, MI) was placed over the vial

opening and was secured with a center-punched vial lid. Assembled dialysis vials were then placed into the holes and secured using a mesh screen. The peeper was inserted into the wetland soil such that the top row of vials was just below the soil surface and was left in place for 4 to 5 weeks. The peeper was removed on the day of soil sampling. Vials were immediately removed from the device and placed into an airtight container with two AnaeroGen™ sachets (Thermo Scientific, Waltham, MA) to minimize oxygen exposure. Samples were transported to the laboratory for same-day processing in an anaerobic glove bag (Coy Laboratory Products, Grass Lake, MI).

Indicator of Reduction in Soils (IRIS) tubes^{23,24} were also placed at the site in duplicate in February, May and August, and remained in the soil during the same period that the peeper was placed. Upon removal, tubes were rinsed with DI water and photographed for a qualitative measure of reduction state along the soil profile.

During each sampling campaign, triplicate soil samples were collected using a soil core sampler with 2" x 6" plastic liners (AMS, Inc., American Falls, ID). The core sampler was forced 60 cm into the sediment bed with a slide hammer to cut roots and other plant material. The three replicate soil cores were collected approximately one meter apart from each other, avoiding large tree branches or roots. The cores were sealed air-tight upon removal from the natural habitat, to protect the redox state of samples. Samples were then transported to the lab on ice, for immediate processing.

Surface gas flux measurements

Surface gas flux measurements were performed within 24 hours of soil and porewater sampling using PVC static flux chambers connected to a cavity ring-down spectroscopy-based

gas analyzer (Model G2508, Picarro, Santa Clara, CA). Vinyl tubing (1/4" ID) was used to connect the chamber to the analyzer creating a closed loop system with gas circulating at a flowrate of $230 \text{ mL} \cdot \text{min}^{-1}$. Surface flux was estimated using the slope of the gas concentration versus time data in conjunction with information about the volume and geometry of the system. Triplicate measurements were taken from three distinct chambers, giving a total of nine measurements per sampling campaign. Chambers and tubing were flushed with air between each measurement. Measurements were conducted mid-day between the hours of 10:00 and 14:00.

Porewater chemical analysis

Anions including organic acids (C2-C6, including C4 & C5 isoforms), nitrite, nitrate, sulfate, phosphate, and chloride were analyzed with ion chromatography (IC) on a Dionex ICS 5000+ equipped with an IonPac AS11HC analytical and AS11 guard column at 30°C . Eluent consisted of generated KOH (Dionex EG-5 & Dionex EGCIH KOH, Thermo Fisher Scientific) at a flow rate of $1.5 \text{ mL} \cdot \text{min}^{-1}$. An eluent concentration gradient of 1-60 mM KOH was applied by first running 7 minutes at 1 mM, increasing at (i) $1.56 \text{ mM} \cdot \text{min}^{-1}$ for 9 minutes, then (ii) $1.67 \text{ mM} \cdot \text{min}^{-1}$ for 9 minutes, and eventually (iii) $3.75 \text{ mM} \cdot \text{min}^{-1}$ for 8 minutes to 60 mM, where concentration was held for 90 seconds. Concentration was then decreased as a step function back to 1 mM KOH for 2 minutes before the end of the run. Component concentrations were externally calibrated ($0.5\text{-}25 \text{ mg} \cdot \text{L}^{-1}$) and validated every 10 samples.

Ammonium, total and ferrous iron, magnesium, and calcium were measured on the Gallery discrete analyzer using off-the-shelf Gallery reagent kits (Cat. No: 984362/3, 984326, 984706, 984358, and 984361, Thermo Scientific, Waltham, MA). Dissolved sulfide was determined using the Cline method²⁵ modified for a 96-well plate; samples were first stabilized and diluted in zinc acetate to a final concentration of 12.5 mM in the anaerobic glove bag before

immediate addition of assay reagent, incubation, and absorbance measurement (670 nm) on a plate reader (CLARIOstar Plus, BMG Labtech, Cary, NC). Porewater pH was measured using a standard pH probe (Thermo Scientific, Waltham, MA).

Metabolomics

Metabolomics was performed on soil cores from the May sampling. Cores were extruded from their plastic sleeves onto aluminum foil, with care taken to maintain layering structure. Cores were then divided into three equal sections and 10 g from each section was sampled into separate 15mL Falcon tubes which were frozen at -80C and shipped on dry ice to the Joint Genome Institute (JGI) for metabolomics processing along with empty tubes to act as extraction blanks.

For nonpolar metabolites, 12 mL of 100% methanol was added to each sample in a 50 mL tube following by a brief vortex then sonication in an iced water bath for 30 minutes. Samples were centrifuged (5 min at 5000 rpm), then supernatant transferred to new 50 mL tubes and dried in a SpeedVac (SPD130DLX-115, Labconco). Next, 4 mL methanol were added to the dried extract, vortexed, sonicated 10 minutes in an iced water bath, centrifuged, then supernatant transferred to smaller 5 mL tubes, dried in a SpeedVac then stored at -80 C until ready for LC-MS. Extraction of polar metabolites was performed similarly but using 100% LC-MS grade water instead of methanol, and instead of drying in a SpeedVac, water extracts were frozen then lyophilized dry (FreeZone 12L, #711212010, Labconco). For each type of extraction, 3 empty tubes were also extracted similarly to serve as extraction control samples.

In preparation for LC-MS/MS analysis, dried extracts were resuspended in 150 uL of 100% methanol containing a mixture of isotopically labeled internal standards, centrifuge-

filtered (0.22 μ m hydrophilic PVDF, Ultrafree MC-GV, #UFC30GV25, Millipore), then transferred to an LC-MS vial with glass insert. LC-MS was performed using Thermo Orbitrap mass spectrometer with electrospray ionization (ESI) source coupled to an Agilent 1290 UHPLC for chromatographic separation. Polar metabolites were detected using normal phase chromatography on 3 μ L injection volumes of both the methanol and water extracts, with LC-MS/MS methods as described previously²⁶ with a Thermo QExactive HF mass spectrometer. Nonpolar metabolites were detected using reverse phase chromatography on 3 μ L injection volumes of only the methanol extracts, with LC-MS/MS methods as described previously²⁷ with a Thermo Exploris 120 mass spectrometer. Briefly, MS1 spectra was collected in both positive and negative ionization modes, with MS2 fragmentation spectra collected on the most intense ions if not already fragmented in the previous 4-7 seconds. Sample order was randomized by replicate, with an injection blank of 100% methanol interspersed between each sample and/or internal standard only injection every 3rd sample and quality control mix every 10-15 samples.

To identify metabolites, targeted analysis was performed by comparing 3 criteria, m/z, retention time and fragmentation spectra, from features in experimental data to those from an in-house database of compound standards run using the same LC-MS/MS methods. The highest level of confidence was achieved, Exceeds Level 1²⁸ for features with m/z less than 5 ppm from theoretical, retention time less than 0.5 min and fragmentation spectra matched that of the standard. Mismatching fragmentation spectra invalidated an identification. When fragmentation spectra were not collected but m/z and retention time were acceptable, this reduced confidence to a Level 1. For untargeted analysis, features (chromatographic peaks with unique m/z and RT)

were found using MZmine²⁹ then putative identifications and feature-based networking made using GNPS³⁰ workflows.

Peak heights from the targeted polar metabolite run were analyzed for significant correlations with soil depth according to the Statistical analysis section below. Compounds were excluded if their maximum peak height was less than 2x the maximum peak height of the extraction controls.

DNA extraction and metagenomic sequencing

Soil cores were extruded from their plastic sleeves onto aluminum foil, with care taken to maintain layering structure. Cores were then divided into three equal sections which were sampled into separate 15mL Falcon tubes where they were homogenized with a metal spatula and 250mg was weighed into bead tubes for duplicate extractions. Bead tubes with soil were then frozen at -80°C until extraction using the DNeasy PowerSoil Pro kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The concentration of DNA was quantified using the Qubit™ dsDNA High Sensitivity Assay (Thermo Fisher Scientific, Waltham, MA, USA). Duplicate DNA extractions were mixed one to one by volume and sent to JGI for metagenomic sequencing.

Plate-based DNA library preparation for Illumina sequencing was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Kapa Biosystems library preparation kit. 200 ng of sample DNA was sheared to 500 bp using a Covaris LE220 focused-ultrasonicator. The sheared DNA fragments were size-selected by double-SPRI (solid-phase reversible immobilization) and then the selected fragments were end-repaired, A-tailed, and ligated with Illumina compatible sequencing adaptors from IDT containing a unique molecular

index barcode for each sample library. The prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1.5 reagent kits, S4 flowcell, following a 2x151 indexed run recipe.

Metagenomics data processing

Filtered reads from each season were co-assembled with MetaHipMer 2 version 2.1.0.90-g7529d1f-master [command: mhm2.py -k 21,33,55,77,99 --post-asm-align --post-asm-abd]³¹ on 1,000 KNL nodes on the NERSC Cori system. Contigs smaller than 500 bp were removed. Alignment information was determined by mapping reads to the assembly reference with BBtools³² version 38.95 [command: bbmap.sh -Xmx450g nodisk=true interleaved=true ambiguous=random mappedonly=t trimreaddescriptions=t usemodulo=t fast=t]. Coverage was determined by running BBTools version 38.95 [command: pileup.sh]. The IMG/M pipeline³³ was used to annotate the assembled contigs, and prokaryotic bins were predicted using metaBAT³⁴.

Metagenome bins from each of the four co-assemblies were de-replicated using dRep³⁵ with default settings. MAG taxonomy was assigned using GTDB-Tk v2.3.2.³⁶ MAG relative abundance was estimated using CoverM³⁷ to map sequencing reads from each sample library to the shared MAG set. MAGs were annotated and metabolism was summarized using DRAM v0.1.2³⁸ as implemented in KBase³⁹. DRAM annotations were downloaded and used for functional analysis.

For gene-based analysis, taxonomic and functional profiles were retrieved from the JGI IMG/M portal using the Abundance Profile tool. For functional profiles, the estimated gene

copies of KEGG KOs were used and center-log₂ ratio (clr) normalized on a per sample basis based on the method from ALDEx2⁴⁰. For taxonomic profiles, the estimated gene copies for order-level GTDB taxonomy at 60+ % identity were used and normalized using relative abundance.

Metaproteomics

Proteomics was performed on the soil samples from the October sampling. Cores were processed as above and sent for protein extraction and analysis at the University of Oklahoma. Soil samples were suspended in lysis buffer (containing 10 mM tris-HCl, 4% SDS and 10 mM dithiothreitol), boiled for 5 min and further disrupted by sonication for a 2 min 10% pulse 5 times. The supernatant was collected by centrifugation at 14,000 g for 10 min. Proteins were then precipitated by TCA (trichloroacetic acid) overnight and pelleted by centrifugation. Protein pellets were washed with ice-cold acetone three times and resuspended in guanidine buffer. Protein concentrations were quantified by Bicinchoninic Acid Assay. Twenty mg samples of proteins were processed using FASP (filter-aided sample preparation) and digested by Trypsin/LysC mixture. Each sample was analyzed using 2D-LC-MS/MS (two-dimensional liquid chromatography-tandem mass spectrometry) on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, USA) at the IDeA National Resource for Quantitative Proteomics. Tryptic peptides were separated into 46 fractions on a 100 × 1.0 mm Acquity BEH C18 column (Waters) using an UltiMate 3000 UHPLC system (Thermo) with a 50 min gradient from 99:1 to 60:40 buffer A:B ratio under basic pH conditions, then consolidated into 12 super-fractions. Each super-fraction was then separated by reverse phase XSelect CSH C18 2.5 um resin (Waters) on an in-line 120 × 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 60 min gradient from 98:2 to 65:35 buffer A:B ratio.

Eluted peptides were ionized by electrospray (2.4 kV) followed by mass spectrometric analysis on an Orbitrap Fusion Tribrid mass spectrometer (Thermo). MS data were acquired using the FTMS analyzer in profile mode at a resolution of 240,000 over a range of 375 to 1500 m/z. Following HCD activation, MS/MS data were acquired using the ion trap analyzer in centroid mode and normal mass range with normalized collision energy of 28–31% depending on charge state and precursor selection range. Protein identification and quantification followed established procedures. Briefly, mass spectrometry spectra were searched using SiproS Ensemble against a protein database constructed from the metagenome. Raw search results were filtered to achieve 1% FDR (false discovery rate) at the peptide level, which was estimated using the target-decoy approach. Peptide identifications are assigned to protein or protein groups in accordance with the parsimonious rule. To avoid ambiguity in data analysis, protein groups were excluded from biological analysis. Protein quantification was achieved through Intensity-based label-free analysis using ProRata. Protein abundances were quantified by the total peak height of all quantified peptides from a protein, normalized against the average total of all data sets.

Statistical analysis

Differences in abundance were assessed across seasons and depths using the Kruskal-Wallis H-test followed by multiple test correction using the Benjamini-Hochberg method, with Dunn's test used as a post-hoc analysis. Abundance correlations with depth were assessed using the Spearman correlation coefficient with permutation testing to estimate p-values. For the permutation testing, the depth-label of each sample was randomly permuted and correlation coefficients were calculated to generate a null-distribution. P-values from the true-data correlation coefficients were estimated based on this null-distribution followed by multiple test

correction. Distance-based PERMANOVA and ANOSIM tests were used to evaluate sample groupings based on season and/or depth.

Results

Site geochemistry varies across depth and season

During the study period, the freshwater wetland field site experienced seasonal fluctuations in water level, temperature, and plant cover (Fig. 3-1a, c). The water table was highest during the spring and summer months and lowest in the winter and fall, ranging from 6.1 m to 6.7 m above sea level; this lake-wide water table pattern is artificially maintained for flood control during the rainy winter season. The water table was above the surface at each sampling except for February where the surface mud was exposed and plant growth was minimal. In line with water table variations, IRIS tubes exhibited the sharpest gradient in reduction state when the water table was high and were less reduced when the water table was low (Fig. S 3-1). Surface water temperatures near the site displayed a typical seasonal pattern, ranging from 6.6 °C in the winter to 24.2 °C in late summer.

Surface CO₂ and CH₄ fluxes varied significantly across seasons (Kruskal-Wallis H-test: $p < 0.0001$). CO₂ flux was significantly higher in October and February compared to May and August timepoints (Dunn's test: $p < 0.01$) (Fig. 3-1e and Table S 3-2). Conversely, the CH₄ flux in August and May was significantly higher than October and February (Dunn's test: $p < 0.001$ for August and $p < 0.05$ for May) (Fig. 3-1f and Table S 3-2). No N₂O flux was observed at any of the timepoints (Fig. S 3-1b and Table S 3-2).

Overall porewater chemical composition varied with both depth and season (PERMANOVA and ANOSIM $p < 0.0001$) (Fig. 3-1b). NH₄⁺ concentration consistently increased with depth from $148 \pm 176 \mu\text{M}$ in the top 20 cm to $312 \pm 103 \mu\text{M}$ from 40 to 54 cm.

The rate of NH_4^+ increase with depth was sharper during the May and August sampling months when the water table was high. The concentration of NO_3^- was generally below $5 \mu\text{M}$ with the exception being in February where some deeper samples reached $10 \mu\text{M}$. NO_2^- was not detected in any of the samples. $\text{H}_2\text{S}/\text{HS}^-$ concentrations displayed a trend similar to NO_3^- , with the highest concentrations occurring in February in deeper layers, where they reached up to $132 \mu\text{M}$. SO_4^{2-} concentrations decreased with depth from $168 \pm 124 \mu\text{M}$ in the top 20 cm to $8.94 \pm 17.9 \mu\text{M}$ from 40 to 54 cm, and were appreciably higher in February reaching up to $465 \mu\text{M}$ at the surface. Concentrations of organic acids, including acetate and propionate, were generally low except in February and May where they were sporadically observed in high concentrations below 20 cm (up to 2.09 mM acetate and $350 \mu\text{M}$ propionate). The divalent cations, Ca^{2+} and Mg^{2+} consistently increased in concentration with depth. Once again, the exception to this trend occurred in February where the concentration of Ca^{2+} and Mg^{2+} was similar across the full depth profile. Porewater pH slightly increased with depth from 7.56 ± 0.25 in the top 20 cm to 7.81 ± 0.27 from 40 cm to 54 cm (Fig S 3-1a).

Targeted metabolomics was performed on soil cores from the May sampling campaign to investigate differences in water-extractable soil metabolites across depth sections. The composition of metabolites was relatively similar in the top and middle sections while the bottom samples had distinct metabolite profiles (PERMANOVA and ANOSIM $p < 0.05$) (Fig. 3-2b). Numerous compounds associated with purine metabolism (e.g. xanthine, 3-methyladenine, 5-methylthioadenosine, uric acid, inosine, xanthosine, hypoxanthine, adenine, and guanosine) increased in abundance towards the surface (Fig. 3-2a) as did DNA extraction yields, suggesting an increase in overall nucleotide metabolism (Fig. S 3-1c). Other metabolites that were negatively correlated with depth included amino acids (e.g. N-acetyl-methionine, lysine,

methionine, and N-acetyl-alanine), quinoline, dimethylbenzimidazole (a component of vitamin-B₁₂), pterin (involved in cofactor biosynthesis), and sn-glycero-3-phosphocholine (involved in phospholipid synthesis).

Fewer compounds increased in abundance with depth. Exceptions included multiple amino sugars (e.g. N-acetyl-glucosamine, glucosamine, N-acetyl-galactosamine, and N-acetyl-mannosamine), phenolic compounds (e.g. 3- and 4-hydroxybenzoic acid and 3-(2-hydroxyphenyl) propanoic acid), glucuronic acid, and 2-hydroxypyridine which were positively correlated with soil depth (Fig. 3-2c).

Depth but not season shapes microbial community composition and functional potential

Metagenome sequencing was conducted on triplicate depth-sectioned soil cores from each season. Libraries averaged 986 ± 222 million reads with 39.6 ± 8.6 % of reads being assembled into 1.57 ± 0.51 gigabase assemblies comprised of 1.52 ± 0.43 million scaffolds.

The order-level taxonomic abundance and functional composition based on KEGG gene orthologs (KOs) was significantly influenced by soil depth (Taxonomy and KOs: ANOSIM $p < 0.005$ and PERMANOVA $p < 0.001$) but not sample season (Taxonomy and KOs: ANOSIM and PERMANOVA $p > 0.5$) (Fig. 3-3a, b). No taxonomic orders or KOs were significantly different across seasons while 282 orders and 4139 KOs were significantly different across depths (Kruskal Wallis H-test: adjusted $p < 0.05$).

Of the top 20 most abundant orders, 15 were significantly correlated with sample depth. Burkholderiales, Rhizobiales, Methylococcales, Chromatiales (all within the phylum Psuedomonadata), Geobacterales, Myxococcales, and Mycobacteriales were negatively correlated with depth, i.e. increased in abundance towards the surface. Conversely,

Anaerolineales, Methanomicrobiales, Syntrophales, Syntrophorhabdadales, Dehalococcoidales, Chloroflexales, and Methanosarcinales were positively correlated with depth. Other prominent orders like Desulfobacterales, Syntrophobacterales, Desulfobaccales, Methanotrichales, and Bacteroidales were abundant but not significantly correlated with depth (Fig. 3-3c).

Many KOs and their associated metabolic modules exhibited significant depth correlations. For example, KEGG modules for cytochromes bd, cbb₃, and bc1 were more abundant in the top layer, along with modules for thiosulfate and sulfide oxidation, denitrification, and dissimilatory nitrate reduction to ammonia (DNRA). Carbon metabolisms enriched near the surface included the glyoxylate cycle, D-galactonate degradation via the De Ley-Doudoroff pathway, and purine degradation to urea (Fig. 3-3d). Additionally, modules involved in the degradation of aromatic compounds—such as benzoate, salicylate, anthranilate, catechol, cumate, and toluene—showed a negative correlation with depth.

The opposite trend, where metabolic module abundance increased with depth, was observed for pyruvate oxidation, methanogenesis, glycolysis, the non-oxidative phase of the pentose phosphate pathway, galactose degradation, and the methylaspartate cycle (Fig. 3-3e). Depth-associated increases were noted for carbon fixation pathways, including the Wood-Ljungdahl pathway and the reductive citrate cycle. Additionally, modules involved in the biosynthesis of UDP-N-acetylglucosamine, UDP-galactose, cobalamin, and purines were more prevalent at greater depths. Unlike the surface-enriched degradation of aromatic compounds via mono- and dioxygenases, benzoyl-CoA degradation via benzoyl-CoA reductase was more abundant in deeper soil. More broadly, archaeal-specific modules, such as those for thiamine, coenzyme A, and F420 biosynthesis, also increased in abundance with depth.

Genome-resolved metagenomics reveals depth-preferences of MAGs and their functional potential

Combined assemblies from each sampling campaign and de-replication were used to create a unified set of metagenome-assembled genomes (MAG) for analysis. The resulting 429 MAGs had an average completeness of $86 \pm 6.9\%$ and contamination of $3.1 \pm 2.3\%$ with genome sizes of 3.4 ± 1.8 Mbp and scaffold N50s of 26.3 ± 42.1 kbp (Fig. S 3-3). MAGs were assigned to 45 different phyla with the most common being Chloroflexota (n=57), Acidobacteriota (n=56), Pseudomonadota (n=37), Bacteroidota (n=29), and Planctomycetota (n=24).

MAG relative abundance was estimated by mapping reads from each individual sequencing library to the set. To focus the analysis on the most prevalent MAGs, we filtered out those with a median relative abundance of less than 0.01%, leaving 239 remaining. Ten taxonomic classes showed significant negative correlations with depth (surface-preferring MAGs), 26 classes were positively correlated with depth (bottom-preferring MAGs), and 19 classes showed no significant depth correlation (neutral MAGs) (Fig. 3-4).

Methanogen MAG classes were all bottom-preferring. Methanomicrobia MAGs were the most abundant and belonged to the family Methanoregulaceae or JACTUA01, and were likely hydrogenotrophic. Of the Methanosarcinia MAGs, three belonged to the family Methanotrichaceae and one was assigned to Methanoperedenaceae. While the Methanotrichaceae are putative acetoclastic methanogens, the Methanoperedens may be an anaerobic methanotroph (ANME-2d) capable of coupling methane oxidation to nitrate reduction⁴¹. Finally, the Thermoplasmata class contained MAGs from the family Methanomassiliicoccaceae and UBA10834. The Methanomassiliicoccaceae harbored genes for the activation of methanol (*mtaBC*) and the UBA10834 for the activation of methylamine and trimethylamine (*mtmB* and

mttB) as substrates for methanogenesis. Many other classes displayed a high completion fraction for methanogenic pathways due to the shared enzymology with other anaerobic pathways.

The capacity for sulfur and nitrogen-based energy metabolism varied across the MAG set. Dissimilatory sulfate reduction potential was present across the diverse set of Desulfobacterota MAGs, while the presence of the dissimilatory sulfite reductase in Gammaproteobacterial MAGs was confined to putative sulfur-oxidizers which use the enzyme in the reverse direction ⁴². The oxidation of thiosulfate via the SOX complex was restricted to Alpha- and Gammaproteobacterial MAGs which tended to have either surface or neutral depth preference. Denitrification and DNRA genes were also most prevalent in surface and neutral MAGs, particularly in the classes UBA9217 (phylum Nitrospirota), Gammaproteobacteria and Actinomycetia.

Certain respiratory enzymes displayed distinct stratification among depth preferences while others were evenly spread. For instance, c-type cytochromes were largely restricted to surface and neutral MAGs while the high-affinity bd-type was found among all depth preferences. V/A-type ATPases were only found in bottom and neutral MAGs, mirroring the distribution of archaea. Hydrogenases were most prevalent in Methanomicrobia (*ech* and *frh*), DSWW01 (phylum Desulfobacterota_F) (*ech* and *frh*), and Verrucomicrobiae (*hox* and *ech*). The RNF complex, which couples the oxidation of reduced ferredoxin to the reduction of NAD⁺ while translocating Na⁺, was also found across depth-preferences. The panel of respiratory enzymes was notably lacking from RBG-13-43-22 (phylum Desulfobacterota), Bathyarchaeia, and Aenigmataarchaeia MAGs suggesting a reliance on fermentation.

Pathways for central carbon metabolism were similarly distributed across depth preferences, with some MAGs having broad central carbon metabolism and some with a limited

repertoire. For example, Terriglobia and Actinomycetia showed a high level of pathway completion across modules, while MSB-5A5 (phylum Zixibacteria), Aenigmatarchaea, and 448-113 MAGs harbored few genes for central carbon metabolism. Capacity for galactose and D-galacturonate degradation, pentose phosphate pathway, were widespread, in contrast fewer MAGs harbored genes for the glyoxylate cycle or acetyl-coA pathway.

The potential to perform various short chain fatty acid (SCFA) and alcohol conversions was found across depth preferences, most often in Myxococcia, Alphaproteobacteria, BSN033 (phylum Desulfobacterota_F), Thermoanaerobaculia, and Bacteroidia MAGs. Glycoside hydrolase and polysaccharide lyases were numerous in Actinomycetia, Terriglobia, Verrucomicrobiae, and Vicinamibacteria which are all surface or neutral MAGs. Glycoside hydrolases were abundant in the bottom-preferring Bacteroidia and Calditrichia. The potential for benzoyl-CoA degradation was common in the phylum Desulfobacterota along with Polarisedimentocolia and UBA4802 (phylum Spirochaetota) with no clear depth preference. Genes for cobalamin (vitamin B₁₂) biosynthesis from both precorrin-2 and cobinamide were found across depth-preferences particularly in Desulfobacterota, Halobacteriota, and Actinomycetota MAGs.

Metaproteomics highlights active microbial taxa and functions

Metaproteomic analysis of the October soil samples revealed key microbial taxa and their actively expressed proteins. Based on median peak height, the most abundant taxa included Burkholderiales, Methanotrichales, Methylococcales, Syntrophobacterales, Rhizobiales, and Methanomicrobiales (Fig. 3-6a). Burkholderiales were more abundant in the top and middle sections, while Methanotrichales were dominant in the bottom section. Overall, the distribution of order-level taxa in the metaproteome aligned well with gene-based metagenomics data,

although some rankings differed—for example, Methanotrichales ranked second in the proteome but 13th in the metagenome.

The most abundant proteins were associated with transport, cell structure and maintenance, and energy metabolism. Outer membrane porins and transporters for iron complexes, branched-chain amino acids, long-chain fatty acids, and sugars were frequently identified. Key proteins for cellular structure and function—including the chaperonin GroEL, archaeal S-layer proteins, and elongation factors—were also prevalent. In terms of energy metabolism, proteins related to methanogenesis, methane oxidation, and sulfur reduction were highly abundant (Fig. 3-6b). Interestingly, a protein involved in anaerobic carnitine metabolism (CaiB) was frequently detected from Syntrophobacterales, Desulfobaccales, and Pseudomonadales.

Linking specific proteins to microbial taxa provided insights into active metabolic processes. For instance, the expression of methyl-coenzyme M reductase (Mcr) confirmed active methanogenesis by Methanotrichales, Methanomicrobiales, and Methanomassiliicoccales. Substrate-specific expression patterns included acetate-utilizing Methanotrichales (via Acs/CdhC), hydrogen/formate-utilizing Methanomicrobiales (via FrhG, MvhD, and Fdh), and methanol-utilizing Methanomassiliicoccales (via MtaBC) (Fig. 3-6c). Expression of methane monooxygenase and methanol dehydrogenase was detected in Methylococcales, Rhizobiales, and Burkholderiales.

Burkholderiales exhibited a wide metabolic repertoire, with expression of alcohol, formate, and malate dehydrogenases, the full suite of sulfur oxidation proteins, complete denitrification pathways, and numerous transporters—highlighting the metabolic versatility of this order and of Gammaproteobacteria more broadly. Consistent with MAG-based analysis,

cytochromes associated with oxygen respiration—particularly the high-affinity *cbb₃*-type—were restricted to Alpha- and Gammaproteobacteria.

In relation to aromatic compound metabolism, we identified 6-oxocyclohex-1-ene-carbonyl-CoA hydrolase (Oah) in Syntrophales, Syntrophorhabdales, and Desulfatiglandales, and 2-aminobenzoate-CoA ligase (AbmG) in Syntrophobacterales and Desulfobaccales. Oah plays a role in benzoate degradation via the benzoyl-CoA pathway, while AbmG is involved in anthranilate (2-aminobenzoic acid) degradation. Finally, detection of ammonia monooxygenase (Amo) and nitrite oxidoreductase (Nxr) from Nitrospirales suggests the presence of complete ammonia oxidizers (comammox) in the wetland system.

Discussion

Wetland microbiomes play a crucial role in regulating soil organic matter⁴³ and GHG fluxes, including CO₂⁴⁴, CH₄ (28 times the global warming potential of CO₂)^{2,13}, and N₂O (265 times the global warming potential of CO₂)⁴⁵. As climate change intensifies, wetland ecosystems will increasingly experience disturbances such as drought, flooding, and saltwater intrusion⁶. Understanding how microbial communities respond to these perturbations is essential for predicting future GHG fluxes—specifically, whether wetlands will become net sources or sinks of GHGs, thereby either exacerbating or mitigating climate change.

In this study, we integrated multiple approaches to examine the biogeochemistry of a lacustrine wetland, tracking how microbial community composition, function, and metabolite profiles varied with depth and season—two factors that offer insights into how wetlands may respond to environmental disturbances. For instance, fluctuations in water levels at the site provided a proxy for drought and flooding conditions, while soil depth reflected a gradient of redox state and nutrient availability. By capturing both temporal and spatial variability, our study

links microbial dynamics to a comprehensive biogeochemical dataset, offering valuable insights into wetland ecosystem function.

Site geochemistry varied significantly across seasons and soil depths, shaping CH₄ and CO₂ flux dynamics. CH₄ emissions peaked in August when both temperature and the water table were high but dropped to their lowest levels in October, despite persistently warm temperatures, likely due to the preceding decline in the water table. In contrast, CO₂ fluxes were highest in October and February, coinciding with low water table conditions. High concentrations of organic acids in February suggest thermodynamic constraints on microbial activity at low temperatures, as these substrates are typically readily metabolized. At other times, acetate concentrations remained around 50 μM—near the minimum threshold for utilization by *Methanotherix* acetoclastic methanogens⁴⁶. The interplay between water table depth^{44,47} and soil organic matter (SOM) decomposability⁴⁸ has been linked to the temperature dependence of CH₄ and CO₂ emissions in wetlands, which may explain the observed patterns.

Metagenomes and proteomes revealed potential for methanogenesis from hydrogen, formate, acetate, and methanol with hydrogenotrophic Methanoregulaceae and acetoclastic Methanotrichaceae emerging as dominant methanogens. For methane oxidation, aerobic Methylococcales (genus *KS41*) were most prevalent, but we also detected the anaerobic *Methanoperedens*. These findings align with a recent meta-analysis of freshwater wetland core microbiomes, which identified *Methanoregula*, *Methanotherix*, and *KS41* as common across sites, while *Methanoperedens* appeared only in high-CH₄ emitting environments⁴⁹. Notably, our study adds proteomic evidence of their activity in the wetland.

Porewater analysis revealed active sulfur and nitrogen cycles, with potential implications for carbon dynamics and GHG emissions. Despite the site's freshwater nature, significant concentrations of sulfate and sulfide were detected, particularly in February. We recovered a diverse array of Desulfobacterota MAGs with metabolic capacities extending beyond sulfate reduction to include denitrification, DNRA, and benzoyl-CoA reduction—consistent with the versatility described in a large-scale genome comparison of this phylum⁵⁰. Sulfate-reducing microorganisms exhibit notable metabolic flexibility, enabling them to function either as sulfate reducers or as syntrophs in association with H₂-utilizing organisms. This adaptability allows them to both compete and cooperate with methanogens, thereby influencing CH₄ emissions^{51,52}. This dynamic was demonstrated in microcosm incubations of soil from this site, where sulfate additions significantly decreased CH₄ production⁵³. We also identified numerous sulfur-oxidizing Alpha- and Gammaproteobacteria with the potential to couple sulfur compound oxidation to denitrification and carbon fixation, demonstrating the intertwined nature of these nutrient cycles.

The consistent accumulation of NH₄⁺ in deep layers is a common feature of anoxic sediments driven by the breakdown of nitrogen-containing organic matter, such as proteins^{11,54}. In contrast, the low nitrogen concentrations at the surface suggest a tight coupling between nitrification and denitrification or assimilation. In terms of nitrification, we identified a *Nitrospira* comammox from the *Palsa-1315* genus. This genus has also been detected in brackish and freshwater wetland metatranscriptomes, where it contributed a majority of transcripts related to the reverse tricarboxylic acid (rTCA) cycle, the carbon fixation pathway used by *Nitrospira*⁵⁵. Together, this suggests that *Nitrospira* comammox may play a key role in nitrification across

wetland ecosystems, possibly driven by a preference for oligotrophic and low oxygen conditions⁵⁶⁻⁵⁹.

We did not observe any N₂O flux from our site, indicating a high microbial capacity for N₂O reduction or minimal production. Indeed, denitrification capacity was prevalent in metagenomes and proteomes. Furthermore, consistently low NO₂⁻ and NO₃⁻ concentrations suggest that any oxidized nitrogen was in high demand as an alternate electron acceptor. Ammonia-oxidizing archaea (AOA) have also been linked to elevated wetland N₂O emissions⁴⁵. Thus, the absence of AOA and the presence of comammox *Nitrospira* in our system may further explain the lack of detectable N₂O flux.

Microbial community composition varied far more with soil depth than with season, highlighting an unexpected stability across time points. Depth is a well-known driver of microbial community structure and influences how microbes respond to environmental perturbations, with deeper layers typically exhibiting reduced or altered responses⁶⁰⁻⁶⁶. While microbial communities are often considered sensitive to disturbance²¹, our findings suggest that natural seasonal variation had a minimal impact compared to the persistent environmental gradient imposed by soil depth. Similarly, a seven-year multi-omic study of a thawing permafrost peatland observed remarkable microbiome stability despite significant temporal changes⁶⁷. Surface soil communities may be more accustomed to fluctuations, allowing them to adjust their metabolism accordingly, whereas deeper-dwelling microbes are adapted to low-energy conditions and remain largely insulated from disturbance. In fact, it has been proposed that frequently disturbed ecosystems select for metabolically flexible generalists, which contribute to community stability by adjusting their metabolic strategies⁶⁸. This raises the question: what is the amount and rate of disturbance that would be required to significantly alter

the wetland microbial community? Answering this would provide critical insights into how wetland soil carbon cycling may change in future climate scenarios.

We identified numerous microbial taxa and metabolic functions correlated with soil depth. At the surface, we observed an increased abundance of metabolically versatile Gammaproteobacteria capable of respiring oxygen and nitrate as well as contributing to methane and sulfur oxidation. In contrast, deeper soil layers exhibited an increasing abundance of archaea and strict anaerobes with energy metabolisms centered on methanogenesis, acetogenesis, sulfate-reduction and fermentation. Similar depth-dependent trends have been observed in metagenomic studies of Swiss forest soils, where archaeal abundance increased significantly with depth⁶⁶, as well as in a restored wetland, where genes for aerobic respiration were negatively correlated with methanogenesis and acetogenesis¹⁹. A broader analysis of metagenomes across ecosystems also revealed strong functional stratification based on oxygen availability⁶⁹. That these structuring trends are detectable both at the meter scale within a single site and across thousands of independently collected samples highlights their fundamental role in shaping complex microbiomes and the functions they carry.

We detected a high abundance of metabolites related to purine metabolism in the surface soil, along with an enrichment of genes associated with purine degradation and an overall increase in DNA concentration. Since purine bases are essential components of DNA and RNA, we hypothesize that extracellular DNA may be a prevalent biofilm component in the surface layer, potentially serving as an important and understudied nutrient source for the microbial community⁷⁰. Furthermore, the shift towards de novo purine biosynthesis in deeper layers emphasizes that purine cycling may be unique to more active surface soils.

In deeper soil layers, we observed an increase in amino sugars and phenolic compounds. Amino sugars, key components of microbial cell walls, are often used as markers of necromass and can constitute a significant fraction of soil organic carbon (SOC) in deep soils^{71–73}. However, it remains unclear whether amino sugar utilization is inhibited in these layers or if microbes in deeper soils tend to form thicker cell walls containing more of these compounds. We also found higher concentrations of certain phenolic compounds in deeper layers, underscoring their role in wetland soil carbon preservation^{74,75}. While the degradation of plant-derived polyphenols in anoxic wetland soils has been demonstrated⁷⁶, our functional metagenomics suggests distinct metabolic strategies across depths, with surface layers having greater access to mono- and dioxygenase enzymes, which are more effective at breaking down aromatic rings, whereas deeper layers rely on the anaerobic benzoyl-CoA reductase pathway⁷⁷. Additionally, deeper soils experience more stable environmental conditions than surface layers, remaining consistently anaerobic and largely insulated from temperature fluctuations, water table changes, and surface nutrient inputs. Under these conditions, the breakdown of complex carbon is often thermodynamically constrained, further influencing carbon cycling dynamics in wetland soils^{11,78}.

The findings of this study reinforce that soil depth exerts a stronger influence on microbial community composition than seasonal fluctuations, likely due to persistent redox gradients and nutrient stratification. A key question arising from this work is whether wetland microbial communities exhibit threshold responses to environmental disturbance—where incremental changes in hydrology could lead to abrupt shifts in community composition and GHG fluxes. Identifying these tipping points will require targeted experiments manipulating disturbance frequency, magnitude, and duration, particularly in relation to water table recession

and rewetting cycles. Future research should aim to identify tipping points for microbial succession and biogeochemical shifts by leveraging microcosms, controlled laboratory reactors, and isolate-based physiology studies to precisely manipulate key environmental variables. The MAGs and geochemical data presented here provide valuable inputs for genome-scale metabolic and thermodynamic models, improving predictions of carbon cycling in wetlands. Together, these approaches could inform wetland management strategies aimed at enhancing carbon sequestration and mitigating GHG emissions. Advancing our understanding of microbial mechanisms in wetland ecosystems remains essential for improving predictions of global carbon dynamics and developing effective climate change mitigation strategies.

Acknowledgements

This work was funded by the US Department of Energy under award number DE-SC0020356. Metagenomics and metabolomics support was provided by the Joint Genome Institute under Proposal ID 507192, titled “Integrating single-cell wetland microbiome structure, function, and activity to ecosystem-scale biogeochemical fluxes”. The work (proposal: <https://doi.org/10.46936/10.25585/60000490>) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xml1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

Data and code availability statement

All metagenomics data can be accessed via the JGI Genome Portal under Proposal ID 507192. All code used for data processing and plotting is available from the project’s github repository (https://github.com/zflink/Wetland_multi-omics).

References

1. Nahlik, A. M. & Fennessy, M. S. Carbon storage in US wetlands. *Nat Commun* **7**, 1–9 (2016).
2. Saunois, M. *et al.* The global methane budget 2000-2017. *Earth Syst Sci Data* **12**, 1561–1623 (2020).
3. Zhang, Z. *et al.* Emerging role of wetland methane emissions in driving 21st century climate change. *Proc Natl Acad Sci U S A* **114**, 9647–9652 (2017).
4. Beillouin, D. *et al.* A global meta-analysis of soil organic carbon in the Anthropocene. *Nat Commun* **14**, (2023).
5. Fluet-Chouinard, E. *et al.* Extensive global wetland loss over the past three centuries. *Nature* **614**, 281–286 (2023).
6. Candry, P., Abrahamson, B., Stahl, D. A. & Winkler, M. K. H. Microbially mediated climate feedbacks from wetland ecosystems. *Global Change Biology* vol. 29 5169–5183 Preprint at <https://doi.org/10.1111/gcb.16850> (2023).
7. Zhang, Z. & Furman, A. Soil redox dynamics under dynamic hydrologic regimes - A review. *Science of the Total Environment* vol. 763 Preprint at <https://doi.org/10.1016/j.scitotenv.2020.143026> (2021).
8. Wilpiseski, R. L. *et al.* Soil Aggregate Microbial Communities: Towards Understanding Microbiome Interactions at Biologically Relevant Scales. *Appl Environ Microbiol* **85**, (2019).
9. Candry, P., Godfrey, B. J. & Winkler, M. K.-H. Microbe-cellulose hydrogels as a model system for particulate carbon degradation in soil aggregates. *ISME Communications* **4**, (2024).
10. Johnston, J. T. *et al.* Increasing aggregate size reduces single-cell organic carbon incorporation by hydrogel-embedded wetland microbes. *ISME Communications* **4**, (2024).
11. Orsi, W. D., Schink, B., Buckel, W. & Martin, W. F. Physiological limits to life in anoxic subseafloor sediment. *FEMS Microbiology Reviews* vol. 44 219–231 Preprint at <https://doi.org/10.1093/FEMSRE/FUAA004> (2020).
12. Calabrese, S., Garcia, A., Wilmoth, J. L., Zhang, X. & Porporato, A. Critical inundation level for methane emissions from wetlands. *Environmental Research Letters* **16**, 044038 (2021).
13. Delwiche, K. B. *et al.* FLUXNET-CH4: A global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth Syst Sci Data* **13**, 3607–3689 (2021).

14. Dalcin Martins, P., Frank, J., Mitchell, H., Markillie, L. M. & Wilkins, M. J. Wetland Sediments Host Diverse Microbial Taxa Capable of Cycling Alcohols. *Appl Environ Microbiol* **85**, 1–15 (2019).
15. Dalcin Martins, P. *et al.* Abundant carbon substrates drive extremely high sulfate reduction rates and methane fluxes in Prairie Pothole Wetlands. *Glob Chang Biol* **23**, 3107–3120 (2017).
16. Angle, J. C. *et al.* Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat Commun* **8**, (2017).
17. Hartman, W. H. *et al.* Multiple microbial guilds mediate soil methane cycling along a wetland salinity gradient. *mSystems* **9**, (2024).
18. Bueno de Mesquita, C. P., Hartman, W. H., Ardón, M. & Tringe, S. G. Disentangling the effects of sulfate and other seawater ions on microbial communities and greenhouse gas emissions in a coastal forested wetland. *ISME Communications* **4**, (2024).
19. He, S. *et al.* Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. *mBio* **6**, 1–15 (2015).
20. Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
21. Allison, S. D. & Martiny, J. B. H. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* **105**, 11512–11519 (2008).
22. Louca, S. *et al.* Function and functional redundancy in microbial systems. *Nat Ecol Evol* **2**, 936–943 (2018).
23. Rabenhorst, M. C. & Burch, S. N. Synthetic Iron Oxides as an Indicator of Reduction in Soils (IRIS). *Soil Science Society of America Journal* **70**, 1227–1236 (2006).
24. Castenson, K. L. & Rabenhorst, M. C. Indicator of Reduction in Soil (IRIS). *Soil Science Society of America Journal* **70**, 1222–1226 (2006).
25. Cline, J. D. Spectrophotometric determination of hydrogen sulfide in natural water. *Limnol Oceanogr* **14**, 454–458 (1969).
26. B. Louie, K. *et al.* JGI/LBNL Metabolomics - Standard LC-MS/MS ESI Method - Polar HILIC-Z v1. Preprint at <https://doi.org/10.17504/protocols.io.kxygxydwk18j/v1> (2024).
27. B. Louie, K. *et al.* JGI/LBNL Metabolomics - Standard LC-MS/MS ESI Method - Nonpolar C18 v1. Preprint at <https://doi.org/10.17504/protocols.io.ewov19r27lr2/v1> (2024).

28. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221 (2007).
29. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
30. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **17**, 905–908 (2020).
31. Hofmeyr, S. *et al.* Terabase-scale metagenome coassembly with MetaHipMer. *Sci Rep* **10**, (2020).
32. Bushnell, B. BBTools. <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>.
33. Chen, I. M. A. *et al.* The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res* **51**, D723–D732 (2023).
34. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, (2015).
35. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal* **11**, 2864–2868 (2017).
36. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
37. Aroney, S. T. N. *et al.* CoverM: Read alignment statistics for metagenomics. (2025).
38. Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**, 8883–8900 (2020).
39. Arkin, A. P. *et al.* KBase: The United States department of energy systems biology knowledgebase. *Nat Biotechnol* **36**, 566–569 (2018).
40. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
41. Haroon, M. F. *et al.* Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**, 567–570 (2013).
42. Loy, A. *et al.* Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ Microbiol* **11**, 289–299 (2009).

43. Yarwood, S. A. The role of wetland microorganisms in plant-litter decomposition and soil organic matter formation: A critical review. *FEMS Microbiol Ecol* **94**, 1–17 (2018).
44. Chen, H., Xu, X., Fang, C., Li, B. & Nie, M. Differences in the temperature dependence of wetland CO₂ and CH₄ emissions vary with water table depth. *Nat Clim Chang* **11**, 766–771 (2021).
45. Bahram, M. *et al.* Structure and function of the soil microbiome underlying N₂O emissions from global wetlands. *Nat Commun* **13**, (2022).
46. Jetten, M. S. M., Stams, A. J. M. & Zehnder, A. J. B. Methanogenesis from acetate: a comparison of the acetate metabolism in *Methanotrix soehngeni* and *Methanosarcina* spp. . *FEMS Microbiol Lett* **88**, 181–198 (1992).
47. Zou, J. *et al.* Rewetting global wetlands effectively reduces major greenhouse gas emissions. *Nat Geosci* **15**, 627–632 (2022).
48. Hu, H. *et al.* Relative increases in CH₄ and CO₂ emissions from wetlands under global warming dependent on soil carbon substrates. *Nat Geosci* **17**, 26–31 (2024).
49. Bechtold, E. K. *et al.* Metabolic interactions underpinning high methane fluxes across terrestrial freshwater wetlands. *Nat Commun* **16**, 944 (2025).
50. Langwig, M. V. *et al.* Large-scale protein level comparison of Deltaproteobacteria reveals cohesive metabolic groups. *ISME Journal* **16**, 307–320 (2022).
51. Pester, M., Knorr, K. H., Friedrich, M. W., Wagner, M. & Loy, A. Sulfate-reducing microorganisms in wetlands - fameless actors in carbon cycling and climate change. *Front Microbiol* **3**, (2012).
52. Plugge, C. M., Zhang, W., Scholten, J. C. M. & Stams, A. J. M. Metabolic flexibility of sulfate-reducing bacteria. *Front Microbiol* **2**, (2011).
53. Wang, D. *et al.* Metaproteomics-informed stoichiometric modeling reveals the responses of wetland microbial communities to oxygen and sulfate exposure. *NPJ Biofilms Microbiomes* **10**, (2024).
54. Reuter, H., Gensel, J., Elvert, M. & Zak, D. Evidence for preferential protein depolymerization in wetland soils in response to external nitrogen availability provided by a novel FTIR routine. *Biogeosciences* **17**, 499–514 (2020).
55. Grüterich, L. *et al.* Assessing environmental gradients in relation to dark CO₂ fixation in estuarine wetland microbiomes. *Appl Environ Microbiol* **91**, (2025).
56. Vilardi, K. *et al.* Co-Occurrence and Cooperation between Comammox and Anammox Bacteria in a Full-Scale Attached Growth Municipal Wastewater Treatment Process. *Environ Sci Technol* **57**, 5013–5023 (2023).

57. Gottshall, E. Y. *et al.* Sustained nitrogen loss in a symbiotic association of Comammox Nitrospira and Anammox bacteria. *Water Res* **202**, (2021).
58. Dimitri Kits, K. *et al.* Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* **549**, 269–272 (2017).
59. Van Kessel, M. A. H. J. *et al.* Complete nitrification by a single microorganism. *Nature* **528**, 555–559 (2015).
60. Dove, N. C. *et al.* Depth dependence of climatic controls on soil microbial community activity and composition. *ISME Communications* **1**, (2021).
61. Eilers, K. G., Debenport, S., Anderson, S. & Fierer, N. Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biol Biochem* **50**, 58–65 (2012).
62. Dove, N. C., Torn, M. S., Hart, S. C. & Taş, N. Metabolic capabilities mute positive response to direct and indirect impacts of warming throughout the soil profile. *Nat Commun* **12**, (2021).
63. Lamit, L. J. *et al.* Peatland microbial community responses to plant functional group and drought are depth-dependent. *Mol Ecol* **30**, 5119–5136 (2021).
64. Zhao, H., Zheng, W., Zhang, S., Gao, W. & Fan, Y. Soil microbial community variation with time and soil depth in Eurasian Steppe (Inner Mongolia, China). *Ann Microbiol* **71**, (2021).
65. Diamond, S. *et al.* Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat Microbiol* **4**, 1356–1367 (2019).
66. Frey, B. *et al.* Shotgun Metagenomics of Deep Forest Soil Layers Show Evidence of Altered Microbial Genetic Potential for Biogeochemical Cycling. *Front Microbiol* **13**, (2022).
67. Cronin, D. R. *et al.* Stable states in an unstable landscape: microbial resistance at the front line of climate change. *bioRxiv* (2025) doi:10.1101/2025.02.07.636677.
68. Chen, Y. J. *et al.* Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. *ISME Journal* **15**, 2986–3004 (2021).
69. Flinkstrom, Z., Bryson, S., Candry, P. & Winkler, M.-K. H. Metagenomic clustering links specific metabolic functions to globally relevant ecosystems. *mSystems* **9**, (2024).
70. Sharma, D. K. & Rajpurohit, Y. S. Multitasking functions of bacterial extracellular DNA in biofilms. *Journal of Bacteriology* vol. 206 Preprint at <https://doi.org/10.1128/jb.00006-24> (2024).

71. Ni, X. *et al.* A quantitative assessment of amino sugars in soil profiles. *Soil Biol Biochem* **143**, (2020).
72. Kallenbach, C. M., Frey, S. D. & Grandy, A. S. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nat Commun* **7**, (2016).
73. Kästner, M., Miltner, A., Thiele-Bruhn, S. & Liang, C. Microbial Necromass in Soils—Linking Microbes to Soil Processes and Carbon Turnover. *Frontiers in Environmental Science* vol. 9 Preprint at <https://doi.org/10.3389/fenvs.2021.756378> (2021).
74. Freeman, C., Ostle, N. & Kang, H. An enzymic “latch” on a global carbon store. *Nature* **409**, 149–149 (2001).
75. Fenner, N. & Freeman, C. Woody litter protects peat carbon stocks during drought. *Nat Clim Chang* **10**, 363–369 (2020).
76. McGivern, B. B. *et al.* Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat Commun* **12**, (2021).
77. Boll, M., Löffler, C., Morris, B. E. L. & Kung, J. W. Anaerobic degradation of homocyclic aromatic compounds via arylcarboxyl-coenzyme A esters: Organisms, strategies and key enzymes. *Environ Microbiol* **16**, 612–627 (2014).
78. Boye, K. *et al.* Thermodynamically controlled preservation of organic carbon in floodplains. *Nat Geosci* **10**, 415–419 (2017).

Figures

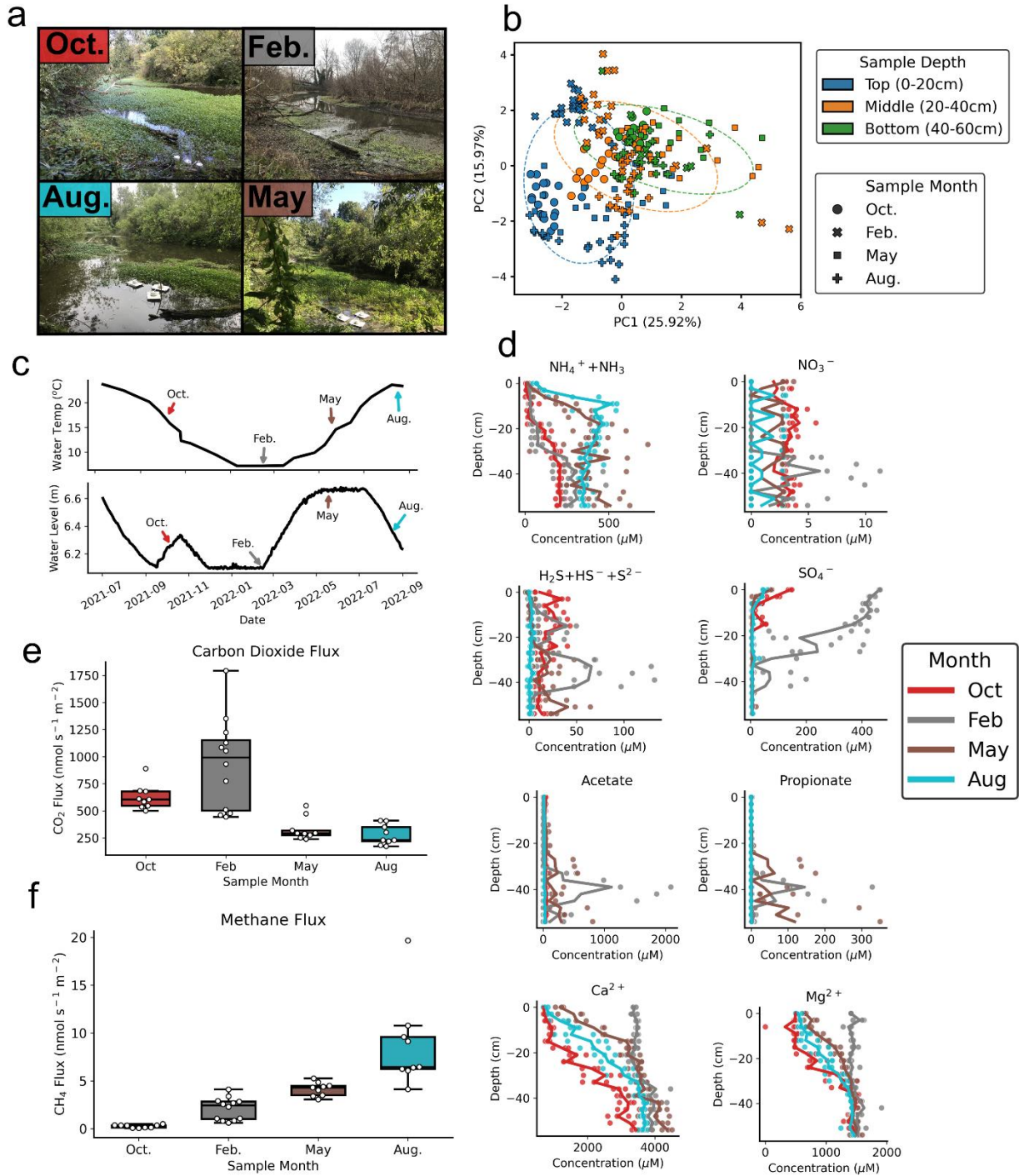


Figure 3-1 – Study site geochemical characterization.

Pictures of the site at each of the sampling months (a). Water temperature and elevation over the course of the study year (c). CO₂ (e) and CH₄ (f) surface fluxes from each month. PCA of porewater metabolite measurements (b). A selection of porewater metabolite depth-profiles from each month (d). Points indicate distinct samples and lines show mean of triplicate samples at each depth.

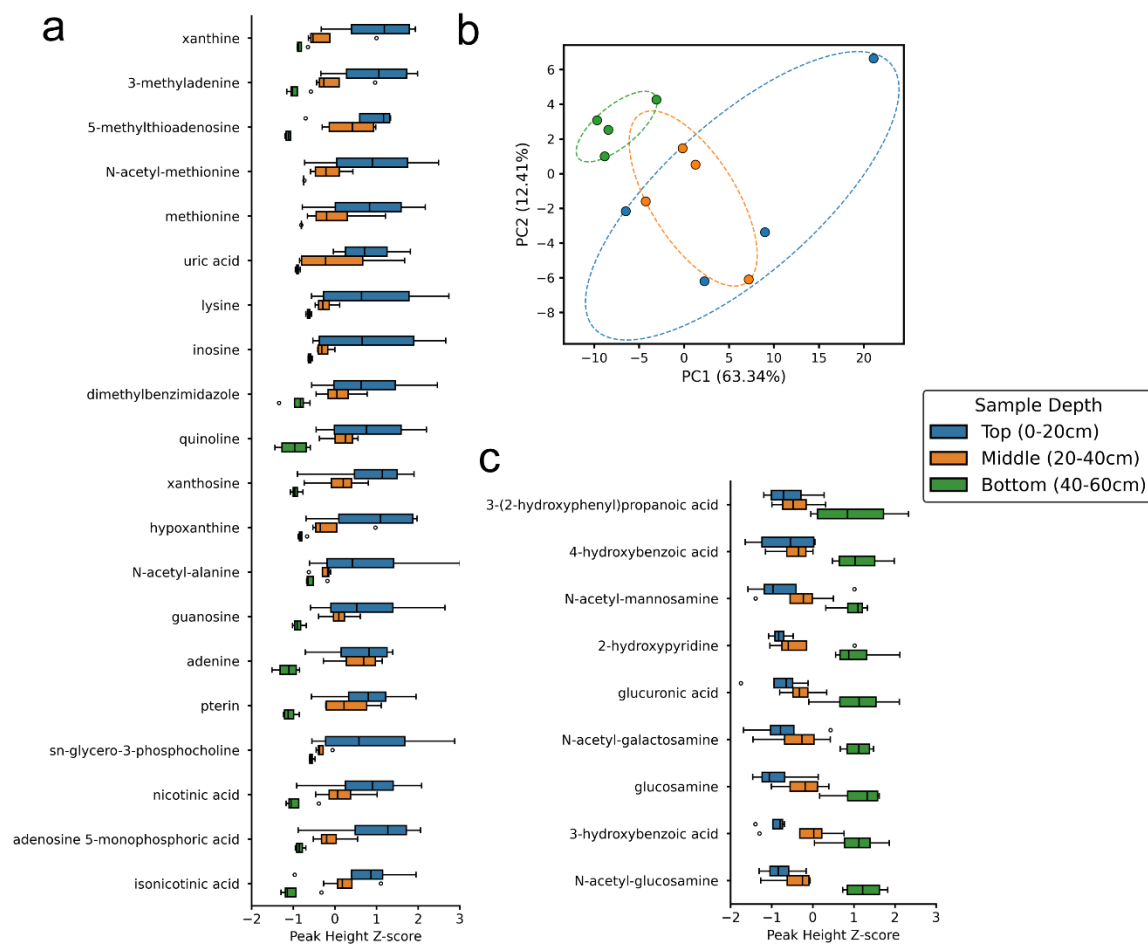


Figure 3-2 – Targeted metabolomics of wetland soil sections from May.

Z-score normalized peak heights of compounds most negatively (a) and positively (c) correlated with soil depth. PCA ordination of measured metabolites colored by depth (b).

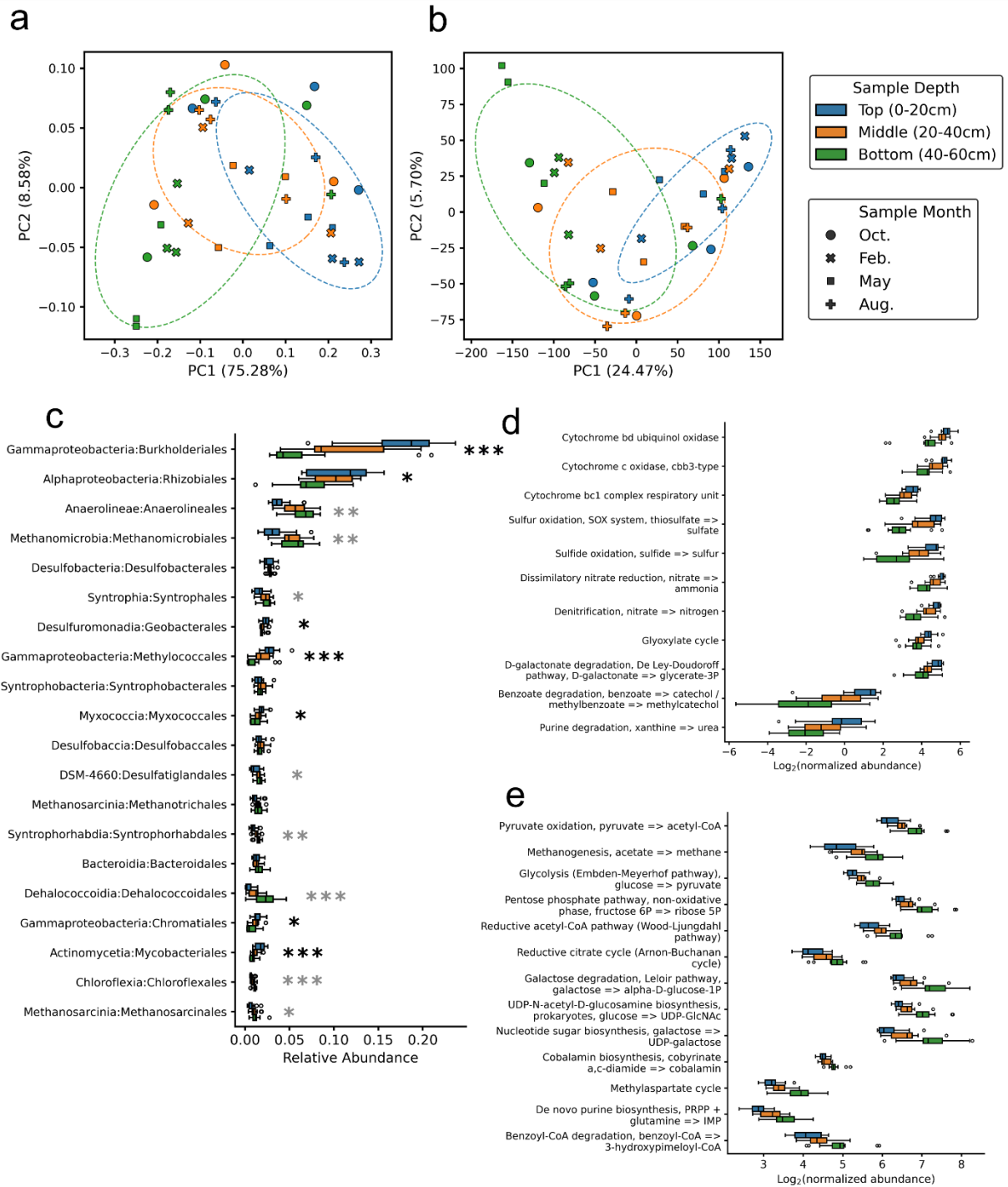


Figure 3-3 – Gene-centric metagenomics of wetland soil cores.

PCoA ordination of metagenome samples based on order-level taxonomic abundance (a) and KO gene ortholog abundance (b). Top 20 most abundant orders based on overall median abundance,

separated by soil depth section (c). Asterisks indicate the FDR adjusted significance level of the Spearman correlation coefficient (*** : $p < 0.001$, ** : $p < 0.01$, * : $p < 0.05$), black asterisks indicate orders negatively correlated with depth and gray indicate orders positively correlated with depth. Select KEGG modules with median abundance significantly negatively (d) and positively (e) correlated with depth.

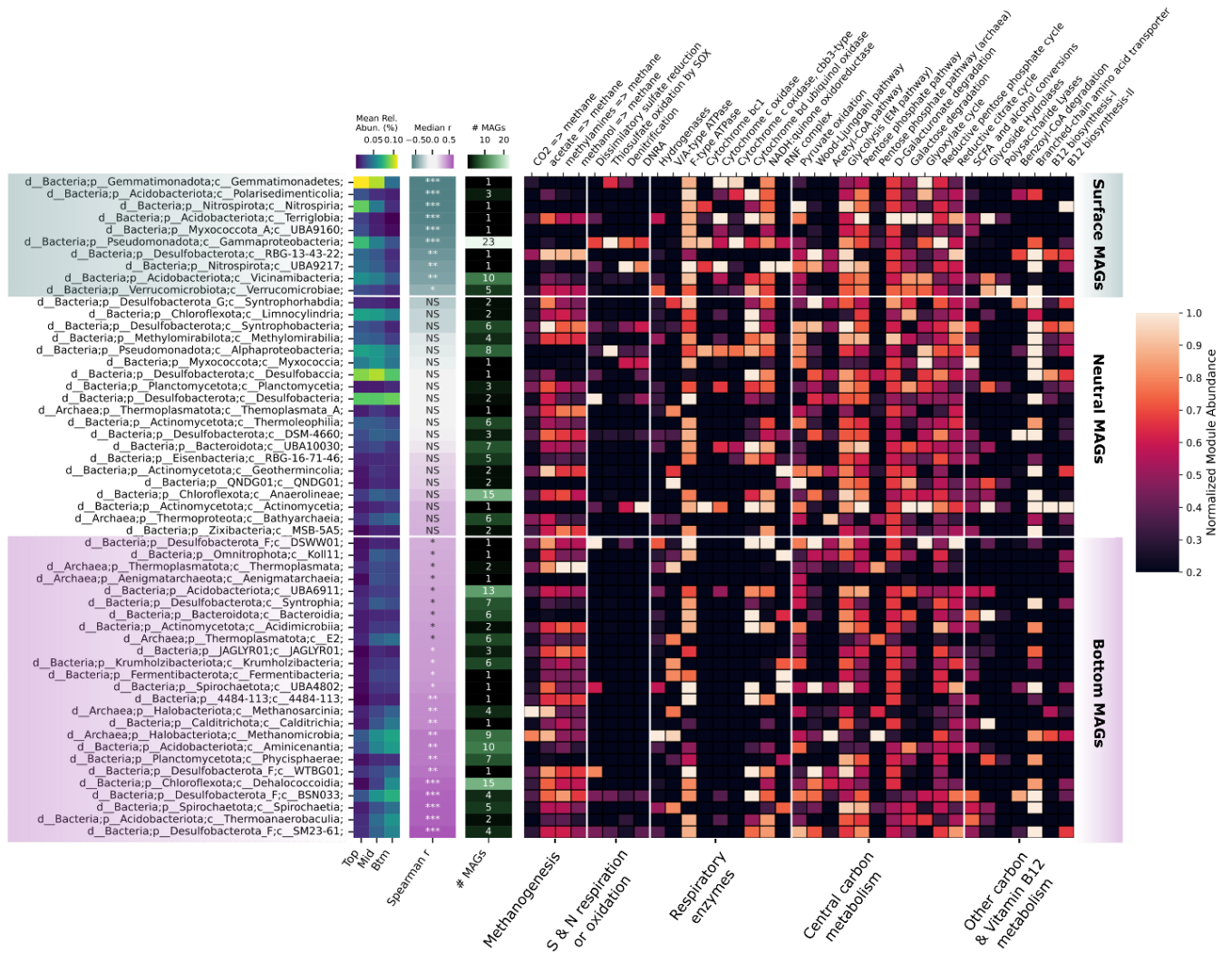


Figure 3-4 – Genome-centric metagenomics of wetland soil cores.

Summary of the most abundant MAGs (median relative abundance greater than 0.01%) grouped by class and sorted by depth abundance correlation. Heatmaps from left to right show mean relative abundance at each depth based on read mapping, median Spearman correlation coefficient with FDR corrected significance (***: $p < 0.001$, **: $p < 0.01$, * : $p < 0.05$), number of MAGs assigned to each class, and average functional module abundance in each class with columns normalized by the maximum value for module completion. Teal and pink

highlighted rows correspond to MAG classes with significantly negative and positive depth correlations, respectively.

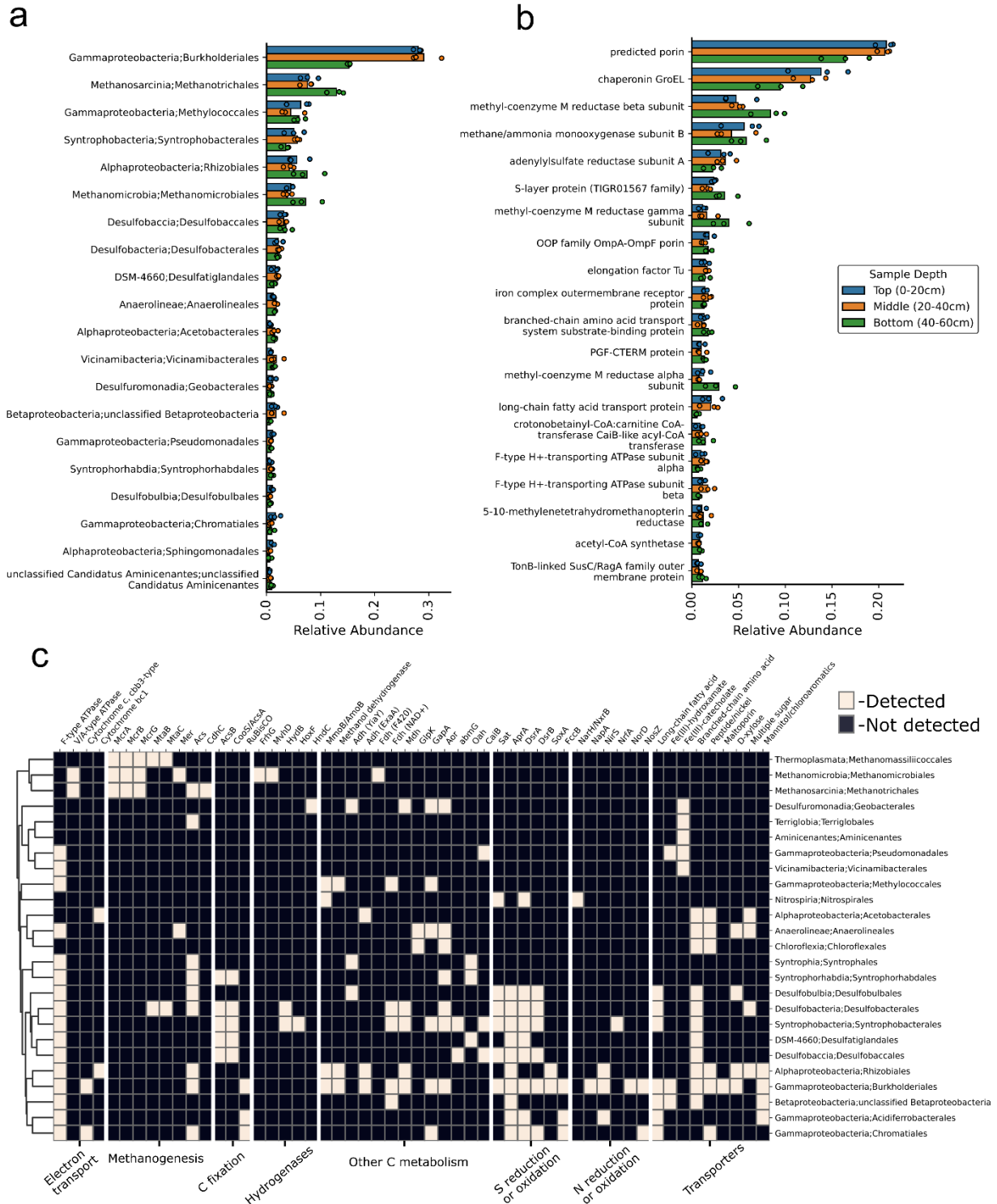


Figure 3-5 – Abundant proteins and taxa based on metaproteomics.

Relative abundance of detected proteins assigned to different taxonomic orders (a) and gene products (b). Clustermap of detected proteins and their order-level taxonomic assignment normalized to show which proteins were most prevalent within a particular order (c). Gene product name abbreviations: methyl-coenzyme M reductase (*McrABG*), methanol-5-hydroxybenzimidazolylcobamide Co-methyltransferase (*MtaB*), methanol corrinoid protein (*MtaC*), 5-10-methylenetetrahydromethanopterin reductase (*Mer*), acetyl-CoA synthetase (*Acs*), acetyl-CoA decarbonylase/synthase complex subunit beta (*CdhC*), acetyl-CoA synthase (*AcsB*), carbon-monoxide dehydrogenase catalytic subunit (*CooS/AcsA*), ribulose-bisphosphate carboxylase large chain (RuBisCO), coenzyme F420 hydrogenase subunit gamma (*FrhG*), F420-non-reducing hydrogenase iron-sulfur subunit (*MvhD*), [NiFe] hydrogenase large subunit (*HydB*), bidirectional [NiFe] hydrogenase diaphorase subunit (*HoxF*), NADP-reducing hydrogenase (*HndC*), methane/ammonia monooxygenase subunit B (*MmoB/AmoB*), alcohol dehydrogenase (*YiaY*), alcohol dehydrogenase (cytochrome c) (*ExaA*), formate dehydrogenase subunit beta (*Fdh* (F420)), formate dehydrogenase major subunit (*Fdh* (NAD⁺)), malate dehydrogenase (*Mdh*), glycerol kinase (*GlpK*), glyceraldehyde 3-phosphate dehydrogenase (*GapA*), aldehyde:ferredoxin oxidoreductase (*Aor*), 2-aminobenzoate-CoA ligase (*abmG*), 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase (*Oah*), crotonobetainyl-CoA:carnitine CoA-transferase CaiB-like acyl-CoA transferase (*CaiB*), sulfate adenylyltransferase (*Sat*), adenylylsulfate reductase subunit A (*AprA*), dissimilatory sulfite reductase (*DsrAB*), sulfur-oxidizing protein (*SoxA*), sulfide dehydrogenase flavoprotein (*FccB*), nitrate reductase/nitrite oxidoreductase beta subunit (*NarH/NxrB*), nitrate reductase NapA (*NapA*), nitrite reductase (NO-forming) (*NirS*), nitrite reductase (cytochrome c-552) (*NrfA*), nitric oxide reductase (*NorD*), nitrous-oxide reductase (*NosZ*).

Supplementary tables and figures

Table 3-1 – Bulk wetland soil characteristics.

SOIL PROPERTY	MEAN ± STDEV
WATER CONTENT (%)	72.8 ± 2.5
LOSS ON IGNITION (%)	21.0 ± 2.8
SILT (%)	86.3 ± 1.3
CLAY (%)	13.7 ± 1.3

Table 3-2 – Surface greenhouse gas fluxes at each sampling timepoint.

Values are mean ± standard deviation with units of $\text{nmol s}^{-1} \text{m}^{-2}$.

MONTH	CO ₂ FLUX (NMOL S ⁻¹ M ⁻²)	CH ₄ FLUX (NMOL S ⁻¹ M ⁻²)	N ₂ O FLUX (NMOL S ⁻¹ M ⁻²)
OCTOBER	598 ± 111	0.26 ± 0.16	-0.01 ± 0.10
FEBRUARY	876 ± 394	1.97 ± 1.09	-0.04 ± 0.40
MAY	315 ± 100	3.94 ± 0.68	0.04 ± 0.39
AUGUST	279 ± 92	8.72 ± 4.61	-0.04 ± 0.46
COMBINED	544 ± 340	3.59 ± 3.83	-0.01 ± 0.36

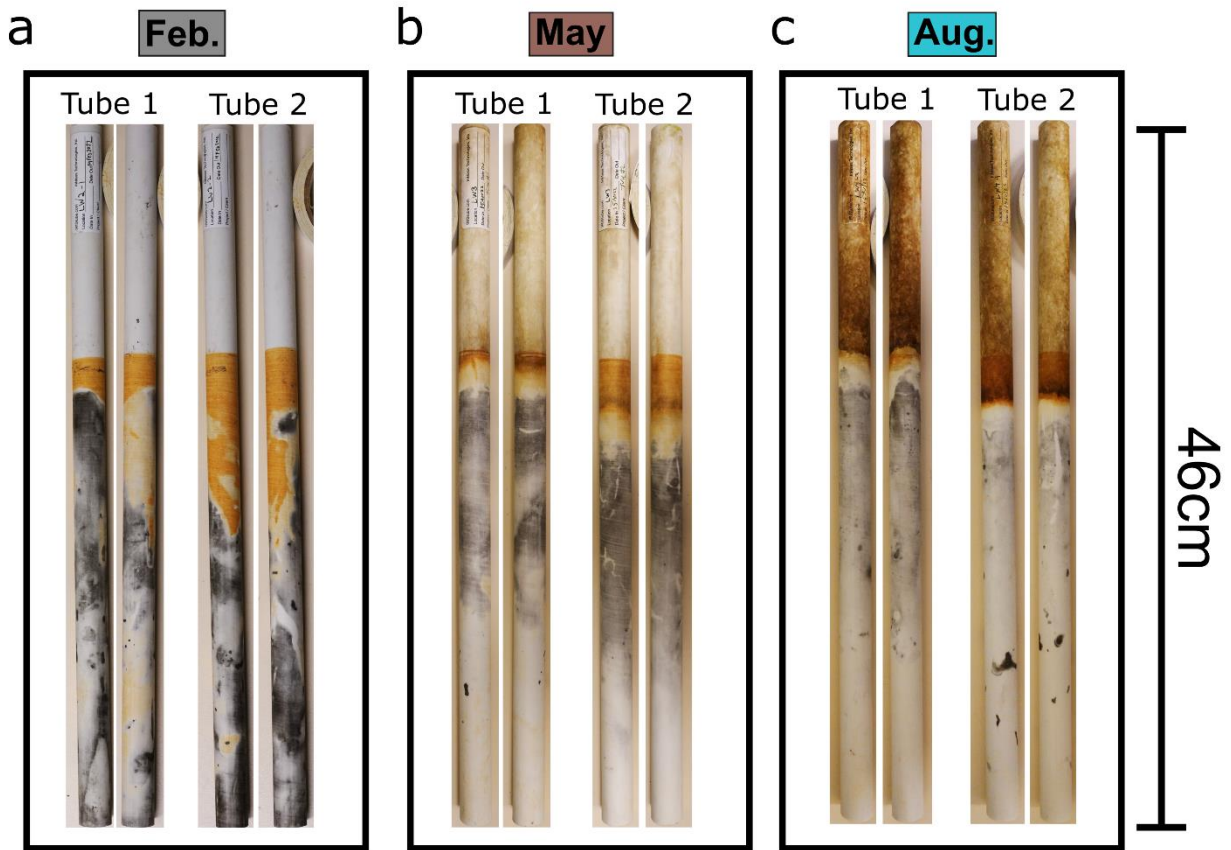


Figure S 3-1 – Indicator of Reduction in Soils (IRIS) tubes from three seasons.

The tubes were coated in an iron oxide coating (orange/red) and inserted into the soil. Reduction of the iron oxide coating leads to solubilization and reveals the white PVC tube underneath. The black areas are likely due to the formation of iron sulfide.

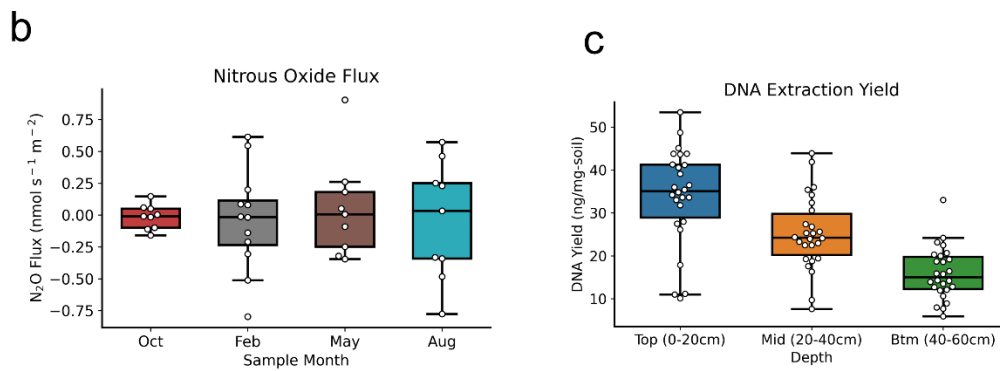
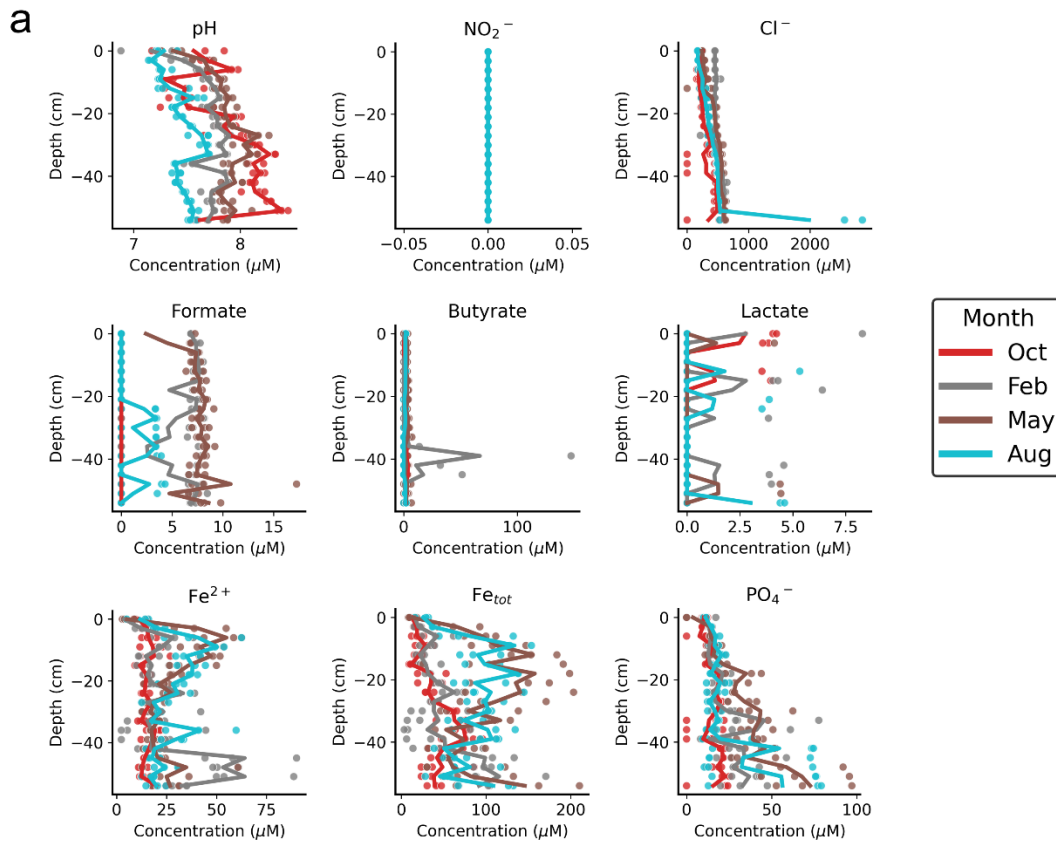


Figure S 3-2 – Additional measurements of site geochemistry.

Porewater metabolites (a), surface N_2O flux (b) and DNA extraction yields (c).

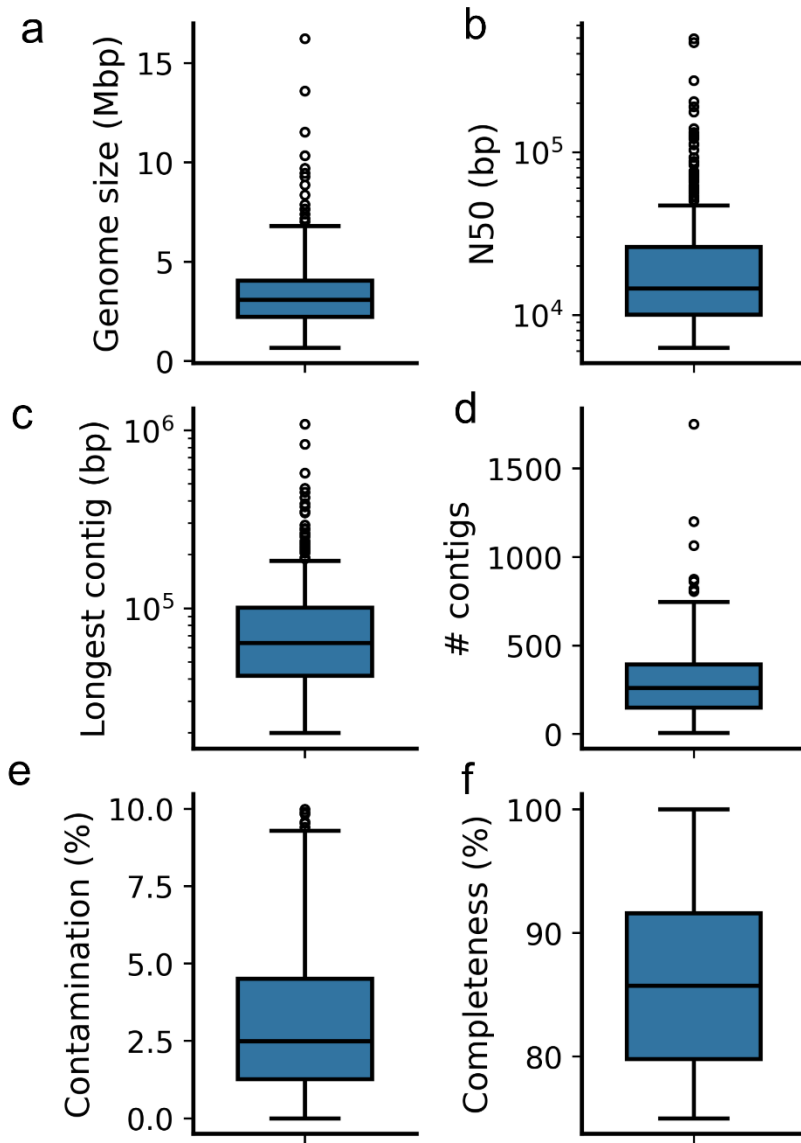


Figure S 3-3 – MAG quality statistics.

Boxplots showing the distribution of genome sizes (a), N50 (b), longest contig size (c), number of contigs (d), genome contamination (e) and completeness (f).



Figure S 3-4 – Proteins detected in different taxa.

Cumulative peak heights of the most abundant proteins detected in different taxa across all samples.

Chapter 4 – N-acetylglucosamine degradation by a wetland microbial community under varying oxygen concentrations

Zachary Flinkstrom^{1†}, Britt Abrahamson^{1†}, Demetrios Teodosiadis¹, Fischer Brunzell¹,
Korena Mafune¹, Mari-Karoliina H. Winkler¹

¹ Department of Civil and Environmental Engineering, University of Washington, Seattle, WA

† These authors contributed equally to this work

Abstract

N-acetylglucosamine (GlcNAc) — an amino sugar monomer — is a crucial component of cell wall polymers chitin and peptidoglycan, making it one of the most abundant organic compounds across domains and a primary component of soil organic carbon (SOC). While much is known about GlcNAc utilization by aerobic isolates, little is known about the assembly of microbial communities around GlcNAc and how GlcNAc recycling — a mineralization process explicitly linking the carbon and nitrogen cycle — influences SOC formation. This study investigated how wetland microbial communities mineralize GlcNAc into nitrogen and carbon and utilize GlcNAc derived metabolites under aerobic and anaerobic conditions. Two low-dilution rate chemostats were used to enrich GlcNAc recycling wetland microbial communities in an oligotrophic environment. Oxygen perturbations were applied throughout the enrichment to simulate periodic drought and flooding events while investigating their impact on metabolic rates, biomass yield, pathway selection, and greenhouse gas (methane, nitrous oxide, and carbon dioxide) production. The chemostats demonstrated distinct carbon and nitrogen metabolite dynamics, variations in microbial community composition, and transcriptomic responses. Despite temporary changes, communities ultimately reverted to prior states, highlighting how community resilience and versatility informs necromass turnover dynamics in natural environments. This study also highlights how GlcNAc recycling links the carbon cycle with nitrification and denitrification.

Introduction

Wetlands are critical carbon sinks that sequester 20-30% of global soil organic carbon (SOC) despite occupying 5-8% of land area¹. To ensure the long-term stability of this carbon reservoir, it is essential to understand how microbially mediated processes contribute to carbon mineralization and SOC formation. Traditionally, SOC formation was thought to arise from the progressive decomposition of complex plant litter to humic substances resistant to decomposition². However, microbial necromass, i.e. organic remnants of cellular decay, have been increasingly recognized as vital components of SOC³. The net rate of cellular decay (production), necromass consumption (recycling), and mineral stabilization contribute to SOC formation through the accumulation of necromass biomarkers⁴. It remains unclear if necromass recycling significantly contributes to carbon and nitrogen mineralization.

The amino sugar N-acetylglucosamine (GlcNAc) is the primary component of the cell wall polymers chitin and peptidoglycan found in fungal and prokaryotic biomass, respectively³. Amino sugars have emerged as a preferred necromass biomarker in sediments due to the ubiquity of GlcNAc across microbial domains. Fragmented cell wall residues — including (GlcNAc) — are an enriched component of necromass that comprise approximately 7% of SOC yet are a negligible fraction of active biomass³⁻⁶. Consequently, the recycling of the necromass biomarker GlcNAc represents an understudied mineralization process explicitly linking the carbon and nitrogen cycle. In our depth-resolved metabolomics study of wetland soils (Chapter 3), we observed that GlcNAc and other amino sugars were more abundant in deeper soil layers and another study found amino sugars to comprise an increasing fraction of soil organic matter in deeper layers⁷. The accumulation of GlcNAc with depth suggests this globally abundant amino sugar may become increasingly recalcitrant under anaerobic conditions, potentially due to

reduced microbial metabolic activity and altered decomposition pathways in oxygen-depleted environments.

While knowledge about GlcNAc utilization by complex microbial communities under environmentally relevant conditions remains sparse, the genomic and physiological aspects of GlcNAc utilization in bacterial isolates has been extensively studied. The pathway, facilitated by genes in the *nag* operon, involves four steps: transport into the cell, phosphorylation of GlcNAc to GlcNAc-6P, deacetylation to GlcN-6P, and deamination/isomerization to Fructose-6P which enters glycolysis⁸. This pathway has been studied in bacteria like *E. coli*⁹, *Aeromonas*^{10,11}, *Shewanella*⁸, *Bacteroidota*¹², *Vibrionaceae*¹³ and *Candidatus Marinimicrobia*¹⁴. However, community assembly around GlcNAc, the influence of oxygen on GlcNAc utilization, and whether there are GlcNAc utilizing specialists are less understood.

In this study, two low-dilution rate chemostats were used to enrich GlcNAc recycling microbial communities from wetland sediments under aerobic and anaerobic conditions to mimic oligotrophic wetland environments. We hypothesized that anoxic conditions would significantly constrain GlcNAc utilization rates and yields. Oxygen perturbations were applied throughout the enrichment to simulate periodic drought and flooding events mimicking those caused by regional weather patterns. We hypothesized that periodic oxygen perturbations would select for a resilient microbial community dominated by facultative anaerobes, capable of rapid adaptation to oxygen fluctuations. The chemostats demonstrated distinct carbon and nitrogen metabolite dynamics, variations in microbial community composition, and transcriptomic responses, highlighting the importance of necromass recycling as a significant nitrogen and carbon mineralization mechanism in the wetland environments.

Materials and Methods

Wetland cell extraction for reactor inoculum

A complex microbial community was extracted from a wetland soil sample taken from an urban lacustrine wetland (University of Washington Arboretum, Seattle, WA, USA - 47° 38' 31.12'' N, 122° 17' 47.01'' W). Samples were taken using a 2"x4" soil corer (AMS, American Falls, ID, USA) driven down 60 cm. Cores were transported to the lab on blue ice and stored at 4°C until processing. Cells were extracted in an anaerobic chamber (Coy Lab Products, Grass Lake, MI, USA) to minimize potential oxygen toxicity. Soil was first suspended (10% w/v) in modified lacustrine medium, followed by sieving (2 mm) to remove coarse particles. The sieved soil suspension was then paddle blended (Stomacher® 400 Circulator, Seward, UK) for 3 x 1 min at 250 rpm with 1-minute rests between runs. Samples were subsequently centrifuged at 200 g for 2 minutes to remove coarse organics and filtered through a 20 µm filter (Whatman™ Grade 41, Cytiva, Marlborough, MA, USA) to remove eukaryotes and fungal material.

Reactor design and operation

Two chemostats were operated at 20 °C with GlcNAc (5 mM influent concentration) as the sole carbon source. Custom 2.7 L glass chemostats were designed for this study. Chemostats were filled with 1.45 L of sterile GlcNAc lacustrine medium and inoculated with 50 mL of wetland sediment extract. The operating working volume of each chemostat was maintained at 1.5L with a 1.2 L headspace. Anoxic conditions were maintained by continuously sparging the chemostats with nitrogen through a 0.2-micron Teflon filter (Millipore-Sigma, St. Louis, MO, USA) at a flowrate of 20 mL·min⁻¹ with a mass flow controller (Alicat scientific, Tucson, AZ, USA) maintained between 1-2 psig. Aerobic conditions were maintained by continuously

sparging the chemostat with air at a flowrate of 27 mL·min⁻¹. After inoculation, the chemostats were mixed without for 24 hours.

The chemostats were fed with lacustrine freshwater media containing 5mM GlcNAc using a drip-feed at a rate of 7.5 mL·h⁻¹ for the remainder of the experiment (0.12 day⁻¹ dilution rate; 8.3 day hydraulic retention time). The liquid volume within the reactor was maintained with an effluent pump. The pH was continuously monitored with a pH probe (Hamilton, Reno, NV, USA) and automatically controlled at 7.5 ± 0.1 using 1M HCl and 1M NaOH. Dissolved oxygen probes (Hamilton, Reno, NV, USA) were used to continuously monitor dissolved oxygen concentrations. Reactors were continuously stirred at approximately 500 rpm using magnetic stir bars and stir plates. The walls of the bioreactors were regularly scraped to avoid the accumulation of biofilms.

The lacustrine media was based on DSMZ 503 (anaerobic freshwater medium) with concentrations updated to reflect averaged porewater measurements of the top 60 cm of sediment at the wetland sampling site. The basal medium composition was as follows (in 1000mL MQ water): GlcNAc (5mM), 6.4 mg (NH₄)₂SO₄ (0.12mM), 35.9 mg CaCl₂·2H₂O (0.24mM), 33.2 mg MgCl₂·6H₂O (0.16mM), 1 mL SL-10 trace element solution, 1 mL selenate-tungstate (Se-W) solution (Table 4-2). The SL-10 trace element solution consisted of (in 1000 mL MQ water): 10 mL 7.7M HCl, 1.5 g FeCl₂·4H₂O, 0.07 g ZnCl₂, 0.15 g MnCl₂·4H₂O, 0.006 g H₃BO₃, 0.19 g CoCl₂·6H₂O, 0.002 g CuCl₂·2H₂O, 0.024 g NiCl₂·6H₂O, 0.036 g Na₂MoO₄·4H₂O. The Se-W solution contained (in 1000 mL MQ water): 0.5 g NaOH, 0.003 g Na₂SeO₃·5H₂O, 0.004 g Na₂WO₄·2H₂O. After autoclaving, heat-sensitive media components were aseptically added as follows (in 1000mL MQ water): 0.3 mL 1M KHCO₃, 0.4 mL 1M NaH₂PO₄, 1.2 mL 0.5M Na₂HPO₄, 0.55 mL 0.5% (NH₄)₂S, and 0.1 mL 7-vitamin solution. The 7-vitamin solution

consisted of (in 100 mL MQ water): 0.1 g vitamin B12, 0.08 g p-aminobenzoic acid, 0.02 g D(+)-Biotin, 0.2 g nicotinic acid, 0.1 g Ca-pantothenate, 0.3 g pyridoxine hydrochloride, 0.2 g thiamine-HCl·2H₂O.

Oxygen perturbations occurred on day 68 and 187 when reactor behavior was stable. Each perturbation lasted 9 and 8 days, respectively, which corresponds to approximately one hydraulic retention time. The perturbation was induced by switching the inlet sparging gases. Anaerobic conditions were maintained in the previously aerobic chemostat with nitrogen and aerobic conditions were maintained in the previously anaerobic chemostat with air.

Sampling and chemical analysis

Liquid and gas samples were collected two to three times per week. Liquid samples were collected from a three-way sampling port attached to the influent and effluent lines. Cell concentrations were estimated from a biomass using optical density at 600 nm measured with a 96-well plate reader. Samples for metabolite analysis were centrifuged at 21500 g for 20 minutes and stored at -20°C. A pH microelectrode (Orion; ThermoScientific) was used to measure pH prior to sample storage. Ammonium, nitrite, and nitrate concentrations were measured from reactor effluent and influent using a Gallery Discrete Analyzer using off-the-shelf assay kits. Concentrations of GlcNAc, organic acids, sugars, and alcohols were measured by high-performance liquid chromatography (HPLC) with an Aminex column and UV detection. Cell dry weight measurements were performed by filtering 10 mL of effluent onto pre-dried and weighed PES filters which were then dried at 105°C to constant weight.

Gas production rates were measured using cavity ring-down spectroscopy (Picarro model G2508) by connecting the analyzer to the reactor headspace and recirculating the gas to achieve

steady-state measurements. Gas production rates were then calculated using information about sparging gas flow rate and the volume of the setup.

DNA extraction and metagenomic sequencing

DNA was extracted from biomass pellets stored at –80C using the DNeasy PowerBiofilm kit (Qiagen) according to the manufacturer's instructions. DNA concentrations were quantified using the Qubit HS dsDNA assay. Samples were sent to the Joint Genome Institute for sequencing.

Plate-based DNA library preparation for Illumina sequencing was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Kapa Biosystems library preparation kit. 200 ng of sample DNA was sheared to 500 bp using a Covaris LE220 focused-ultrasonicator. The sheared DNA fragments were size selected by double-SPRI and then the selected fragments were end-repaired, A-tailed, and ligated with Illumina compatible sequencing adaptors from IDT containing a unique molecular index barcode for each sample library. The prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using NovaSeq XP V1.5 reagent kits, S4 flowcell, following a 2x151 indexed run recipe.

RNA extraction and metatranscriptomic sequencing

Biomass pellets for RNA extraction were immediately frozen on dry ice and stored at –80C to prevent degradation. RNA was extracted using the RNeasy PowerBiofilm kit (Qiagen) according to the manufacturer's instructions. An additional DNase treatment step was performed

to ensure no residual DNA remained. RNA was quantified using the Qubit BR RNA assay and RNA quality was assessed using a BioAnalyzer before sending samples to the Joint Genome Institute for library preparation sequencing.

Metagenomics and metatranscriptomics data analysis

Filtered metagenomic reads were assembled with metaSPAdes v3.15.2¹⁵ and filtered metatranscriptomic reads were de-novo assembled with rnaSPAdes v4.0.0¹⁶. Contigs smaller than 200 bp were removed. Alignment information was determined by mapping reads to the assembly reference with BBtools¹⁷ version 38.95 [command: bbmap.sh -Xmx450g nodisk=true interleaved=true ambiguous=random mappedonly=t trimreaddescriptions=t usemodulo=t fast=t]. Coverage was determined by running BBTools version 38.95 [command: pileup.sh]. The IMG/M pipeline¹⁸ was used to annotate the assembled contigs, and prokaryotic bins were predicted using metaBAT¹⁹.

For gene-based analysis, taxonomic and functional profiles were retrieved from the JGI IMG/M portal using the Abundance Profile tool. For functional profiles, the estimated gene copies of KEGG KOs were used and center-log₂ ratio (clr) normalized on a per sample basis based on the method from ALDEx2²⁰. For taxonomic profiles, the estimated gene copies for order-level GTDB taxonomy at 60+ % identity were used and normalized using relative abundance.

Results

Carbon and nitrogen cycling dynamics reflect O₂-dependent shifts in community function

Two planktonic wetland microbial communities were enriched under anaerobic and fully oxic conditions to quantitatively link necromass mineralization rates with community structure in response to fluctuating oxygen availability. N-Acetylglucosamine (GlcNAc) served as the sole carbon, nitrogen, and energy source with no dissolved electron supplied. All substrates were continuously provided with a dilution rate of 0.005 h^{-1} (Table 4-1). Fully anoxic and aerobic conditions were ensured by continuous nitrogen (N_2) and air sparging, respectively. Dissolved oxygen was maintained above 6.5 mg/L ($>75\%$ air saturation) under aerobic conditions, and oxygen (O_2) served as the sole electron acceptor. Two ~ 8.5 -day drought-flooding perturbations were simulated on day 67.9 and 188.8 by switching the air/ N_2 supply to the anaerobic and aerobic chemostats. The chemostats were reverted to their prior condition upon terminating the perturbation. These experimental transitions provided insights into how shifting oxygen concentrations influences community function, structure, and stability in wetland-like environments.

Each chemostat progressed through distinct operational phases, including start-up, steady-state, transient disturbances, and recovery. Fermentation dominated the two steady-states (Phase 2 & 4) in the anaerobic chemostat, as indicated by their high biomass-specific production rate of acetate and propionate (Fig. 4-1A; Table 4-1). Acetate was the primary GlcNAc mineralization product and was gradually decreased throughout the initial 5 phases of the enrichment. Following the second O_2 perturbation (Phase 6), a sharp increase in CH_4 production marked the transition to methanogenesis (Phase 7), which coincided with the rapid consumption of acetate and propionate. Biomass-specific CO_2 and biomass production increased during the O_2 perturbation (Phase 3 & 5) suggesting the anaerobic microbial community enhanced respiration and growth in response to O_2 availability. Transient anaerobic features emerged in the aerobic

chemostat during N₂ perturbations (Fig 4-1B). Notably, acetate accumulated in response to the two N₂ pulses (Phase 3 & 5) demonstrating the community shifted from heterotrophic respiration to fermentative metabolisms. Surprisingly, CH₄ was detected throughout the first 4 phases of the aerobic GlcNAc degrading enrichment—albeit at rates below $0.02 \pm 0.018 \mu\text{mol}\cdot\text{C}\cdot\text{mg}_{\text{DCW}}^{-1}\cdot\text{h}^{-1}$ —and was no longer observed at the beginning of the second steady-state. As observed in the anaerobic chemostat, CO₂ production increased with O₂ availability. However, the biomass production rates remained consistently higher in the aerobic chemostat, underscoring the increased yield of heterotrophic respiration relative to fermentation.

While both chemostats initially mineralized GlcNAc to ammonium (NH₄⁺), the trajectory of nitrogen transformation significantly differed between the anaerobic and aerobic microbial communities. Nitrogen cycling was largely inactive in the anaerobic chemostat (Fig. 1A), as indicated by the constant NH₄⁺ production rate and absence of other nitrogen species. In contrast, the aerobic chemostat displayed dynamic nitrogen transformations throughout the enrichment (Fig 1B). Before the first N₂ perturbation, NH₄⁺ was stoichiometrically oxidized to nitrite (NO₂⁻) with a steady-state biomass-specific ammonia-oxidation rate of $3.612 \pm 0.601 \mu\text{mol}\cdot\text{N}\cdot\text{mg}_{\text{DCW}}^{-1}\cdot\text{h}^{-1}$ and transitioned to complete nitrification after the N₂ fluctuation. This transition was characterized by increasing nitrate (NO₃⁻) production as NO₂⁻ production decreased in Phase 5 and 7. Nitrous oxide (N₂O), a potent greenhouse gas, was produced at higher rates in the aerobic chemostat at steady-state and exhibited significant spikes immediately after all O₂ fluctuations. This suggests constitutive N₂O production by the nitrifying population and that it responds to changes in O₂ availability. The largest N₂O pulse occurred immediately after the onset of second N₂ perturbation, when NO₃⁻ was the predominant nitrogen species. Nitrate-reduction by facultative partial denitrifiers likely contributed to this massive spike in N₂O. The

different N₂O dynamics reflects the expression of distinct N₂O production pathways in the aerobic and anaerobic chemostat.

Microbial community response

Microbial cell densities were higher in the aerobic chemostat compared to the anaerobic chemostat throughout the study period (Fig. 4-1). In general, the ODs were more stable in the anaerobic chemostat, while the aerobic chemostat experienced fluctuations particularly around regular cleaning of biofilms from the reactor vessel walls. The perturbations caused an increase in OD in the anaerobic chemostat and a decrease in the aerobic chemostat.

The two reactors maintained distinct taxonomic and functional compositions throughout the experiment, as confirmed by both PERMANOVA and ANOSIM analyses ($p < 0.001$; Fig. 4-3). Prior to perturbation, the anaerobic reactor was primarily dominated by Gammaproteobacteria, Clostridia, and Bacteroidia, whereas the aerobic reactor was enriched in Alphaproteobacteria, Gammaproteobacteria, and Bacteroidia (Fig. 4-2A, C). Minor taxa also differed between reactors: the anaerobic reactor harbored Fusobacteriia, Bacilli, Desulfuromonadia, and Desulfovibrionia, while the aerobic reactor contained low-abundance populations of Verrucomicrobia and Fimbriimonadia. At the family level, Rhodocyclaceae (Gammaproteobacteria) steadily increased up to 60% abundance in the anaerobic reactor, while Devosiaceae (Alphaproteobacteria) was the dominant family in the aerobic reactor though its abundance declined in the second half of the experiment (Fig. S 4-5).

During the pre-perturbation phase, the anaerobic reactor exhibited significantly higher taxonomic diversity compared to the aerobic reactor (Fig. S 4-4A, B). Both systems, however, experienced pronounced shifts in microbial composition in response to their respective oxygen

perturbations. In the anaerobic reactor, perturbations led to increases in Alphaproteobacteria and Bacteroidia and decreases in Clostridia and Fusobacteriia (Fig. 4-2A, B). Conversely, the aerobic reactor responded to perturbation with declines in Alphaproteobacteria and increases in Clostridia and Gammaproteobacteria; notably, Bacilli became more prominent during the second perturbation (Fig. 4-2C, D).

Following the first perturbation, the anaerobic reactor showed a decline in both taxonomic and functional diversity, while diversity in the aerobic reactor remained stable or increased. This contrast became especially pronounced during the second perturbation, during which the aerobic reactor exhibited significantly higher diversity than the anaerobic counterpart. Recovery trajectories also differed between reactors. In the aerobic reactor, the taxonomic composition of transcripts rapidly returned to pre-perturbation states, while the anaerobic reactor failed to revert to its original community structure (Fig. 4-3B). Despite these taxonomic differences, both reactors exhibited similar trends in the recovery of functional composition (Fig. 4-3D).

Over the course of the experiment, Gammaproteobacteria became increasingly dominant in the anaerobic reactor, while Verrucomicrobiae steadily increased in the aerobic reactor. Notably, Bacteroidia remained relatively stable in both systems across all operational phases. Ultimately, both reactors underwent substantial shifts in taxonomic and functional composition, culminating in distinct endpoint communities. While the aerobic reactor experienced a more pronounced shift in taxonomic structure (Fig. 4-3A), both reactors showed comparable changes in functional composition (Fig. 4-3C).

Differences in metabolic pathways between reactors and across operational phases

Clear differences in amino sugar metabolic pathways were observed between the two reactors. The *ngcE* ABC transporter for GlcNAc was significantly more abundant in the aerobic reactor, while the *nagE* PTS transport system was more abundant in the anaerobic reactor (Fig. 4-4). Similarly, *gspK* (glucosamine kinase) and *dac* (GlcNAc deacetylase) were enriched in the aerobic reactor, whereas *nagB* (glucosamine-6-phosphate deaminase) was more abundant in the anaerobic reactor (Fig. 4-4). Both systems exhibited active transcription of chitinases, notable given that chitin is a polymer of GlcNAc. A schematic representation of GlcNAc pathway differences is shown in Figure 5, highlighting that enzymes favored in the anaerobic system require less ATP. Specifically, the PTS transport system utilizes phosphoenolpyruvate (PEP) to simultaneously transport and phosphorylate GlcNAc, while the *ngcE* ABC transporter consumes one ATP for transport and another for subsequent phosphorylation.

Distinct patterns also emerged in one-carbon (C1) metabolism. As expected, genes involved in methanogenesis were significantly enriched in the anaerobic reactor metagenomes and metatranscriptomes, including *mcr*, *mvh*, *cdh*, *mta*, and *mtt*, suggesting methane production from hydrogen, acetate, methanol, and methylamines. Surprisingly, methanogenesis-related transcripts were also detected in the aerobic reactor during the perturbation phase. Methanol dehydrogenases (*xox* and *mxh*) were enriched in the aerobic reactor, although their abundance increased in the anaerobic reactor following the oxygen perturbation (Fig. 4-4).

Utilization of organic acids and alcohols also varied by reactor. Genes for acetate, propionate, and butyrate kinases, as well as ethanolamine utilization, were more abundant in the anaerobic reactor metagenomes, with corresponding transcript levels increasing in the aerobic reactor during the N₂ perturbation. Conversely, a propanol-preferring alcohol dehydrogenase, L-lactate dehydrogenase, and acetoacetate decarboxylase were more prevalent in the aerobic

reactor, with their expression also stimulated by oxygen perturbation in the anaerobic reactor. Numerous other alcohol dehydrogenases were expressed with reactor-specific preferences.

Cytochrome terminal oxidases were generally more abundant in the aerobic reactor and were upregulated in response to the N₂ perturbation. In contrast, sodium-pumping respiratory complexes (*rnf*, *nqr*) and hydrogenases (*hnd*, *ech*, *mvh*) were significantly more abundant in the anaerobic reactor.

The key ammonia oxidation gene, *amo*, was more abundant in the aerobic reactor, although its overall abundance was low due to its specialist function. Transcripts for *nar* (nitrate reductase)/ *nxr* (nitrite oxidoreductase) and *nirK* (nitrite reductase) were significantly more abundant in the aerobic reactor, whereas *nrf* (associated with dissimilatory nitrate reduction to ammonia) and *nas* (assimilatory nitrate reductase) transcripts were enriched in the anaerobic reactor. *Nor* (nitric oxide reductase) was more abundant in the anaerobic reactor metagenomes. Notably, denitrification genes showed little transcriptional response to perturbations, with the exception of a decrease in *nap* (nitrate reductase) expression in the aerobic reactor and an increase in *nos* (nitrous oxide reductase) expression in the anaerobic reactor (Fig. 4-4).

Discussion

This study examined how the ubiquitous structural amino sugar N-acetylglucosamine (GlcNAc) was mineralized by wetland microbial communities in aerobic and anaerobic low-dilution-rate chemostats, to better understand necromass turnover in wetlands. Transient perturbations were used to probe community responses to shifts in oxygen availability—an analog for natural drought and flooding events. GlcNAc was readily mineralized under all conditions, but downstream metabolite dynamics and microbial community composition differed

markedly between the two reactors. The anaerobic reactor, dominated by Gammaproteobacteria, showed steadily increasing methane concentrations and declining levels of organic acids, suggesting the gradual establishment of acetoclastic methanogenesis and resilience to transient oxygen exposure. In contrast, the aerobic reactor, dominated by Alphaproteobacteria, mineralized GlcNAc completely to carbon dioxide. However, under nitrogen gas (N₂) perturbations, organic acids rapidly accumulated and oxidized nitrogen was depleted, indicating a metabolic shift from aerobic respiration to fermentation and denitrification. Over time, the aerobic system developed complete nitrification, while the anaerobic reactor consistently retained residual ammonium in the effluent, highlighting a tight coupling between carbon and nitrogen cycling.

This study highlights how amino sugars, such as GlcNAc, serve as important links between the carbon and nitrogen cycles, with their breakdown feeding into nitrification, denitrification, dissimilatory nitrate reduction to ammonium (DNRA), and assimilation pathways. We detected transcripts for ammonia monooxygenase from *Nitrosomonas oligotropha*, an ammonia-oxidizing bacterium known to be enriched under low-ammonia conditions^{21,22}. In the aerobic reactor, nitrite and nitrate produced via nitrification were rapidly denitrified during the N₂ perturbation, resulting in sharp spikes in nitrous oxide emissions. Elevated N₂O emissions were also consistently observed during aerobic phases, potentially due to ammonia oxidation²³ or aerobic denitrification—a process recently reported in nitrification-inhibited chemostats exposed to frequent oxygen fluctuations²⁴. Supporting the possibility of aerobic denitrification, we observed high expression of denitrification-related gene transcripts across all oxygen conditions.

We observed oxygen-dependent variation in the preferred gene systems for GlcNAc utilization. Under anaerobic conditions, GlcNAc transport was primarily mediated by the *nagE* phosphotransferase system (PTS), whereas the *ngc* ABC transporter was more prevalent under aerobic conditions. The PTS system couples GlcNAc transport with phosphorylation, driven by phosphoenolpyruvate, offering an energetically efficient mechanism²⁵. In contrast, the ABC transporter relies on ATP for transport and does not confer simultaneous phosphorylation²⁶. However, it provides a key advantage in its higher affinity for GlcNAc ($K_M = 0.5 \mu\text{M}$ vs. $5 \mu\text{M}$ for the PTS system) and its dual activity for transporting chitobiose (GlcNAc_2)^{27,28}. Thus, the PTS system may be favored in anaerobic, energy-limited environments due to its efficiency, while the ATP-dependent, high-affinity ABC transporter may be advantageous under aerobic conditions, where organisms can afford greater energy expenditure to outcompete others for substrate, thereby reducing residual GlcNAc concentrations. Indeed, this dichotomy between ABC and PTS-mediated transport has been explored in the context of marine bacteria, suggesting that ABC transporters allow for oligotrophic organisms to attain nano-molar half-saturation constants²⁹.

The aerobic and anaerobic reactors followed distinct microbial succession trajectories, despite being inoculated with the same wetland soil extract. This underscores the powerful selective pressure oxygen exerts on microbial communities. We hypothesized that alternating oxygen conditions might drive convergence between systems; however, despite some functional and taxonomic overlap, the communities remained distinct. Both reactors favored facultative anaerobes, but only the anaerobic system could support aero-tolerant or transiently surviving strict anaerobes, which may explain the observed decline in Shannon diversity. Still, the rapid return of methanogenic activity suggests that some strict anaerobes endured transient oxygen

exposure, as has been demonstrated in wetland sediments³⁰. Methane and low levels of *mcr* transcripts detected in the aerobic reactor during the N₂ phase further suggest the persistence of methanogens, likely protected in biofilms or microsites.

Fluctuating redox conditions, common in wetland environments, shape microbial activity and carbon cycling. Our chemostat systems serve as valuable models for probing these dynamics. While GlcNAc was readily mineralized under all conditions, aerobic systems had higher energy availability that likely supported greater investment in high-affinity transporters and hydrolytic enzymes. Although we studied the GlcNAc monomer, it is typically present in complex polymers such as chitin or peptidoglycan, which require additional enzymatic breakdown—an energy-intensive process that is prone to “cheaters”^{31,32} and may be limited under anaerobic conditions. This energetic constraint along with limits on transporter affinity could help explain the accumulation of necromass and amino sugars in deeper, anoxic wetland soils.

The recognition that microbial necromass forms a substantial component of stable soil organic matter (SOM) has reshaped our understanding of microbial contributions to long-term carbon storage³³. These findings extend beyond wetlands, with implications for marine systems, wastewater treatment, and anaerobic digestion^{34,35}. Ultimately, understanding how microbial biomass is recycled—and under what conditions carbon is retained or lost—requires both taxonomic insight and physiological characterization, which chemostat systems are well-suited to deliver.

Future work should expand on these findings by investigating larger biopolymers like chitin, assessing how hydrolysis demands influence community assembly, or testing how

resource limitation and starvation phases affect necromass recycling. Together, these efforts can move us closer to a predictive understanding of microbial contributions to SOM formation and persistence. Such insight is critical for designing strategies to enhance soil carbon sequestration and support climate-resilient ecosystem management.

Acknowledgements

This work was funded by the US Department of Energy under award number DE-SC0020356.

References

1. Candry, P., Abrahamson, B., Stahl, D. A. & Winkler, M. H. Microbially mediated climate feedbacks from wetland ecosystems. *Glob Chang Biol* **29**, 5169–5183 (2023).
2. Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).
3. Glaser, B., Turrión, M.-B. & Alef, K. Amino sugars and muramic acid—biomarkers for soil microbial community structure analysis. *Soil Biol Biochem* **36**, 399–407 (2004).
4. Buckeridge, K. M., Creamer, C. & Whitaker, J. Deconstructing the microbial necromass continuum to inform soil carbon sequestration. *Funct Ecol* **36**, 1396–1410 (2022).
5. Salas, E. *et al.* Reevaluation and novel insights into amino sugar and neutral sugar necromass biomarkers in archaea, bacteria, fungi, and plants. *Science of The Total Environment* **906**, 167463 (2024).
6. Ni, X. *et al.* The vertical distribution and control of microbial necromass carbon in forest soils. *Global Ecology and Biogeography* **29**, 1829–1839 (2020).
7. Ni, X. *et al.* A quantitative assessment of amino sugars in soil profiles. *Soil Biol Biochem* **143**, (2020).
8. Yang, C. *et al.* Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *Journal of Biological Chemistry* **281**, 29872–29885 (2006).
9. Plumbridge, J. A. Repression and induction of the *nag* regulon of *Escherichia coli* K-12: the roles of *nagC* and *nagA* in maintenance of the uninduced state. *Mol Microbiol* **5**, 2053–2062 (1991).

10. Li, S. W., He, H., Zeng, R. J. & Sheng, G. P. Chitin degradation and electricity generation by *Aeromonas hydrophila* in microbial fuel cells. *Chemosphere* **168**, 293–299 (2017).
11. Stumpf, A. K., Vortmann, M., Dirks-Hofmeister, M. E., Moerschbacher, B. M. & Philipp, B. Identification of a novel chitinase from *Aeromonas hydrophila* AH-1N for the degradation of chitin within fungal mycelium. *FEMS Microbiol Lett* **366**, (2019).
12. Cottrell, M. T. & Kirchman, D. L. *Natural Assemblages of Marine Proteobacteria and Members of the Cytophaga-Flavobacter Cluster Consuming Low-and High-Molecular-Weight Dissolved Organic Matter*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* vol. 66 <https://journals.asm.org/journal/aem> (2000).
13. Riemann, L. & Azam, F. Widespread N-acetyl-D-glucosamine uptake among pelagic marine bacteria and its ecological implications. *Appl Environ Microbiol* **68**, 5554–5562 (2002).
14. Katayama, T. *et al.* A Marine Group A isolate relies on other growing bacteria for cell wall formation. *Nat Microbiol* **9**, 1954–1963 (2024).
15. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).
16. Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* **8**, (2019).
17. Bushnell, B. BBTools. <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>.
18. Chen, I. M. A. *et al.* The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res* **51**, D723–D732 (2023).
19. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, (2015).
20. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
21. Bollmann, A. & Laanbroek, H. J. Continuous culture enrichments of ammonia-oxidizing bacteria at low ammonium concentrations. *FEMS Microbiol Ecol* **37**, 211–221 (2006).
22. Limpiyakorn, T., Kurisu, F., Sakamoto, Y. & Yagi, O. Effects of ammonium and nitrite on communities and populations of ammonia-oxidizing bacteria in laboratory-scale continuous-flow reactors. *FEMS Microbiol Ecol* **60**, 501–512 (2007).
23. Prosser, J. I., Hink, L., Gubry-Rangin, C. & Nicol, G. W. Nitrous oxide production by ammonia oxidizers: Physiological diversity, niche differentiation and potential mitigation

- strategies. *Global Change Biology* vol. 26 103–118 Preprint at <https://doi.org/10.1111/gcb.14877> (2020).
24. Roothans, N. *et al.* Aerobic denitrification as an N₂O source from microbial communities. *ISME Journal* **18**, (2024).
 25. McCoy, J. G., Levin, E. J. & Zhou, M. Structural insight into the PTS sugar transporter EIIC. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 577–585 (2015).
 26. Thomas, C. & Tampé, R. Structural and Mechanistic Principles of ABC Transporters. *Annu Rev Biochem* **89**, 605–636 (2020).
 27. Wang, F., Xiao, X., Saito, A. & Schrepf, H. *Streptomyces olivaceoviridis* possesses a phosphotransferase system that mediates specific, phosphoenolpyruvate-dependent uptake of N-acetylglucosamine. *Molecular Genetics and Genomics* **268**, 344–351 (2002).
 28. Saito, A. & Schrepf, H. Mutational analysis of the binding affinity and transport activity for N-acetylglucosamine of the novel ABC transporter Ngc in the chitin-degrader *Streptomyces olivaceoviridis*. *Molecular Genetics and Genomics* **271**, 545–553 (2004).
 29. Norris, N., Levine, N. M., Fernandez, V. I. & Stocker, R. Mechanistic model of nutrient uptake explains dichotomy between marine oligotrophic and copiotrophic bacteria. *PLoS Comput Biol* **17**, (2021).
 30. Angle, J. C. *et al.* Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat Commun* **8**, (2017).
 31. Allison, S. D. Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecol Lett* **8**, 626–635 (2005).
 32. Frankena, J., Van Verseveld, H. W. & Stouthamer, A. H. Substrate and energy costs of the production of exocellular enzymes by *Bacillus licheniformis*. *Biotechnol Bioeng* **32**, 803–812 (1988).
 33. Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* vol. 528 60–68 Preprint at <https://doi.org/10.1038/nature16069> (2015).
 34. Guo, H. *et al.* Reconsidering hydrolysis kinetics for anaerobic digestion of waste activated sludge applying cascade reactors with ultra-short residence times. *Water Res* **202**, (2021).
 35. Datta, M. S., Sliwerska, E., Gore, J., Polz, M. F. & Cordero, O. X. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat Commun* **7**, (2016).

The anaerobic (A) and aerobic (B)) GlcNAc degrading, wetland sediment chemostats were started under carbon limitations for 55 days (Phase 1). Steady-state was achieved by the anaerobic and aerobic chemostat on day 14.78 and 38.8, respectively. The chemostats were maintained at SS (Phase 2) for over 29 days (>5 generation times) and were perturbed with O₂/N₂ for 8.9 days on day 67.9 (Phase 3; grey). Afterwards, the reactors underwent a 27.1 day transition (Phase 4) and achieved a new steady-state (Phase 5) on day 103.9. The chemostats were perturbed again for 8.2 days on day 188.8 (Phase 6; grey) and transitioned to a final steady-state (Phase 7). Biomass-specific production rates of formate, acetate, propionate, biomass, CO₂, CH₄, NH₄⁺, NO₂⁻, NO₃⁻, and N₂O production were measured throughout the enrichment by dividing the molar production rate by dry cell weight. Rates are separated into carbon, nitrogen, biomass, and gas production rates with points represent individual production rate measurements. Lines denote the 3-point rolling average production rate with shading representing the standard deviation. Phases and sequencing timepoints (triangles) are marked above each figure. Samples for metagenomic (Phase 1-7) and metatranscriptomic (Phase 1-4) sequencing were collected before (pink), during (purple), and after (orange) each O₂/N₂ perturbation. Additional samples were collected during start-up (yellow).

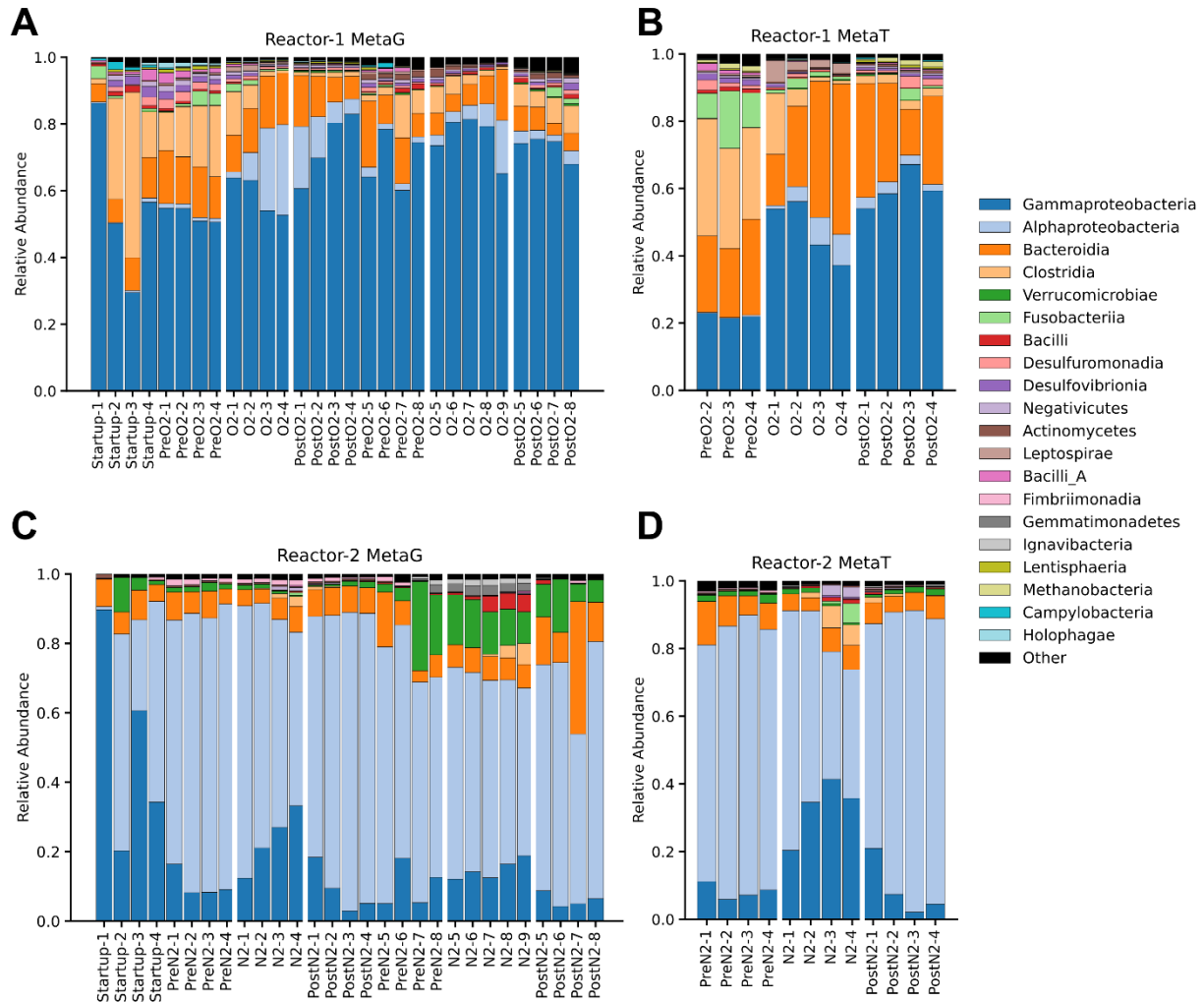


Figure 4-2 – Class-level taxonomic abundance across reactors and operational phase.

Relative abundance of the top 20 most abundant classes in Reactor 1 (Anaerobic) metagenomes, i.e. MetaG (A) and metatranscriptomes, i.e. MetaT (B) and Reactor 2 (Aerobic) metagenomes (C) and metatranscriptomes (D). Operational phase is detailed on the x-axis labels. Perturbation phases (O₂ for Reactor 1 and N₂ for Reactor 2) are highlighted by the vertical white lines.

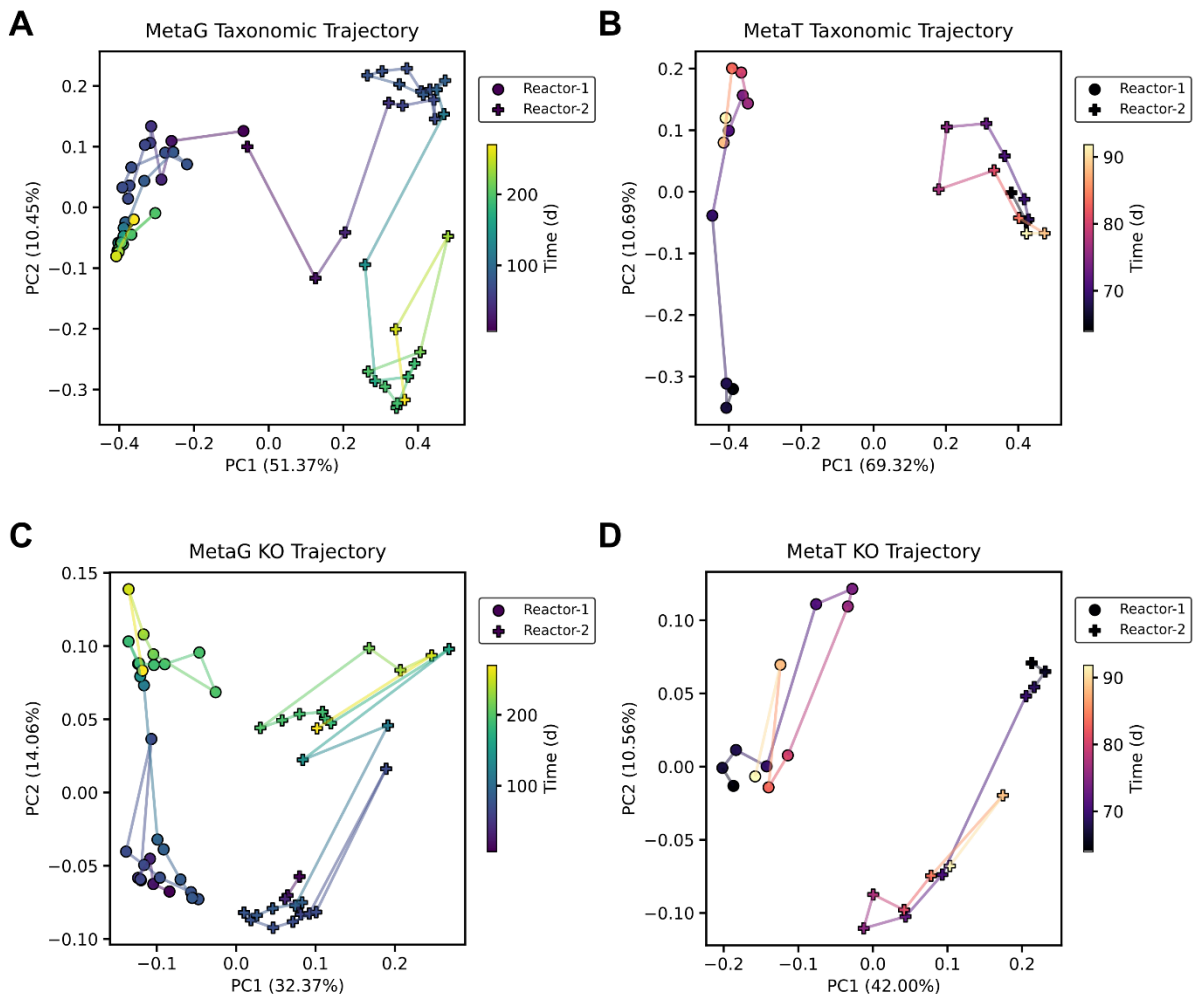


Figure 4-3 – Trajectory of taxonomic and functional composition.

PCoA plots of family-level taxonomic composition of metagenomes, i.e. MetaG (A) and metatranscriptomes, i.e. MetaT (B). PCoA plots of functional composition based on KO abundance for metagenomes (C) and metatranscriptomes (D). Plots are separated by reactor and colored based on sample date with time-adjacent samples connected with lines.

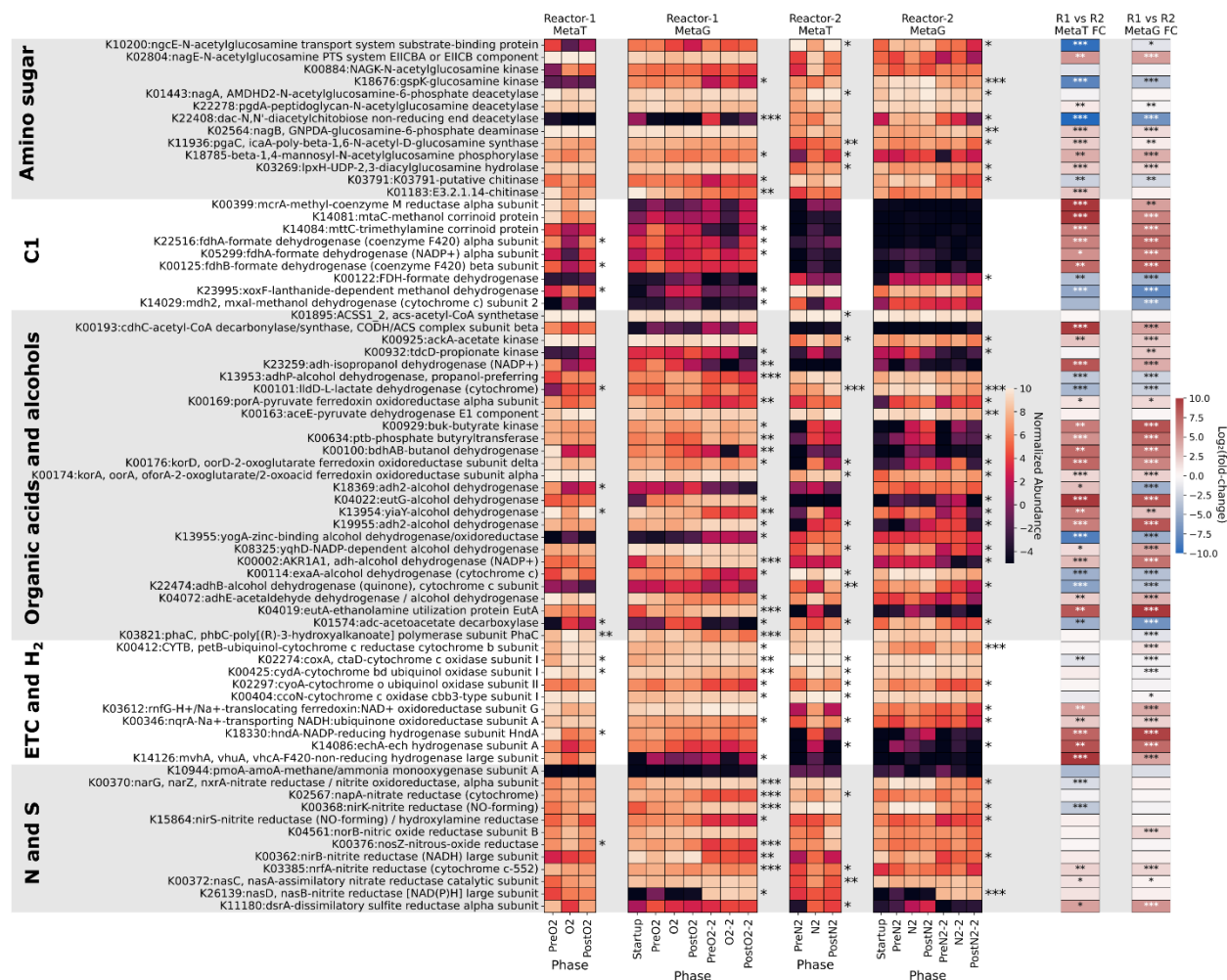


Figure 4-4 – Functional gene variation across reactors and phases.

Mean abundance of key gene orthologs in each operational phase across reactor metagenomes (MetaG) and metatranscriptomes (MetaT). Orthologs are organized by metabolism category as denoted by the gray shading. Asterisks next to heatmaps denote orthologs with abundances that were significantly different across phases in that particular reactor. Mean fold-change between Reactors 1 and 2 is shown in the right two columns, with asterisks denoting significantly different gene orthologs (* : $p < 0.05$, ** : $p < 0.01$, and *** : $p < 0.001$).

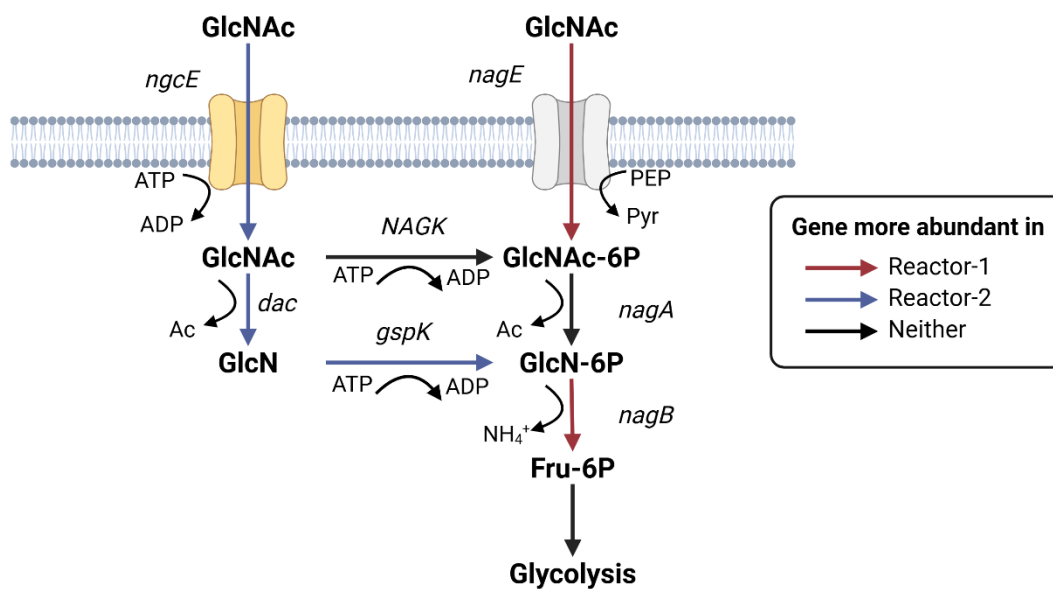


Figure 4-5 – N-acetylglucosamine utilization pathways.

Schematic representation of different N-acetylglucosamine utilization pathways with gene abbreviations shown in italics. Pathway arrows are colored based on observed differences in metagenomic and metatranscriptomic abundance. Abbreviations: N-acetylglucosamine (GlcNAc), N-acetylglucosamine-6-phosphate (GlcNAc-6P), glucosamine (GlcN), glucosamine-6-phosphate (GlcN-6P), fructose-6-phosphate (Fru-6P), acetate (Ac), phosphoenolpyruvate (PEP), pyruvate (Pyr).

Supplementary Material

Table 4-1 – Averaged biomass-specific carbon and nitrogen production rates of the anaerobic (R1) and aerobic (R2) chemostats.

Biomass-specific production rates were measured over the total duration of each phase (1: start-up; 2: steady-state; 3: perturbation; 4: transition; 5: steady-state; 6: perturbation; 7: steady-state). These rates were calculated by dividing the production rate by the measured cell dry weights. The perturbations have been highlighted in grey.

Compound	Reactor	Phase							Unit
		1	2	3	4	5	6	7	
NH ₄ ⁺	R1	6.65 ±0.58	7.75 ±0.82	4.88 ± 1.00	6.44 ±1.10	6.74 ±0.70	4.40 ±0.90	5.75 ±0.62	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	2.49 ±2.43	0.0 ±0.0	1.59 ±1.27	1.50 ±1.18	0.0 ±0.016	1.87 ±1.51	0.17 ±0.66	
NO ₂ ⁻	R1	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	1.83 ±1.85	3.61 ±0.60	0.39 ±1.10	0.85 ±0.72	2.02 ±1.32	0.24 ±0.50	0.50 ± 0.40	
NO ₃ ⁻	R1	0.0 ±0.0	0.0 ±0.0	0.0 ± 0.002	0.0 ±0.001	0.0 ±0.002	0.0 ±0.001	0.0 ±0.001	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	0.13 ±0.22	0.34 ±0.22	0.0 ±0.0	0.34 ±0.36	2.96 ±0.92	0.87 ±1.66	4.22 ±1.63	
N ₂ O	R1	0.0 ±0.001	0.001 ±0.0	0.01 ±0.01	0.0 ±0.002	0.001 ±0.0	0.0 ±0.003	0.002	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	0.09 ±0.09	0.02 ±0.01	0.02 ±0.04	0.11 ±0.01	0.03 ±0.05	1.58 ±3.10	0.04 ±0.06	
Biomass-N	R1	2.57 ±0.51	2.39 ±0.37	4.93 ±1.18	3.75 ±1.33	2.62 ±0.45	4.38 ±1.10	3.06 ±0.84	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	4.13 ±1.15	5.06 ±1.42	5.03 ±0.90	5.08 ±1.19	4.40 ±1.23	4.67 ±1.20	3.70 ±1.03	
GlcNAc-N	R1	-10.4 ±0.73	-10.8 ±0.36	-9.71 ±1.86	-9.57 ±1.87	-8.61 ±1.83	-8.96 ±0.60	-10.0 ±0.60	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	-9.18 ±0.82	-8.27 ±0.56	-9.22 ±2.59	-8.13 ±1.90	-7.06 ±2.23	-8.81 ±0.62	-9.09 ±0.49	
N-balance	R1	-1.05 ±3.54	-2.24 ±1.33	1.35 ±0.79	0.83 ±2.63	-0.48 ±2.80	0.25 ±0.40	-2.155	μmol·mgDCW ⁻¹ ·h ⁻¹
	R2	0.23 ±3.85	1.85 ±0.88	-2.17 ±2.42	-0.14 ±2.30	2.62 ±2.64	2.19 ±6.24	-0.57 ±1.49	
Formate	R1	1.17 ±2.99	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	0.0 ±0.0	0.0 ±0.0	0.13 ±0.27	0.0 ±0.0	0.0 ±0.0	1.51 ±1.71	0.0 ±0.0	
Acetate	R1	43.78 ±6.52	50.10 ±1.70	13.9 ±18.12	23.31 ±13.6	26.15 ±7.34	10.16 ±11.4	6.47 ±7.08	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	0.15 ±0.65	0.0 ±0.0	7.56 ±7.43	3.29 ±8.66	0.0 ±0.0	10.14 ±9.6	1.01 ±4.10	
Propionate	R1	5.36 ±2.72	10.77 ±1.75	1.73 ±3.70	10.51 ±6.44	11.63 ±3.15	1.92 ± 3.58	9.38 ±3.21	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	0.0 ± 0.0	0.0 ±0.0	1.47 ±1.73	0.51 ±1.37	0.0 ±0.0	0.63 ±0.90	0.05 ± 0.27	
CO ₂	R1	24.78 ±4.25	19.13 ±1.03	60.94 ±19.1	23.33 ±12.8	24.79 ±6.49	34.96 ±11.2	39.45 ±6.59	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	43.00 ±17.9	40.92 ±4.81	27.15 ±7.01	44.15 ±12.6	39.32 ±5.76	30.79 ±9.28	41.58 ±8.99	
CH ₄	R1	1.37 ±1.21	3.54 ±0.71	0.21 ±0.20	2.95 ±1.87	10.95 ±5.83	1.15 ±1.95	27.02 ±12.2	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	0.02 ±0.02	0.003 ±0.0	0.042 ±0.04	0.01 ±0.03	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	
Biomass-C	R1	12.84 ±2.53	11.93 ±1.87	24.67 ±5.91	18.73 ±6.65	13.08 ±2.22	21.89 ±5.46	15.3 ± 4.20	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	20.65 ±5.74	25.28 ±7.08	25.15 ±4.46	25.42 ±5.92	22.02 ±6.16	23.34 ±6.02	18.49 ±5.13	
GlcNAc-C	R1	-83.23 ±5.8	-86.56 ±2.9	-77.7 ±14.9	-76.6 ±14.9	-68.9 ±14.6	-71.66 ±4.7	-80.21 ±4.8	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	-73.41 ±6.6	-66.12 ±4.5	-73.8 ±20.8	-65.0 ±15.2	-56.5 ±17.8	-70.49 ±4.9	-72.74 ±3.9	
C-balance	R1	-14.0 ±16.3	-10.2 ±12.8	23.80 ±24.1	-3.72 ±15.0	-6.29 ±20.4	-6.86 ±21.2	-21.4 ±34.4	μmol·C·mgDCW ⁻¹ ·h ⁻¹
	R2	-36.0 ±36.3	-27.2 ±27.8	-12.2 ±18.3	2.47 ±25.97	-16.2 ±29.6	-6.90 ±13.6	-34.7 ±25.6	

Table 4-2 – Modified lacustrine media composition

Compound	Concentration (mM)
GlcNAc	5
(NH ₄) ₂ SO ₄	0.05
NaH ₂ PO ₄	0.04
Na ₂ HPO ₄	0.06
KHCO ₃	0.3
CaCl ₂ * H ₂ O	0.244
MgCl ₂ * 6 H ₂ O	0.16
SL-10	1x
Se-W	1x
7-Vit	1x
HEPES	10

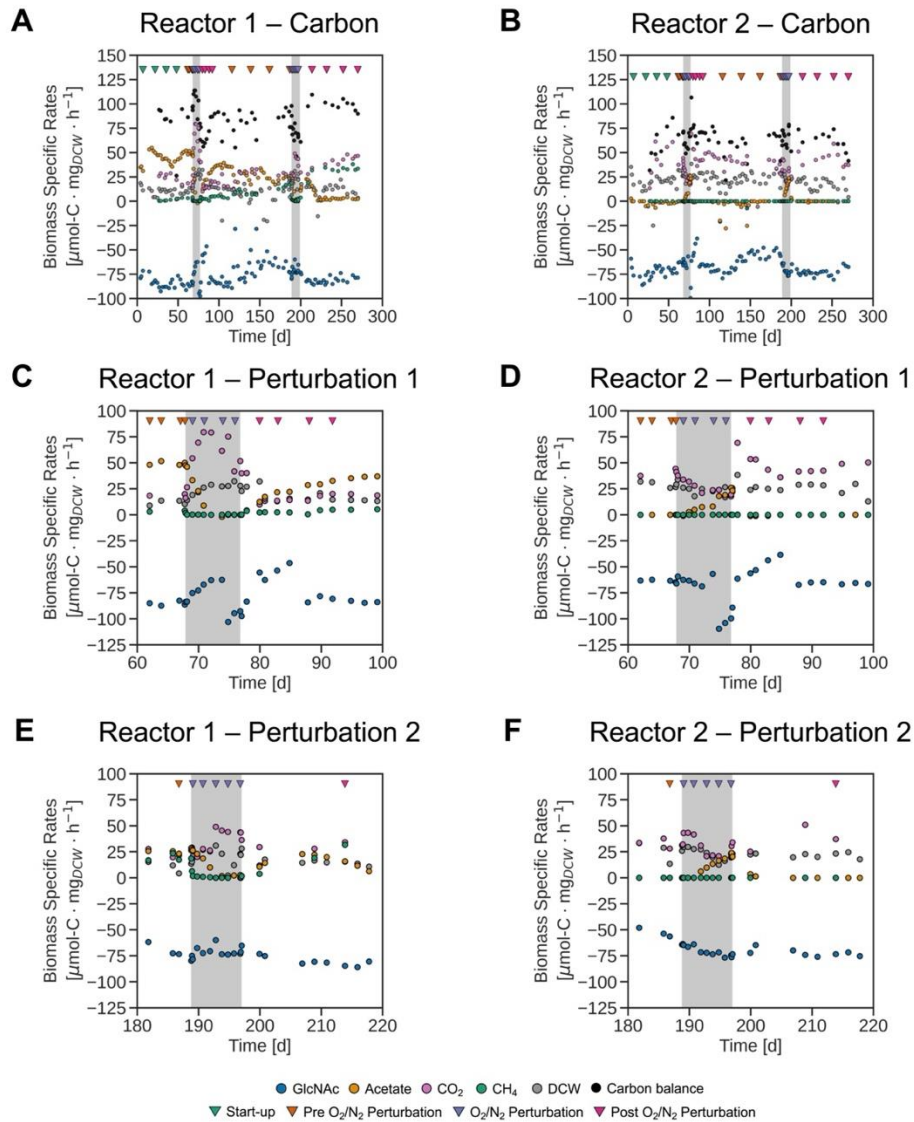


Figure S 4-1 – Biomass-specific carbon production rates.

Biomass-specific carbon production rates ($\mu\text{mol-C} \cdot \text{mg}_{\text{DCW}}^{-1} \cdot \text{h}^{-1}$) were measured over the entire operational period for the anaerobic (A) and aerobic (B) GlcNAc degrading, wetland sediment chemostats. Steady-state was achieved by the anaerobic and aerobic chemostat on day 14.78 and

38.8, respectively. The chemostats were maintained at steady-state for over 29 days (>5 generation times) and were perturbed with O₂/N₂ for 8.9 days on day 67.9 (C-D; grey;). Afterwards, the reactors underwent a 27.1 day transition and achieved a new steady-state on day 103.9. The chemostats were perturbed again for 8.2 days on day 188.8 (E-F; grey) and transitioned to a final steady-state. Biomass-specific production rates were measured throughout the enrichment by dividing the molar production rate by dry cell weight. Samples for metagenomic and metatranscriptomic sequencing (triangles) were collected before (pink), during (purple), and after (orange) each O₂/N₂ perturbation. Additional samples were collected during start-up (yellow).

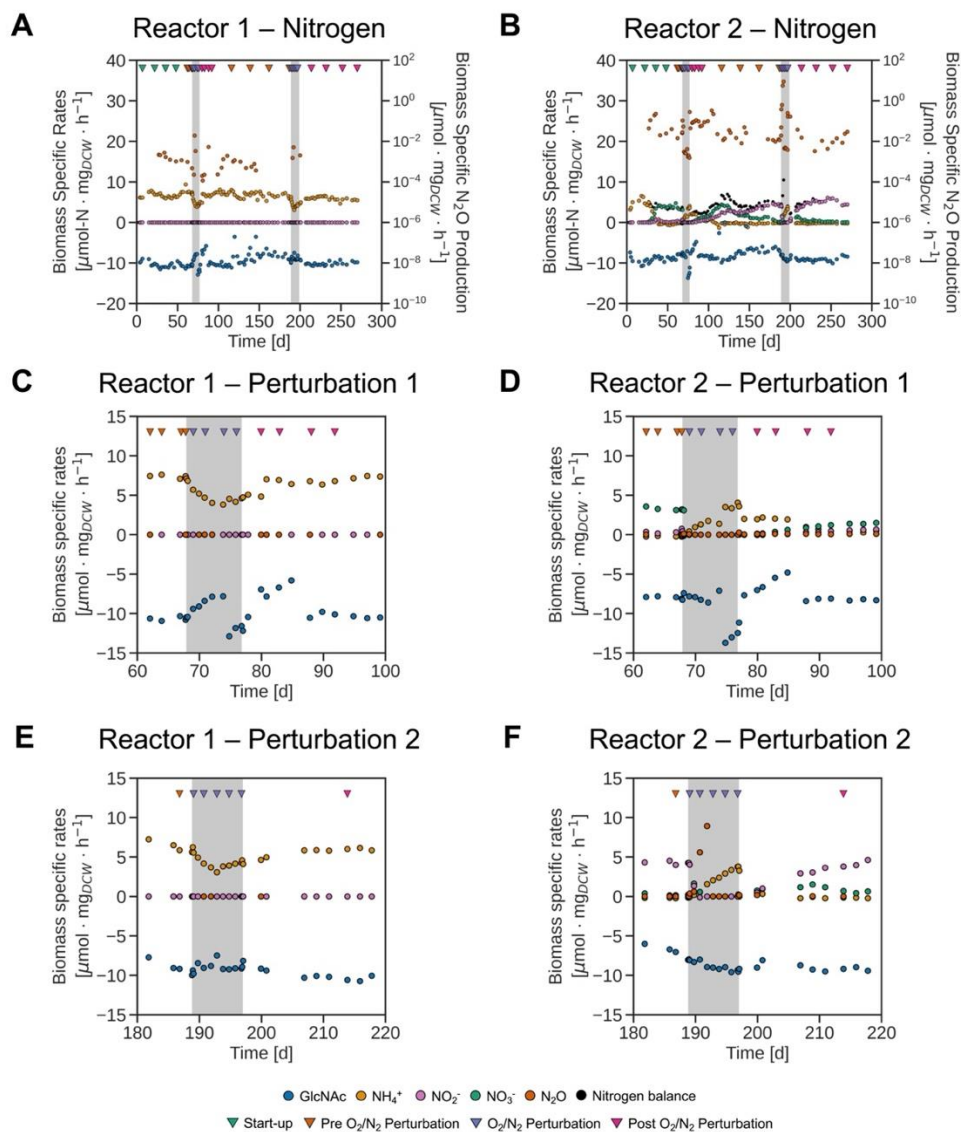


Figure S 4-2 – Biomass-specific nitrogen production rates.

Biomass-specific nitrogen production rates ($\mu\text{mol} \cdot \text{mg}_{\text{DCW}}^{-1} \cdot \text{h}^{-1}$) were measured over the entire operational period for the anaerobic (A) and aerobic (B) GlcNAc degrading, wetland sediment chemostats. Steady-state was achieved by the anaerobic and aerobic chemostat on day 14.78 and 38.8, respectively. The chemostats were maintained at steady-state for over 29 days (>5 generation times) and were perturbed with O_2/N_2 for 8.9 days on day 67.9 (C-D; grey;).

Afterwards, the reactors underwent a 27.1 day transition and achieved a new steady-state on day 103.9. The chemostats were perturbed again for 8.2 days on day 188.8 (E-F; grey) and transitioned to a final steady-state. Biomass-specific production rates were measured throughout the enrichment by dividing the molar production rate by dry cell weight. Samples for metagenomic and metatranscriptomic sequencing (triangles) were collected before (pink), during (purple), and after (orange) each O₂/N₂ perturbation. Additional samples were collected during start-up (yellow).

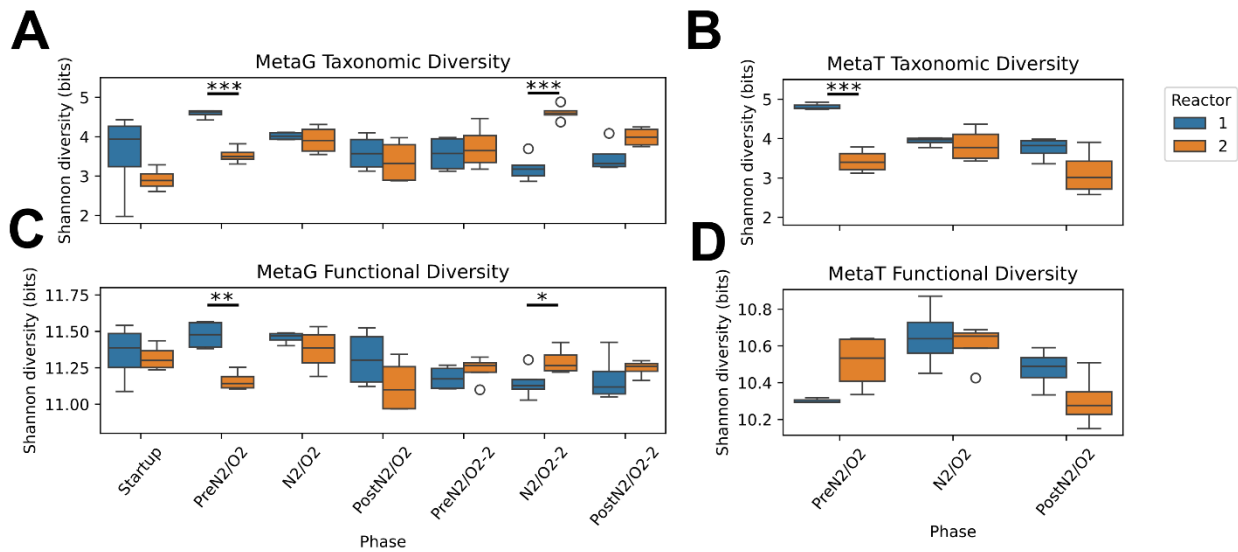


Figure S 4-3 – Taxonomic and functional diversity.

Shannon diversity of reactor metagenomes (A, C) and metatranscriptomes (B, D) based on family-level taxonomic composition (A, B) and Kegg ortholog abundance (C, D). Statistical comparisons were made between reactors in each phase with significant differences noted with asterisks (* : $p < 0.05$, ** : $p < 0.01$, and *** : $p < 0.001$).

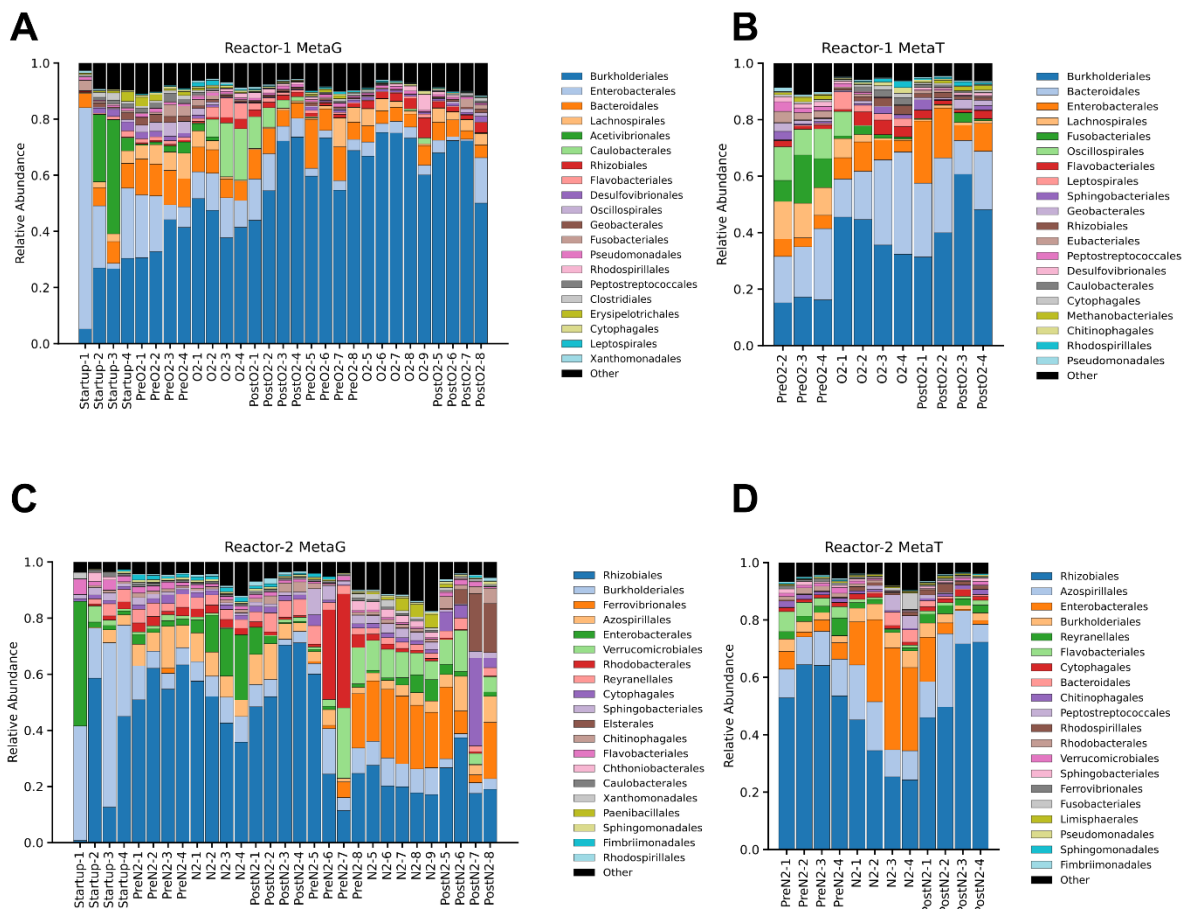


Figure S 4-4 – Order-level taxonomic composition across reactors and operational phases.

Relative abundance of the top 20 most abundant orders in Reactor 1 metagenomes, i.e. MetaG (A) and metatranscriptomes, i.e. MetaT (B) and Reactor 2 metagenomes (C) and metatranscriptomes (D). Operational phase is detailed on the x-axis labels.

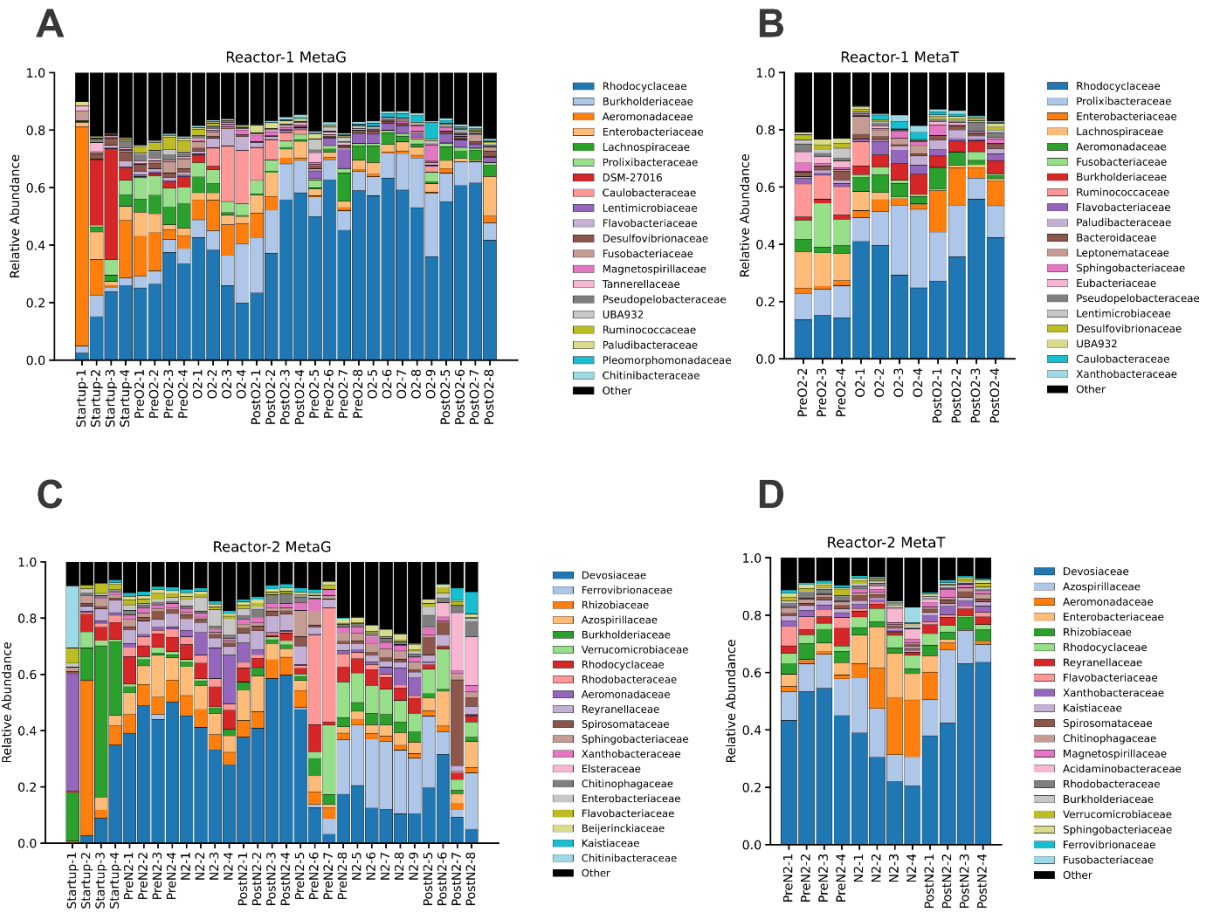


Figure S 4-5 – Family-level taxonomic composition across reactors and operational phases.

Relative abundance of the top 20 most abundant families in Reactor 1 metagenomes, i.e. MetaG

(A) and metatranscriptomes, i.e. MetaT (B) and Reactor 2 metagenomes (C) and

metatranscriptomes (D). Operational phase is detailed on the x-axis labels.

Chapter 5 – Other scientific contributions

This chapter highlights additional scientific contributions that complement the broader themes of the dissertation but fall outside the central focus of how oxygen gradients impact microbial nutrient cycling. These contributions include enrichment studies from wetland soils, a genome-centric metagenomic analysis of anaerobic wastewater granule microbial communities, and the isolation and characterization of a nitrite-oxidizing bacterium from a nitrate-contaminated subsurface.

The enrichment studies use culture-based approaches to investigate microbial guilds involved in specific carbon and nitrogen cycling processes. In particular, they examine polyhydroxyalkanoate (PHA) accumulation in methanotrophs under varying methane and ammonium concentrations, and utilize bioreactors to identify lactic acid-driven chain-elongating microorganisms. This work complements Chapter 4, which also uses enrichment strategies to explore N-acetylglucosamine utilization by wetland microbial communities.

“Identifying microbial functional guilds performing cryptic organotrophic and lithotrophic redox cycles in anaerobic granular biofilms” builds on the findings from Chapter 2 by analyzing and clustering the functional gene content of individual metagenome-assembled genomes (MAGs) recovered from anaerobic granular biofilms. While Chapter 2 focused on the functional potential of the microbial community as a whole, this work zooms in on the individual genomes that make up the community, offering new insights into how genome-resolved metagenomics can enhance our understanding of anaerobic nutrient cycling.

In *“Comparative Physiological and Genomic Characterization of a Nitrobacter vulgaris Strain from a Nitrate-Contaminated Subsurface,”* a novel strain of nitrite-oxidizing bacteria is

characterized to investigate how its physiological traits compare to existing members of the *Nitrobacter* genus. This analysis explores how environmental context—specifically, high nitrate and nitrite concentrations—influences microbial tolerance and function. The genome of this strain was sequenced and analyzed alongside other *Nitrobacter* genomes, revealing, notably, the presence of genes for nitrous oxide reduction. The ecological and biogeochemical implications of this unexpected finding are discussed.

Abstracts and key figures are presented below.

Comparative physiological and genomic characterization of a novel *Nitrobacter vulgaris* strain from a nitrate-contaminated subsurface

Zachary Flinkstrom¹, Kristopher A. Hunt¹, Britt Abrahamson¹, Pierce Harvell², Xiangpeng Li², David A. Stahl¹, Wei Qin², Mari-Karoliina H. Winkler¹

¹ Department of Civil and Environmental Engineering, University of Washington, Seattle, WA

² School of Biological Sciences, Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA

Prepared for submission to *Applied and Environmental Microbiology*

Abstract

Nitrite-oxidizing bacteria (NOB) represent a crucial node in global nitrogen cycle. By catalyzing the second step of nitrification—the oxidation of nitrite to nitrate to generate energy for growth—NOB activity controls the fate of nitrite (NO₂⁻) in aerobic environments. Despite thriving in diverse environments including soils, freshwater, marine ecosystems, subsurface habitats, and water treatment systems, organisms capable of nitrite oxidation are predominantly confined to *Nitrobacter*, *Nitrospina*, *Nitrobacter*, *Nitrotoga*, and other specific lineages. The

genus *Nitrobacter*, recognized for its facultative heterotrophic metabolism, is often associated with high-nitrogen environments. Here, we report the isolation and physiological characterization of a novel strain, *Nitrobacter vulgaris* MLSD-S22, from a nitrate- and heavy-metal-contaminated subsurface. Growth inhibition experiments revealed that MLSD-S22 and the type *N. vulgaris* strain Z exhibited similar sensitivities to nitrite, nitrate, and chloride, with nitrite being the most inhibitory. Microrespirometry demonstrated that the *N. vulgaris* strains and *Nitrobacter winogradskyi* Nb-255 possessed higher affinities for nitrite and oxygen than previously reported for *Nitrobacter* NOB, suggesting some potential to compete in low-substrate environments. Long-read DNA sequencing provided a complete genome for MLSD-S22, revealing two plasmids and a fully intact nitrous oxide reduction operon—an unexpected feature for *Nitrobacter*. This discovery points towards an unrecognized mechanism of nitrogen loss in this genus and raises questions about the contribution of *Nitrobacter* NOB to denitrification. These findings broaden the physiological and genomic diversity of *Nitrobacter*, offering new insights into their adaptation strategies and potential contributions to nitrogen loss.

Importance

Nitrobacter species play a key role in nitrogen cycling, yet their physiological adaptations and genomic diversity remain underexplored. This study characterizes *Nitrobacter vulgaris* MLSD-S22, isolated from a nitrate- and heavy-metal-contaminated subsurface, revealing unexpected genomic and metabolic traits. Higher nitrite and oxygen affinities compared to previous reports indicate some *Nitrobacter* are more competitive in low-substrate environments than previously believed. Notably, the presence of a complete nitrous oxide reduction operon suggests a potential contribution to denitrification. These findings challenge assumptions about *Nitrobacter* metabolism and have implications for modeling nitrification and

nitrogen loss in natural and engineered systems. By expanding *Nitrobacter* ecophysiology, this work provides a foundation for future research on the metabolic flexibility of *Nitrobacter* and their contributions to nitrogen transformations in diverse environments.

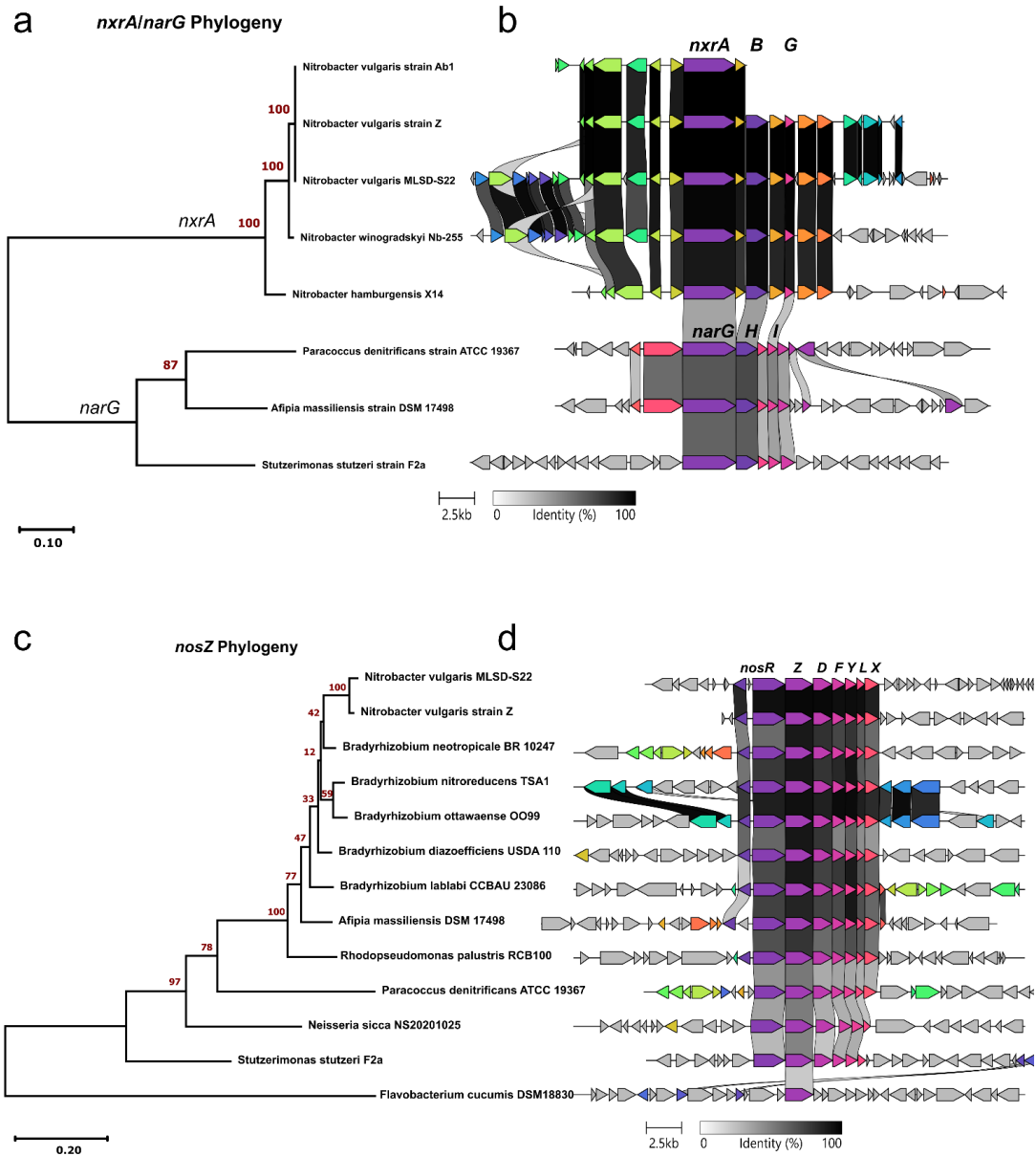


Figure 5-1 – Phylogeny and operon structure of nitrite oxidation and nitrous oxide reduction in *Nitrobacter vulgaris* MLSD-S22.

Maximum likelihood phylogenetic tree of *nxrA* and *narG* from *Nitrobacter* and related denitrifiers (a). Operon structure of *nxrA/narG* used in tree construction (b). Maximum likelihood phylogenetic tree of *nosZ* from *Nitrobacter vulgaris* and related denitrifiers (c). *Nos*

operon structure for genes used in the tree (d). Colors and shaded links in (b) and (c) represent pairwise amino acid similarity.

Resource availability governs polyhydroxyalkanoate (PHA) accumulation and diversity of methanotrophic enrichments from wetlands

Zachary Flinkstrom^{a†}, Yujin Kim^{b†}, Pieter Candry^{a*}, Mari-Karoliina H Winkler^a, Jaewook Myung^{b*}

^a Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA

^b Department of Civil and Environmental Engineering, KAIST, Daejeon, Republic of Korea

† These authors contributed equally to this work and share the first authorship

Peer reviewed and published in *Front Bioeng Biotechnol* **11**, (2023)

<https://doi.org/10.3389/fbioe.2023.1210392>

Abstract

Aquatic environments account for half of global CH₄ emissions, with freshwater wetlands being the most significant contributors. These CH₄ fluxes can be partially offset by aerobic CH₄ oxidation driven by methanotrophs. Additionally, some methanotrophs can convert CH₄ into polyhydroxyalkanoate (PHA), an energy storage molecule as well as a promising bioplastic polymer. In this study, we investigate how PHA-accumulating methanotrophic communities enriched from wetlands were shaped by varying resource availability (i.e., C and N concentrations) at a fixed C/N ratio. Cell yields, PHA accumulation, and community composition were evaluated in high (20% CH₄ and 10 mM NH₄⁺) and low resource (0.2% CH₄ and 0.1 mM NH₄⁺) conditions simulating engineered and environmental settings, respectively. High resource availability decreased C-based cell yields, while N-based cell yields remained

stable, suggesting nutrient exchange patterns differed between methanotrophic communities at different resource concentrations. PHA accumulation was only observed in high resource enrichments, producing approximately $12.6 \pm 2.4\%$ (m/m) PHA, while PHA in low resource enrichments remained below detection. High resource enrichments were dominated by *Methylocystis* methanotrophs, while low resource enrichments remained significantly more diverse and contained only a minor population of methanotrophs. This study demonstrates that resource concentration shapes PHA-accumulating methanotrophic communities even at fixed C/N ratios. Together, this provides useful information to leverage such communities in engineering settings as well as to begin understanding their role in the environment.

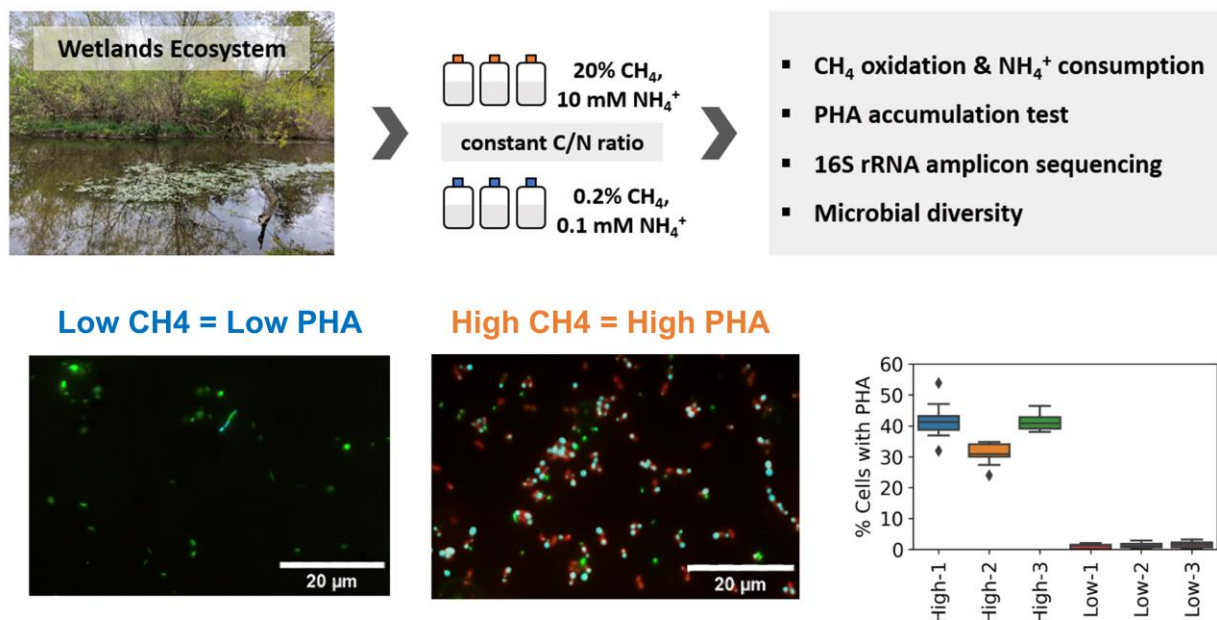


Figure 5-2 – Resource availability governs polyhydroxyalkanoate (PHA) accumulation and diversity of methanotrophic enrichments from wetlands.

Graphical abstract detailing study design and key results. Methanotrophic enrichments were established from wetland soil under high and low methane and ammonium concentrations. PHA accumulation was only observed in the high resource condition.

Wetlands harbor lactic acid-driven chain elongators

Pieter Candry^{1,#,*}, Zachary Flinkstrom¹, Mari-Karoliina Henriikka Winkler¹

¹ Civil and Environmental Engineering, University of Washington, 201 More Hall, Box 352700, Seattle, WA 98195-2700, USA

Correspondence to: Pieter Candry, Civil and Environmental Engineering, University of Washington, 201 More Hall, Box 352700, Seattle, WA 98195-2700, USA; phone: +1-206-556-8282; E-mail: pcandry@uw.edu

* Current Address: Laboratory of Systems and Synthetic Biology, Wageningen University & Research. 6708 WE, Wageningen, The Netherlands; E-mail: pieter.candry@wur.nl

Peer reviewed and published in *Microbiol Spectr* 12, (2024)

<https://doi.org/10.1128/spectrum.02105-23>

Abstract

Wetlands are globally significant carbon storage hotspots. Recent research has suggested that microbially derived metabolites may contribute to soil organic matter formation. Identifying pathways driving formation of such metabolites is critical to understand the global impact of wetland carbon cycling. Here, we evaluate the presence of chain elongating organisms converting 2-3 carbon compounds (i.e., lactic and acetic acid) to medium-chain carboxylic acids (i.e., six-carbon caproic acid) in wetland soils. We demonstrate the enrichment of a lactic acid driven chain elongating community from wetland soils producing a mixture of butyric and caproic acid. The enriched community was dominated by Clostridiaceae, Ruminococcaceae, and Lachnospiraceae, three families with known chain elongators. Amplicon sequencing identified 3 Ruminococcaceae and 1 Clostridiaceae zero-radius OTU (zOTU) that were (i) present in the soil, (ii) enriched over 1% relative abundance in the bioreactor, and (iii) were closely related to known chain elongators. Moreover, close relatives of the 3 Ruminococcaceae zOTU were also observed in several other wetland microbiomes. From this observation, we conclude that close relatives of known chain elongators, potentially capable of lactic acid driven MCCA production themselves, are present in wetland soils. This observation may have implications for our understanding of carbon cycling and storage in wetland ecosystems.

Importance

Wetlands are globally significant carbon cycling hotspots that both sequester large amounts of CO₂ as soil carbon as well as emit a third of all CH₄ globally. Their outsized role in

the global carbon cycle makes it critical to understand microbial processes contributing to carbon breakdown and storage in these ecosystems. Here, we confirm the presence of chain-elongating organisms in freshwater wetland soils. These organisms take small carbon compounds formed during the breakdown of biomass and turn them into larger compounds (six to eight carbon organic acids) that may potentially contribute to the formation of soil organic matter and long-term carbon storage. Moreover, we find that these chain-elongating organisms may be widely distributed in wetlands globally. Future work should identify these organisms' contribution to carbon cycling in wetlands and the potential role of the products they form in carbon sequestration in wetlands.

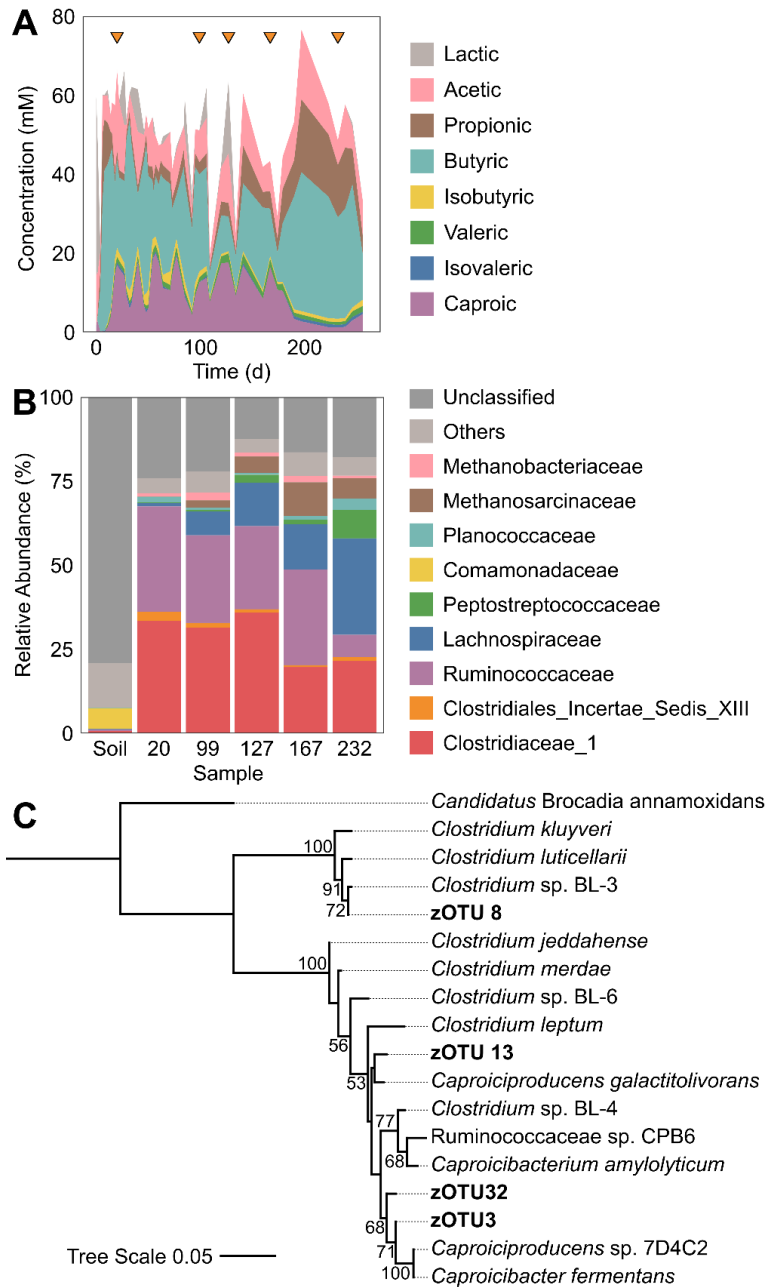


Figure 5-3 – Enrichment of a lactic acid-driven chain elongation community from wetland soil.

CSTR effluent product profiles (A) with orange triangles indicating days sampled for community characterization. Relative abundance of community members classified at the family-level (B).

Phylogenetic tree of the V4-V5 region of the 16S rRNA gene for zOTU of interest along with

relevant relatives (C). zOTU of interest (i) were detected in the wetland soil inoculum, (ii) had a relative abundance of at least 1% at any time during reactor operation, and (iii) are related to known chain elongators. Bootstrap values greater than or equal to 50% are shown as percentages at each node. Scale indicates substitutions per nucleotide position.

Identifying microbial functional guilds performing cryptic organotrophic and lithotrophic redox cycles in anaerobic granular biofilms

Zachary Flinkstrom^{1*†}, Samuel J. Bryson^{1,2†}, Bojan Pelivano¹, Pieter Candry^{1,3}, Kristopher Hunt¹, Mari-Karoliina H. Winkler¹

¹ Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, 98101, USA

² Phase Genomics, Seattle, WA, 98109, USA

³ Laboratory of Systems and Synthetic Biology, Wageningen University & Research. 6708 WE, Wageningen, The Netherlands

† These authors contributed equally to this work

*Correspondence to: Zachary Flinkstrom, Civil and Environmental Engineering, University of Washington, 201 More Hall, Box 352700, Seattle, WA 98195-2700, USA; E-mail: zflink@uw.edu

In review at *PLOS One*

Abstract

Granular biofilms used in anaerobic digester systems contain diverse microbial populations that interact to hydrolyze organic matter and produce methane within controlled environments. Prior research investigated the feasibility of utilizing granular biofilms obtained from an anaerobic digester to remove nitrate without the addition of exogenous electron donors. These granules possessed a unique structure of alternating light and dark iron sulfide and pyrite rich layers that potentially served as both an electron source and sink, linking carbon, nitrogen, sulfur,

and iron cycles. To characterize the functional roles of diverse microbial populations enriched within these layered biofilms, we analyzed metagenomes obtained from three different granules. Comparisons between the functional gene content of forty metagenome assembled genomes (MAGs) identified phylogenetically cohesive functional guilds. Each of these functional MAG clusters was assigned to specific steps in anaerobic digestion (hydrolysis, acidogenesis, acetogenesis, and methanogenesis) and anaerobic respiration (denitrification and sulfate reduction). Comparisons with metagenomes derived from a variety of natural and engineered ecosystems confirmed that the enriched denitrifying bacteria were similar to populations typically found in wetlands and biological nitrogen removal systems. Analysis of read alignments to individual genes within the forty MAGs identified conserved genomic features that were representative of the functions that distinguished functional guilds. Overall, this research illustrates the utility of functional based classification of microorganisms for characterizing ecosystem functions and highlights the potential application of engineered ecosystems to serve as experimental models for complex natural ecosystems.

Importance

Microbial communities in natural and engineered systems play a crucial role in breaking down organic matter through various pathways, depending on the availability of oxygen and other compounds that can accept electrons, such as nitrate, sulfate, and iron. Under anaerobic conditions, these processes are carried out by specialized groups of microorganisms responsible for hydrolysis, acid production (acidogenesis), acetate production (acetogenesis), and methane formation (methanogenesis). When alternative electron acceptors like nitrate are present, some microorganisms can also perform anaerobic respiration. In this study, we used metagenomic sequencing to recover and analyze genomes from anaerobic digester granules treating nitrate-

contaminated water. By clustering microbial genomes based on their gene content, we identified distinct functional groups and their roles in the breakdown of organic matter and nutrient cycling. We also compared these genomes to other metagenomic datasets from wetlands, groundwater, and anaerobic digesters, to assess how well the findings apply across different ecosystems and identify functions conserved across environments. This research demonstrates the value of using functional clustering of microbial genomes to understand ecosystem processes. It also highlights how engineered microbial communities can serve as simplified models for studying the complex interactions found in natural systems.

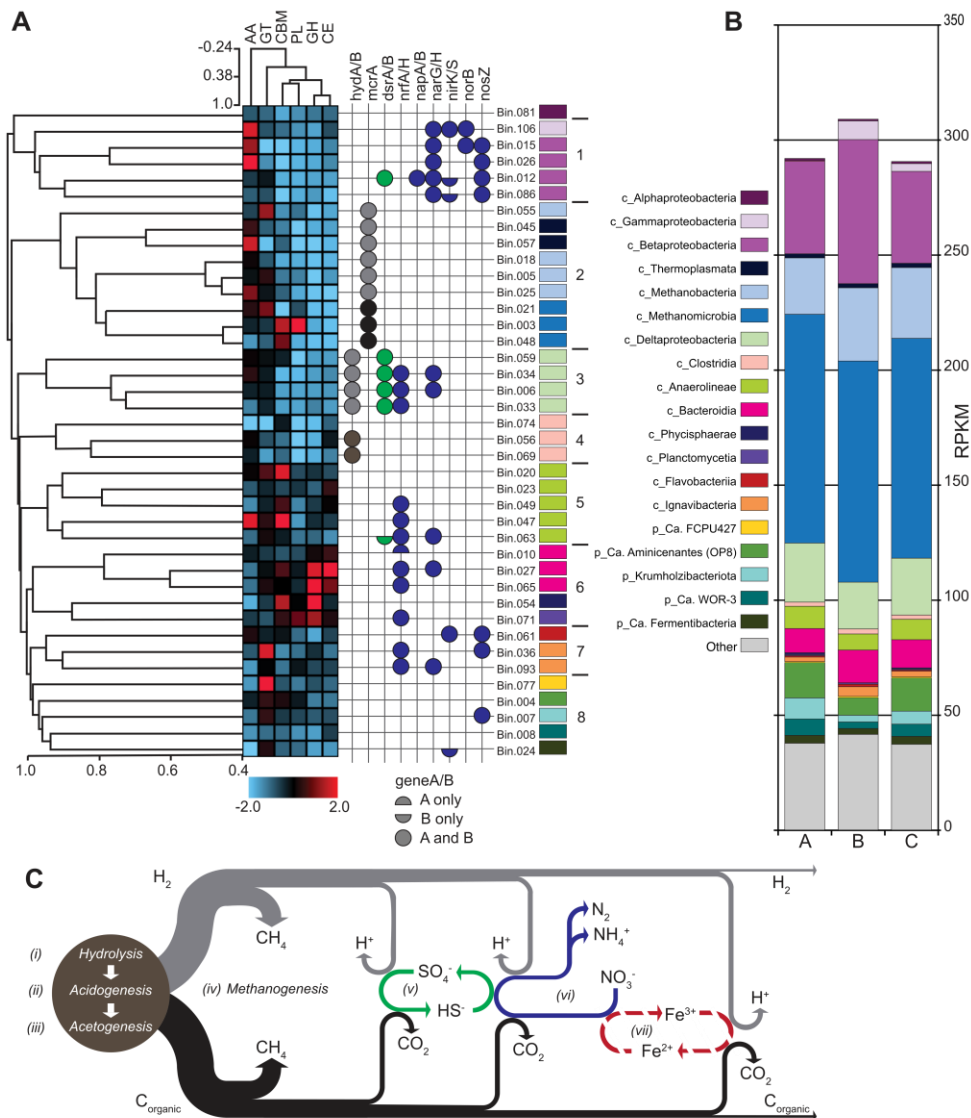


Figure 5-4 – Functional and taxonomic characterization of granular biofilm MAGs.

(A left to right) The dendrogram presents clustering based on Pearson correlation of normalized KEGG ortholog (KO) annotation counts for 40 MAGs followed by genomic characteristics and taxonomic assignments. The heatmap depicts CAZyme counts (log transformed, normalized, and scaled to a range of -2 to 2) for each category: AA-auxiliary activities, GT-glycosyl transferase, CBM-carbohydrate binding motif, PL-polysaccharide lyase, GH-glycoside hydrolase, and CE-

carbohydrate esterase. The grid indicates the presence of specific functional genes, hydrogenase (*hydAB*), methyl coenzyme M reductase (*mcrA*), dissimilatory sulfite reductase (*dsrAB*), formate dependent nitrite reductase (*nrfAH*), periplasmic nitrate reductase complex (*napAB*), dissimilatory nitrate reductase (*narGH*), nitrite reductases (*nirK* and *nirS*), nitric oxide reductase (*norB*), and nitrous oxide reductase (*nosZ*). The color of each gene column corresponds to the redox processes (i-vii) in panel C. The taxonomic assignment for each MAG is designated by the same colors as in panel B. (B) Stacked bar-chart representing the microbial community structure based on RPKM values assessed for each recovered bin for each of the three Illumina sequence libraries obtained for each granule, columns A, B, and C. Each bar includes the summed RPKM values for all bins for each assigned taxonomy. “Other” includes bins assigned to low abundance taxa, taxa without a representative high-quality MAG, and all un-binned contigs. (C) Diagram of potential biologically mediated redox processes (i-vii) occurring within the granules.

Chapter 6 – Conclusions and perspective

Conclusions

The overarching goal of this dissertation was to investigate how oxygen gradients influence the composition and function of wetland microbial communities. To achieve this, we combined field observations, laboratory bioreactor experiments, and metagenomic analyses to connect microbial-scale processes with ecosystem-scale patterns of nutrient cycling. A particular focus was placed on understanding how oxygen availability regulates greenhouse gas emissions and the stability of soil organic carbon in wetland environments. The central findings of this research are summarized below.

Chapter 2 assessed the utility of metagenomic functional gene profiles for distinguishing microbial communities across diverse environmental contexts. This foundational step was critical for evaluating whether functional profiles could serve as reliable indicators of microbiome function and distinguish microbiomes under different environmental conditions—an approach used throughout subsequent chapters. An analysis of metagenomic datasets from the JGI IMG/M database revealed that functional profiles are indeed powerful discriminators. Unsupervised clustering identified three broad functional clusters—aerobic terrestrial, photic aquatic, and anaerobic—with oxygen and sunlight availability emerging as dominant structuring forces. The anaerobic cluster encompassed a wide range of environments, including digesters, sediments, and animal gut samples, highlighting the strong deterministic role of anaerobiosis in shaping microbial communities.

These findings informed our interpretation of microbial responses to oxygen availability in subsequent chapters. In particular, we hypothesized that under drought conditions, increased oxygen intrusion into wetland soils would be a key driver of functional change in microbial

communities. This justified the use of functional gene profiles as reliable indicators of community shifts in both the field and bioreactor studies.

Chapter 3 presents a seasonal, depth-resolved biogeochemical and multi-omics characterization of a freshwater wetland. This study integrated porewater chemistry, greenhouse gas fluxes, and metagenomic and metabolomic profiling over four seasons and three soil depths. The key finding was that soil depth, rather than seasonality, was the strongest driver of microbial community composition and function. The surface soil was enriched with metabolically versatile Gammaproteobacteria capable of aerobic respiration, nitrate reduction, and sulfur and methane oxidation. In contrast, deeper soil layers hosted microbial taxa—such as Anaerolinea and Methanomicrobia—engaged in strictly anaerobic processes like methanogenesis, acetogenesis, and fermentation.

Depth-resolved metabolomics revealed corresponding trends in metabolic signatures, with purine nucleotides and amino acids enriched near the surface, and amino sugars and phenolic compounds accumulating at depth. Despite clear seasonal fluctuations in porewater chemistry and gas fluxes, microbial community composition remained remarkably stable across time within each depth. This suggests strong spatial structuring of communities by persistent redox gradients, with surface communities composed of metabolically flexible generalists and deeper communities composed of highly efficient specialists adapted to thermodynamically constrained conditions. This led to a central question: How much disturbance—such as oxygen intrusion—and for how long is required to shift these communities?

Chapter 4 directly investigates that question using a chemostat bioreactor enrichment approach with a model substrate, N-acetylglucosamine (GlcNAc), to explore necromass utilization under different oxygen conditions. Two bioreactors seeded with wetland soil were

maintained under fully aerobic and fully anaerobic conditions. To test disturbance response, oxygen concentrations were inverted in two pulse experiments. As expected, the aerobic reactor produced only CO₂ and oxidized nitrogen species (nitrate, nitrite, and nitrous-oxide), while the anaerobic reactor generated organic acids, ammonium, CO₂, and methane. The aerobic system demonstrated tight coupling between carbon and nitrogen cycles with GlcNAc breakdown feeding into nitrification and denitrification pathways.

Microbial community composition differed significantly between the two systems, even after both had experienced a combination of aeration and N₂ sparging. While taxonomic diversity increased in the aerobic reactor over time, the opposite trend occurred in the anaerobic reactor as strict anaerobes were lost. We observed differences in GlcNAc transport with high affinity ABC transport preferred under aerobic conditions and the more efficient PTS transport preferred under anaerobic conditions. It's possible that the decrease in high-affinity transporters under anaerobic conditions may lead to the observed accumulation of amino sugars in deeper layers of the wetland.

Perspective

The deeper I ventured into microbial ecology during my PhD, the more I became struck by an apparent paradox: microbial communities display a striking determinism at certain levels, yet retain an almost infinite complexity at finer scales. At their core, microbiomes are extraordinarily complex systems—but somehow, the emergent properties of metabolism and functional gene abundance can be remarkably similar across systems. This raises a philosophical question: if the same patterns emerge time and again, what is the value in continuing to study them? Perhaps the answer lies in the question itself—understanding why and how these patterns

emerge is among the most fundamental pursuits in microbial ecology. This work attempts to elaborate on that premise.

Throughout this dissertation, we saw that oxygen availability is a powerful driver of microbial community composition. This trend was consistent across thousands of metagenomes from diverse environments (Chapter 2), along a soil depth gradient in a single wetland (Chapter 3), and within bioreactors seeded with the same wetland soil but exposed to different oxygen regimes (Chapter 4). Though the contexts varied, the emergent patterns—both taxonomic and functional—remained consistent. While the influence of oxygen is not a novel insight in itself, the real intrigue lies in understanding when and how a microbial community transitions between oxygen-defined states.

In the face of environmental change—particularly in sensitive ecosystems like wetlands—these transitions, or tipping points, become critically important. Interestingly, our work did not observe clear tipping points in community composition, either in response to seasonal changes in the field or in response to perturbations in our bioreactor experiments. However, despite the taxonomic stability, the metabolic outputs of these communities shifted dramatically. A vivid example occurred when the aerobic bioreactor turned anoxic, leading to rapid denitrification of available nitrate and nitrite and causing a sharp spike in nitrous oxide emissions. This transition unfolded over mere hours, suggesting a rapid metabolic shift among existing community members rather than a structural reorganization of the community.

These kinds of transient responses to environmental fluctuations offer a promising direction for future research. Studies that vary the duration or frequency of disturbances could help define the thresholds—or limits—of microbial flexibility and community succession. Kinetic modeling approaches can also play a key role, enabling exploration of broader parameter

spaces. For instance, with defined growth parameters for different microbial guilds, we can simulate competitive dynamics and coexistence under varying conditions. Synthetic microbial communities offer an important experimental middle ground between *in silico* models and natural ecosystems, enabling a more mechanistic understanding of community function. Applied to a consistent origin community—like a freshwater wetland—these different experimental systems can begin to illuminate how complexity gives rise to emergent ecosystem behavior.

It's easy, as microbiologists, to get lost in the intricacies of genes, pathways, and interactions. But zooming out to the human-scale context is crucial. For example, improving wetland management for carbon sequestration and greenhouse gas mitigation depends on our understanding of microbial carbon processing. Based on our findings and broader literature, managing water table levels appears to be a particularly promising strategy. It's also a feasible intervention, given that hydraulic control infrastructure already exists in many places—including our study site on Lake Washington, where water levels are regulated by a system of locks. In winter, water levels are drawn down to prevent flooding, but this also exposes saturated soils to air, leading to increased CO₂ emissions. Maintaining a higher water table during these periods could reduce aerobic decomposition and help preserve soil organic matter.

All of this reinforces a central message: context matters in microbial ecology. Greater consideration should be given to the broader environmental and practical contexts in which microbes operate. That means studying microbiomes in environmentally relevant model systems—like our low-dilution chemostats that mimic oligotrophic conditions—or in settings where mechanistic insights can be applied to other systems, particularly in engineered environments. While new 'omics technologies continue to expand our view of microbial life, we must remain committed to hypothesis-driven research. These tools generate hypotheses at an

unprecedented rate; we now have a backlog in need of careful testing. I hope future research—and funding—leans more heavily into the experimental validation of these hypotheses through culture-based approaches. That is how we will make real progress toward predictive models of microbial communities and their role in biogeochemical processes.

Curriculum Vitae

ZACH FLINKSTROM

zflink@uw.edu | [LinkedIn](#)

Seattle, WA, USA

Education

- PhD** Environmental Engineering, University of Washington, Seattle, WA Exp. June 2025
Advisor: Prof. Mari Winkler
- BS** Bioengineering, Northeastern University, Boston, MA 2018
Minor in Mathematics

Research and Work Experience

- Brown and Caldwell**, Seattle, WA June 2025
Environmental Engineer
- Winkler Lab**, University of Washington 2020–2025
Graduate Research Assistant
- Soil Fertility and Nutrient Recycling**, University of Copenhagen 2024
Visiting PhD student (Valle Fellowship)
- Editas Medicine**, Cambridge, MA 2018–2020
Research Associate, Discovery Biology
- Pinto Lab**, Northeastern University 2016 – 2018
Undergraduate Researcher
- Moderna Therapeutics**, Cambridge, MA 2017
Protein Sciences Co-op

Publications

Flinkstrom, Z., Bryson, S., Candry, P., Winkler, M.K.H. (2024). Metagenomic clustering links specific metabolic functions to globally relevant ecosystems. *mSystems*, 9 (8).
<https://doi.org/10.1128/msystems.00573-24>

Wang, D. et al. Metaproteomics-informed stoichiometric modeling reveals the responses of wetland microbial communities to oxygen and sulfate exposure. *npj Biofilms Microbiomes* 10, 55 (2024).
<https://doi.org/10.1038/s41522-024-00525-5>

Qin, W. et al. (2024). Ammonia-oxidizing bacteria and archaea exhibit differential nitrogen source preferences. *Nature Microbiology*, 9, 524-536. <https://doi.org/10.1038/s41564-023-01593-7>

Candry, P., Flinkstrom, Z., Winkler, M.K.H. (2023). Wetlands harbor lactic acid-driven chain elongators. *Microbiology Spectrum*. <https://doi.org/10.1128/spectrum.02105-23>

Kim, Y., Flinkstrom, Z., Candry, P., Winkler, M.K. H., & Myung, J. (2023). Resource availability governs polyhydroxyalkanoate (PHA) accumulation and diversity of methanotrophic enrichments from wetlands. *Frontiers in Bioengineering and Biotechnology*, 11. <https://doi.org/10.3389/fbioe.2023.1210392>

Flinkstrom, Z. et al. Soil depth shapes microbial community composition and function in a freshwater wetland. (In prep).

Flinkstrom, Z. et al. N-acetylglucosamine degradation by a wetland microbial community under varying oxygen concentrations. (In prep).

Flinkstrom, Z. et al. Comparative physiological and genomic characterization of a *Nitrobacter vulgaris* strain from a nitrate contaminated subsurface. (In prep).

Flinkstrom, Z. et al. Identifying microbial functional guilds performing cryptic organotrophic and lithotrophic redox cycles in anaerobic granular biofilms. (Submitted)

Wang, D. et al. Emergent higher-order microbial interactions revealed by comparative metabolic modeling of synthetic communities with varying species compositions. (Submitted)

Conference Posters and Presentations

Flinkstrom Z. and Winkler M. (2024). Understanding and harnessing microbes for climate change mitigation. Graduate Climate Conference.

Vuono D., Flinkstrom Z., et al. (2023). Continuous Measurements of NOB Growth through the application of photoconduction in a Microbial Growth Intervalometer (MAGI). International Conference on Nitrification and Related Processes (ICoN8).

Wang D., Flinkstrom Z., et al. (2022). Experimental models bridging single cell-to-ecosystem scales to evaluate climate-wetland feedback mechanisms. DOE Genomic Sciences Program PI Meeting.

Candry P., Flinkstrom Z., et al. (2021). Bridging Scales: Conceptualizing Microbe-Climate Links in Wetland Ecosystems. DOE Genomic Sciences Program PI Meeting.

Dass A., Diner B., Nayak R., Flinkstrom Z., et al. (2020). Dual AAV-based 'Knock-out-and-replace' of *RHO* as a Therapeutic Approach to Treat *RHO*-associated Autosomal Dominant Retinitis Pigmentosa. American Society of Gene & Cell Therapy (ASGCT).

Diner B., Fusco R., King E., Dass A., Flinkstrom Z., et al. (2019). Towards a CRISPR/Cas-based therapy for herpetic eye diseases. International Herpesvirus Workshop (IHW).

Gincley B., Dhru S., Worden H., Flinkstrom Z., Goodman G., Miller K., Pinto A., Mueller A., Lannin T., Belthangady C. (2019). ARTIMiS: Autonomous Real-Time Image-based Platform for Microbial Detection. Association of Environmental Engineering and Science Professors (AEESP).

Teaching and Service

- Teaching assistant in charge of lab section for CEE 352: Introduction to Microbial Principles in Environmental Engineering (Fall 2022, 2023, & 2024)
- Chair of UW CEE Graduate Student Advisory Board (2023-2025)

Technical Skills

- Python, C++, R, Bash, machine learning, statistics, data visualization, bioinformatics (metagenomics, 16S amplicon analysis), quantitative image analysis
- Microbial physiology, bioreactor operation and modeling, GHG flux measurement, plasmid design and cloning, qPCR, CRISPR-Cas