NORTHWESTERN UNIVERSITY

Metagenomics Guided Design of Wastewater Bioprocesses

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

For the degree

DOCTOR OF PHILOSOPHY

Field of Chemical and Biological Engineering

By

James Sheridan Griffin

EVANSTON, ILLINOIS

September 2018

ABSTRACT

Understanding and controlling microbial community assembly is critical to developing novel bioprocesses for nutrient and energy recovery from wastewater and preparing for global climate change. As molecular biology tools and DNA sequencing improve, microbial ecologists can progress from answering qualitative questions about "who is there, and what are they doing?" to rationally designing microbial communities for specific functions. Despite our increasing ability to monitor microbial community phylogenetic and functional composition, parsing the impact of community interactions, abiotic factors, dispersal limitations, and stochastic factors in complex microbial consortia remains a major challenge. This thesis first lays out important fundamental concepts and contemporary 'omics and computational tools in microbial ecology. Next it investigates community assembly mechanisms at regional and watershed scales and then applies these concepts to designing a microbial electrochemical cell (MEC) based bioprocess and investigating MEC community structure.

Chapters 1 and 2 introduce the background and methods for molecular microbial ecology, respectively. In chapter 3, a longitudinal field study was used to test for regional synchrony in population dynamics in activated sludge bioreactors. Using a nested experimental approach, longitudinal sampling of multiple reactors per site across several regional plants was performed. Community beta diversity over time and between sites was compared and temperature driven seasonal variation was the dominant process observed. This trend was consistent across multiple community scales in activated sludge systems and visible in nitrifier populations. In chapter 4, the amplicon sequencing, multivariate ordination and variance partitioning approaches developed previously were applied to study soil hydrology driven community assembly in an intensively managed landscape. In the final chapter, ongoing work integrating native prairie microbial

ecology with a novel longitudinal and spatially resolved groundwater monitoring system is discussed. The focus of these sections is demonstrating the impact of hydrology & soil moisture on community structure.

The next two chapters are devoted to lab-scale bioreactor studies of microbial electrochemical cells for H_2O_2 production. Exoelectrogenic biofilms were enriched in 3D-printed bioreactors and biological acetate oxidation was coupled to H_2O_2 electrosynthesis via oxygen reduction. Buffer composition, electrode composition and hydraulic residence time were optimized for H_2O_2 titer and pH. The produced H_2O_2 was utilized for tandem catalytic sulfoxidation of a model aromatic thioether compound with a heterogeneous niobium(V)-catalyst, and demonstrated similar activity to commercially produced H_2O_2. While this last finding is expected, the process demonstrates a novel approach for wastewater valorization through continuous production of H_2O_2 and its immeditate use as a selective oxidant in aqueous conditions for green chemistry applications. Finally, a genome-centric metagenomic approach is used to investigate the relationship between electron donor availability and MEC community structure, an open community assembly question of relevance for MEC scale-up. The methodologies applied in this thesis have applications for rationally designing other bioprocesses based on underlying microbial community assembly rules as well as investigating physiological niches of keystone microbes for existing processes.

Thesis Advisor

Dr. George F Wells

Final Examination Committee

Dr. Neda Bagheri

Dr. Justin Notestein

Dr. Keith Tyo

ACKNOWLEDGEMENTS

I'd like to thank my advisor, George Wells for all of his support and mentorship over the past several years. He has taught me everything I know scientifically speaking but more importantly has demonstrated consistency, communication and collaborative skills that will help me for the rest of my life. Starting a lab with you, Alex and Han was one of the most stressful and rewarding things I have ever done and I can't really imagine grad school in another environment at this point.

I would like to thank my committee members Professors Justin Notestein, Neda Bagheri and Keith Tyo for their help and feedback along the way. Justin Notestein, thank you for your encyclopedic knowledge of chemistry and your ability to translate it in ways even I can understand. To all of the amazingly talented professors I collaborated with, especially: Kim Gray, Aaron Packman, John Kelly, Jack Gilbert, and Hyung-Sool Lee, thank you for your ideas, time, and resources. To the excellent core facility directors at Argonne and UIC (Stefan Green), thank you for your help over the years. To everyone at the MWRD, especially Geeta Rijal, thank you for helping get our lab off the ground and for helping me see things from a very non-academic perspective.

Each of my labmates has contributed to my thesis by making the time spent in graduate school productive and enjoyable. Gao Han, your worth ethic and determination are extremely inspiring. Without you to push me I would not have accomplished nearly as much. Thank you for thanklessly and independently leading so many aspects of our lab (instrumentation, method development, reactor design and operation) over the past five years. Alex Rosenthal, your mentorship, positive and humble attitude and joy of tinkering in lab have been immensely

valuable to me and I'm extremely glad you helped shape the lab during your time here. Morgan Petrovich and Paul Roots: watching you grow and persevere during your time here has been a pleasure and I wish you the best for your theses and beyond. A special thanks to all of the undergrads I've mentored (Loren, Mark, Jordan, Ihsan), this work wouldn't exist without your contributions.

To all of my friends, you are truly amazing and inspiring. Special thanks to Paul Jaffe & Brian Basham, you are some of the kindest, most down to earth people I've ever met and I've looked up to you both over the past decade. Nick Grabenstein and Chris Varenhorst, thank you for letting me crash on your couches when I visit San Francisco and so much more. To the communities I've been a part of (Bexley, Beta, Dore, StarCats, Burning Man, Aloft, etc.), thanks for pushing me to the edge [athletic, emotional, legal, institutional or otherwise]. I feel more alive having shared a time and a place with each of you and felt your passion.

To my partners over the years, thank you for everything. You represent the ideal aspects of what I hold most dear. Mary Wang, you are the reason I'm writing this at all. Meeting you changed my life when it desperately needed to and I owe you so much for that. For now: "que sera, sera". To Natalie, thank you for opening my eyes to a side of myself that I never knew.

To everyone who has ever given me advice or feedback, or turned me down for an opportunity because I wasn't ready, thank you for making me better.

Lastly, to my parents: thank you for providing me with everything I needed to get where I am today. I am so grateful for the sacrifices you made, the example you led, and the education you enabled me to get.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Integrated Microbial Resource Management

1.1 INTRODUCTION

The evolution of microbial redox processes has reshaped the earth over its 4.5 billion year history. Microbial life drives many of Earth's biogeochemical cycles and makes up approximately 15% of all biomass on Earth.[1-3] In nature, microbes are present in diverse communities, but most industrial biotechnology processes use monocultures of model organisms like *E. coli* or *S. cerevisiae*. To address global challenges like access to clean water and food, climate change, and the spread of antibiotic resistance, it is critical to develop our ability to understand and engineer robust and resilient microbial communities.[4]

The development of the activated sludge process for wastewater treatment was one of the first applications of microbial community engineering and one of the most important public health innovations of the 20th century. Although simple, this process contributed to a dramatic decrease in waterborne illness in the developed world and still forms the basis for most industrialized forms of biological wastewater management. First industrialized in the 1910's, the process combines sewage aeration with biomass retention to select for microbial communities that rapidly oxidize organics and ammonia.[5] Activated sludge processes select for rapidly growing aerobic heterotrophs, resulting in high-energy requirements due to aeration (approximately 3% of domestic electricity production)[6] and production of large quantities of waste biomass that must be further treated. Despite these limitations, activated sludge processes demonstrate how environmental pressure can be utilized to select for and enrich beneficial communities of microbes at scale.

New environmental biotechnology processes are still developed via long-term selective pressure and selective biomass retention, but the functional microbes that catalyze these processes are no longer treated as a black box. Recently, the term *microbial resource*

management was coined to describe the process of integrating microbial physiology, ecological theory, molecular tools and mathematical modeling to engineer microbial communities for specific functions.[4] Critical to microbial resource management is understanding what makes microbial communities resistant (insensitive to environmental changes), and resilient (able to recover after disturbance), as shown in Figure 1.1.[7] Studies across many engineered consortia demonstrate that microbial communities contain ecologically redundant populations and are inherently dynamic, even under undisturbed conditions or periods of functional stability.[8, 9] Thus, differentiating between natural variability during functional stability and functionally relevant response to disturbance is essential for process monitoring and improvement.[10]

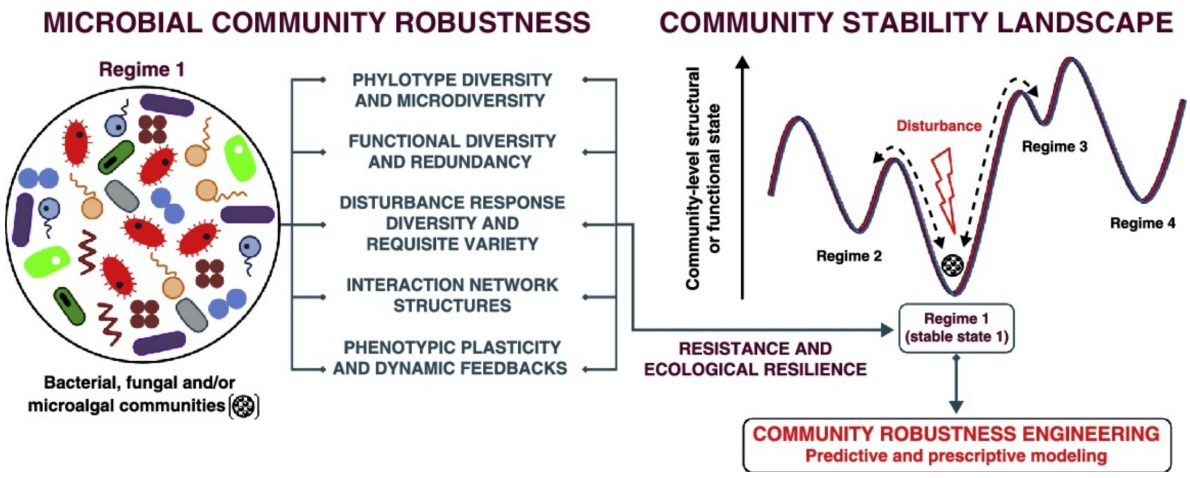

Figure 1.1 Conceptual overview of predicting community robustness and stability for engineered microbial communities. Reproduced from reference.[10]

Microbial resource management has been driven in part by next-generation sequencing (NGS). The development of high-throughput amplicon sequencing, analysis platforms and comprehensive phylogenetic databases has enabled low-cost population monitoring in engineered systems.[11-13] Meanwhile, whole genome sequencing (WGS) and metagenomics have

aided discovery and characterization of functional organisms such as exoelectrogens, commamox, and anammox organisms.[14-16] Recently, genome-resolved metagenomics has enabled direct reconstruction of genomes from environments, allowing detailed functional descriptions of the uncultivated organisms found in many environmental biotechnology processes. The success of microbial resource management approaches is evidenced in the commercialization of new wastewater bioprocesses based on metabolism of as-yet-cultivated anammox bacteria that catalyze direct anoxic conversion of ammonium to nitrogen gas (e.g. DEMON),[17] and enhanced biological phosphate removal processes based on enrichment of as-yet-uncultivated polyphosphate accumulating organisms (PAOs).[18] Microbial resource management has also been employed to improve understanding of anaerobic digestion processes and struvite ($MgNH_4PO_4 \cdot 6H_2O$) production.[8, 19]

While ecological theory can guide process operation and optimization, process design can be further improved by considering opportunities for resource recovery and evaluating waste management in a holistic or integrated manner. Traditional environmental biotechnology process development has focused on minimizing the cost of treating individual waste streams, however many of the processes described above have the potential to generate valuable co-products such as biogas or fertilizers. Life cycle analyses of existing wastewater resource recovery processes found that many are consistently beneficial when incorporating environmental benefits and considering broad system boundaries.[20] While life cycle analyses can draw arbitrary system boundaries, in practice current planning processes emphasize direct costs to utilities. *Integrated resource management* planning methodologies that consider economic, social and environmental impacts for a broad set of stakeholders are needed in order to fully utilize new resource recovery options.[21]

Microbial electrochemical cells (MECs) are a promising technology that employ a mixed culture microbial community for sustainable wastewater treatment for waste carbon recovery and a useful platform technology for integrated resource management.[22] MECs utilize anode respiring bacteria or *exoelectrogens* that use extracellular electron transfer (EET) to reduce solid extracellular substrates such as metals or electrodes.[23] Current generated this way can be used for electricity in microbial fuel cells or can be coupled to a variety of other electrochemical reactions such as oxygen reduction to form hydrogen peroxide.[24] In addition, MECs are particularly well suited for studying microbial community adaptation and metabolism as produced current can be directly related to metabolic flux through EET pathways.[25] While these systems are highly versatile, their technoeconomic viability is currently limited by power density and difficulty maintaining performance in pilot-scale systems treating real waste.[26, 27] MECs scale up challenges have been attributed to exoelectrogen maintenance and out-competition of methanogens, presenting a unique opportunity to explore both novel process design as well as community ecology.[28]

1.2 Background

1.2.1 Microbial Community Assembly

Microbial ecology is fundamentally concerned with understanding how and why diversity arises in microbial communities. Early microbiologists suggested that microbes faced fewer dispersal limitations than macroscopic organisms and were inherently cosmopolitan, leading to the maxim that "everything is everywhere, but the environment selects".[29] This strict environmental determinism dominated microbiology theory during the 20th century but was challenged by the arrival of neutral theory and molecular community fingerprinting techniques.[30, 31] Neutral ecological models state that ecological diversity is driven by stochastic factors such as

immigration, birth and death, rather than ecological processes. Concurrent with the development of neutral theory, molecular fingerprinting techniques based on PCR amplification of genes encoding ribosomal RNA followed by profiling with terminal restriction fragment length polymorphism (T-RFLP), or denaturing gradient gel electrophoresis (DGGE) profoundly shifted the focus of microbial ecology from culturing towards analysis of molecular data.[30]

Molecular fingerprinting techniques enabled rapid characterization of microbial diversity within complex communities. DNA fingerprinting and later amplicon sequencing work revealed that microbial communities are highly dynamic, functionally redundant and diverse but globally resemble communities in similar environments.[32, 33] Due to the inability to correlated phylogenetic gene marker abundance with community function, high sequencing error, and the inherent noisiness of individual populations, early fingerprinting studies of microbial ecosystems including those in wastewater suggested that species distributions were consistent with neutral theory predictions.[34-36] These findings led to a debate over the extent to which stochastic (neutral) versus deterministic (environmental selection) processes influence microbial community assembly. More recent theoretical approaches such as the metacommunity framework have emphasized that multiple community assembly mechanisms including selection, dispersal, historical and neutral factors each play a role in an environment-dependent context.[37]

Environmental bioreactors have proven to be useful model systems for testing and developing microbial ecology theory and have informed our understanding of community assembly mechanisms,[9, 38, 39] resilience and stability.[40, 41] Engineered environments possess well-defined physical boundaries and clearly phylogenetically delineated functional guilds responsible for nutrient removal and resource recovery processes.[42] Microbial resource management approaches to wastewater management have begun to recognize and account for ecological

processes, many of which are shared between engineered and other environments like freshwater lakes.[38, 43-45] Repeatable community dynamics have now been demonstrated in both lab- and full-scale bioreactor communities[8, 46], as well as in regional lakes.[47, 48] These patterns, termed 'regional synchrony' or 'temporal concordance' can be linked to exogenous drivers like regional weather patterns, known as the *Moran Effect*, or due to time-dependent processes during startup.

As microbial diversity becomes better characterized and community assembly is better understood, microbial ecologists can begin to move beyond the task of merely quantifying microbial diversity and instead focus on the mechanisms underlying diversity rational design of microbial communities for specific functions.[49] One approach is improved experimental design aimed towards hypothesis testing. Beyond this, shotgun metagenomic sequencing that does not rely on PCR amplification of marker genes has enabled a number of discoveries missed by phylogenetic approaches. Several metagenomic studies of bioreactors have found higher rates of functional stability (defined in terms of coverage of major pathways) compared with phylogenetic variability.[43, 50] In addition, metagenomic sequencing has identified previously unknown organisms like comammox organisms (complete ammonia oxidizers), that are phylogenetically classified as nitrite oxidizing bacteria, but which contain a functional ammonia oxidation pathway due to a horizontal gene transfer event.[15] These results demonstrate the utility of 'omics approaches that more directly capture biological relevant information. Thus far, there are limited studies using transcriptomics, proteomics, and metabolomics in mixed communities[51], however increased feasibility of these methods will improve efforts to predict and design stable community function.

1.2.2 Microbial Electrochemical Cells for Resource Recovery

Successful application of microbial resource management strategies can enable recovery of reusable water, energy, and materials such as bioplastics, fertilizers or fuels from wastewater.[21] Energy is currently recovered in the form of methane-rich biogas via anaerobic digestion (AD) and used in combined heat and power systems. This approach is mature but only technoeconomically viable for large scale treatment of concentrated waste streams.[52] Recently, alternative approaches such as ethanol or hydrogen fermentation have been investigated.[52] Microbial electrochemical cells (MECs), including microbial fuel cells and microbial peroxide producing cells (MFCs and MPPCs, respectively) are another promising approach. As a resource recovery platform, MECs have several advantages compared with anaerobic digestion. They can operate at lower temperatures than AD processes, as most archaeal methanotrophs are thermophiles. In addition, they can handle much lower organic loading rates than AD processes.[52] MECs can be designed to recover carbon via direct electricity, hydrogen, or hydrogen peroxide production from municipal wastewater.[53]

Despite these potential advantages, MECs are limited by low volumetric power densities as well as scale-up and feasibility challenges. Extensive work has been done characterizing the impact of design variables including cell configuration, electrode materials, inoculum source, carbon source and concentration and enrichment strategies on MEC performance.[54, 55] Key parameters controlling current density and electron recovery efficiency are the cell internal resistance, degree of exoelectrogen enrichment, and in the case of H_2O_2-producing cells, stability of the end product.[56, 57] Internal resistance within electrochemical cells can be broken down into individual contributions from the electrodes, electrolyte and separator resistances or by mechanisms such as ohmic resistance or activation losses (Figure 1.2). Activation losses are caused primarily by the kinetics of the biological or electrochemical reactions themselves.

Ohmic losses on the other hand represent resistance within the anode biofilm or electrodes, electrolyte or separator membrane.[58]

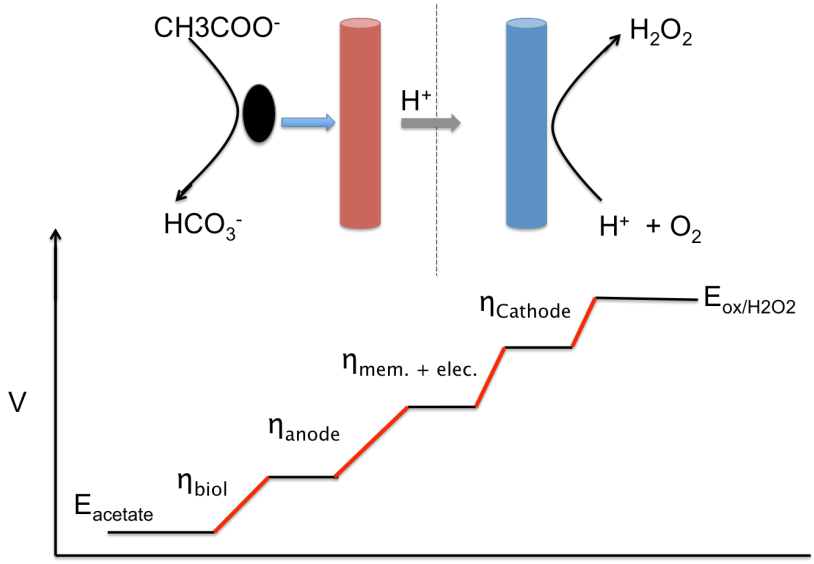

Figure 1.2 Schematic of potential loss (overpotentials) (η) within a two-chamber H2O2 producing MEC. ηbiol is the loss owing to extracellular electron transfer. ηanode, ηmem and ηcathode are the overpotentials due to interfacial resistance at the anode, electrolyte and membrane resistance and cathode resistance, respectively.

To minimize internal resistance, MECs are typically designed with minimal spacing between electrodes, high buffering capacity and high electrolyte conductivity. A variety of carbon based electrode materials including graphite, carbon cloth or felt, and 3D structures such as reticulated vitreous carbon (RVC), carbon fiber brushes and carbon nanotube (CNT)-coated sponges have been investigated as anodes.[59] Micro-scale carbon fiber anode MFCs fed acetate have demonstrated power densities of up to 6.9W/m^2 on the basis of anode surface area.[58] Despite the interest in anode materials, several studies have reported that cathodic overpotential is a larger contribution to resistance than anodic losses.[58]

Translating lab-scale performance gains into performance at scale remains difficult. Stable exoelectrogen enrichment and outcompetition of methanogens in real waste streams is a

major challenge. The highest recorded MEC power density was over 1kW/m^3 in a 2.5 mL reactor[60]; however a recently reported 1000L pilot-scale MFC reached a maximum current density of only 7.4 A/m^3 and most (86%) removed COD was converted to methane.[28] In addition, MEC power output is inherently variable and low voltage. Series-connected MEC stacks have been developed to increase voltage, but voltage reversals have been observed.[61]

Another approach to converting electrical current into useful products is electrochemical synthesis. Hydrogen production is the most thoroughly characterized approach but recently, hydrogen peroxide (H_2O_2) production has been explored.[62] Hydrogen peroxide has a number of uses including disinfection and bleaching, and a life cycle analysis found that high-performing MPPCs were environmentally and economically favorable compared to AD processes.[27] Cathode conditions are especially important in MPPCs due to the tendency of produced hydrogen peroxide to be further reduced to water at the anode or degraded in solution, leading to low coulombic efficiencies (<25%)[63] or high required energy inputs.[64] Several strategies to stabilize produced hydrogen peroxide are possible. Transition metal ions are known to promote hydrogen peroxide breakdown, so chelating agents such as EDTA and DTPA could be used to inhibit this mechanism.[65] H_2O_2 stabilizing agents such as acetanilide, phenol, tin or nitrates are another potential route.[65] H_2O_2 stability varies greatly by pH, so effective management of pH increase in the catholyte is another approach to improving H_2O_2 yield. Finally, optimizing the catholyte residence time near the cathode to reduce electrocatalytic reduction is another route to limit degradation. Although hydrogen peroxide yield and stability in MPPCs has been pointed out as a challenge, little work investigating these methods has been done to date. Another approach to minimizing H_2O_2 degradation is immediate on-site utilization. In accordance with integrated

resource management approaches, MEC-derived H_2O_2 may be a cost-effective oxidant for lignin or for chemical synthesis via tandem catalysis.

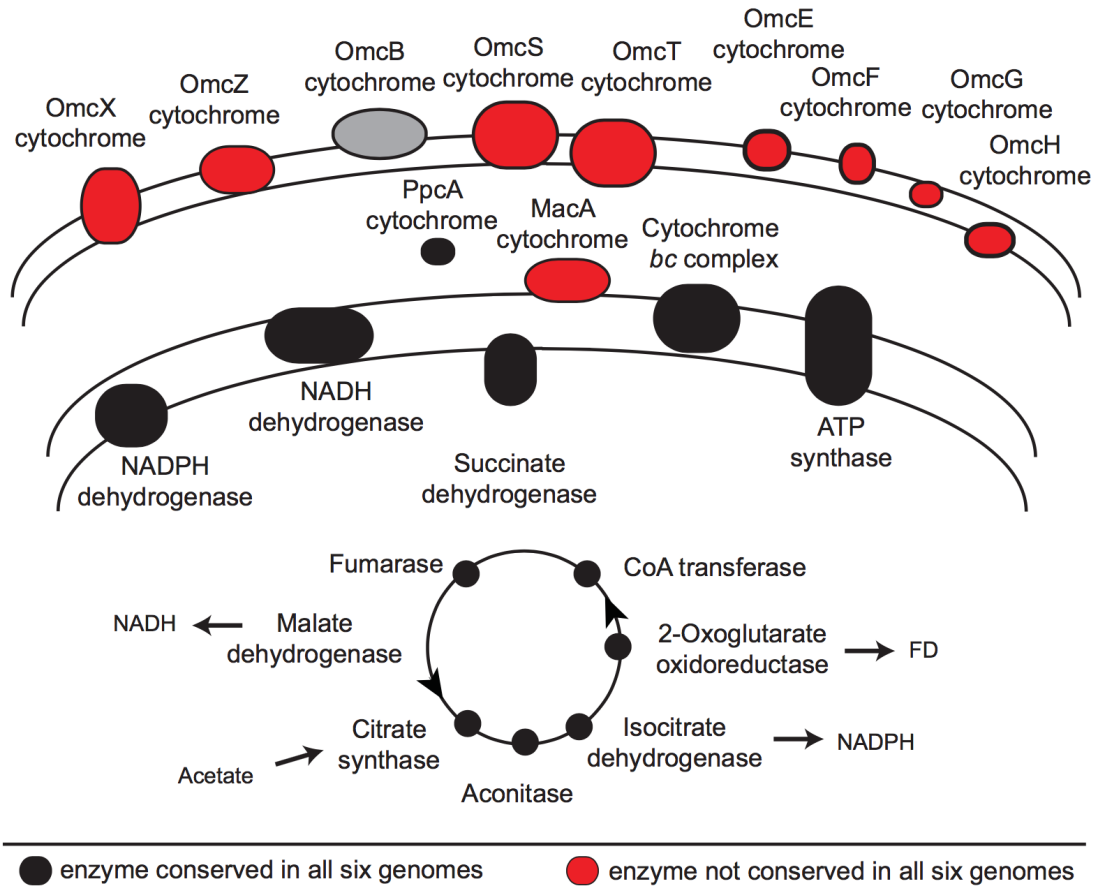

Figure 1.3 Energy metabolism in six *Geobacter sulfurreducens* genomes. ATP is generated via TCA oxidation. Periplasmic and outer membrane cytochrome proteins necessary for EET are shown. The enzymes are colored black if there were orthologs for every subunit in all of the species and red if there were not. OmcB is shown in gray because there positional but not sequence-based orthologs. Reproduced from reference.[14]

1.2.3 Exoelectrogen Metabolism and Biofilm Communities

Microbial electrochemical cells generate current from microbial extracellular electron transfer (EET). EET is a metabolic adaptation to environments lacking soluble electron acceptors

such as oxygen. By reducing insoluble substrates such as Fe(III) outside the cell, exoelectrogens can utilize fermentation byproducts such as acetate for energy production in subterranean anaerobic environments. The most well characterized exoelectrogens are *Geobacter spp.*, members of the family *Geobacteraceae* within the *Deltaproteobacteria* class. Although there are other known exoelectrogens (e.g. *Shewanella* spp.), *Geobacter* are typically the dominant exoelectrogens in lab-scale acetate-fed MECs possibly due to their high affinity for acetate as well as their ability to grow thick conductive biofilms.[66] Extracellular electron transfer in *Geobacter* is thought to proceed via TCA cycle oxidation, followed by electron transport along a complicated outer membrane transport chain and into a conductive pili network (Figure 1.3).[14]

Despite thorough characterization of Geobacter physiology and EET mechanisms, practical implementation of MECs is limited by our ability to design and maintain microbial consortia capable of converting complex organics into electrical current. In addition, there are fundamental knowledge gaps about how complex organic carbon sources influence the structure and function of anode biofilm communities. Previous studies have shown that carbon source strongly influences *Geobacter* enrichment and performance in MECs.[54, 67, 68] *G Sulfurreducens,* a well-studied exoelectrogen found in MECs, can only use acetate or hydrogen as electron donors.[69] Performance in systems treating real waste streams is lower than in lab-scale systems because most exoelectrogens rely on syntrophic interactions with fermenters to break down longer volatile fatty acids (VFAs) into useable electron donors (Figure 1.4). These communities tend to be much more diverse and contain fermenters and homo-acetogenic bacteria that produce acetate as well as methanogens and exoelectrogens that compete for acetate.[70]

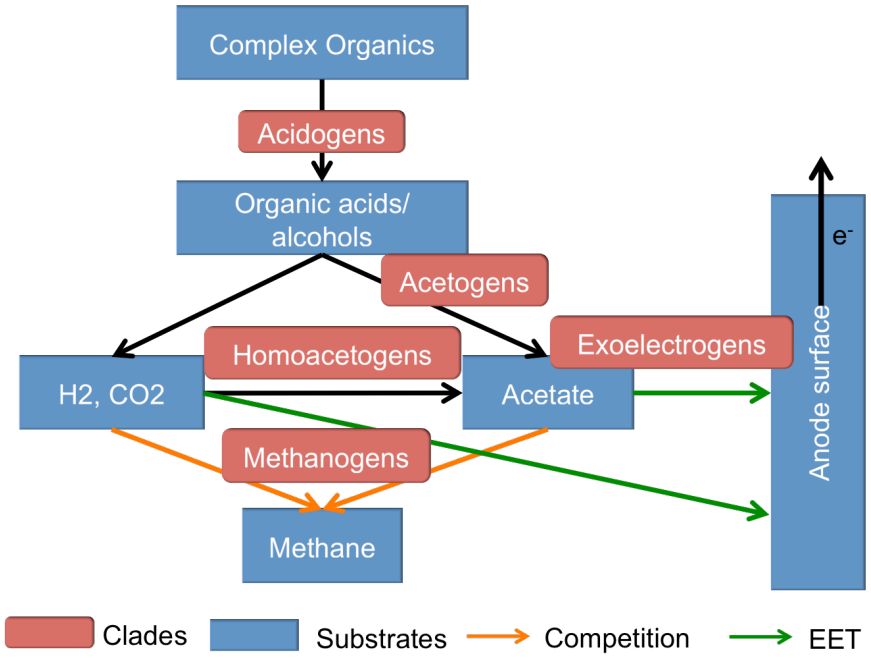

Figure 1.4 Conceptual map of syntrophic interactions in MFCs treating complex organic substrates. Community diversity is critical to maintain performance and metabolic flexibility but also presents challenges in terms of competition for acetate.

Characterizing both the exoelectrogen populations as well as the flanking community is critical to improve scale-up and design of MECs. Much of what we know about Geobacter and exoelectrogen metabolism comes from pure culture experiments; however it is now possible to directly investigate functional populations within enrichment cultures using genome-centric metagenomics[51]. Metagenomic studies offer the potential to study previously unknown diversity and performance variation[71] and investigate dominant EET mechanisms in MECs. In addition, few studies examining the prevalence of the more then 100 outer membrane cytochromes (OMC) in the *Geobacter* pangenome have been performed in MEC populations.[72]

Beyond exoelectrogen physiology, the functional roles of flanking (non-exoelectrogenic) community members are poorly understood. Community-scale differences in flanking community function could be used to explain the effectiveness of enrichment strategies. Little is

known about how electron donor availability influences the abundance of genes associated with steps in the syntrophic fermentation process for acetogenesis such as formyltetrahydrofolate synthetase, (FTHFS)[73] and unwanted processes such as methanogenesis, e.g. methyl-coenzyme M reductase (MCR).[74] Gene abundance within these supporting pathways could be used to inform new strategies for selective enrichment of exoelectrogenesis over methanogenesis in the presence of complex organics.

1.3 THESIS OVERVIEW

The work in this thesis attempts to improve our understanding of microbial community assembly mechanisms and develop resource recovery processes using this knowledge. The first two works focus on field-scale studies of microbial diversity in engineered and natural environments. The last two works focus on development of microbial electrochemical cell-based resource recovery and investigation of metabolic processes in these systems.

In chapter one, important microbial ecology frameworks and community assembly mechanisms are introduced. Known exoelectrogenic metabolic pathways and biofilm community structure is described. Current microbial electrochemical cell approaches and performance are reviewed. Chapter two describes the computational methods used to analyze amplicon and metagenomic next generation sequencing data. A brief review of algorithms and software tools is provided, followed by protocols and best practices for using the sequencing pipelines developed for the following work.

The third chapter discusses regional population synchrony in full-scale activated sludge microbial communities.[46] Synchronous population dynamics imply deterministic community assembly driven by regional exogenous environmental conditions. This finding is relevant to both environmental biotechnology practitioners and microbial ecologists because it demonstrates

a fundamental similarity between natural aquatic ecosystems and activated sludge bioreactors and provides a theoretical basis for rational design and engineering of microbial communities via operational parameters. Previous work had found mixed evidence for deterministic community assembly in full-scale activated sludge systems.[35] Our work demonstrates predictable and consistent seasonal fluctuations across four regional plants and many scales (OTU, phylum, alpha and beta diversity). Moreover, these predictable fluctuations were evident in a key functional nitrite oxidizing bacterial population.

The fourth chapter extends the community assembly mechanism work described above into natural soil and freshwater environments. The chapter is devoted to a study of microbial diversity in an intensively managed landscape at the watershed scale.[75] Further work integrating a novel soil moisture monitoring system and derived hydrological model of a tallgrass prairie with spatially resolved community structure data is presented in the future work section of chapter seven. These environments represent extremes in terms of hydrological flow in Midwestern soils, from well-drained agricultural soils to native marshes. Together, these works focus on the impact of subsurface flow on community structure via dispersal and transport dependent environmental selection.

Chapter five discusses development of a lab-scale process for carbon recovery via microbial electrochemical cells.[76] After enriching an exolectrogenic biofilm from a wastewater community, bioelectrochemical H_2O_2 production was optimized by tuning buffer and electrode composition. Using this approach, a citric acid-phosphate buffered MPPC obtained a competitive H_2O_2 concentration (3.1 g L^{-1}) at a low energy input (1.6 Wh g^{-1} H_2O_2) and pH (10) compared with previous results in the literature. While previous work had focused on H_2O_2 production only, we developed a two-step approach using the H_2O_2 for heterogeneous catalysis, focusing on

selective aqueous sulfoxidation of model compounds over niobium(V)-silica catalysts. We achieved 82% conversion of 50 mM 4-hydroxythioanisole to 4-(methylsulfinyl)-phenol with 99% selectivity with a 0.5 mol% catalyst loading in 100 minutes in aqueous media. Immediate consumption of H_2O_2 overcomes both storage challenges and unlocks higher value applications than disinfection or bleaching.

The sixth chapter discusses amplicon and metagenomic sequencing of exoelectrogenic biofilms enriched with different carbon sources. A major challenge of scaling up MECs remains the metabolic limitations of exoelectrogens. Because most carbon in common waste feedstocks is not readily biodegradable or accessible by exoelectrogens, it is necessary to convert this into volatile fatty acids (VFAs) or acetate. Previous attempts to characterize MEC biofilms enriched with different electron donors have relied on amplicon sequencing of the 16S rRNA gene, which cannot directly identify functional differences, with the exception of a few known monophyletic functions and pathways. A genome-centric metagenomic pipeline was developed and tested to identify and functionally annotate population bins for communities enriched under different conditions.

The final chapter summarizes the overall conclusions and discusses future work. Looking ahead, a major future direction could be characterizing the response of different exoelectrogenic communities to electron donor perturbations, via both short-term (transcriptomic) and longer-term (metagenomic) adaptations. Collectively, this thesis expands our capabilities to harness ecological principles for bioprocess design and carbon recovery. In addition, careful design of field studies and integration of novel sampling & monitoring approaches can greatly improve our ability to identify dominant assembly processes in complex environments and answer questions about what gives rise to biological diversity. By understanding the degree to which selection,

dispersal and stochasticity drive assembly under reactor conditions, as well as the physiological niche of target organisms, novel resource recovery processes can be scaled up and deployed more readily. Leveraging the ongoing NGS revolution, we can improve existing bioprocesses and develop new approaches based on microbial resource management strategies.

CHAPTER TWO

Metagenomic Analysis Methods and Trends

A key technique used in the work presented in this thesis is the analysis of next-generation sequencing (NGS) data. Since the completion of the human genome project, sequencing costs have rapidly decreased, enabling many of the approaches presented here.[77] Due to the pace of technological innovation, a comprehensive review of state-of-the-art approaches is likely to be outdated as soon as it is published, but many such reviews exist.[78-81] Although other approaches such as Minion nanopore sequencing[82] and Pacific Biosciences' Single Molecule Real Time Sequencing (SMRT) exist,[83] the short-read sequencing-by-synthesis (SBS) approach developed by Illumina is the most widespread NGS method due to its low-cost and high accuracy. Sequencing-by-synthesis uses a repeated extension–termination–cleavage–extension cycle to produce millions to billions of ~100-300 base pair reads per run.[84, 85]

Below, I summarize state-of-the-art methods for amplicon clustering and metagenomic assembly and describe workflows I have published on github for these bioinformatics tasks. I focus on analyzing sequencing data produced by Illumina's Miseq and Hiseq platforms. Many of the tools described below were designed with the technical specifications of Illumina data in mind: high average base quality, short individual reads, and very low rates of indel errors compared to other methods.

2.1 16S Amplicon Analysis

Amplicon sequencing refers to targeted sequencing of PCR amplified genes of interest. Compared to whole-genome sequencing approaches it is much more cost-effective when targeting a small number of genes. One of its most common is profiling microbial community diversity via 16S rRNA gene sequencing.[12, 86] The 16S gene codes for the small subunit of the prokaryotic ribosome, and has long been used as a universal phylogenetic marker gene for bacteria and archaea.[87] Due to the functional secondary structure of the rRNA, there are a

number of highly conserved regions within the 16S gene interspersed with hypervariable regions that can be sequenced with short-read sequencing approaches.[88]

A number of amplicon clustering and analysis tools have been developed including Mothur[89], QIIME[13], UPARSE/Vsearch[90, 91], DADA[92], and Swarm.[93] Generally speaking, all of these methods follow an approach similar to that in Figure 2.1, although the exact order of and boundaries between steps differs by implementation. The most active areas of development have been in clustering approaches.

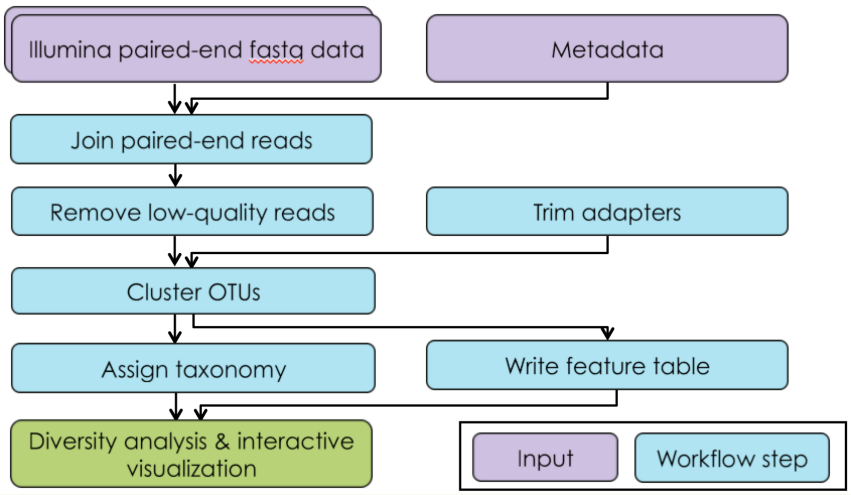

Figure 2.1 Amplicon sequence processing workflow. Input files (fastq files and sample metadata table) are shown in purple, and workflow analysis steps are shown in blue. The output of the pipeline is a feature table (sample x OTU), phylogenetic tree, OTU taxonomy table, and several standard alpha and beta diversity metrics.

Early clustering algorithms (e.g. Mothur and QIIME) used reference-based clustering or *de novo* approaches with arbitrary global thresholds (canonically 97% sequence similarity) and randomly seeded agglomerative methods to cluster sequences into operational taxonomic units (OTUs). Reference-based clustering inherently ignores novel biodiversity, whereas the *de novo* clustering approaches introduce input-order dependency and fail to differentiate closely related strains and sequencing errors. To address these challenges, new open source *de novo* clustering

algorithms that prioritize abundant sequences (e.g. Vsearch)[91], or utilize denoising to resolve sub-OTU level diversity into exact "amplicon sequence variants" (ASVs) (e.g. DADA2) have been developed.[92] Workflow implementations of and documentation for Vsearch- and DADA2 based pipelines are available at github.com/jimbopants/amplicon_processing and -/DADA2_processing, respectively.

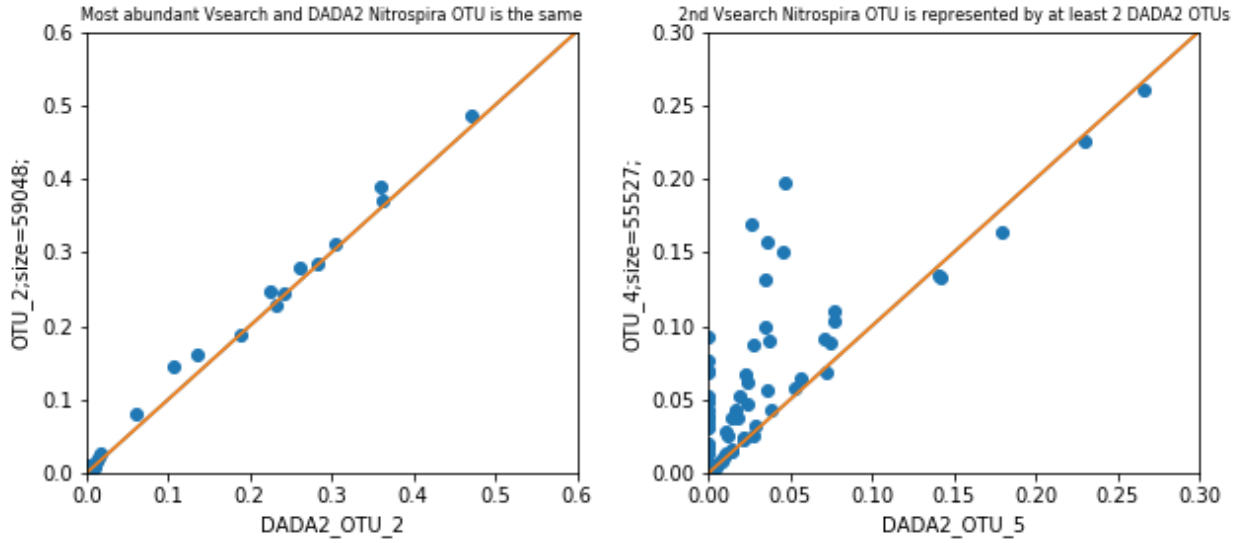

Figure 2.2 Comparison of cluster abundances produced by DADA2 (x-axis) and Vsearch (y-axis) corresponding to the same representative sequence. (A) All samples (blue dots) show nearly identical abundance (45 degree line). (B) This Vsearch OTU is represented by at least 2 ASV clusters produced by DADA2, as indicated by the large number of samples with OTU abundance > ASV abundance.

Comparing the two approaches, the Vsearch OTU clustering approach is faster due to DADA2's error model training. In addition, the DADA2 developers recommend using run-specific error cutoffs, requiring preliminary visualization of average per-base error profiles with something like FastQC. I compared *Nitrospira* affiliated OTUs and ASVs generated from a bioreactor time-series and found that most of the abundant ASVs had 1:1 correspondence with a single OTU, but occasionally Vsearch agglomerated multiple ASVs into a single OTU (Figure 2.2). Benchmarking studies performed by others show that ASV clustering approaches

outperform OTU-based methods.[94] Future amplicon sequencing projects are strongly encouraged to use ASV methods based on DADA2 or the recently developed Deblur.[95]

2.2 Ecology Metrics & Statistical Methods

Alpha and Beta Diversity metrics offer a way to further reduce the complexity of amplicon sequencing datasets into interpretable results. Alpha diversity approximates the diversity within a sample, whereas beta diversity summarizes compositional differences between samples. Common alpha diversity metrics include richness (number of species) as well as metrics weighted by evenness (Shannon entropy) and phylogenetic differences (Faith's PD).[96, 97] Testing for differences in diversity metrics between groups of communities is typically performed with Analysis of Variance (ANOVA), a multi-group extension of a t-test that compares intra-group to inter-group mean and variation. Alpha diversity can be linearly modeled as a function of environmental parameters or gradients using standard regression approaches.[98]

While a number of macroecological beta diversity metrics are used in practice, one of the most widespread and useful metrics for classifying microbial communities based on amplicon sequencing data is Unifrac distance. Unifrac distance is calculated from the ratio of the intersection to the union of a phylogenetic tree spanning two communities.[99] Weighted and generalized Unifrac are more statistically powerful approaches that incorporate species abundance differences as well as presence-absence data. In general, weighted Unifrac is the most powerful for detecting changes in abundant species whereas generalized Unifrac captures changes in moderately abundant community members while retaining the ability to detect changes in rare and abundant taxa.[100] Beta diversity measures are inherently pairwise and require multivariate statistical approaches. Statistical significance of categorical groupings can be tested using Permutational Analysis of Variance (PERMANOVA) and Analysis of Similarities

(ANOSIM), non-parametric tests of group variance and similarity, respectively. Correlation between community dissimilarity and spatial, temporal, or environmental distance matrices can be detected via Mantel tests or Procrustes analysis.[101]

Ordination methods like principal component analysis (PCA) are exploratory data analysis techniques used to visualize differences in community structure by reducing data dimensionality onto a set of uncorrelated principal components. In addition to unconstrained approaches like PCoA, constrained ordination techniques like redundancy analysis (RDA) can be used to visualize community composition changes correlated with explanatory environmental variables. RDA can be thought of as a multiple linear regression process followed by PCA of the linear model coefficients. RDA ordination only captures community variation explained by the explanatory matrix considered, skewing visual results if composition is not well correlated to measured covariates. This feature can also be exploited by comparing multiple RDA models containing different environmental parameters to identify community variance uniquely explained by subsets of parameters in a process known as partial RDA.[102] A close analog of this process known as distance-based RDA, implemented as the "capscale" function in Vegan, allows extension of this process to distance matrices (e.g. Unifrac distances) as well as raw observation data.

Excellent online and text references for numerical ecology methods are available.[103, 104] Software packages for calculation, visualization and statistical analysis of ecological metrics include Quantitative Insights Into Microbial Ecology (QIIME)[13], the R Vegan[105] and Phyloseq[106] packages, and Python scikit-bio package.

2.3 Genome Binning From Shotgun Sequencing Approaches

While amplicon sequencing drove many early NGS microbiome projects[107, 108], phylogenetic marker genes surveys paint an incomplete picture of microbial diversity. In addition to technical challenges like chimeric 16S sequences[109], 16S rRNA copy number variation and intra-genome heterogeneity[110], marker gene studies are incapable of directly informing function. Horizontal gene transfer, evolutionary drift, and mobile genetic elements make extrapolating function from single marker genes challenging, although several tools have been developed to do so.[111]

Recent whole genome sequencing (WGS) or "shotgun metagenomic" approaches have been developed to sequence community DNA directly, without amplifying 16S genes. Among these approaches, genome binning or genome-centric metagenomics have recently become more popular. Genome binning is the process of assembling and sorting short reads into population "bins". Ideally these bins represent a circular genome from a single cell, but in practice they are often a number of assembled scaffolds from closely related organisms. Most binning strategies revolve around using genetic signatures- GC content or tetranucleotide frequency- or differential coverage across multiple samples.

I implemented the genome-binning pipeline shown in Figure 2.3 and available at github.com/jimbopants/metagenomics_assembly. It is comprised of read assembly, read mapping, genome binning, bin QC, and functional annotation steps, described below. Tool selections were initially based on recommendations from the Critical Assessment of Metagenomic Interpretation (CAMI) project in which method performance was compared across a gold standard dataset.[112]

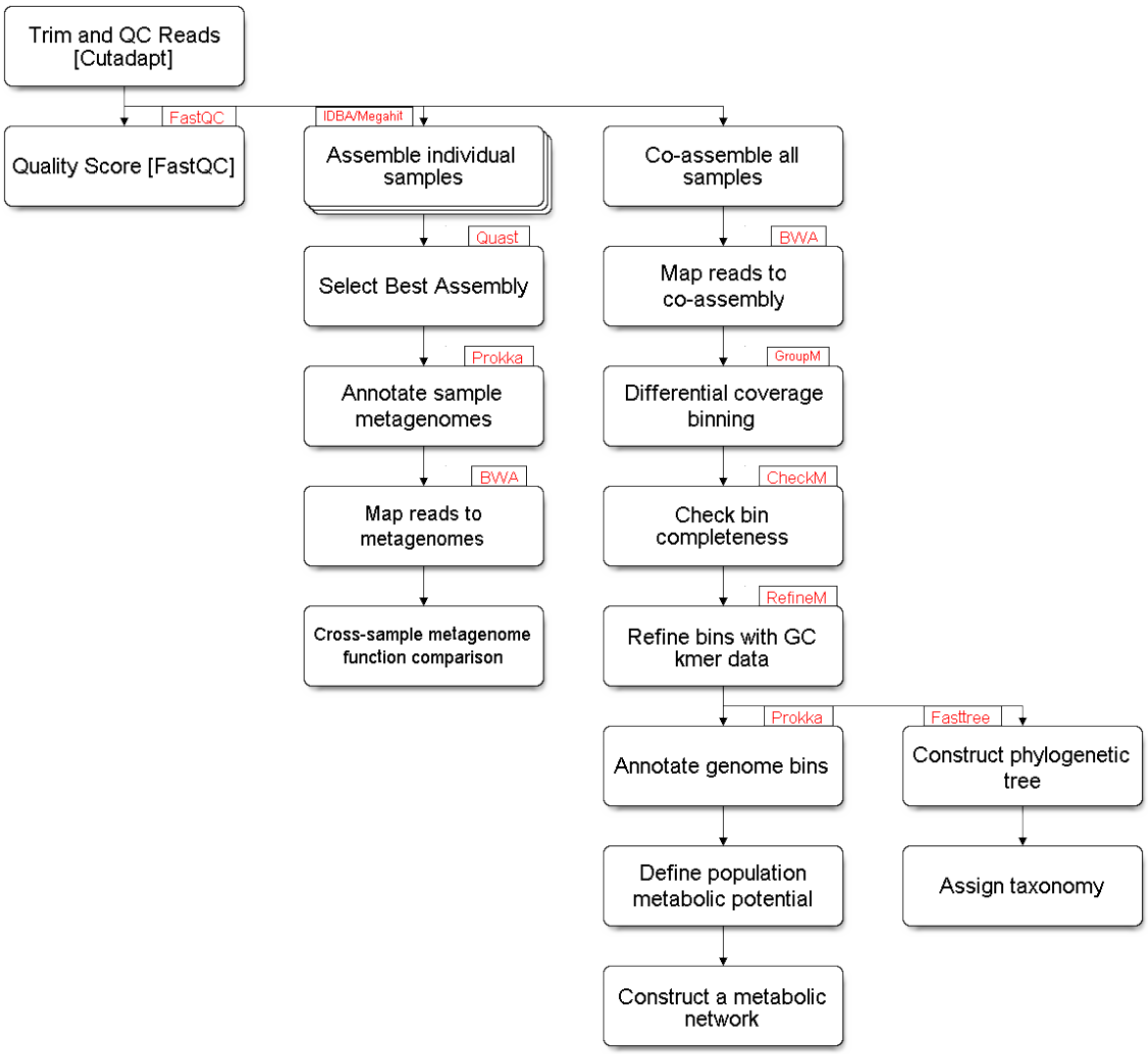

Figure 2.3 Genome-centric metagenomic analysis pipeline developed during this thesis. Steps are outlined in boxes and specific software tools are shown in red.

Read Assembly: *De novo* assembly of metagenomic data is computationally intractable with global sequence alignment methods due to their $O(n^2)$ complexity with the number of sequences. Current genome assembly approaches split reads into shorter kmers (>20 bp) and construct de Bruijn graphs by matching kmer prefix and suffixes.[113] A 2017 benchmark of open-source metagenomic assemblers found that MetaSpades, IDBA, and Megahit performed best but recommends comparing the assembly quality (N50 and % assembled) from several methods with

Quast.[114-118] If using differential coverage based genome binners, co-assembly of data from all samples is necessary to infer contig coverage profiles. Sample co-assembly is necessary for differential coverage binning but significantly increases assembly time.

Read mapping: This step generates contig coverage profiles by sample. Reads from each sample are aligned to the contigs using Burrows-Wheelers Transform based alignment methods such as BWA or Bowtie .[119, 120] Afterwards, coverage profiles can be generated using BEDTools.[121]

Genome Binning: Metagenomic binning methods remain an active area of research. Some of the best performing tools available today are CONCOCT, Metabat and Maxbin.[122-124] All of these approaches incorporate both differential coverage and genomic signatures such as tetranucleotide frequency. Bin quality is assessed via CheckM on the basis of inclusion of ubiquitous essential single copy genes.[125] "Completeness" measures the fraction of essential single copy genes present in a bin whereas "contamination" measures the number of repeats for these genes. Metrics reported by CheckM cover a small fraction of a total genome and extrapolating these metrics genome-wide may produce overly optimistic quality estimates. Standard quality thresholds for evaluating genomes have been published as well.[126] Beyond the individual methods themselves, a number of bin refinement tools exist that improve predictions by integrating bins from binning ensembles.[127-129] Annotation consists of open reading frame (ORF) detection via Prodigal[130], followed by classification against protein homolog families with HMMER[131].

2.4 Containerization of Bioinformatics Software

Bioinformatics analysis often depends on a variety of 3rd party software tools implementing sequential data manipulation steps. Many of these tools originate in academia,

where developers are rewarded for high impact papers, not necessarily software design and deployment practices. As a result, as new algorithms and tools are disseminated via research papers, users may be faced with significant time investments spent installing and maintaining 3rd party software pipelines.[132] Each step in these pipelines may have incompatible dependencies and formats, require user modifications to interface with up- or downstream tools, and require replacement as time goes on. This leads to tradeoffs between following up to date best practices and avoiding time wasted installing and fixing software in the user's environment.

A current trend in bioinformatics software is the "containerization" of software using platforms like Docker. Docker containers are pre-built and encapsulated Linux environments that can be packaged with bioinformatics tools already installed. They are similar to virtual machines (VMs) but require less overhead to run. Docker containers are a promising approach to standardizing bioinformatics tools and minimizing time wasted on systems administration and debugging. Two repositories for bioinformatics-related Docker images are BioContainers[133] and BioBoxes.[134] At the time of this thesis, BioContainers contains > 2,000 validated containers including many of the tools discussed in this chapter. Benchmarking studies of common NGS analysis steps show that Docker containers do not add significant overhead compared to locally installed software.[135]

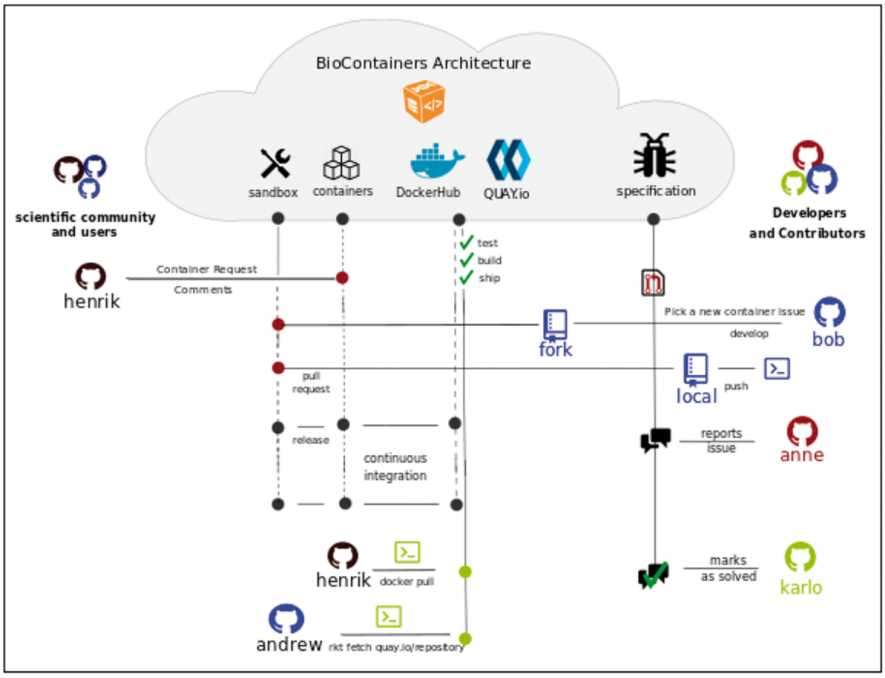

Figure 2.4 Software deployment and usage with BioContainers. Reproduced from reference.[133]

Northwestern's Quest High-Performance Computing (HPC) Cluster is in the process of installing and supporting Singularity, (personal communication) a Docker-like container platform designed to work with HPC resource schedulers.[136] As a general rule, checking container repositories prior to installing software locally should improve efficiency. Northwestern metagenomics researchers may also benefit from running analyses via kbase[137], a Joint Genome Institute project hosting many metagenomics tools, or purchasing resources on Amazon Web Services (AWS) for rapid deployment and benchmarking of new pipelines via Docker. Looking forward, Docker distribution of NGS tools is likely to become more widespread as it eases burdens on both developers and users.

CHAPTER THREE

Microbial Community Assembly In Full-Scale Activated Sludge Bioreactors

Material in this chapter is based on the published work:

Griffin, J.S. and Wells, G.F., 2017. *The ISME journal*, *11*(2), p.500.

3.1 ABSTRACT

Seasonal community structure and regionally synchronous population dynamics have been observed in natural microbial ecosystems but have not been well documented in wastewater treatment bioreactors. Few studies of community dynamics in full-scale activated sludge systems facing similar meteorological conditions have been done to compare the importance of deterministic and neutral community assembly mechanisms. We subjected activated sludge weekly samples from six regional full scale bioreactors at four wastewater treatment plants obtained over one year to Illumina sequencing of 16S rRNA genes, resulting in a library of over 17 million sequences. All samples derived from reactors treating primarily municipal wastewater. Despite variation in operational characteristics and location, communities displayed temporal synchrony at the individual OTU, broad phylogenetic affiliation and community-wide scale. Bioreactor communities were dominated by 134 abundant and highly regionally synchronized OTU populations that accounted for over 50% of the total reads. Non-core OTUs displayed abundance dependent population synchrony. Alpha diversity varied by reactor but showed a highly reproducible and synchronous seasonal fluctuation. Community similarity was dominated by seasonal changes, but individual reactors maintained minor stable differences after one year. Finally, the impacts of mass migration driven by direct biomass transfers between reactors was investigated, but had no significant effect on community similarity or diversity in the sink community. Our results show that population dynamics in activated sludge bioreactors are consistent with niche driven assembly guided by seasonal temperature fluctuations.

3.2 INTRODUCTION

Mixed-culture activated sludge systems are the most important form of modern wastewater treatment.[138] Consistent system performance relies on a complex microbial

community to remove organic carbon and nutrients in the face of dynamic environmental conditions. In addition to their importance for protecting environmental and public health, engineered environments such as activated sludge have been shown to be valuable environments to study fundamental microbial ecology phenomena[42], in part due to their physical partitioning into ecological "islands" with well defined and monitored ecosystem functions. Wastewater microbiology has informed our understanding of community assembly mechanisms[9, 39, 139], resilience and stability[40, 41], and identified novel microbes and metabolic pathways[15, 140]. Translating this knowledge into effective strategies for 'microbial resource management'[4] remains a challenge due to the high diversity and dynamics in activated sludge bioreactors, even during times of functional stability.

In both natural and engineered microbial ecosystems, the extent to which stochastic versus deterministic processes influence microbial community assembly is still debated. Traditional niche community assembly theory predicts that deterministic processes such as regional meteorological conditions and operational differences control assembly, whereas neutral theory predicts that trophically similar community members are ecologically similar and that stochastic processes such as immigration, birth and death lead to community differences.[31, 36] Community assembly theories such as the metacommunity framework incorporate both niche processes, such as species sorting and dispersal limitations, as well as neutral processes.[37] Previous studies have shown that deterministic factors explain much of the variation in activated sludge community structure within the same bioreactor, but neutral models can explain some aspects of community assembly, such as rare taxa dynamics.[39] However, the factors controlling the relative contributions of different community assembly mechanisms are still poorly understood.

Longitudinal studies of full-scale bioreactors have found that microbial communities appear to display continuous rather than cyclical succession patterns in activated sludge systems.[43, 139] Many of these studies focused on single reactors, making it challenging to identify whether observed associations were repeatable and generalizable. In contrast, seasonal population dynamics have been shown in a variety of marine and freshwater environments.[44, 45] The apparent differences between community assembly in natural and engineered environments[141] have been ascribed to more highly controlled environments in engineered systems studied to date; however, seasonal differences in ecosystem function (performance) including nutrient removal are common.[43]

Spatially correlated population fluctuations or 'regional synchrony'[47] and synchronized shifts in community structure or 'temporal concordance'[48] are hallmarks of deterministic community assembly mechanisms. Concordance and synchrony dictate how broadly we can generalize from community studies. Regional population synchrony[142] is explained by similar regional weather patterns, known as the "Moran Effect", immigration between communities and historical factors[143]. Synchrony has been demonstrated during startup of lab-scale reactors[8] but whether the same processes control dynamics in full-scale reactors is an open question. Regional population synchrony would provide further evidence for deterministic control of WWTP microbial community dynamics. To date, limited replication from full-scale bioreactors undergoing similar environmental conditions have made it difficult to identify synchrony in engineered systems.

The primary objectives of this work were to test whether deterministic factors such as seasonal variability drove microbial community assembly and whether assembly mechanisms affected rare and abundant OTUs differently in regional activated sludge bioreactors. We

analyzed local communities within a plant that shared an immigration (influent) source as well as geographically distinct communities within a region to determine the relative importance of extrinsic environmental factors and immigration. Temporal population synchrony was quantified as a means of identifying repeatable deterministic responses of individual OTUs to regional environmental dynamics. We examined synchrony at multiple scales, from whole community metrics of alpha and beta diversity, to dynamics of individual OTU populations between reactors. We hypothesized that, similar to natural aquatic environments, seasonal variation would be a major determinant of community composition and diversity in activated sludge and would outweigh stable differences between plants. Further, we hypothesized that abundant and rare OTUs would exhibit different dynamics, with abundant OTUs tending to exhibit synchronized seasonal blooms in different reactors due to seasonal factors such as temperature and rare or transient OTUs displaying significantly less synchrony.

3.3 Results and Discussion

3.3.1 Ecosystem function and operating conditions

Operating data, environmental conditions, and effluent quality for all six reactors are shown in Table 3.1. Influent flow rate varied from 9 million gallons per day (MGD) at Hanover to 219 MGD at O'Brien. Influent composition, nutrient concentrations, and effluent quality were similar in all four plants; however average solids retention time (SRT) varied from 6.4 days in Egan North to 17.5 days in Hanover 1. Average sludge volume index, a measure of settleability and indicator for filamentous bulking problems, varied from 76 ml/g in Kirie to 206 ml/g in Hanover 1. Monthly summaries of environmental and operating parameters are shown in Supplementary Table S3.1. Water temperature varied seasonally from a low of 13° C in March 2015 to 20° C in October 2014. Influent Total Kjeldahl Nitrogen varied seasonally with a peak

concentration of 32.8 mgN/L in November and a minimum concentration of 25.9 mgN/L in May, but nitrification efficiency remained above 94% throughout the year. Average SVI was highest in February, March and June when one or more plants experienced bulking events.

3.3.2 Taxa showed consistent seasonal dynamics across plants

Despite relatively stable performance throughout the year, core OTU populations in different reactors were highly dynamic and regionally synchronized. OTU abundances were highly skewed, with a small number of OTUs constituting most of the sequences recovered (Figure 3.1) A 'core' community of 134 OTUs was found in all reactors at all time points. These core OTUs made up 51% of the total reads. A total of 599 OTUs were constitutively present in one or more plants but not all, suggesting that individual plants harbored some stable differences in composition throughout the year.

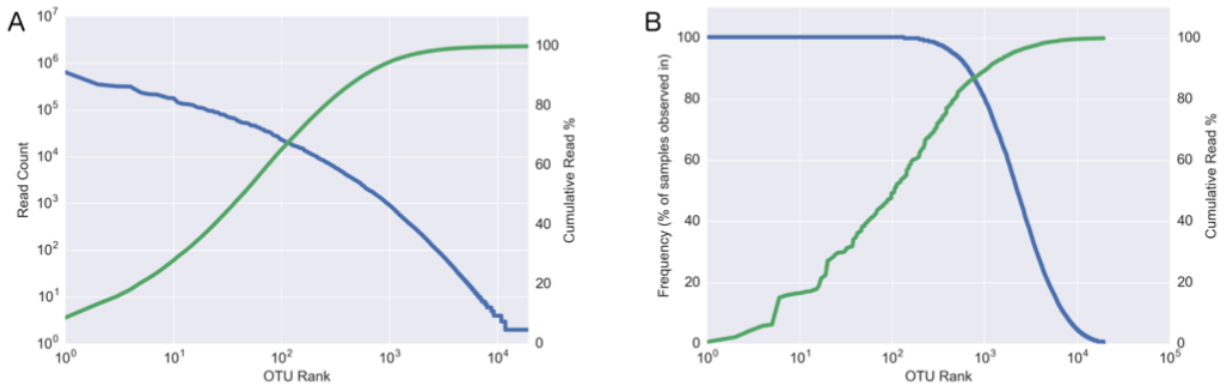

Figure 3.1 Rank abundance curves sorted by OTU raw abundance (A) and OTU frequency (B). 134 OTUs were core to every time point, and accounted for over 51% of total reads. Similarly, the top 100 OTUs by absolute read abundance accounted for over 60% of all reads.

Figure 3.2A shows synchrony coefficients for core OTUs compared to a bootstrap distribution of randomly selected OTU populations. Core OTUs were much more synchronized

than randomly compared taxa (Student's T-test = 13.1, $p < 0.001$), consistent with niche driven selection influenced by regional environmental factors. Non-core OTUs present in at least 20% of samples displayed enriched synchrony as well (Student's T-test = 31, $p < 0.001$). In addition, there was a significant correlation between synchrony and OTU occurrence frequency (Pearson $R = .55$, $p < 0.001$) and average rank abundance (Spearman $R = -.61$, $p < 0.001$) (Figure 3.2B & C). Common OTUs displayed synchronized changes in abundance between reactors whereas rare OTUs tended to be uncorrelated. The apparent discrepancy between abundant and rare OTU synchrony suggests that seasonal effects may have a larger influence on shaping abundant OTU dynamics whereas operating conditions or neutral factors may play a larger role in shaping the rare microbiome of full-scale activated sludge bioreactors.

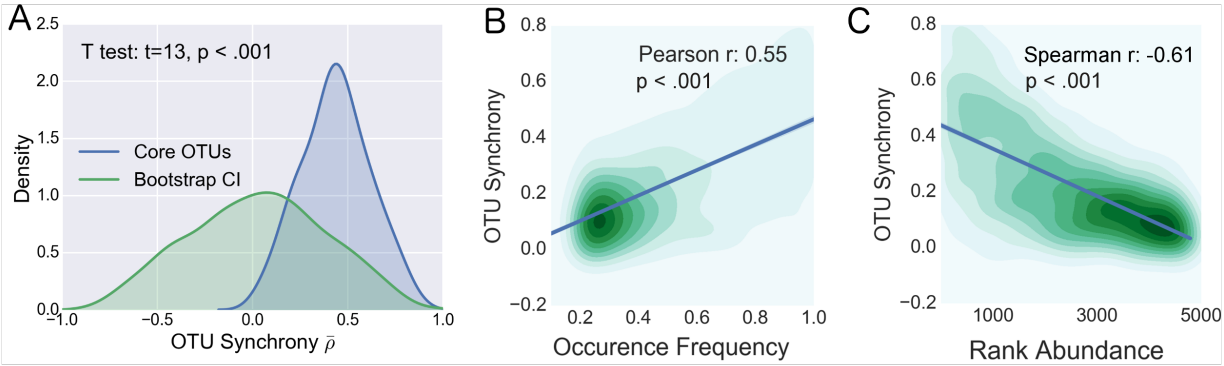

Figure 3.2 Regional OTU synchrony in activated sludge communities. (**a**) Density plot of regional synchrony for 'core' OTUs (present in all samples) vs a bootstrap sample of randomly selected OTU time-series pairs. Density plots of synchrony as a function of observed frequency and average abundance are shown in **b** and **c**. Most OTUs were relatively rare and uncorrelated between plants, but a small number of frequent OTUs were highly synchronized.

We further examined whether populations linked by plant specific factors (e.g. nutrient load and operational set points) were more synchronized than those that shared regional

environmental variables only (e.g. water temperature, precipitation). Mean synchrony coefficients for core OTU populations in the two pairs of bioreactors sharing influent, Egan South and North and Hanover 1 and 7, were compared to average population synchrony values across all fifteen reactor pairs. Egan populations were the most synchronized (average $\rho = .62$ for Egan compared to .44 across all reactors). Hanover populations were less synchronized (average $\rho = .42$) than the average across all regional reactors, likely owing to environmental differences imposed by the anoxic selector present in Hanover 1.

Next we investigated whether synchronous OTU dynamics gave rise to similar shifts in broader phylogenetic groups. Class level abundances averaged by season and plant are shown in Figure 3.3. Classes with significantly different abundances in winter (December-February) and summer (June-August) are indicated by p-values in the legend. Relative standard deviations of class abundances within weeks are shown in Supplementary Table 3.2. Dominant taxa were consistent between plants, and within-week variation between reactors was low for most classes. However, class abundances varied throughout the year, in agreement with high rates of temporal OTU synchrony. *Betaproteobacteria* were the most abundant class throughout the year, followed by *Saprospirae*, *Delta-*, *Alpha-* and *Gammaproteobacteria*. *Betaproteobacteria* were most dominant during winter, making up on average 44% of the total reads, but were significantly less abundant (Student's T-test = 22.1, $p < .001$) and accounted for only 27% of total reads in summer. Student's T-tests comparing class abundances in samples collected in summer and winter revealed that 11 of the 15 most abundant classes were significantly more abundant during summer than winter after Bonferroni Correction (Student's T-test, all $p < .0001$). Among these, *Nitrospira* were on average 84% more abundant in summer than winter samples.

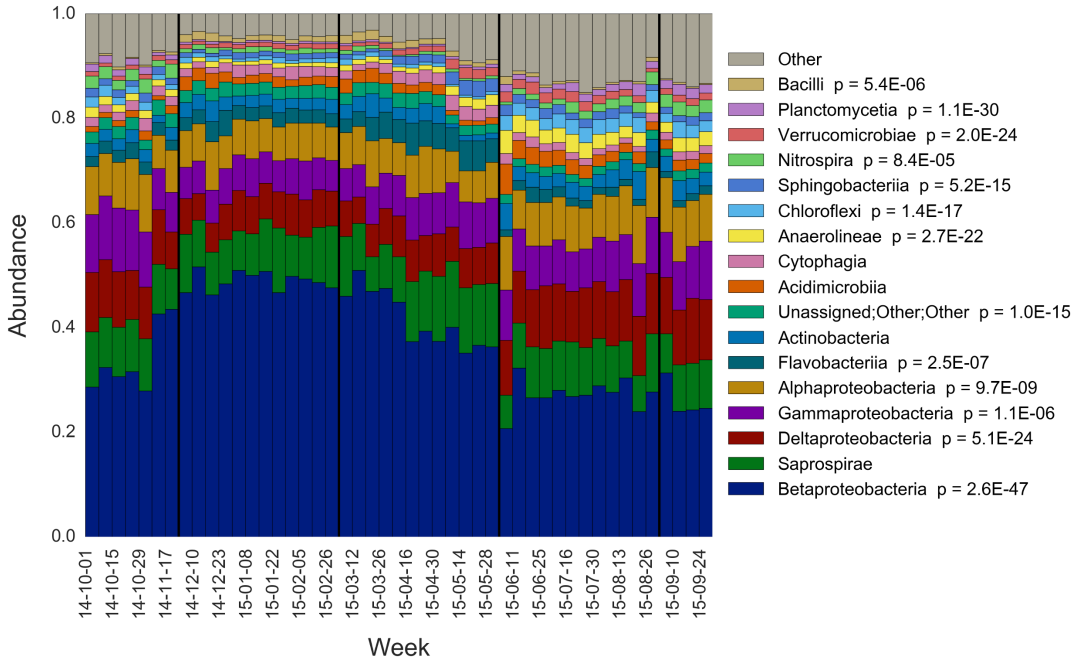

Figure 3.3 Average abundance of the top 15 most abundant bacterial classes at each time point. Sampling date is shown on the *x* axis. Black lines represent cutoffs for different seasons starting from Fall 2014. *Betaproteobacteria* were the most dominant class at all time points, but increased in relative abundance between winter and summer. *P*-values for classes with significantly different abundance in winter and summer are shown in the legend.

3.3.3 Alpha diversity is driven by seasonal temperature fluctuations

Population synchrony and shifts in broad phylogenetic abundance were accompanied by repeatable and putatively seasonal alpha diversity patterns. Faith's Phylogenetic Diversity and Shannon Diversity were used to quantify richness and evenness and phylogenetic diversity throughout the sampling period (Figure 3.4). One-way ANOVA revealed statistically significant differences between seasons for both phylogenetic (R_{ANOVA} = 93.23, P < 0.001) and Shannon (R_{ANOVA} = 42.84, P < 0.001) diversity. For both metrics, diversity peaked in fall between September and November and reached a minimum in all reactors in March or December. Diversity in all plants gradually increased during spring and summer, such that there was no statistically significant difference in Shannon diversity (Student's T-test, T = .98, P = .33)

between September 2015 and October 2014 samples. Pairwise Pearson correlation coefficients of alpha diversity between plants confirmed synchronized changes in community diversity over time based on both Faith's Phylogenetic Diversity (mean R = .65± .11, all p < .001) and Shannon Diversity (mean R = .54 ± .16, all p < .001). In addition to reproducible and synchronized seasonal variation in alpha diversity, ANOVA revealed that individual reactors harbored stable differences in diversity throughout the year (R_{ANOVA} = 11.4, P <0.001). Kirie was the most diverse community in 29 time points (60%), and Egan South and O'Brien were the least diverse in 31% and 39% of all samples, respectively.

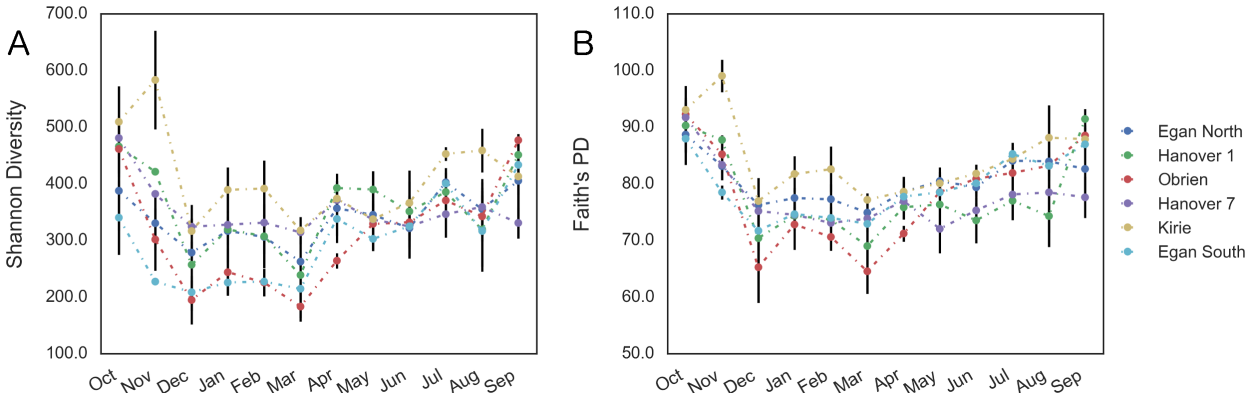

Figure 3.4 Time series of (**a**) Shannon and (**b**) Faith's PD alpha diversity for all six reactors. Data shown are average of all samples taken during each month. Alpha diversity was highest in October and November, and lowest in December and March.

Multiple linear regression was used to identify the unique fraction of variation in Shannon Diversity that could be linked to environmental and operational factors (Supplementary Table S3.3). Regressors included regional environmental gradients (influent water temperature, chloride concentration and precipitation) that varied seasonally, local environmental conditions (BOD, N, P influent concentration) and plant location, operational parameters (SVI, SRT, and

MLSS) and performance indicators (effluent nitrite concentration, BOD, N and P removal). The relative importance of each variable was assessed by averaging sequential sum squared error over all orderings of regressors to account for potential colinearity between regressors. Together, these factors explained 49% of the variation in Shannon diversity.

Plant identity and temperature were the most important predictors of Shannon diversity and explained 25% and 12% of the variation respectively. One caveat of interpreting relative importance of regressors is that predictors with higher variance tend to have inflated effect sizes[144]. Despite this, temperature is an important driver of community structure in many microbial ecosystems and is likely to be important in nutrient and oxygen rich environments like activated sludge. In addition, increased SVI, which is typically associated with enrichment of filamentous bacteria, was linked to decreased Shannon diversity. Operational parameters that could be most easily controlled, solids retention time and suspended solids concentration, accounted for less than 4% of total variation.

3.3.4 Community similarity was controlled by region-wide factors

Principal Coordinates Analysis (PCoA) of weighted unifrac community distance revealed seasonal clustering of communities (Figure 3.5). Mantel correlograms (Figure 3.6) comparing weighted unifrac distance and time lag between samples were created to assess periodic changes in community structure, as well as conserved differences between reactors. Dissimilarity was plotted as a function of time for samples originating from the same reactor, samples taken from separate reactors in the same plant with similar influent, and samples taken from different plants. At low temporal lag, communities from the same reactor were more autocorrelated than communities from other reactors or plants. As time lag between samples began to increase, average dissimilarity increased for all groups of reactor pairs. However, the

beta diversity difference between intra- and inter- reactor sample pairs diminished. Average dissimilarity peaked at a lag of 28 weeks. The mean dissimilarity between samples taken 6 and 51 weeks apart was comparable, indicating that both continual succession as well as seasonal effects controlled overall community dynamics.

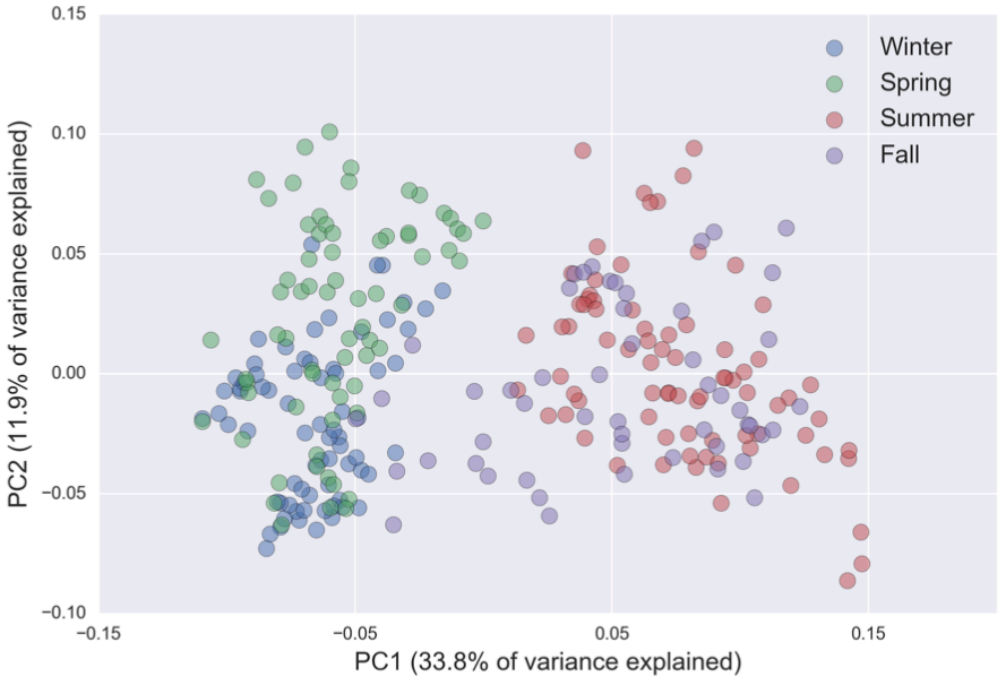

Figure 3.5 Principal Coordinate Analysis of all samples, colored by season of origin. Samples roughly grouped by season with winter and spring together and summer and fall grouped.

On average, samples taken from different reactors within the same plant were more similar to each other than those taken from different plants at every time point, suggesting that influent composition or plant-to-plant difference in operating set points also influenced community structure. Mantel tests were used to calculate the correlation between weighted unifrac distance and temporal distance. The total time between samples and the day number lag between samples (182 days for samples half a year apart and 0 for samples exactly a year apart) were calculated for each pair of samples. The day number lag was more correlated with

differences in community structure between all communities ($R_{pearson}$= .47, P < 0.001) than absolute time difference (R = .36, P < 0.001).

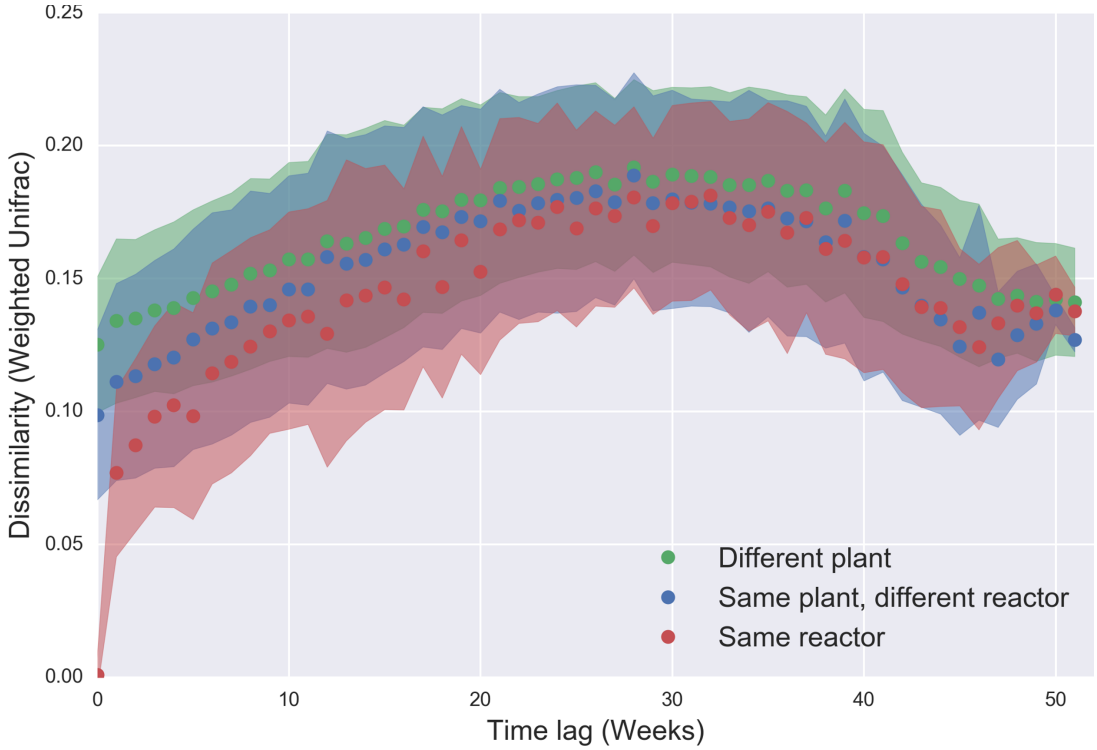

Figure 3.6 Temporal correlograms of time lag between samples versus dissimilarity (weighted unifrac distance) showing the mean and standard deviation (shaded area) for each week. Colors represent whether the compared samples originate from the same reactor (red circles), different reactors at the same plant (blue squares) or different plants (green triangles). The day of year lag between samples correlated well with weighted unifrac distance ($R_{pearson}$= 0.47, P<0.001).

To better visualize seasonal community structure, weighted unifrac trajectories for each reactor were plotted individually, with points colored by season of origin and connected over time by lines (Figure 3.7). In general, samples from the same season grouped together, with communities tending to be highly similar to one another in winter and summer, and transition between these clusters in fall and winter. The overall placement of the seasonal clusters along the principal coordinates was similar across reactors. In addition, a strong annual cycle was observed

for all of the reactors studied, and samples taken in fall of 2015 closely resembled those taken from the fall of 2014. PERMANOVA was used to test the significance of bacterial community groupings by reactor and season. Grouping by both season (Pseudo-$F_{PERMANOVA}$ = 49.15, P < 0.001) and reactor (Pseudo-$F_{PERMANOVA}$ = 11.61, P < 0.001) showed significantly different community structure.

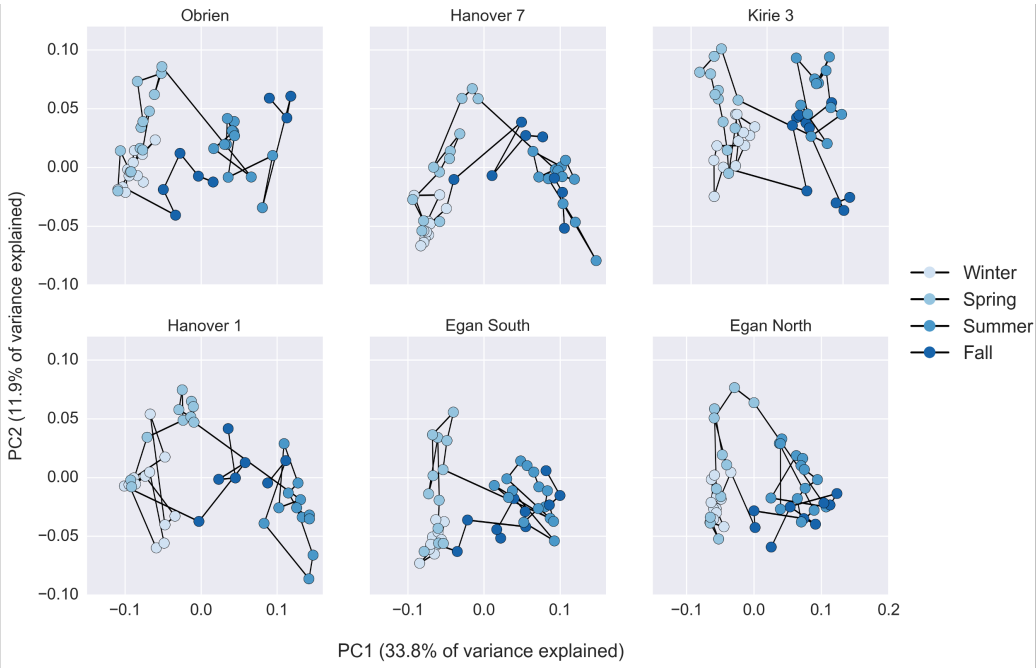

Figure 3.7 Principal coordinate analysis of weighted unifrac distances between samples. Colors represent season of origin and lines connect samples taken on consecutive sampling trips.

Using partial redundancy analysis (partial RDA) the beta diversity that could be uniquely explained by regional, local, operation and performance regressors was calculated (Supplementary Table S3.4). Weighted unifrac distances were constrained by each set of explanatory variables after partialing out the other sets of variables. Variance that was explained by one or more set of variables is listed as covariation. Together, 52% of total variation was

explained. Similar to alpha diversity, temperature was the largest individual driver of community dissimilarity and accounted for above 10% of the total variation in beta diversity. Over half of the total explained variance (28%) was controlled by covariation between one more variables. The largest portions of covariation were between temperature, local and performance factors due to seasonal fluctuations in influent nutrient load and effluent quality (Supplementary Table S3.4). Taken together with the PERMANOVA results, reactor communities in the system were highly controlled by deterministic factors.

Finally, direct sludge transfers between plants were used to quantify the impact of mass effects on community similarity. During the sampling period, waste activated sludge was transferred between Kirie, Egan North and Egan South on six occasions to in an attempt to improve settling during SVI upsets. We compared weighted unifrac distance between source and sink communities collected the week before and after transfers to find evidence for colonization of the sink community by the source. Transfer sludge source and sink reactor, mass, and pre- and post-transfer weighted unifrac distance are shown in Table 3.1. No significant change in similarity in source and sink communities was evident before and after reseeding, (Student's T-test = .39, P = .71) suggesting that sink communities were resistant to colonization or invasion at tested levels of immigration or effects were too transient to see with our sampling strategy.

Table 3.1 Impact of mass effects on community WU distance

Date	Source	Sink	Transfer mass %[1]	Pre-transfer WU	Post-transfer WU
12/15/14	Egan_North	Egan_South	5.34	0.086	0.073
12/23/14	Egan_North	Egan_South	4.19	0.073	0.079
1/8/15	Kirie	Egan_South	10.82	0.108	0.118
2/5/15	Kirie	Egan_North	12.23	0.117	0.104

3/4/15	Kirie	Egan_North	8.47	0.104	0.102
3/10/15	Kirie	Egan_North	9.17	0.102	0.091

[1] As a percent of total solids load in the sink reactor

Student T-Test of pre vs. post transfer: t=.39 p=.71

3.3.5 Seasonal changes in ecosystem function

Lastly, we examined whether synchronous population shifts could explain seasonal differences in nitrification performance and bulking event frequency observed in these plants. Operating data showed that several plants experienced seasonal increases in effluent nitrite residual during winter between 2010 and 2014 (Figure 3.8). Nitrification is traditionally thought to be a two-step process in which ammonia oxidizing bacteria (AOB) such as *Nitrosomonas* convert ammonia to nitrite and nitrite oxidizing bacteria (NOB) such as *Nitrospira* further oxidize nitrite to nitrate[145]. We found that *Nitrospira* abundance decreased by nearly 70% from October to April (~2.3% to .75%) (Figure 3.9A) before rebounding in the summer. Hanover 1 was the only plant that did not experience a decrease in *Nitrospira*, possibly due to higher SRT preventing washout. In contrast, the dominant genus of AOB in each reactor, *Nitrosomonas,* showed little seasonal variability (Figure 3.9B).

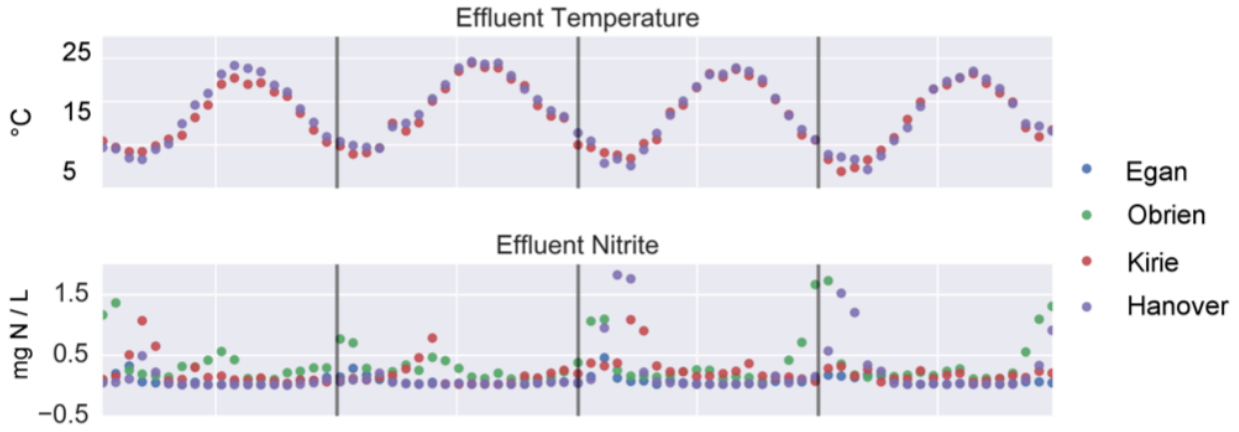

Figure 3.8 Seasonal temperature (top) and effluent nitrite concentration (bottom) over a four-year period showing seasonal effluent nitrite residual in multiple plants.

During our time-series monitoring, several bulking events lasting longer than a week were investigated. Bulking sludge is a common process disturbance caused by blooms of filamentous or bulking bacteria such as *Microthrix* or foaming *Actinomycetes.*[146] During the startup phase of MLE operation, Hanover 1 had a 3 day average SVI > 150 for a period of 8 months from December 2014 – August 2015. Hanover 7 had several instances of elevated SVI in November 2014, and May, June and July 2015. Egan North had a bulking event from February – Mid March 2015 and again from May – June 2015. Based on bulking and filamentous bacteria (BFB) previously identified using 16S rRNA amplicon-based sequencing[147] we identified 11 genera putatively linked to poor settling characteristics with abundance >1% in at least 1 sample, including *Microthrix*, *Thiothrix*, *Caldilinea, Trichococcus*, *Rhodococcus*, *Haliscomenobacter*, *Gordonia*, *Kouleothrix*, *Mycobacterium*, *Tetrasphaera*, and *Isosphaera*.

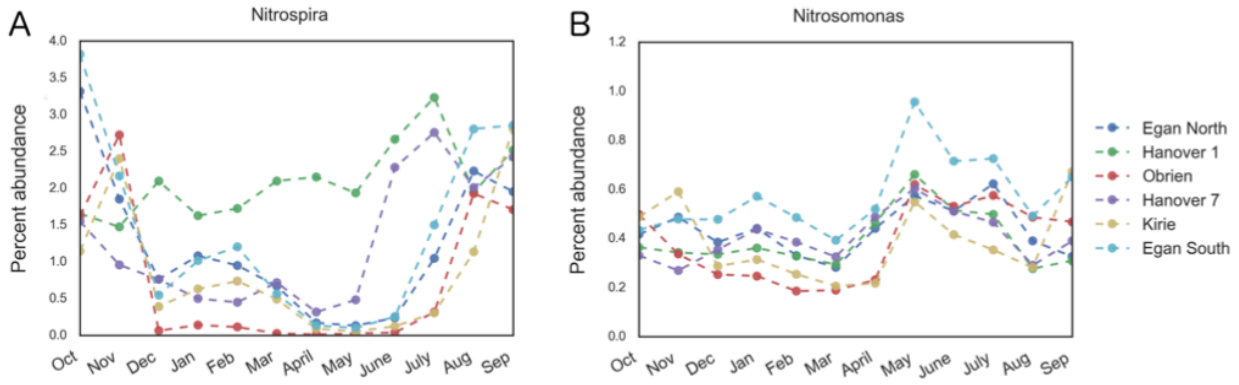

Figure 3.9 Seasonal variability in Nitrospira and Nitrosomonas at genus level. All but one reactor (Hanover 1) exhibited strong seasonal variability in Nitrospira abundance and relatively stable Nitrosomonas abundance.

The most common taxa in terms of average abundance were *Thiothrix* (2.2%), *Kouleothrix* (1.3%), *Caldilinea* (1.2%), and *Microthrix* (.8%); however, the BFB were highly dynamic (Figure 3.10). Of the observed bulking bacteria, *Caldilinea, Mycobacterium, Haliscomenobacter & Gordonia* primarily displayed seasonal variation and synchrony between plants but did not correlate with SVI. Elevated *Microthrix* abundance was linked to increased SVI ($R_{pearson}$ = .44, p <0.001), however. *Microthrix (S7A)* and *Thiothrix* (S7B) were inversely correlated with one another ($R_{pearson}$ = -.37, p <0.001), Microthrix tended to dominate in reactors with elevated SVI whereas *Thiothrix* were more abundant in the three reactors that did not experience bulking events. In general, bulking bacteria tended to be more diverse in fall and dominated by a few taxa during the winter and spring.

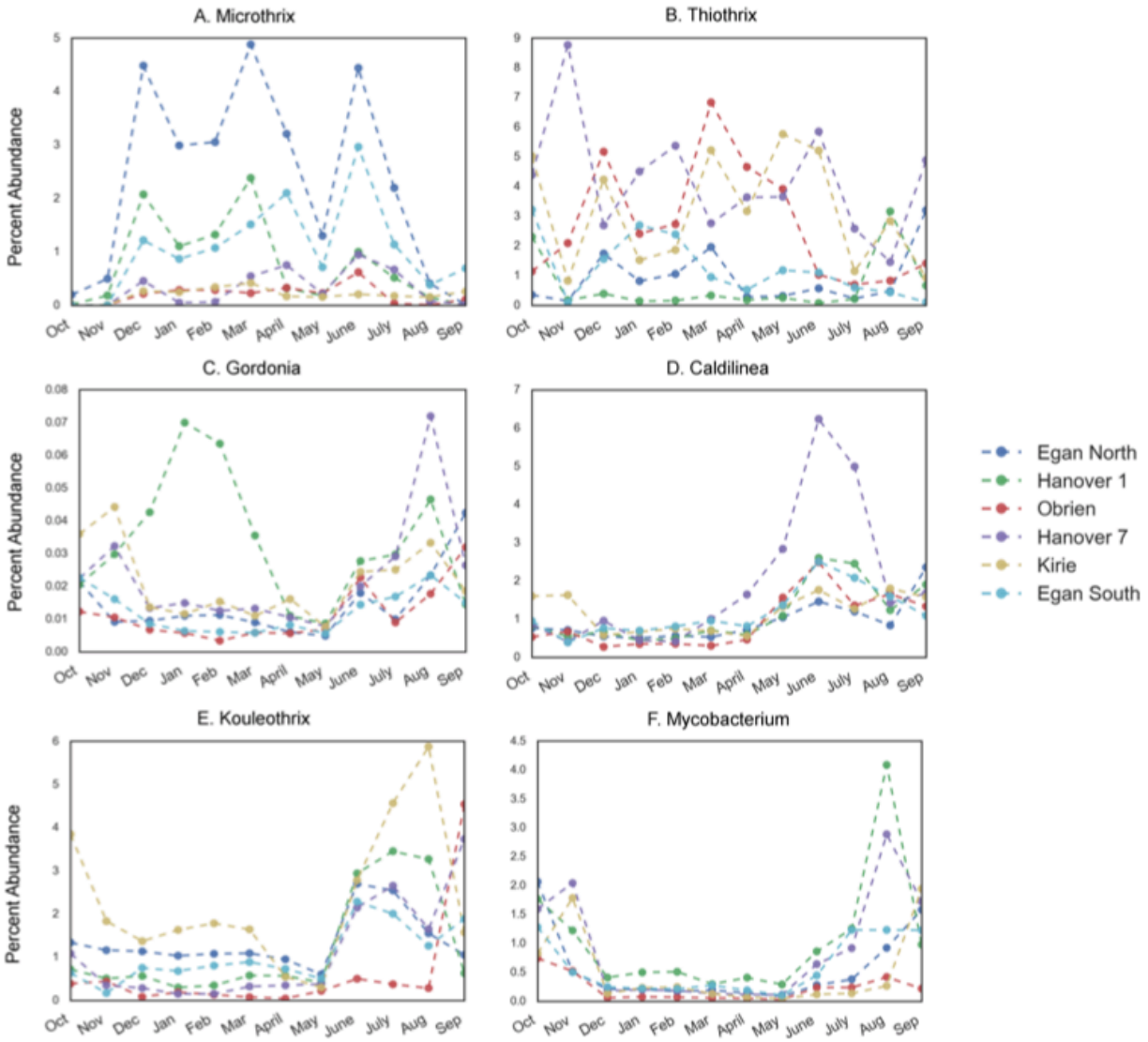

Figure 3.10 Seasonal variation in the top 6 most abundant bulking and filamentous genera. Microthrix (A) and Thiothrix (B) were the two most abundant genera and Microthrix was present during several bulking episodes. Other genera displayed synchronized variation.

3.3.6 Dominant taxa were regionally synchronized

In this study, community dynamics in regional activated sludge bioreactors were monitored to identify the relative contribution of different assembly mechanisms. Our results show that OTU populations are highly regionally synchronized between reactors (Figure 3.2) across several scales. Synchrony in macroscopic ecological systems is explained by deterministic

assembly mechanisms such as spatially correlated environmental conditions[148] and trophic interactions such as predation[149]. In microbial communities, synchrony has been observed in rivers[45] and dispersal limited lakes[48], but this effect has not previously been observed in full-scale activated sludge bioreactors.

The observation of synchrony in dispersal limited reactors combined with the strong effect of temperature on community structure (Supplementary Table S3.4) supports regional temperature fluctuations as a primary driver of community assembly. In addition to temperature, seasonal nutrient variation is typically a driver of regional synchrony in aquatic microbial ecosystems,[47, 150] but we found that influent nutrient concentrations in the studied WWTPs were stable over the course of the year and contributed little to observed variation. While all reactor pairs showed significant synchrony in microbial community dynamics, this effect was highest for core OTU populations in the two Egan bioreactors that shared the same influent as a potential immigration source. Interestingly, population dynamics in the Hanover bioreactors that also shared the same influent were less similar than between plants, likely due to environmental differences caused by the anoxic selector in Hanover 1. Previous studies investigating the impact of immigration from raw influent on activated sludge community structure have found highly variable estimates for the fraction of OTUs shared between influent and activated sludge, from 5-10% OTU overlap[151] to 35%[152]. Communities from linked processes within a single WWTP also show evidence for within-plant immigration[153] further suggesting that both selection and mass effect mechanisms influence community assembly. Although we did not find strong evidence that local immigration affected synchrony, it is possible that synchrony could be explained in part by region-wide changes in influent community structure, in addition to direct effects on the activated sludge community.

Despite high phylogenetic diversity and stable nutrient removal performance, functionally redundant taxa did not appear to randomly dominate different bioreactors. Instead, individual taxa displayed synchronized patterns across multiple plants. High rates of synchrony have previously been observed in the startup of replicate lab-scale anaerobic digesters[8] but it was unclear whether similar assembly mechanisms would drive dynamics in full-scale open engineered environments. A high density of microbial interactions has previously been reported in activated sludge.[154] In principle, these ecological interactions may compound seasonal environmental changes and increase synchrony.

Few studies have quantified differences in synchrony between rare and abundant taxa in activated sludge or other environments. Population synchrony was strongest for core and highly abundant OTUs and was positively correlated with OTU frequency (Figure 3.2). By analyzing synchrony along a continuous gradient of average OTU abundance and observed frequency, our results show that community assembly mechanisms differ along a continuum between "core" and "rare" taxa. Previously, rare taxa have demonstrated higher variability and turnover rates compared to general taxa within activated sludge.[9] In another study, 10% of the total reads found in WWTPs came from OTUs with negative growth rates in activated sludge that are likely inactive or slow-growing.[152] Rare OTUs may be more impacted by dispersal or immigration from influent sources rather than environmental pressure within the reactors.

One limitation of our approach to quantify synchrony and dynamics is the use of 16S rRNA gene abundance, which does not directly inform community functional capabilities. Further, 16S rRNA amplicon and shotgun metagenomic studies do not always yield similar estimates of community diversity and dynamics.[155] However, amplicon sequencing enables higher temporal coverage of communities due to its lower cost. Phylogenetically related

organisms tend to be functionally similar[156] and this has been used to predict functional capabilities for communities from phylogenetic studies.[157] A combined metagenomic and 16S rRNA gene sequencing study of activated sludge revealed that community function was less seasonally variable than phylogenetic abundance.[43]

3.3.7 Diversity and class abundance varied with temperature

Phylogenetic (Faith) and non-phylogenetic (Shannon) measures of alpha diversity were highly synchronized between reactors (Figure 3.4). Notably, in all six reactors surveyed, diversity followed a cyclical trend, dropping sharply near the end of the year and increasing again in the spring. Averaging over orderings of regressors revealed that temperature was the single most important predictor of diversity (Supplementary Table S3.3). Measured environmental and operational parameters explained nearly twice as much variation as plant location. Kim et al. demonstrated a similar decrease in alpha diversity in a single activated sludge bioreactor in winter[9], although the opposite effect was found in marine systems[44].

Diversity has been shown to be critical for productivity, resilience to disturbance and functional stability in so-called "biodiversity ecosystem function relationships".[158] We found that nine of the top fifteen classes reproducibly decreased in abundance across all reactors surveyed while dominance by *Betaproteobacteria* increased during winter (Figure 3.3). These broad-scale shifts in diversity were accompanied by a decrease in nitrifier populations and an increase in bulking and filamentous bacteria (BFB). Shannon diversity was positively correlated with removal efficiency of BOD, nitrogen, and phosphorous, but negatively correlated with sludge volume index (SVI), suggesting that a positive biodiversity ecosystem function relationship exists for treatment communities. Elevated SVI is a process disturbance caused by blooms of BFB such as *Microthrix* or foaming *Actinomycetes.*[159] Decreased alpha diversity is symptomatic

of bulking, as BFB accounted for more than 10% of all sequencing reads during some of the bulking episodes. It may also indirectly lead to susceptibility to bulking by opening niches for BFB, which typically bloom under low nutrient availability, temperature and dissolved oxygen conditions.[160]

Nitrospira- a key nitrite-oxidizing bacteria- displayed much higher abundance in summer than in winter in all activated sludge bioreactors. Seasonal decline of *Nitrospira* may be related to increased effluent nitrite observed across several previous winters in the plants we surveyed (Figure 3.8), although effluent nitrite accumulation was not apparent during the time series described here. Nitrifier growth rates are strongly influenced by temperature and higher nitrite residuals have been observed in winter in nitrifying activated sludge bioreactors in previous studies.[161]

3.3.8 Community similarity was driven by temperature and season

In addition to cyclical alpha diversity, community similarity was temporally concordant. Mantel Correlograms (Figure 3.6) revealed a cyclical trend in weighted unifrac distance between samples. Strikingly, beta diversity peaked just over half a year apart, and communities sampled one year apart were as similar as those sampled 6 weeks apart on average. Differences between plants were evident at low temporal distance but disappeared in samples from different seasons. These differences reemerged for samples taken a year apart, and it is possible that plant-wide idiosyncrasies such as influent composition or distinct operational conditions (e.g. SRT) lead to consistent year after year differences in population structure. More longitudinal studies are necessary to test this hypothesis. Constrained multivariate analyses revealed that, similar to alpha diversity, temporal variation in community structure was primarily driven by changes in temperature (Supplementary Table S3.4).

Seasonal community succession appears to be a general feature of aquatic microbial communities [141]. It has been shown across multiple years in marine[162] and freshwater environments.[163] However, in the limited studies available continual succession has been more frequently reported for activated sludge systems.[164] The authors speculated that due to the controlled nature of WWTPs, intrinsic factors play a stronger role than extrinsic factors in shaping communities. Contrary to previous studies of activated sludge communities, continual drift accounted for only a minor portion of variation in community composition in our study. It should be noted, however, that deconvoluting short-term fluctuations and long-term seasonal variation is not possible with a single year of data, and more work is needed to identify whether synchronous OTU patterns repeat in following years, similar to multi-year patterns observed in other environments.[165]

One mechanism that could explain observed regional synchrony and seasonal variability in WWTPs is large temperature fluctuations combined with biomass wasting. Because solids are continuously wasted during operation, bacteria must maintain a growth rate greater than the inverse of the SRT to be maintained to stay within the reactor. Lower temperatures reduce prokaryotic growth rates and may wash out slow growing microbes at low solids retention times (SRT). Cold weather plants can operate at SRTs as high as 30 days[152] and it is possible that the relatively short SRTs in our reactors (8.0-13.9 days on average) contributed to the strong seasonal taxonomic shifts.

3.4 CONCLUSION

Regionally synchronized patterns of biogeography have been observed in a variety of microbial ecosystems, but the importance of these processes in highly managed ecosystems such

as activated sludge is not well known. Our primary objective in this study was to quantify microbial population synchrony as a means of clarifying the relative importance of deterministic seasonal environmental factors and stochastic processes on bacterial community succession in activated sludge. A clear seasonal pattern of microbial community structure was evident, from community-wide diversity metrics to dynamics of dominant OTUs between plants, and temperature was the primary driver of changes in alpha and beta diversity. In contrast, neutral factors such as immigration from continuous shared influent and intentional waste activated sludge reseeding attempts did not significantly alter beta diversity between communities. Core OTUs present in every sample were highly abundant and strongly synchronized between reactors, while less common OTUs tended to fluctuate more randomly, suggesting they were driven by local differences in operating conditions or stochastic processes. Key functional groups such as nitrite oxidizing bacteria displayed repeatable seasonal differences in fully aerated systems. Our results demonstrate the importance of seasonal variability on microbial consortia and the influence of deterministic community assembly mechanisms in wastewater treatment bioreactors.

3.5 METHODS

3.5.1 Plants, sampling, and sequencing

Activated sludge biomass samples were collected between October 2014 and September 2015. Six activated sludge bioreactors at four full-scale wastewater treatment plants (Hanover Park, Egan, Kirie and O'Brien Water Reclamation Plants) were sampled weekly for one year except during plant shutdowns. All but one of the reactors operated as fully aerated activated sludge systems performing BOD removal and nitrification. The Hanover 1 reactor operated with a Modified Ludzack-Ettinger process with an anoxic zone for denitrification, followed by an

aerobic zone for nitrification. Two independent reactors each at Hanover Park and Egan with separate return activated sludge systems and single reactors at Kirie and O'Brien were sampled, yielding a total of six independent activated sludge reactors sampled. Operational and environmental parameters were monitored according to standard methods[166] (Table S3.1). Plants were located in the USA in Chicago, Illinois suburbs with Hanover Park in Hanover Park, Egan in Schaumburg, Kirie in Des Plaines, and O'Brien in Skokie, Illinois.

For each weekly sample, 50ml mixed liquor grab samples were collected from near the inlet and outlet of the reactor and combined as a single time point. 1.5ml aliquots of activated sludge collected from the inlet and outlet of the reactor were centrifuged at 10,000g to pellet biomass, rinsed twice in 1ml of TAE buffer, and extracted according to the MP Bio Soil DNA extraction kit protocol (MP Biomedicals, Santa Ana CA). Amplification was performed in two steps using the Fluidigm Biomark: Multiplex PCR strategy. First, the hypervariable V4 16s region of the rRNA gene was amplified using the Earth Microbiome Project primer set with forward primer CS1-515f (5'- GTGCCAGCMGCCGCGGTAA) and reverse primer CS2-806r (5'-GGACTACHVGGGTWTCTAAT). Two 20 μL independent PCR reactions were performed per DNA extract, using Epicentre Premix F (Epicentre, Madison WI), Expand Hi-Fidelity Taq (Roche Diagnostics, Indianapolis IN), 200 nM primer, and 100 ng of genomic DNA in a Biorad T100 thermal cycler (Bio-Rad, Hercules CA) at 95 °C for 5 minutes followed by 28 cycles of: 95 °C (30 s), 55 °C (45 s), and 68 °C (30 s) and a final elongation step at 68°C for 7 minutes.

Amplicons from replicate PCRs were pooled for the second round of PCR and labeled with barcodes unique for each sample using Accuprime Supermix (ThermoFisher, Carslbad CA), 50 μM forward and reverse primers (Fluidigm, South San Francisco CA), and 1 μL of template from the first round of PCR at 95 °C for 5 minutes followed by 8 cycles of of: 95 °C (30 s), 60°C

(30 s), and 68 °C (30 s). To minimize over-amplification, PCR products were checked via UV-Vis and gel electrophoresis for smearing of the bands. The total cycle count was kept to 36 cycles per the EMP guidelines. The resulting amplicons were processed with a Qiagen PCR purification kit and sequenced at the University of Illinois Chicago DNA Services Facility. DNA sequencing was performed using a Miseq V2 sequencer (Illumina, San Diego CA) at the University of Illinois Chicago DNA Services Facility. Sequences can be accessed on Genbank (PRJNA317773).

3.5.2 Amplicon Sequence Processing and Quality Control

Paired end Illumina V4 16S rRNA sequences were processed using Vsearch 1.9.1.[167] After merging paired end reads, sequences with more than one expected error, longer than 300 base pairs, or with any unknown nucleotides were discarded. Singletons and likely chimeras were also discarded using default settings in Vsearch. Samples with fewer than 5,000 reads (<.5% of mean) were discarded due to low coverage. Representative sequences from each OTU were aligned using the Greengenes imputed core reference align and PyNast implemented in Quantitative Insights Into Microbial Ecology (QIIME 1.9.0-20140227).[13, 168, 169] After filtering the alignment to remove gaps and hypervariable regions, a phylogenetic tree was built using FastTree.[170] Samples were rarefied to the minimum sequencing depth ten times, and relative abundances were averaged between rarefactions before diversity metrics were calculated. Of the 271 samples collected over 49 weeks, 22,551,652 unique sequences passed quality filtering, and were dereplicated and clustered into 19,171 operational taxonomic units (OTUs) at 97% similarity. After chimera filtering, 17,666,983 of the initial sequences (78%) were mapped back to OTUs.

Alpha and beta diversity metrics were calculated in QIIME using an OTU table rarefied to the lowest sequencing depth (11,542 sequences per sample). A phylogenetic alpha diversity metric, Faith's Phylogenetic Diversity (PD) which weights changes in observed taxa by their phylogenetic distance to the nearest neighbor, and Shannon diversity, ^{1}D, which measures community evenness as well as richness, were used to quantify ecosystem diversity over time in all six reactors. Weighted Unifrac was used to calculate beta diversity between samples.[99] Taxonomy was assigned to OTU representative sequences using uclust and the Greengenes sequence database.[171]

3.5.3 Statistical Analysis

All statistical analyses were performed in Python (2.7) using the skbio package (0.4.0) and R (3.2) using vegan (3.2). Significance of observed alpha diversity trends were assessed using Analysis of Variance (ANOVA). Then, multiple linear regression was used to identify factors that best explained changes in alpha diversity. Measured environmental and operational variables were normalized to unit variance. These were used alongside categorical variables for each plant as regressors to predict alpha diversity. Because the regressors were not independent of each other, relative importance was calculated by averaging sequential sum squared error over all orderings of regressors[172] in the R package, relaimpo[173] resulting in estimates of the variation uniquely explained by individual regressors.

Two distinct methods were used to evaluate the effect of persistent reactor specific differences as well as seasonal and continual change on community similarity (Beta diversity). Permutational analysis of variance (PERMANOVA) was used to evaluate the significance of categorical groupings based on month and season. Mantel tests were used to evaluate the correlation between absolute temporal distance and beta diversity as well as seasonal temporal

distance and beta diversity. Ordination methods (Principal Coordinate Analysis and Redundancy Analysis) were used to visualize community similarity and calculate the fraction of variation in community structure explained by environmental gradients or influent differences. The preceding statistical methods are described in more detail in the supplementary methods.

3.5.4 Regional Synchrony Calculations

Regional OTU synchrony, defined as synchronous changes in individual OTU population in separate communities, was calculated using Pearson correlation.[143] For each OTU, N, the Pearson rank correlation $\rho_{i,j}^{N}$ of its abundance time-series in each reactor pair, (i, j) was calculated. Regional synchrony values for each OTU were calculated by averaging $\rho_{i,j}^{N}$ across all reactor pairs. Due to compositional effects[174], and the autocorrelation of populations over time[143] correlation coefficients will not necessarily be zero for uncorrelated OTUs. To test whether OTU populations in different reactors were more synchronized with each other than expected for independent OTUs, synchrony values were compared to a bootstrap confidence interval obtained from sampling 10,000 random pairs of OTU time-series.

3.6 SUPPLEMENTARY TABLES

Supplementary Table 3.1 Average operating data for 6 bioreactors in this study

		Egan North	Egan South	Hanover 1	Hanover 7a	Kirie	Obrien
Plant and Reactor	Coordinates	42°01'11.4"N 88°02'14.7"W	" "	42°00'00.1"N 88°08'12.7"W	" "	42°01'17.6"N 87°56'13.7"W	42°01'12.1"N 87°42'51.0"W
	Flow rate (mgd)	23.8±4		9.3±3		30.7±12	218.8±37
	Solids Retention Time (days)	6.4±1.2	11.5±4.2	17.5±4.8	14.2±4.1	10.2±2.1	9.7±1.1

	Mixed Liquor Suspended Solids (mg/L)	2243±217	2284±190	2113±282	2346±210	2806±290	2183±155
	Sludge Volume Index (ml/g)	118±28	92±12	206±86	115±24	76±15	89±12
Influent	BOD5 (mg/L)	164.9± 28		146.8± 31.8		138.3± 29.4	126.3± 39.9
	Ammonia (mgN/L)	16.3±2.8		18.1±4.8		16.2±3.8	14.7±2.4
	Total Phosphorous (mgP/L)	5.8±0.9		4.6±1.1		4.2±0.8	3.9±1.1
Effluent	Suspended Solids (mg/L)	4.8±1.2		5.1±2.1		2.2±0.3	7.1±2.2
	BOD5 (mg/L)	2.2±0.3		5.5±3.4		2.7±0.8	9.9±5
	Ammonia (mgN/L)	0.1±0.1		0.7±0.8		0.4±0.2	1.1±0.8
	NO3- and NO2- (mgN/L)	15.9±2.3		Not avail.		9.2±3.1	9.7±1.2
	Total Phosphorous (mgP/L)	2.9±0.8		2.8±0.7		1±0.6	1.5±0.3

[a]Egan and **Hanover** plant-wide, influent and effluent values are the same for both sampled reactors.

Supplementary Table 3.2 Average class level variation within weeks

Phylum	Class	Average Coefficient of Variation[1]
p__Nitrospirae	c__Nitrospira	0.76
p__Actinobacteria	c__Actinobacteria	0.67
p__Firmicutes	c__Bacilli	0.65
p__Chloroflexi	c__Chloroflexi	0.60
p__Actinobacteria	c__Acidimicrobiia	0.51
p__Bacteroidetes	c__Flavobacteriia	0.35
p__Chloroflexi	c__Anaerolineae	0.35
p__Proteobacteria	c__Gammaproteobacteria	0.31
p__Planctomycetes	c__Planctomycetia	0.30
p__Bacteroidetes	c__Cytophagia	0.29

p__Bacteroidetes	c__[Saprospirae]	0.25
p__Verrucomicrobia	c__Verrucomicrobiae	0.24
p__Bacteroidetes	c__Sphingobacteriia	0.23
Unassigned	Other	0.20
p__Proteobacteria	c__Deltaproteobacteria	0.20
p__Proteobacteria	c__Alphaproteobacteria	0.16
p__Proteobacteria	c__Betaproteobacteria	0.14

[1] Coefficient of variation was calculated from mean and standard deviation across all reactors during each time point, then averaged across all weeks to identify clades with asynchronous abundance profiles.

Supplementary Table 3.3 Relative Importance of factors for Shannon Diversity

Predictors	**Percentage**	**Lower 95%**	**Upper 95%**	**Category**	**Correlation coefficient for 1 term model**
Influent chloride	1.44	0.53	4.24	Regional	-14.31
Precipitation	0.37	0.09	2.07	Regional	0.18
Water temperature	11.54	6.57	15.45	Regional	34.13
Influent BOD	0.44	0.36	1.56	Local	2.53
Influent NH4	0.64	0.41	2.43	Local	-1.14
Influent Phos	1.03	0.69	2.42	Local	-5.07
Hanover	3.88	1.24	6.15	Local	21.27
Egan	3.80	1.89	5.53	Local	-32.94
Kirie	9.38	5.00	12.91	Local	77.82
MLSS	1.14	0.76	2.06	Operational	10.23
SRT	1.48	0.54	3.53	Operational	-2.61
SVI	0.99	0.31	2.58	Operational	-8.49
BOD Removal %	4.34	1.34	7.75	Performance	-28.86
Effluent Nitrite Residual	2.45	1.09	4.40	Performance	-19.06
NH4 Removal %	5.18	2.78	7.68	Performance	-22.56
Phos. Removal %	1.22	0.86	2.08	Performance	-6.54
Total variance explained	**49.32**	**24.46**	**82.86**	Total	
Sum of relative importance by category					
Regional	13.79				
Local	19.16				
Operational	3.61				
Performance	13.20				

Supplementary Table 3.4 Partial RDA Variance Partitioning

RDA Variable	Value	Grouping
Influent chloride	0.38	Regional
Precipitation	0.19	Regional
Water temperature	10.14	Regional
Influent BOD	0.45	Local
Influent NH4	0.71	Local
Influent Phos.	0.60	Local
Hanover	1.51	Local
Egan	1.60	Local
Kirie	1.58	Local
MLSS	0.52	Operational
SRT	0.42	Operational
SVI	2.79	Operational
NH4 Removal %	0.30	Performance
Effluent nitrite residual	0.47	Performance
BOD Removal %	0.20	Performance
Phos. Removal %	0.39	Performance
Regional & Local	6.05	Covariation
Regional & Performance	5.58	Covariation
Regional & Operational	3.97	Covariation
Local & Performance	7.52	Covariation
Local & Operational	3.66	Covariation
Performance & Operational	0.79	Covariation
Total variance explained	**49.82**	
Variance explained by category		
Regional	10.70	
Local	6.46	
Operational	3.73	
Performance	1.37	
Covariation	27.56	

CHAPTER FOUR

Hydrology and Connectivity in an Intensively Managed Landscape

Material in this chapter is based on the published work:

Griffin, J. S., Lu, N., Sangwan, N., Li, A., Dsouza, M., Stumpf, A. J., ... & Gilbert, J. A. (2017). *FEMS microbiology ecology*, *93*(10), fix120.

4.1 ABSTRACT

Intensive management increases the rate of nutrient and particle transport within a basin, but the impact of these changes on microbial community assembly patterns at the basin scale is not yet understood. The objective of this study was to investigate how landscape connectivity and dispersal impacts microbial diversity in an agricultural watershed. We surveyed soil, sediment and water microbial communities along the Upper Sangamon River basin (USRB) in Illinois—a 3600 km^2 watershed strongly influenced by human activity, especially landscape modification and extensive fertilization for agriculture. We employed statistical and network analyses to reveal the microbial community structure and interactions across water, soil and sediment media. Using a Bayesian source tracking approach, we predicted microbial community connectivity within and between environments. We identified strong connectivity within environments (up to 85.4 ± 13.3% of sequences in downstream water samples sourced from upstream samples, and 44.7 ± 26.6% in soil and sediment samples), but negligible connectivity across environments, which indicates that microbial dispersal was successful within but not across environments. Species sorting based on media type and environmental parameters was the dominant driver of community dissimilarity. Finally, we constructed OTU association networks for each environment and identified a number of co-occurrence relationships that were shared between habitats, suggesting that these are likely to be ecologically significant.

4.2 INTRODUCTION

Microbes play key roles in global biogeochemical cycles, yet the factors controlling microbial biogeography remain poorly understood.[175, 176] Modern community assembly theories relate microbial diversity patterns to mechanisms such as selection, dispersal, drift and

mutation[177] whose importance varies across different scales and environments. The selection or "species sorting" perspective emphasizes the influence of environmental heterogeneity on community composition. Meta-analyses of global biodiversity have revealed clear species sorting due to environmental selection but also identified unexpected overlap between soil, sediment and freshwater communities.[32] Environmental selection occurring through periodic disturbances such as flooding, seasonal nutrient variation and biofilm erosion often leads to temporal shifts in community structure.[178] In contrast to strictly selection based community assembly views, initial colonization of surface freshwater and biofilm communities occurs through dispersal from upslope soil communities[179] followed by selection processes.[180] Even in established communities, hydrological connectivity can facilitate microbial dispersal through hyporheic exchange between surface and groundwater as well as headwater mixing at stream confluences.[181] Field experiments with model organisms such as *E. coli* have revealed that exchange between streams and underlying sediments can facilitate the dispersal of microbes, but few studies have been done to identify cross-environment dispersal on larger scales.[182]

Intensive agricultural practices alter nutrient and soil particle transport within watersheds, though the impact of these modifications on microbial dispersal is currently unknown. Intensive landscape management has altered the hydrology of the Midwest region of the United States due to altered runoff, high prevalence of underground tile drainage[183], and increased nitrogen loading from agriculture.[184] Tile drainage consists of underground pipes below agricultural fields that quickly drain excess water from soils. It accounts for approximately 80% of runoff in the Upper Sangamon River Basin (USRB) and can increase landscape connectivity between soil and groundwater and accelerate soil erosion.[185] Increased runoff combined with monocultural corn-soybean crop rotations can minimize heterogeneity between habitat patches. Decreased

environmental heterogeneity and faster hydrological transport of microbes can increase dispersal between patches and alter microbial biogeography. In addition to increasing dispersal, this enhanced "landscape connectivity" can reduce nutrient variability across a basin or watershed as well as decrease phylogenetic and functional diversity[186] resulting in lower community resistance and resilience.[187]

The impact of environmental gradients on microbial beta diversity is well documented. Within aquatic ecosystems (e.g., rivers and lakes) and semi-aquatic ecosystems (e.g., floodplains), nutrient loading, hydrology, land use type, and climate change have been shown to shape microbial communities.[179, 188, 189] Similarly, microbial biogeographic patterns in soil are influenced by land use type, pH, soil composition, and climatic conditions across broad geographical scales.[190-192] Microbial dispersal limitations have been reported less frequently[193], but distance decay patterns in the absence of clear environmental gradients suggests that transport as well as historical factors can lead to increased beta diversity.[194] Intermittently flooded river cut banks and floodplains with exposed sediments can facilitate dispersal between soil and aquatic environments but the importance of these areas impact on microbial dispersal and community similarity to surrounding environments is unknown.

Here, we characterize the surface terrestrial (i.e. soil and floodplain sediments) and aquatic microbial communities within the USRB to illustrate the importance of landscape connectivity and agricultural land use in structuring microbial communities. We define connectivity as the ease of dispersal between habitat patches.[195] We used 16S rRNA V4 amplicon sequencing to analyze the microbial communities of diverse environments in the basin, which is part of the Intensively-Managed Landscape Critical Zone Observatory (IML-CZO) in central Illinois. We hypothesized that on the basin level, bacterial communities would be

influenced by both environmental selection as well as dispersal via fluvial transport. We also hypothesized that hydrology-driven dispersal would connect bacterial communities down-river in all three environments. We compared community composition, diversity, and co-occurrence patterns to test these hypotheses. Further, we employed source tracking to quantify the downstream longitudinal connectivity and lateral connectivity between media types. Lastly, we identified environmental gradients linked with community dissimilarity and then compared the relative importance of distance and environmental factors on community dissimilarity within each environment.

4.3 Results and Discussion

4.3.1 Dominant Taxa in the USRB Watershed

The USRB microbiome survey yielded 14 million high-quality 16S rRNA V4 amplicon sequences representing 117,602 distinct OTUs (97% nucleotide identity) from 77 samples. Many taxa were present in multiple environments, and several were ubiquitous across all of the environments in this study, suggesting adaptation to a broad range of conditions or frequent connectivity between environments. We defined ubiquitous taxa as those found in at least 80% of the samples from each environment. We found that at the class level, *Actinobacteria*, *Alphaproteobacteria*, *Betaproteobacteria* and *Deltaproteobacteria* were ubiquitous. At the family level, *Comamonadaceae* and *Burkholderiales* were ubiquitous across all three environments studied.

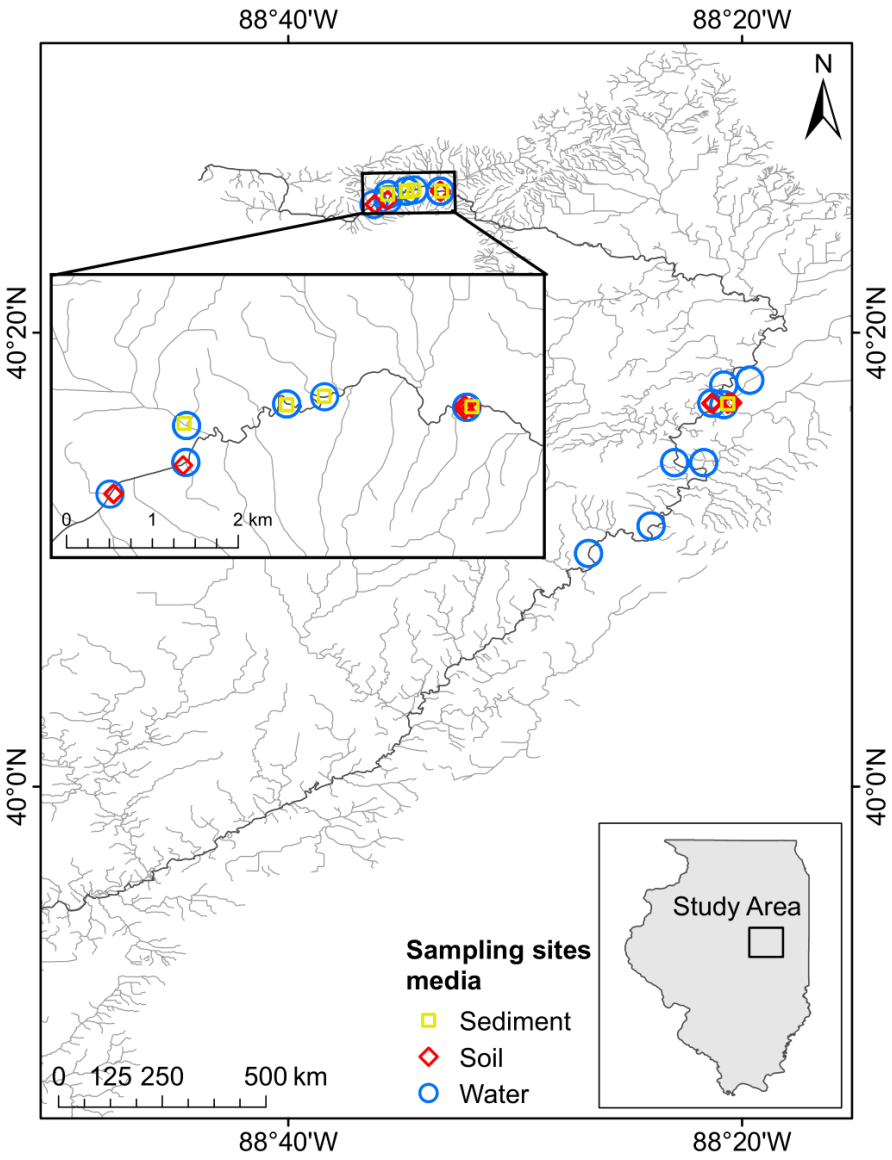

Figure 4.1 Sampling locations within the Upper Sangamon River. Gold squares represent sediment samples, red diamonds are soil samples and blue circles are water samples.

Despite the ubiquity of some organisms, the average relative abundance of dominant bacteria differed significantly between the water, soil, and sediment samples taken in the USRB (Figure 4.2). *Proteobacteria* was the most abundant phylum in all three environments (mean abundance = 36%). As expected, dominant phyla differed between water and soil communities. Surprisingly, similar taxonomic classes dominated sediment and soil communities in this basin.

The ternary diagram in Figure 4.2C shows the fraction of sequences from the nine most common phyla found in all three environments. *Proteobacteria*, *Actinobacteria,* and unclassified reads cluster near the center, indicating similar abundance communities in water, soil and sediment. The majority of *Chloroflexi* and *Bacteroidetes* sequences derived from soil and water samples respectively, and the remaining phyla had similar abundances in soil and sediment though much lower abundances in water samples.

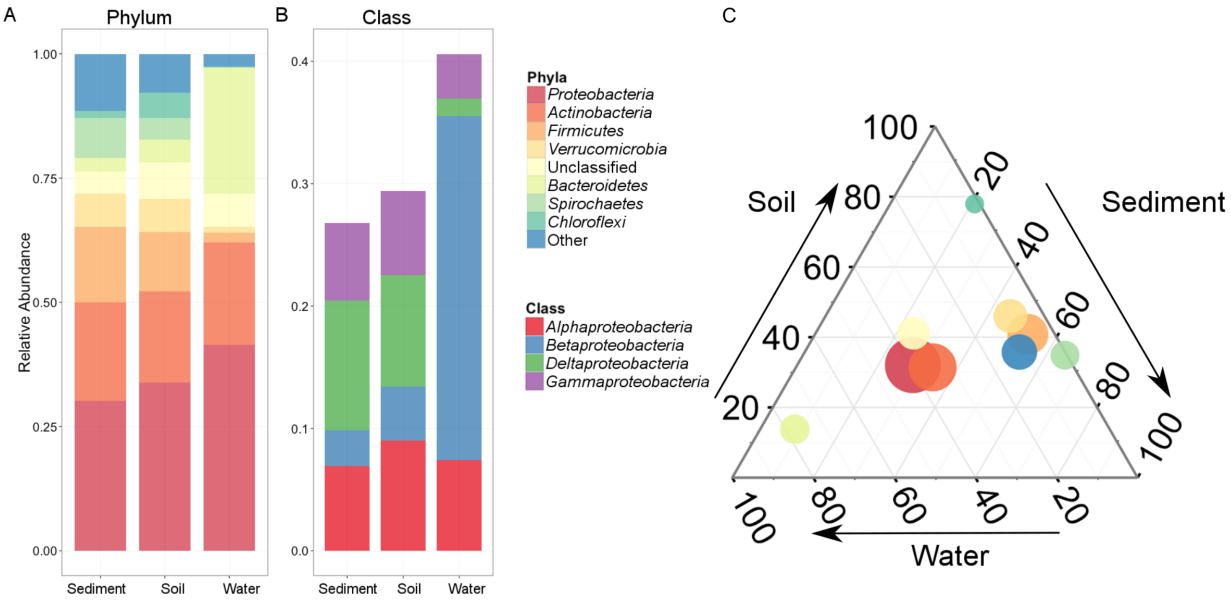

Figure 4.2 The abundance of microbial taxa across environments in the USRB. A) Average relative abundance of dominant phyla in water, soil, and sediment. B) Average relative abundance of Proteobacteria classes in each environment. C) Ternary plot indicating fraction of most abundant phyla from each environment. Circle size represents the average abundance of the phyla, and location represents the fraction of reads from that phyla found in each of the three environment types.

Analyses at finer taxonomic resolution revealed larger differences in community structure. Differential abundance analysis (ANOVA, $P < 0.001$, Figure 4.2B) revealed significantly higher abundances of *Alphaproteobacteria*, *Deltaproteobacteria,* and *Betaproteobacteria* across soil (10.8 ± 8.1%), sediment (14.0 ± 11.3%), and water (33.1 ±

15.0%), respectively. Genus-level differential abundance analysis (ANOVA, $P < 0.001$, Figure S4.2) highlighted 20 genera as differentially abundant between soil, sediment and water samples. Known soil microbes such as *Bacillus* and members of the *Rhizobiales* predominated in terrestrial habitats, whereas water samples had higher abundances of commonly occurring freshwater genera, including *Leptothrix*, *Terrabacter*, and *Polynucleobacter*.

4.3.2 Species sorting controls community assembly between environments

We observed that the microbial communities maintained distinct compositions in terrestrial (soil and sediment) and aquatic environments on the basin scale (Figure 4.2). Our analysis confirmed that media type is a significant factor in shaping microbial community diversity.[32, 196, 197] The clear differences between aquatic and terrestrial communities are not surprising, given the differences between planktonic and attached-growth conditions. Contrarily, the high degree of overlap between soil and sediment communities was surprising. We found that soil and sediment communities could not be distinguished using weighted Unifrac distances at the 97% OTU level (Figure 4.3C). We expected that soil communities would be distinct from sediment communities due to land cover, water saturation, and flooding frequency differences; however, we observed highly similar composition in both environments. Intensive landscape management in the USRB may contribute to this "biotic homogenization" by creating similar soil conditions (eutrophic, artificially graded, and relatively homogeneous) in both the agricultural soils and neighboring stream banks and floodplains. In addition, rapid water discharge through the tile drainage systems increases soil erosion and dispersal of microbes from soil to stream banks in the USRB.[185] Combined with nutrient conditions and higher residence time for terrestrial microbes, these factors could explain the high similarity between soil and sediment communities

Despite compositional similarity between soil and sediment samples, we found that stream bank sediments had higher richness and Shannon diversity than the soil communities studied. Higher alpha diversity in sediment communities may result from the high degree of niche availability afforded by intermittent flooding and connectivity with the river. USRB sediments are flooded during high flow events due to the rapid discharge from stream ditching and agricultural tile drainage. The intermediate disturbance hypothesis predicts that community richness will decrease in environments with too few ecological disturbances because they become dominated by specialized taxa.[198] Intensive landscape management decreases the duration of overland flooding and results in chemostatic nutrient conditions in the watershed[199], which could decrease diversity in soil and aquatic communities. Intermittent connectivity could replenish microbial diversity that would otherwise be lost under such intensive management.

To corroborate the sources of downstream microbes from upstream communities identified previously, we analyzed dispersal limitations using partial redundancy analysis (RDA) and distance decay curves. We compared the impact of environmental variables and geographic distance on aquatic community structure using partial RDA. By examining the variance explained by spatial and environmental differences separately, we identified that location and then nutrient concentrations explained the most variation in aquatic community beta diversity (Table S4.1). This supports our hypothesis that connectivity and transport are key factors that structure water communities, even when compared with species sorting processes. Our distance decay analysis revealed that distance did not explain variation in the soil communities we examined, suggesting that deterministic responses to local environmental factors explained most of the variation in this environment.[197] In contrast, spatial distance was correlated with weighted Unifrac distance for sediment and aquatic communities, implying that unmeasured

environmental factors or stochastic processes, such as dispersal, regulated communities in these two environments. Interestingly, the most abundant (≥1% relative abundance) OTUs displayed stronger distance-decay patterns than rare (<1%) OTUs in these communities, suggesting that assembly processes impacted the rare and abundant biosphere differently.

4.3.3 Microbial community dissimilarity patterns

First we compared community alpha diversity between environments using richness (observed OTUs), as well as Shannon diversity after rarefaction to 10,000 sequences per sample. Both metrics were significantly higher for communities in sediment (Shannon diversity = 440 ± 195) than soil or water (316 ± 311 and 216 ± 541, respectively, ANOVA, $P = 0.027$) (Figure S4.3). Within each environment, site conditions were compared with alpha diversity to identify factors that were correlated with richness. In the water samples, temperature, pH, specific conductivity, nutrient concentrations, and sampling depth were analyzed. Water temperature and pH were negatively correlated with community richness (Pearson $R = -0.66$ and -0.74, respectively, $P < 0.001$) (Figure S4.4). Other environmental parameters did not show significant correlations. Within the terrestrial communities, microbial community richness decreased with depth in the soil ($r = -0.39$, $P < 0.05$), but not with sediment depth.

Beta diversity analysis (weighted UniFrac distance) revealed clear demarcations between environments—soil, sediment, and water, shown in a PCoA ordination plot of all samples in Figure S4.5. Figure 4.3 shows intra-group (e.g. all water vs. water and sediment vs. sediment) and inter-group (e.g. all soil vs. water) weighted UniFrac distances in blue and green, respectively, for each pair of environments. Results of ANOVA tests of inter- vs. intra group similarity are shown above each pair of environments. Panel A and B demonstrate the consistent and distinguishable community structures for terrestrial vs. aquatic communities. Weighted

unifrac distance was higher between than within environments for sediment vs. water samples (ANOSIM R = 0.77, P < 0.001). The mean of inter-group pairwise weighted UniFrac distance was 0.53 and those of intra-group comparisons were 0.39 and 0.29 for water and sediment, respectively, as depicted in the distance histogram. Similar to water and sediment communities, water and soil communities were distinct from each other (ANOSIM R = 0. 74, P < 0.001). In contrast, 3C shows that the soil and sediment consortia had a much lower inter-group dissimilarity (mean = 0.35) that was similar to the corresponding intra-group comparisons (ANOSIM R = -0.14, P = 0.951). These results indicate that microbial communities significantly differ between terrestrial (soil and sediment) and aquatic (water) environments.

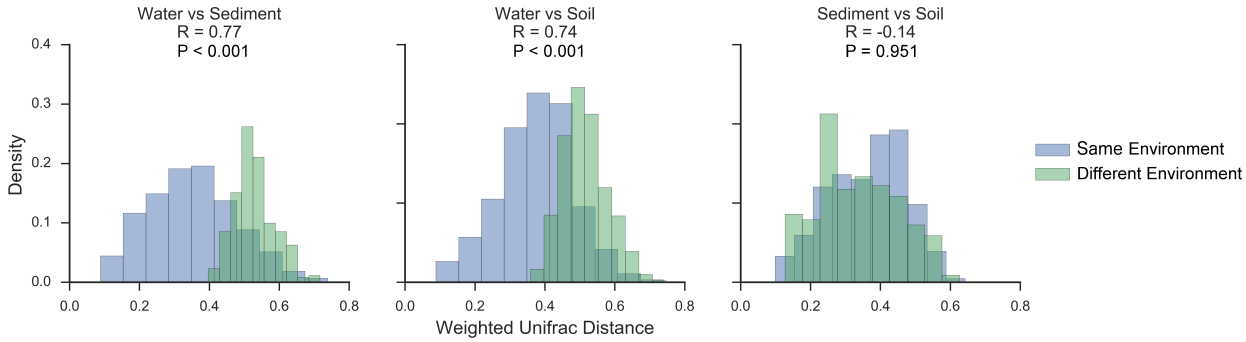

Figure 4.3 Differentiation of terrestrial and aquatic environments. Histograms of weighted UniFrac distances calculated within and between sample media types and associated ANOSIM R-values and P values. "Same Environment" (e.g. both water or both sediment) and "Different Environment" refer to beta diversity for samples originating from the categories above each histogram. (A) Histograms of "Water-Water" and "Sediment-Sediment" in blue and "Water-Sediment" in green are statistically different, suggesting these communities are distinct. (B) Distribution of distances for all within and between water and soil samples. (C) Distribution of distances for all within and between soil and sediment samples showed no statistical significant differences, implying these communities were highly similar.

In addition to environmental species sorting, microbial communities often display spatial autocorrelation or distance decay patterns. Typically, environmental selection is thought to increase distance decay whereas high dispersal minimizes this effect.[177] To test whether dispersal

impacted overall or rare community structure in the basin, we quantified community beta diversity distance decay using weighted Unifrac distances based on all OTUs, as well as based on only highly abundant (≥1% average abundance) or rare OTUs (<1% abundance) (Figure S4.6). Communities differed more across medium type than by spatial distance across the watershed (ANOSIM $R = 0.53$, $P < 0.001$). Interestingly, the soil and sediment communities were more spatially homogenous than the water communities, which was unexpected due to higher dispersal and connectivity in the aquatic communities. We identified significant distance-decay relationships for both abundant (≥1% average abundance) and rare OTUs (<1% abundance) in aquatic communities (Mantel $r = 0.34$ and 0.20, respectively, $P < 0.05$). Abundant OTUs in sediment samples displayed significant distance-decay as well (Mantel $r = 0.39$, $P < 0.05$), though none of the soil community subsets showed distance-decay relationships.

Using partial redundancy analysis (RDA), we calculated the fraction of beta diversity that could be uniquely explained by measured environmental factors and stream-wise distance between the water communities (Table S4.1). We individually constrained weighted Unifrac distances using nutrient concentrations (TKN, nitrate, and ortho-phosphate), other environmental factors (temperature, pH, and conductivity), and a single variable representing location in the drainage network (up- or downstream sites) while partialling out the other variables. Sites in the up- or downstream locations were separated by at most 9 km stream-wise, whereas at least 35 km separated the two groups. The largest single factor in explaining community similarity was the location along the drainage network, which explained 13.6% of weighted Unifrac distance. Nutrient concentrations were the next most important factors with TKN and nitrate explaining 8.2% and 10.0% of community variation, respectively. Covariation between measured variables explained an additional 8.5% of total variation. Similar to distance decay results, community

composition was strongly influenced by location, suggesting that microbial dispersal is important in structuring microbial communities within the USRB.

4.3.4 Microbial connectivity within the USRB

We characterized microbial community connectivity at the basin scale to test whether microbial dispersal influenced community diversity. Most notably, aquatic microbial communities were highly related throughout the watershed (64A). SourceTracker analysis revealed that 85.4 ± 13.3% of the sequences in downstream water samples derived from upstream water sources. However, there was significantly less contribution from upstream soil and sediment to downstream water samples (0.2 ± 0.8%, and 0.02 ± 0.08%, for sediment sources). Source proportions for downstream aquatic communities did not vary with distance from upstream sources (Mantel test, P = 0.99), suggesting that there was strong upstream-downstream connectivity along the drainage network. A large unknown source was observed in all water samples. Both soil and sediment communities (Figure 4.4B and 4.4C) sourced higher proportions of OTUs from upstream soil and sediment (soil—22.3 ± 23.3% from upstream soil, 44.7 ± 26.6% from upstream sediment; sediment—38.49 ± 26.30% from upstream soil, 29.7 ± 30.7% from upstream sediment) than from upstream water (to soil—0.5 ± 11.2% and to sediment—0.1 ± 3.5%).

Notably, the proportion of sequences in sediment sourced from upstream samples decreased with distance from the source (Mantel R = 0.30, P = 0.007 for soil; R = 0.21, P = 0.038 for sediment). No significant spatial patterns were found in upstream soil or sediment sources (Mantel P = 0.373 and 0.688, respectively) to downstream soil communities. Spatial differences in source community proportions could indicate that dispersal limitations prevent connectivity between up- and downstream communities in the sediment. Because hydrological

connectivity, and therefore dispersal between soil samples at different sites, is unexpected, the lack of distance decay may be explained by overall environmental similarity and spatially uncorrelated factors, such as soil depth. Cross environment dispersal of microbial communities due to strong lateral riparian connectivity was not observed. Bayesian estimation indicated very little connectivity between aquatic and terrestrial (soil and sediment) sources.

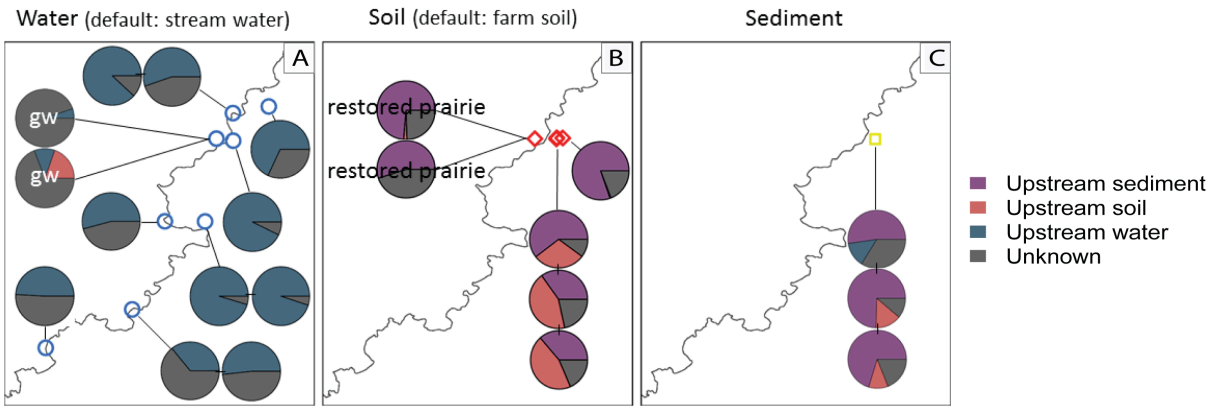

Figure 4.4 Bayesian estimation of source communities for downstream A) water, B) soil, and C) sediment samples. Estimated source contributions from each media type are shown for each sample.

To identify whether ground and tile water could serve as dispersal pathways, we compared ground and tile water community composition with soil, sediment, and river water samples. Ground, tile and, river water communities differed significantly based on weighted Unifrac distances (ANOSIM, $R = 0.88$, $P < 0.001$). These differences were due in part to differentially abundant taxa, including several taxa found only in terrestrial samples and ground- or tile water samples. Notably, nine OTUs were differentially abundant between river and tile or groundwater, including OTUs from phylum *OP3*, families *Oxalobacteraceae* and *Syntrophobacteraceae* and genus *Rhodococcus*. All of the four differentially abundant OTUs observed in groundwater but not river water samples were also found in soil samples and three were found in sediment samples, suggesting that there is frequent dispersal between

terrestrial and groundwater environments. One of the five differentially abundant OTUs that was observed in tile water but not river water samples was also identified in soil samples suggesting that it may have originated from drainage through porous media.

4.3.5 Connectivity influences fluvial microbial community structure

We characterized diversity and connectivity in soil, sediment and water microbial communities within a highly managed and predominantly agricultural basin. Using SourceTracker, we identified a large proportion of the downstream aquatic communities that originated in upstream source aquatic communities in the USRB (85.4 ± 13.3%). Given the hydrological connectivity between these sites, the community similarity between them was unsurprising. Notably, we found that a substantial fraction of the microbial community in downstream terrestrial (soil and sediment) samples could have originated in upstream sediments (contributing up to 38.5 ± 26.3% to downstream terrestrial samples) and soil media (up to 44.7% ± 26.6%). The high overlap between soil and sediment suggests higher-than-expected dispersal between these environments combined with similar environmental pressure, which has led to community homogenization. We observed a low frequency of cross-habitat OTU sourcing, which was especially surprising between sediment and aquatic environments. The sediment samples in this study came from exposed sections of the river cut bank that were dry during sampling. Fully submerged sediments may show higher similarity to the water column itself, especially during late summer.

Previously, downslope inoculation from terrestrial communities has been identified as a major factor in the dispersal of microbial communities in freshwater aquatic environments[180, 200], however there is limited evidence for aquatic seeding of terrestrial communities. Shared OTUs between tile water and terrestrial communities, indicate that tile drainage is one route for

connectivity in this direction within the Sangamon River Basin. The extent of lotic connectivity to surrounding environments is influenced by factors including flooding frequency, drainage rates, stream order, and mixing.[179] Our results indicate that hydrodynamic transport of bacterial cells plays an influential role in fluvial community structure, but cross-habitat dispersal is likely low due to fast drainage through tile drainage networks and intensively managed river flow.

Cross-habitat connectivity and dispersal from terrestrial to aquatic environments occur through a variety of processes including overland transport, mixing of tributary water, discharge of groundwater from cut banks and hyporheic exchange.[201] Few studies have attempted to quantify the impact of landscape management on microbial transport. Lindström and Bergström [202] found that connectivity increased community similarity, but there was no difference in cell transport between lakes connected by ditches or by streams. Intensive management of the USRB, i.e. construction of drainage ditches and tiles, has drained seasonally flooded low-gradient areas and reduced the intensity of flooding. Tile drainage and groundwater discharge appear to connect terrestrial environments to the Sangamon River, as evidenced by OTUs found only in the soil, sediment and ground- or tile water (Table S4.2). Surprisingly, the total abundance of these OTUs was low and we observed limited evidence of lateral riparian exchange between soil or sediment and lotic communities in the USRB.

Overland flooding is another mechanism for cross-habitat dispersal in the USRB. Flooding in the USRB occurs primarily in the spring and early summer. Sampling for this study was performed during late summer when baseflow was lower and it appears that there was not active exchange of organisms between the riparian sediments (streambank and floodplain) and river water during this time. Lowland flooding remains a source of intermittent connectivity in the watershed, but the effects on microbial community composition may be transient. Microbial

community responses to storm events can occur on the order of days[203], and these episodic events would be missed by our sampling regime. More work comparing connectivity in highly managed and natural ecosystems is necessary to identify whether the low rates of cross-environment connectivity in our study are typical outside of intensively managed landscapes. Further, higher frequency sampling in parafluvial sediments could address the potential for these areas to contribute to dispersal between terrestrial and aqueous media.

One limitation of our work is that the SourceTracker algorithm used in this work assumes that sequences present in downstream samples originated from one of several possible upstream source communities. Species sorting or habitat selection presents a confounding factor and potential source of community overlap. Quantifying connectivity during flooding as well as dry conditions and in different types of hydrological networks could help elucidate whether community similarity is a result of connectivity or of habitat selection.

4.3.6 Co-occurrence patterns differed between environments

Based on the observed differences in landscape connectivity and differences between aquatic and terrestrial (soil and sediment) communities, we inferred meta-community co-occurrence networks within and across environments. We used an ensemble correlation metric comprised of compositionality corrected Spearman's and Pearson's correlation coefficients as well as FDR-adjusted p-values to identify significant associations between OTUs and compared interaction frequency and network structure in different environments. The generated meta-community co-occurrence networks captured 4553 relationships (edges) involving 500 OTUs (nodes). Co-occurrence networks for each environment and different strengths of correlation are shown in Supplementary Figure S4.7. Most of the identified relationships had correlation strengths > 0.5, including 1596 in soil communities, 405 relationships in sediment communities,

and 1939 in water communities. Correlations with $\rho > 0.75$ were much less common—200 in soil, 255 in sediment, and 137 in water. Compared to the other networks, the sediment network contained fewer relationships but a higher percentage of strong positive associations (38% of the total edges). The differences in network edge density and frequency of strong associations is surprising but may be due to differences in niche partitioning strength in different habitats. Only one exclusion relationship was identified in water with correlation < -0.5 in both metrics, and it was between OTU 113 (family Microbacteriaceae*)* and OTU 342 (genus *Flavobacterium*). We did not observe negative correlations < -0.5 in either soil or sediment networks

We identified co-occurrence relationships that were consistent across environments (Figure 4.5). The sediment and soil co-occurrence networks displayed the highest overlap with 60 shared relationships. Only 24 relationships were common to the soil and water networks and only 9 relationships were shared between the sediment and water networks. Only a single co-occurrence relationship with a correlation > 0.5 was found to be consistent across all environments, the relationship between OTU 142, (genus *Arthrobacter)* and OTU 867 (genus *Mesorhizobium*). Both genera are common soil bacteria, although their appearance in water samples was unexpected. We then calculated degree for each OTU in each environment. All three networks had long-tailed degree distributions fit by power law functions indicating scale-free network structures and non-random co-occurrence patterns. Notably, the scaling exponent of the power law differed for water (a = 0.61) and soil and sediment (a = 0.44 and 0.43 respectively) OTU networks. Networks in soil and sediment had degree distributions significantly different from those in water (Wilcoxon rank-sum test, $P < 2.6\times10^{-3}$ and $P = 1.6\times10^{-24}$, respectively). Despite their similar power law exponents, soil and sediment

communities also had significantly different degree distributions (Wilcoxon rank-sum test, $P < 5.7\times10^{-10}$), due in part to fewer overall edges in the sediment network.

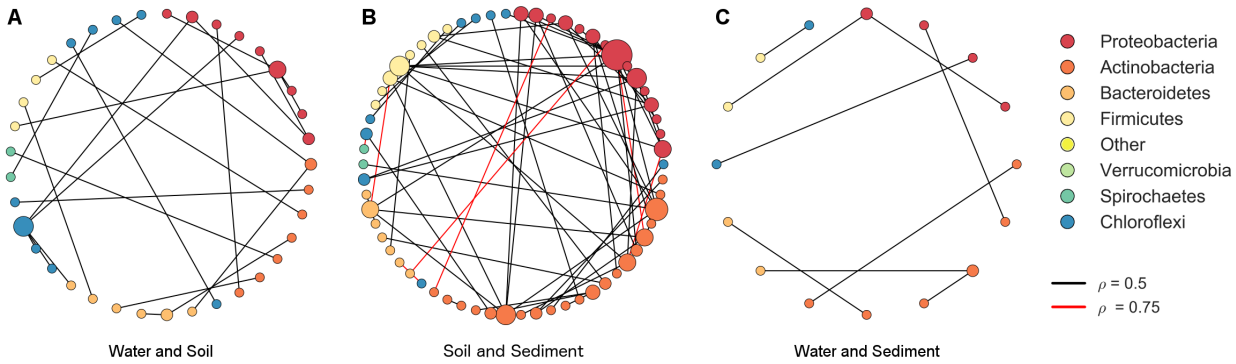

Figure 4.5 Pairwise co-occurrence relationships across environments. The consistent pairwise co-occurrence relationships with ρ values greater than 0.5 were determined A) between water and soil, B) between soil and sediment, and C) between water and sediment. Consistent pairwise co-occurrence relationships with ρ values greater than 0.75 were only found between soil and sediment. We highlighted the OTU nodes coming from the eight most abundant phyla observed in all samples. The bubble size of each OTU node within a network is proportional to the degree of the node.

Modern community assembly theories, such as the metacommunity framework, include both connectivity-driven dispersal processes as well as species sorting mechanisms.[143] In the USRB, the river represents a high dispersal system in which community structure should be heavily influenced by transport. We analyzed co-occurrence networks formed by microbes within and across environments in the USRB for differences between aquatic and terrestrial community structure. Co-occurrence patterns in these networks indicate how the microbial communities are structured by interspecific competition and other mechanisms.[204]

One interesting feature of the association networks we generated was the low prevalance of negative (co-exclusion) relationships. Only a single negative relationship with an FDR corrected $P < 0.05$ and $R < -0.75$ was observed across all three networks. Due to the

compositional and sparse nature of high-throughput sequence data, co-occurrence network construction can suffer from several sources of bias. Sparse datasets may result in a number of "none vs. none" comparisons that can interfere with detected true negative correlations between OTUs. In a recent comparison of different association metrics, Weiss et al. showed correlation metrics differed in their ability to detect certain relationship types. In particular, CoNet (the *ccrepe* package) was poor at detecting competitive (mutual exclusion) relationships compared with Spearman correlation. Despite this bias, when these metrics were combined in an ensemble metric, the overall precision and accuracy improved.

Our results showed that consistent and significant pairwise co-occurrence relationships between OTUs were much more common in soil and water communities compared with sediment communities. The observation of non-random co-occurrence patterns suggests that deterministic processes such as interspecies competition or niche differentiation dominate the structure of communities in terrestrial environments.[205, 206] Biotic interactions should increase the frequency of observed co-occurrence relationships whereas stochastic processes and dispersal should limit the appearance of co-occurrence patterns.[207] Fewer observed associations combined with higher overall alpha diversity in the sediment communities suggest that the sampled sediments were more influenced by stochastic processes and disturbances.

Similar to previously described microbial co-occurrence networks, OTU association networks in all three environments had ed degree distributions, although the scaling coefficients varied between all three environments. Limited consistent co-occurrence relationships were observed across all three media types, which indicates that each environment has distinct microbial communities (Lozupone and Knight, 2007; Nemergut *et al.*, 2011). Just 2.9% of all co-occurrence relationships found in either soil or sediment samples were present in both media

types, and consistent cross environment relationships were less common between the other environments.

Microbial biogeography patterns result from a combination of environmental selection and dispersal within and between environments. In this study, we identified strong signals of landscape connectivity within lotic communities, but we found little dispersal between terrestrial and aquatic environments. Dispersal along the river led to significant distance decay patterns that were evident in analyses via Mantel tests and SourceTracker. These trends persisted even after correcting for measured environmental variation in partial RDA analysis. However, terrestrial communities displayed fewer significant distance decay patterns. Together, our results are consistent with the metacommunity community assembly framework. More work is needed to identify whether dispersal and connectivity play a larger role in shaping communities during the wet season and how communities in more natural hydrological systems behave.

4.4 METHODS AND MATERIALS

4.4.1 Site selection and sampling

We surveyed the microbiome of the USRB, a low-relief 3600-km^2 area located in central Illinois.[185, 208] A map of the basin is provided in Figure 4.1. The USRB is one of three primary field sites of the IML-CZO project. The USRB includes a wide variety of landscape and land cover types that have been highly modified by anthropogenic processes.[209] Over 83% of the USRB land area is row-crop agriculture, primarily farmed for corn and soybeans[210] (Figure S4.1). The remaining 17% of the land area is used for growing grains and specialty crops and cattle grazing (pastures), or covered by deciduous to mixed riparian forest, restored tallgrass prairie, urban land, and woody wetlands.[211] The microbiome survey was conducted over a period

of several months in 2014. An initial set of soil and sediment samples were collected from two cores (CHAM-14-01 and CHAM-14-02) taken at one site on March 7, 2014 (Figure 4.1). Additional soil, sediment, and water samples were collected across the northern half of the USRB from August 21-August 22, 2014. A total of 77 samples were obtained. Sampling locations are shown in Figure 4.1.

In the August 2014 campaign, samples were collected of river water, surface soil, and floodplain sediments from areas having a wide variety of land cover types. Soil samples were collected from the A and B soil horizons that are developed in different parent materials, including windblown silt and sand (loess), river alluvium, glacial outwash, and glacial till.[212, 213] Soil samples were obtained from 10–20 cm deep hand-dug pits, while sediment samples were taken from exposed cut banks along the Sangamon River and tributary streams and water samples were taken directly from their respective sources. Sediment samples included weathered and unweathered (leached and calcareous), sandy to clayey materials interpreted as river alluvium, colluvium, loess, glacial outwash, and glacial till, which form the floodplains, eroded slopes, and upland areas in the USRB.[214, 215]

The soil and sediment samples were grouped into four classes based upon the land cover type at the site: 1) cultivated land for row crops—primarily corn and soybeans; 2) pasture, used for grazing by cattle and goats; 3) restored prairie recently converted from cultivated lands; and 4) riparian deciduous and mixed forest. Water samples were classified as follows: 1) river water from the Upper Sangamon River and its tributaries, 2) groundwater from 6-inch and 12-inch deep lysimeters, and 3) outflow from tile drains. For both soils and sediments, 1 g of material was collected using a sterile scoop from the cleaned off walls in hand-dug pits or vertical faces in cut banks. The sampled material was transferred to a sterile Whirl-Pak bag and immediately

frozen on dry ice. For the river water, a 1L sample was collected for microbial community analysis using a sterilized bucket, then transferred immediately to a sterile 1L glass bottle, and finally transported on dry ice until the samples were filtered through a sterile Millipore 0.22 mM membrane. A separate 1 L water sample was obtained for chemical analyses, performed at the Illinois State Water Survey (ISWS). Soil, sediment, and filtered water samples were stored at -80^{o}C prior to DNA extraction. For the tile water samples, 1L of water was collected directly from the tile drain near the river outlet. Groundwater samples were pumped and collected through the lysimeters.

4.4.2 Water chemistry measurement

For each river and tile water sample, pH and specific conductivity were measured in the field using a Hydrolab HL4 Multiparameter Sonde manufactured by Ott Hydromet. Water samples for nutrient analysis were stored in the field on dry ice and sent to the ISWS for analysis within 24 hours of collection. Water samples were analyzed for total phosphorus (mg P/L), total Kjeldahl nitrogen (TKN, mg/L), dissolved phosphorus (mg P/L), orthophosphate (mg P/L), nitrate (mg N/L), and ammonia (mg N/L) following standard methods.[216] Groundwater collected in the lysimeters was not analyzed for water chemistry due to the limited volume of this sample type.

4.4.3 DNA extraction, 16S rRNA gene amplification and sequencing

Genomic DNA was isolated from soil, sediment, and water filter membrane samples using a modified protocol of the PowerSoil®-htp 96-well Soil DNA Isolation Kit (MO BIO, Laboratories, Inc.). The modified protocol includes a step to heat the lysing solution and sample mixtures to 65^{o}C for 10 min prior to bead-beating. Despite the known biases of amplicon sequencing and the 515F/806R primer set[217], we selected this approach due to the large number

of samples and purpose of detecting broad scale dispersal between environments. Genomic DNA was amplified using the Earth Microbiome Project (EMP) barcoded primer set (515F-806R) targeting the V4 region of the bacterial 16S rRNA gene, and was adapted for the Illumina MiSeq platforms by adding flow-cell adapter sequences and twelve extra bases in the adapter region of the forward amplification primer that support paired-end sequencing. The reverse amplification primer also contained a twelve-base barcode sequence that supports pooling of up to 2,167 different samples in each lane. PCR conditions followed the EMP protocol.[86] Following pooling, amplicons were quantified using PicoGreen (Invitrogen) assay and a plate reader. Once quantified, different volumes of PCR product from each sample were pooled into a single tube to equalize the concentration of amplicons from each sample. This pool was then purified using the UltraClean® PCR Clean-Up Kit (MO BIO) and quantified using Qubit (Invitrogen). After quantification, the molarity of the pool was determined and diluted to 2nM and then the pools were denatured and sequenced together in a single Illumina MiSeq run (150 bp x 2). Raw FASTQ files were submitted to the Short Reads Archive (Accession number SRR5581933).

4.4.4 Sequence data analysis

Paired end reads were quality trimmed to remove sequences with more than 1 error per 100 bases, *de novo* chimera filtered, and processed for operational taxonomic unit (OTU) clustering using the UPARSE pipeline[90] in USEARCH v8[171] at a 97% identity cutoff. A second chimera filtering was performed using the UCHIME algorithm[109] after OTU clustering as a post-processing step based on the ChimeraSlayer reference database from the Broad Microbiome Utilities (version microbiomeutil-r20110519). Taxonomy was assigned to OTUs using the Ribosomal Database Project (RDP) database[11]. Multiple sequence alignment and phylogenetic reconstruction were performed using PyNast and FastTree[169]. The filtered OTU table was

processed to remove samples with fewer than 150 reads, which are usually uninformative and indicative of low-quality reads[218]. Finally, we rarefied samples to 10,000 sequences for downstream alpha and beta diversity analyses.

Alpha and Beta diversity metrics were calculated in Quantitative Insights Into Microbial Ecology (QIIME). Ordination methods (Principal Coordinate Analysis and Redundancy Analysis) and multivariate analyses (ANOVA, PERMANOVA, ANOSIM) were performed in Python v2.7.11 using the "skbio" package (0.4.0). Alpha diversity was calculated using Faith's Phylogenetic Diversity (PD) and Shannon Diversity. Correlations between alpha diversity and soil and water quality measurements were calculated using Pearson correlation between diversity and each physical measurement and correcting for multiple comparisons using FDR.

To assess differences between samples in terms of composition, pairwise Weighted Unifrac distances were calculated from rarefied OTU abundance counts using QIIME's "beta_diversity.py" script. The UniFrac distance is defined as the fraction of a phylogenetic tree consisting of two communities that leads to taxa found in one community but not the other. Weighted UniFrac weights path lengths by OTU abundance before calculating similarity. Beta diversity comparisons were performed with weighted UniFrac distances using ANOSIM implemented in the R package "Vegan" to compare within and between environment distributions of beta diversity values for statistical differences.

Partial Redundancy analysis (partial RDA) was selected to analyze the relationships between weighted Unifrac distances and environmental parameters due to its ability to detect community dissimilarity uniquely explained by different factors. For the partial RDA analysis, a matrix containing Z-score normalized environmental parameters and nutrient concentrations, as well as factors encoding up or downstream location was constructed. Partial RDA was performed

by comparing the variance explained by RDA models containing two groups of parameters separately and together and subtracting the variance explained by each component individually to quantify the covariation between parameters and variation uniquely explained by each group. Mantel tests were used to analyze correlations between weighted UniFrac distances and spatial distances.

Both R and QIIME v1.9[13] were used for detailed downstream analysis, e.g. to estimate the total expected OTU richness from original OTU tables using the R package Vegan specnumber function[105, 219] and to calculate alpha and beta diversity with filtered OTU tables. Abundant (>1% relative abundance) and rare (<1%) OTUs were separated for Mantel spatial analysis to better resolve the spatial patterns of microbes in different environments.

4.4.5 Ternary Diagram and Violin Plot Construction

The ternary diagram in Figure 4.2 shows the fraction of sequences assigned to the nine most abundant phyla observed in the three different environments as well as the relative abundance of each phylum. Each axis is labeled with an environmental media and an arrow representing increasing frequency of sequences found in that environment. For each phylum, its average relative abundance was calculated in each environment. Then its position in the ternary diagram was calculated by normalizing the relative abundance in each environment by the sum of its relative abundance over all environments. For instance, if a phylum had a relative abundance of 50% in the soil and sediment compartments, and 25% abundance in the water component, it would appear at the intersection of 40% on the sediment and soil axes and 20% on the water axis. The area of the circle corresponds to the log normalized average relative abundance of the phylum across all environments.

Violin plots of alpha diversity for each environment were created in Python using the seaborn package. A violin plot is similar to a histogram but continuous rather than discrete and similar to a box-and-whisker plot but contains modal information.

4.4.6 Co-occurrence network construction

Following the recommendations in Weiss, et al. [220] we analyzed co-occurrence patterns within and between environments. We rarefied our dataset to the top 1,000 OTUs by average abundance to remove rare OTUs and divided samples into groups by media type—water, soil, and sediment. We inferred the co-occurrence network in each media type using an ensemble correlation metric composed of Spearman's and Pearson's correlation coefficients, which represent the strengths of co-occurrence between pairs of OTUs.[221] Positive correlations are considered to indicate co-occurrence relationships, while negative correlations indicate either competition or non-overlapping niches. We corrected correlation coefficients obtained with both metrics for compositionality effects using the *ccrepe* package in R[222] and adjusted the p-values for multiple testing using the False discovery rate (FDR) method implemented in the *fdrtool* R package.[223]

We constructed co-occurrence networks with the *igraph* R package.[224] Each OTU is represented by a node, and pairwise relationships (co-occurrence or exclusion) are represented by edges. For our ensemble approach, edges were only included if they exceeded the FDR-adjusted p-value and compositionality corrected ρ cutoffs in both metrics. We investigated these relationships based on the strength of correlation ρ values ≥ 0.5 and 0.75 for co-occurrence, and ρ values ≤ -0.75 and -0.5 for competition.[221] The cutoff of FDR-adjusted p-values was set to 0.05. We then characterized the co-occurrence network topology in terms of node degree

(number of adjacent nodes). Wilcoxon rank-sum tests were performed to test for statistical differences in network degree distribution between aquatic and terrestrial environments.

4.4.7 Microbial community connectivity tests

We employed a Bayesian-based community-wide approach in the software SourceTracker v0.9.8[225] to identify connectivity patterns of microbial communities on the basin scale. Sourcetracker was originally developed to infer the proportion of sequences in a "sink" community originating from multiple "source" communities. It uses Gibbs sampling to assign each sequence in a sink community to a likely source community based on its abundance in each source. We performed source tracking on OTUs present in at least 7 samples (around 10% of all samples). All samples were rarefied to 1,000 sequences, using the default setting of a = 0.001. If <1,000 sequences were available for a sample, then all sequences for the sample were used. To test for longitudinal connectivity along the drainage network, we estimated the contribution from upstream source communities to those downstream. All the samples located upstream of each "sink" sample were considered as valid "source" communities in the analyses. Sources were grouped by sample media (upstream water, soil, and sediment). We evaluated lateral riparian exchange between environments by examining contributions from aquatic microbial communities to microbial communities in soil and sediment, and vice versa.

4.4.8 Statistical analyses

Statistical analyses were performed in R, Python and QIIME. Differential abundance across media types was analyzed using the Kruskal-Wallis test implemented in the QIIME script group_significance.py. ANOVA is a multi-group extension of a t-test that compares within-group to between-group mean and variation. For each taxa, a one-way ANOVA test is conducted to test whether the abundance distribution for that taxa differs between the three environments

(soil, sediment, and water). The reported p-values were corrected for multiple comparisons by the Benjamini-Hochberg FDR procedure. The reported FDR corrected p-values were corrected for multiple comparisons by the Benjamini-Hochberg FDR procedure. Beta diversity comparisons were performed with weighted UniFrac distances using ANOSIM implemented in the R package Vegan. Partial Redundancy analysis (partial RDA) was selected to analyze the relationships between weighted Unifrac distances and environmental parameters due to its ability to detect community dissimilarity uniquely explained by different factors. Mantel tests were used to analyze correlations between weighted Unifrac distances and spatial distances.

4.5 Supplementary Materials

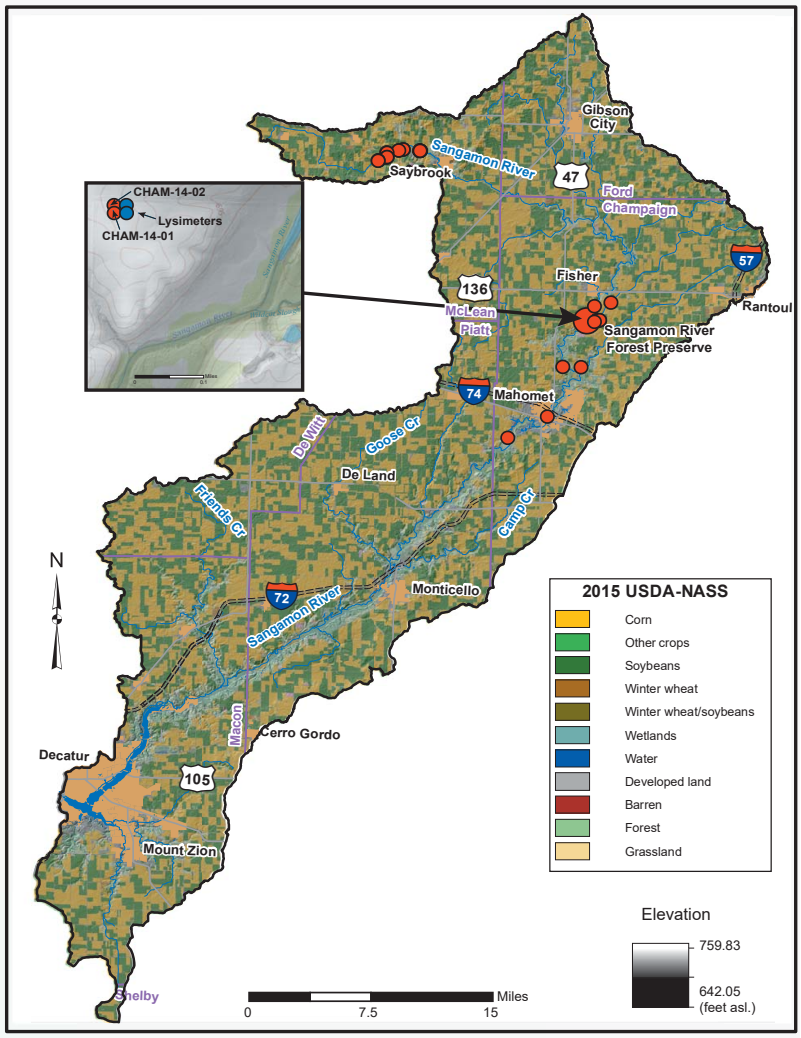

Figure S4.1 Map of 2015 USDA-NASS land cover categories in the Upper Sangamon River Basin. The land use layer is overlain on a hillshade of the digital elevation model of the land surface topography. Sample sites are located by the red dots.

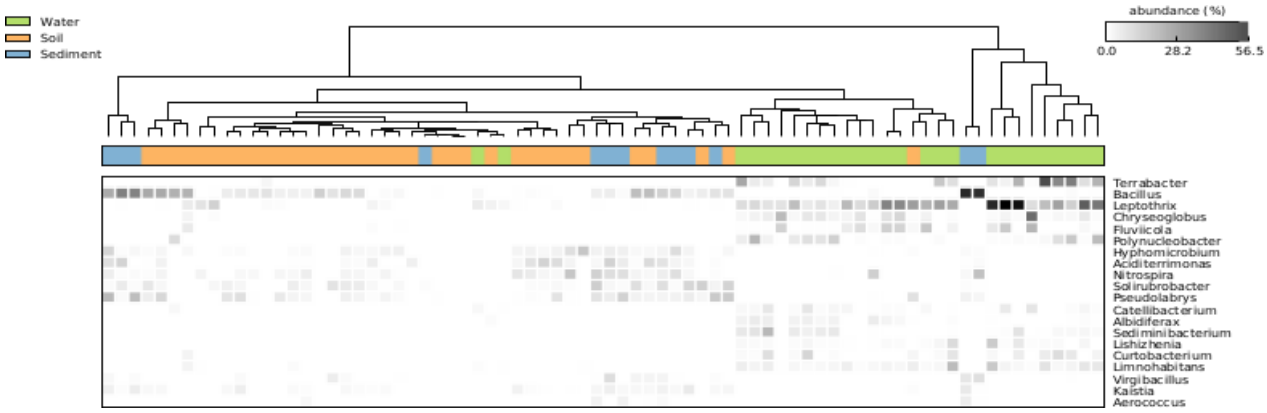

Figure S4.2 Hierarchical clustering of samples based on the abundance of differentially abundant genera in the three media types. The top color bar represents the media type of the sample, and the shade of grey represents the abundance of that genus in each sample. Differentially abundant taxa across the three media types were identified using the group_significance.py script in QIIME and the Kruskal-Wallis test. A higher-resolution version is available in the original manuscript.

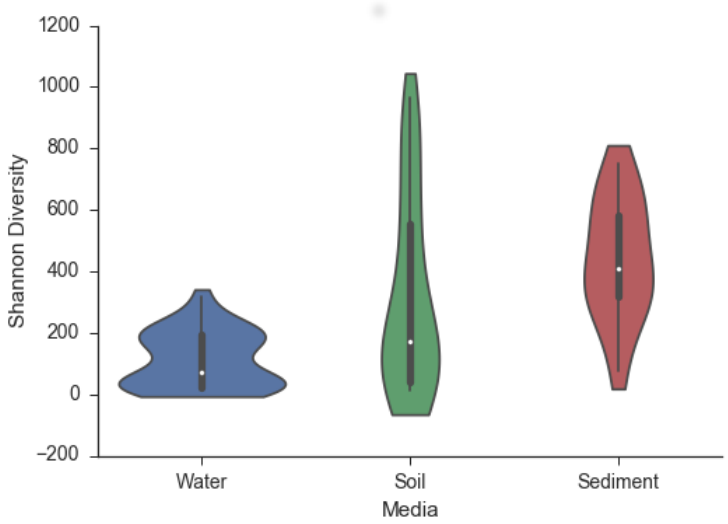

Figure S4.3 Violin plot of Shannon's Diversity Index for all samples grouped by media type. The width of each violin represents a kernel density estimation of the underlying distribution of the Shannon Diversity index.

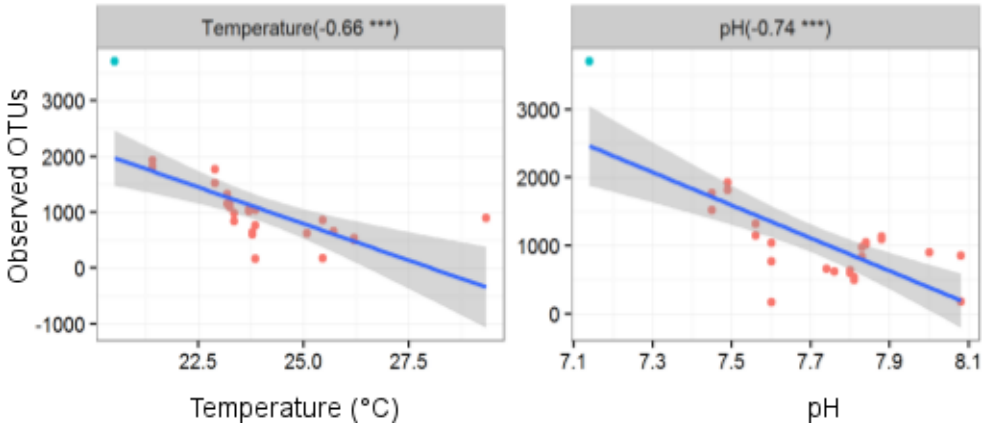

Figure S4.4 Factors significantly correlated with richness (observed OTUs) in water samples. Only significant factors are shown. The R^2 is shown in the subfigure heading.

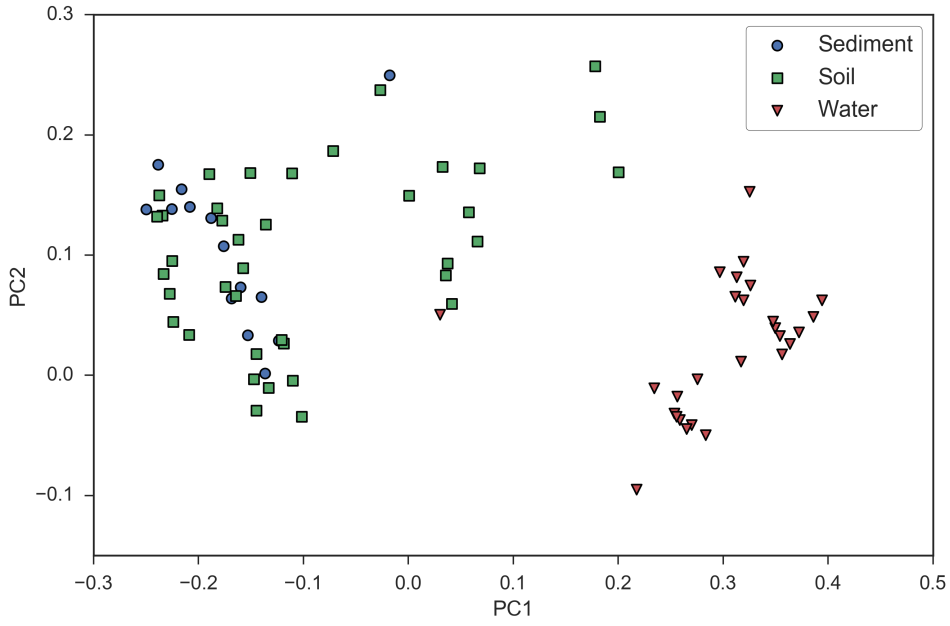

Figure S4.5 Principal Coordinate Analysis of Weighted UniFrac data for all samples. Blue circles are sediment samples, green squares are soil and red triangles are water samples.

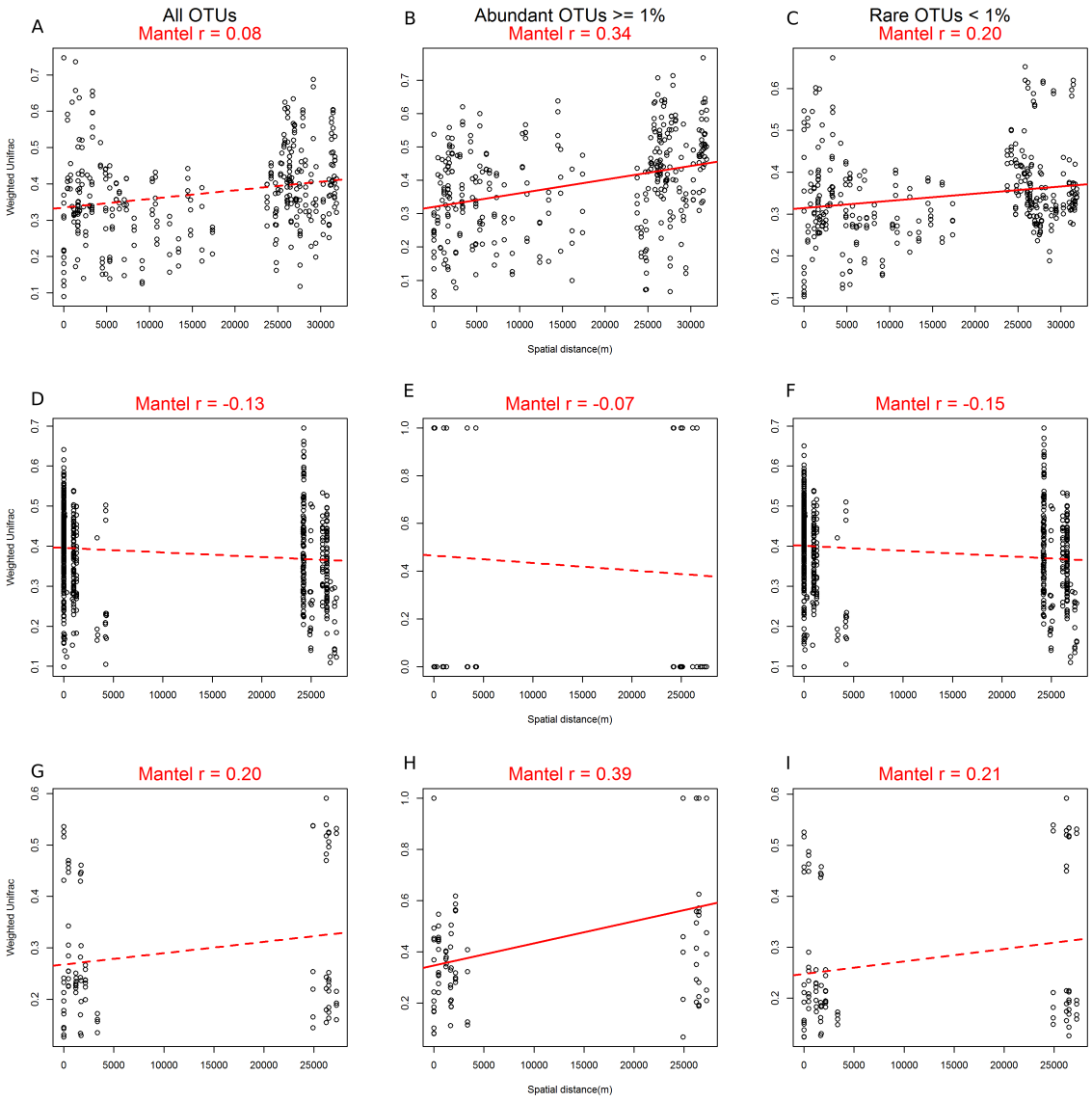

Figure S4.6 The pairwise relationships between weighted Unifrac and spatial distance for all species A), abundant B), and rare species C) in water, same groups in soil D), E), and F), and in sediment G), H), and I) samples. The regression slopes of the linear relationships are shown with solid (statistically significant, Mantel test, 999 permutations, $P < 0.05$) or dashed (not statistically significant) lines. The Mantel r values are reported for each test.

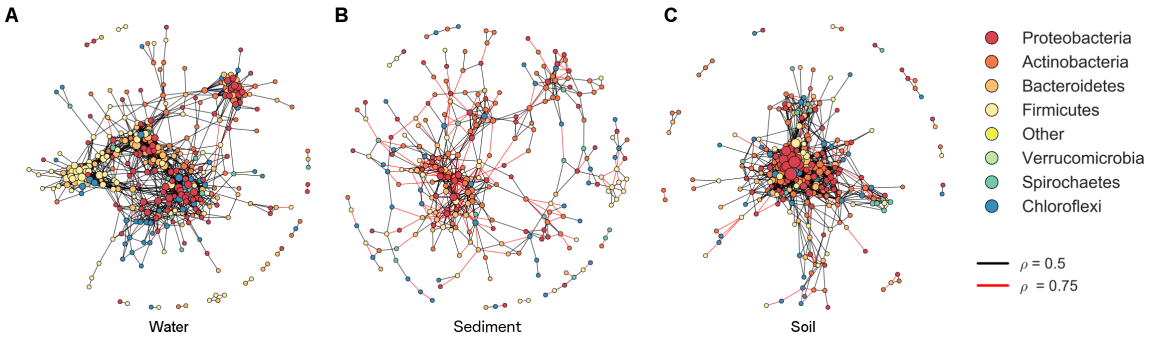

Figure S4.7 Co-occurrence network of interactions between OTUs within water (A), sediment (B), and soil (C). The networks were constructed based on the strengths of ensemble correlation coefficients (Spearman's and Pearson's correlation coefficients both greater than 0.5 or 0.75 for co-occurrence or less than -0.5 for competitive relationships). A cutoff value of less than 0.05 for FDR adjusted p-values in both metrics was used as well. OTU nodes are color coded based on phyla for the eight most abundant phyla observed in all samples.

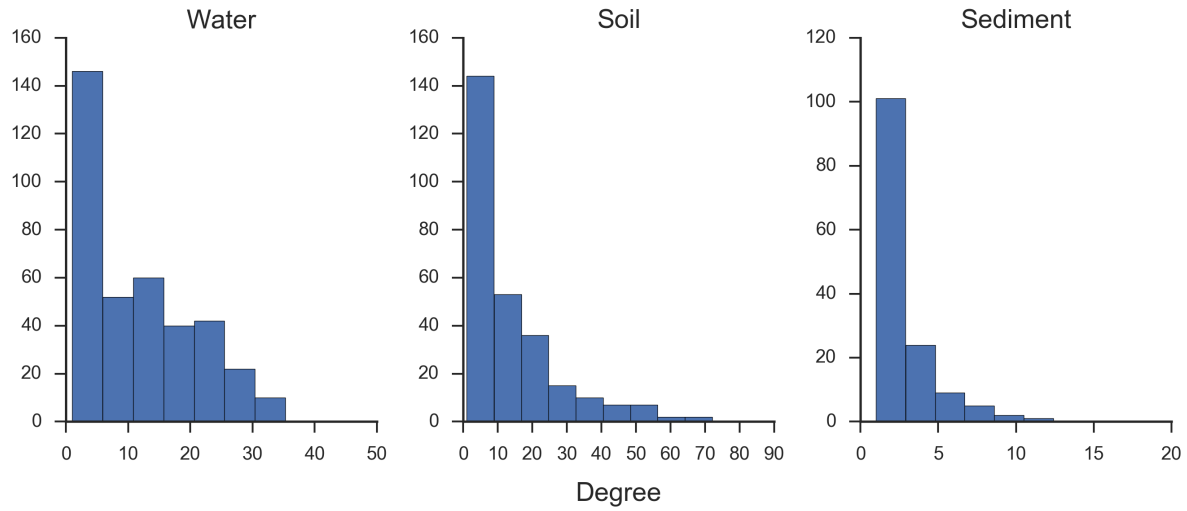

Figure S4.8 The distribution of node degree in water, soil, and sediment.

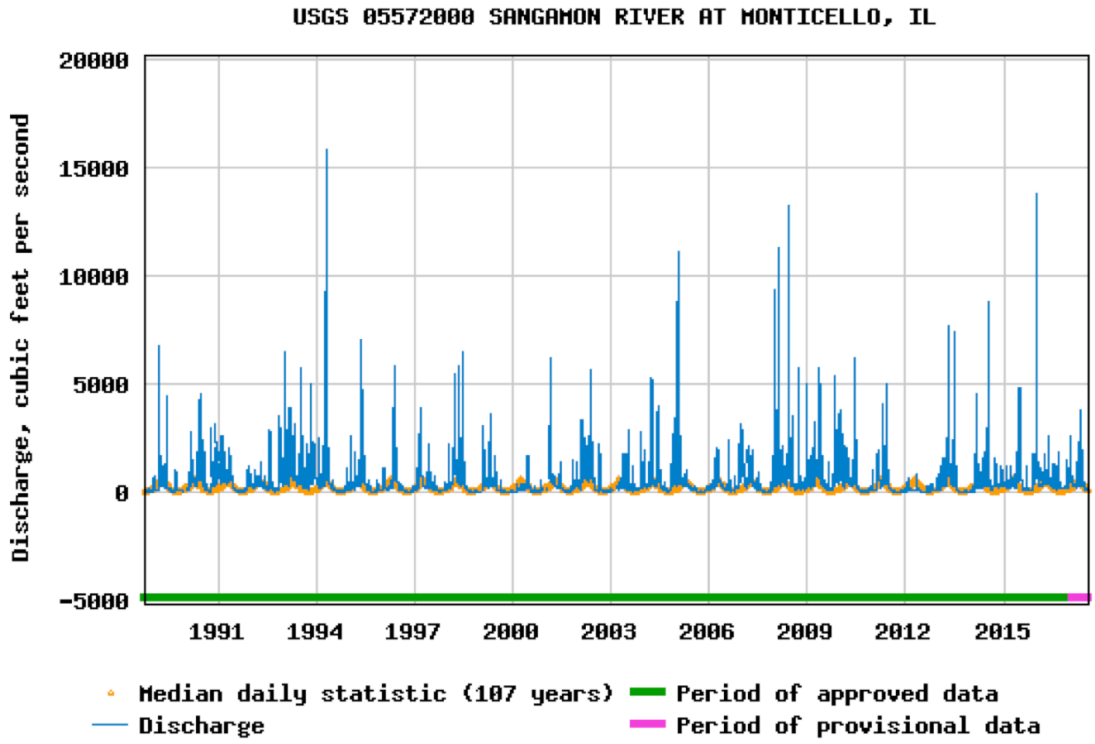

Figure S4.9 30-year flow data from a monitoring station near the downstream sampling locations. Bankfull channel flow is approximately 2000 ft^3 per second based on the channel width at this location.

Supplementary Tables

Table S4.1 Partial RDA Variance Partitioning

Factor	Explained Variance	Grouping
Total Kjeldahl Nitrogen	8.20	Nutrient
Nitrate	9.97	Nutrient
Ortho-phosphate	4.42	Nutrient
pH	1.68	Environmental
Temperature	7.62	Environmental
Conductivity	5.81	Environmental
Upstream	13.60	Spatial
Nutrient & Environmental	6.82	covariation
Nutrient & spatial	0.19	covariation
Environmental and spatial	1.50	covariation
Total	59.80	
By category		

Nutrient	22.59	
Other environmental	15.10	
Spatial	13.60	
Covariation	8.51	
Unexplained	40.20	

Supplementary Table S4.2 Tile and ground water connectivity to terrestrial environments

Reads by Media Type					
OTU	Stream water reads	Ground water mean reads	Tile water mean reads	Soil reads	Sediment reads
OTU_744404	0	0	254	764	0
OTU_578701	0	636.5	0	61	5
OTU_520222	0	177.5	0	36	9
OTU_4429158	0	175.5	0	23	39
OTU_4435984	0	66.5	0	1	0

CHAPTER FIVE

Urban Biorefineries: Resource Recovery Via Bioelectrochemically Derived Hydrogen Peroxide

Material in this chapter is based on the published work:

Griffin, J., Taw, E., Gosavi, A., Thornburg, N.E., Pramanda, I., Lee, H.S., Gray, K.A., Notestein, J.M. and Wells, G., 2018. *ACS Sustainable Chemistry & Engineering*, *6*(6), pp.7880-7889.

5.1 ABSTRACT

In this work, we demonstrate a combined bioelectrochemical and inorganic catalytic system for resource recovery from wastewater. We designed a microbial peroxide producing cell (MPPC) for hydrogen peroxide (H_2O_2) production and used this bioelectrochemically-derived H_2O_2 as a green oxidant for sulfoxidation, an industrial reaction used for chemical synthesis and oxidative desulfurization of transportation fuels. We operated an MPPC equipped with a gas diffusion electrode cathode for six months, achieving a peak current density above 1.4 mA cm^{-2} with 60% average acetate removal and 61% average anodic coulombic efficiency. We evaluated several cathode buffers under batch and continuous flow conditions for solubility and pH compatibility with downstream catalytic systems. During 24-hour batch tests, a phosphate buffered MPPC achieved a maximum H_2O_2 concentration of 4.6 g L^{-1} and a citric acid-phosphate buffered MPPC obtained a moderate H_2O_2 concentration (3.1 g L^{-1}) at a low energy input (1.6 Wh g^{-1} H_2O_2) and pH (10). The MPPC-derived H_2O_2 was used directly as an oxidant for the catalytic sulfoxidation of 4-hydroxythioanisole over a solid niobium(V)-silica catalyst. We achieved 82% conversion of 50 mM 4-hydroxythioanisole to 4-(methylsulfinyl)-phenol with 99% selectivity with a 0.5 mol% catalyst loading in 100 minutes in aqueous media. Our results demonstrate a new and versatile approach for valorization of wastewater through continuous production of H_2O_2 and its subsequent use as a selective green oxidant in aqueous conditions for green chemistry applications.

5.2 INTRODUCTION

Modern wastewater treatment is vital for protecting human and environmental health. However, most design decisions are based on a cost benefit analysis that does not fully value resource recovery. This has resulted in an unsustainable “once-through” model that focuses on

waste removal and disposal. Bioelectrochemical systems (BESs) use a biological catalyst to produce electrical current from wastewater.[22] Most BESs have focused on producing electricity in microbial fuel cell (MFC) configurations; however, there are some examples of cathodic electrochemical synthesis including hydrogen (H_2) or hydrogen peroxide (H_2O_2) production in microbial electrolysis cells (MECs) or microbial peroxide producing cells (MPPCs), respectively.[56, 226] A life cycle assessment of MPPCs found that at scale, MPPC-derived H_2O_2 has nearly 60% lower life cycle greenhouse gas emissions than H_2O_2 produced via the traditional anthraquinone auto-oxidation process.[27] Previous studies on H_2O_2 production in MPPCs[62, 227, 228] have focused on producing alkaline H_2O_2 for disinfection, industrial bleaching or non-specific advanced oxidative processes such as the bio-electro-Fenton process for removal of recalcitrant contaminants from water.[229]

In addition to its proven uses in disinfection and bleaching, H_2O_2 is a promising green oxidant for sustainable chemistry applications because, unlike organic hydroperoxides, it generates water as its only byproduct.[230-232] While the use of H_2O_2 with solid oxide catalysts to selectively oxidize sulfides has been extensively studied,[233-236] there are few examples of selective oxidation in these systems in aqueous conditions, and none involving bioelectrochemically derived peroxide.[237] We recently demonstrated high rates and selectivities in the oxidation of thioanisole and several benzothiophene derivatives by H_2O_2 over a highly dispersed supported niobium(V)-silica (Nb(V)-SiO_2) catalyst in acetonitrile, which outperformed benchmark titania-silica and zirconia-silica materials.[238] Thus, sulfoxidation of a thioanisole derivative over a similarly formulated niobium(V)-silica catalyst was chosen as a model reaction system for proof-of-concept testing of the direct use of bioelectrochemically derived H_2O_2 for selective oxidation in an aqueous buffer (rather than in acetonitrile or other organic solvents).

This approach could be adapted to the epoxidation and/or dihydroxylation of alkenes,[239] the oxidative depolymerization of lignin macromolecules[240], or other H_2O_2-dependent catalytic processes.

H_2O_2 can be generated for these applications via electrocatalytic oxygen reduction. In BESs, anode respiring bacteria anaerobically consume organic matter using an external electrode as a terminal electron acceptor via a metabolic process known as extracellular electron transfer (EET).[241] Current produced during EET can reduce oxygen to water via a four-electron reduction (Equation 1). Oxygen can also be partially reduced to H_2O_2 via a two-electron oxygen reduction reaction that occurs on graphitic cathodes (Equation 2).[242] H_2O_2 can be reduced further, resulting in a net production of water and decreased coulombic efficiency in an MPPC.[243]

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O \;\; E^{0\prime} = 0.81\text{V}^{244} \qquad \text{(Eq. 1)}$$

$$O_2 + 2H^+ + 2e^- \rightarrow H_2O_2 \;\; E^{0\prime} = 0.28\text{V}^{244} \qquad \text{(Eq. 2)}$$

H_2O_2 synthesis in acetate-fed MPPCs is exergonic,[228] although most previous efforts have applied external voltage to improve H_2O_2 synthesis rates and titers at the cost of increasing energy intensity. Rozendal et al. developed the first bioelectrochemical H_2O_2 reactor and achieved concentrations of 1.9 g L^{-1} at an efficiency of 83% and energy input of 0.93 Wh g^{-1} H_2O_2 using an acetate media as feed.[62] Fu et al. demonstrated that H_2O_2 could be generated with a net positive energy production but achieved a maximum concentration of 79 mg L^{-1}.[228]

System design improvements such as minimizing electrode spacing[56] and the use of composite carbon black-graphite-PTFE cathodes[245-247] have led to higher efficiency and lower system overpotentials. Despite recent improvements in system performance and cell design, "pH splitting", or opposing pH shifts at the anode and cathode chambers, remains a major problem limiting MPPC H_2O_2 titer. Proton consumption during oxygen reduction raises catholyte pH and

commonly used ion exchange membranes primarily transport ions other than hydroxide or protons at typical ionic strengths.[248] Elevated cathode pH decreases coulombic efficiency by promoting H_2O_2 decomposition[249] and increases cell overpotential in MPPCs, thus increasing energy input per gram of H_2O_2 produced. In a recent study, 80% of the initial H_2O_2 stored in pH 12.5 NaCl electrolyte solution decomposed in three days.[56] For the latter reason, there is significant motivation for the immediate use of the produced H_2O_2 in a continuous process.

In this study, we present a novel hybrid biological and chemical catalytic approach to wastewater resource recovery via bioelectrochemical H_2O_2 production and subsequent use of the H_2O_2 solution as a green oxidant in chemical synthesis via heterogeneous catalysis. Specifically, we demonstrate thioether sulfoxidation, an industrially relevant process for the synthesis of medicinally relevant sulfoxides and sulfones[250] and for the removal of sulfur compounds from fuels and industrial effluents. First, we optimized MPPC cell design, hydraulic residence time and cathode buffer composition to maximize H_2O_2 concentration at a pH compatible with a silica-based catalyst. We operated an MPPC using a continuous flow cathode for six months and produced H_2O_2 at average effluent concentrations of 3.2 g/L at a 40-hour HRT. Next we characterized the kinetics and selectivity of catalytic sulfoxidation of 4-hydroxythioanisole to 4-(methylsulfinyl)-phenol by H_2O_2 in aqueous buffer solutions and used that to design and operate a system for continuous sulfoxidation using the MPPC effluent.

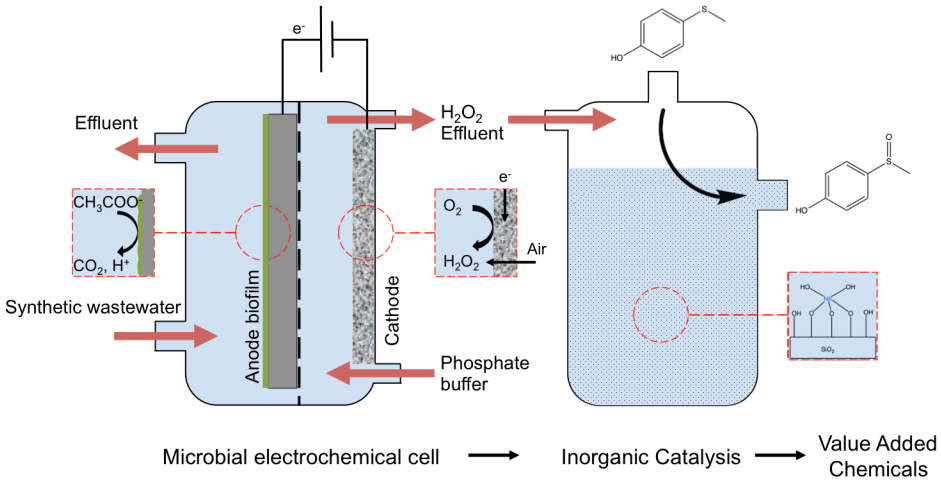

Figure 5.1 Schematic of the hybrid microbial electrochemical cell (MEC) inorganic catalytic system for wastewater valorization. In the MEC stage, acetate or wastewater is consumed by a biofilm in the anode and the resulting current generates H_2O_2 in the cathode. The H_2O_2 rich effluent is used in a heterogeneous catalysis process downstream as a green oxidant.

5.3 Results and Discussion

5.3.1 Microbial Peroxide Producing Cell (MPPC) operation

We operated two MPPCs with a fixed anode potential of -0.3V (vs. Ag/AgCl) and gas diffusion electrodes optimized for H_2O_2 production for six months. Bioelectrochemical current began after 5 days and rapidly entered an exponential growth phase. Current density peaked 10 days after inoculation at 1.4 mA cm^{-2} but declined to an average of 0.8 ± 0.3 mA cm^{-2} for the duration of the experiment. Long-term cell current for one cell is shown in Figure 5.8. Consistent with previously reported long-term BESs, differences in current maxima before and after replacing media were apparent after short downtime periods.[50] Current density varied during operation due to recurring batch acetate removal experiments, feed interruptions, and maintenance breaks during operation. After startup, average acetate removal was 8.7 ± 2.7 mg cm^{-2} day^{-1} (normalized to anode surface area), with a anodic coulombic efficiency of 61 ± 22% (Figure 5.9). During operation, attached growth on the sides of the reactor was observed, likely

due to the growth of microaerobic bacteria that could grow under the hypoxic reactor conditions (<0.1 ppm dissolved oxygen). Periodic removal of attached growth performed before the batch tests indicated by the arrows in Figure 5.9 led to higher coulombic efficiencies and lower overall acetate removal rates.

During MPPC operation, we optimized cathodic H_2O_2 production in batch experiments. MPPCs equipped with carbon black-based gas diffusion cathodes and a 100 mL catholyte reservoir were operated for 24 hours at an anodic setpoint of -0.3V vs Ag/AgCl using several buffers to evaluate overall pH rise and cathodic coulombic efficiency. Results from a 24-hour batch test performed in a 200 mM phosphate buffer are shown in Figure 5.2. Average current density was 1.2 mA cm^{-2}, and the resulting final pH and H_2O_2 concentration were 13.2 and 4610 mg L^{-1}, respectively. While the H_2O_2 concentration was promising, the high cathodic pH led us to consider alternative buffers that would be compatible with downstream supported metal oxide catalysts.

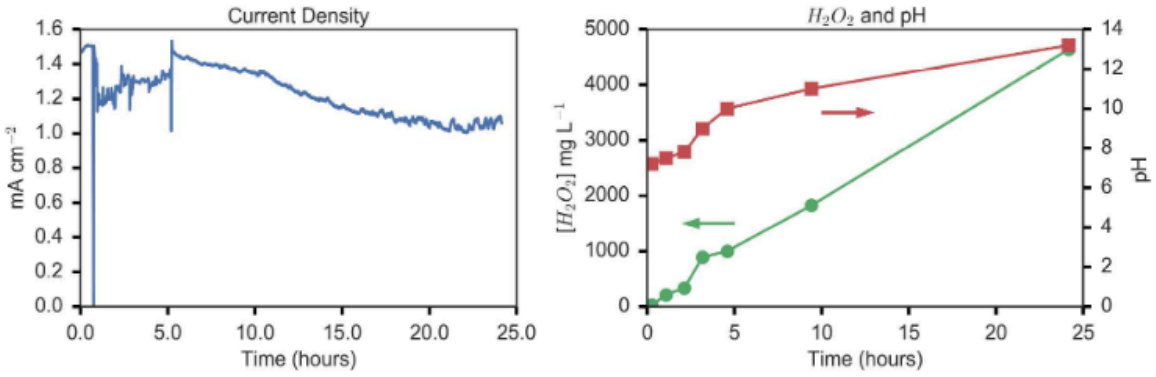

Figure 5.2 Representative plot of a 24-hour batch assay performed in a 200 mM sodium phosphate buffer.

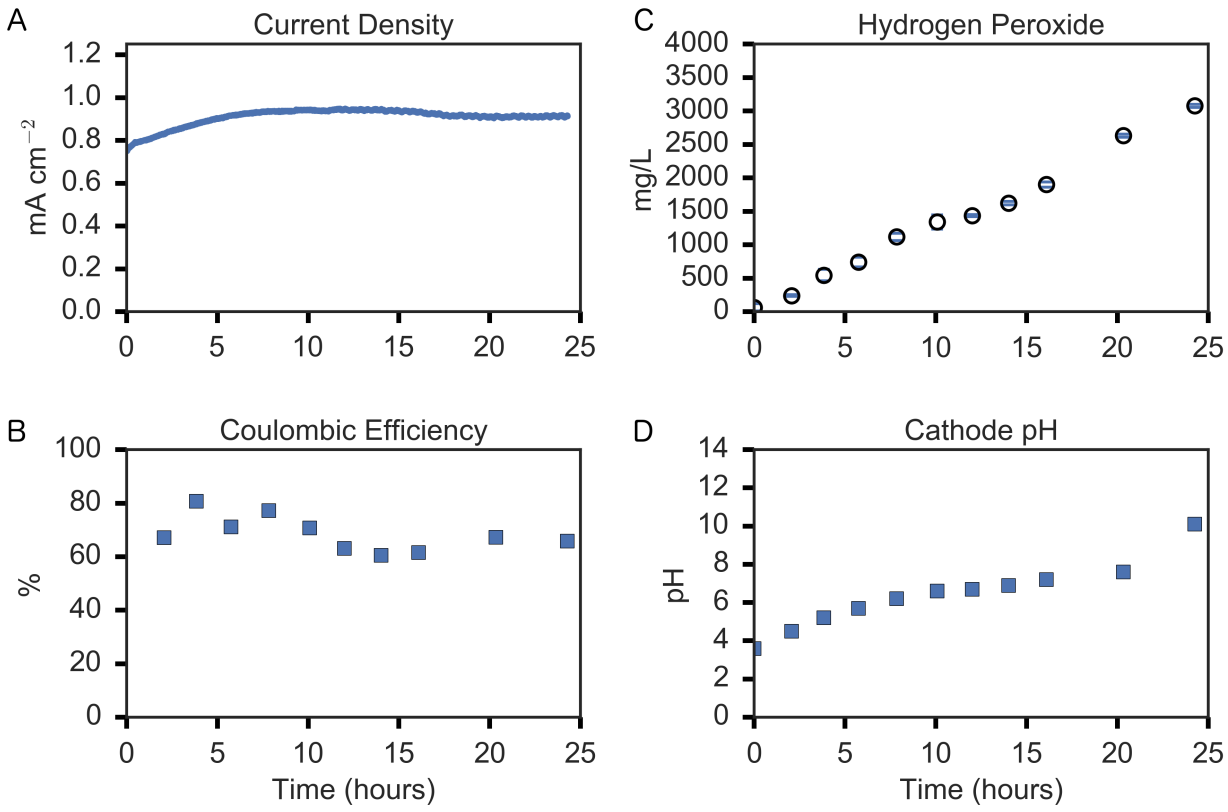

Figure 5.3 H_2O_2 production during a 24 hour batch operation. (A) Cell current was stable over the course of the experiment. (B) H_2O_2 titer reached 3.1 g L^{-1}. Blue error bars represent standard deviation based on three technical replicates. (C) Average cathodic coulombic efficiency was 66%. (D) Catholyte pH increased slowly throughout the batch until spiking to 10.1 when the buffer capacity was exhausted.

Data from a typical run using a pH 3 citric acid/sodium phosphate buffer (132 mM $C_6H_8O_7$: 68 mM Na_2HPO_4) are shown in Figure 5.3. Average current over the course of 24 hours was 0.91 ± 0.4 mA cm^{-2} and anodic coulombic efficiency, defined as the ratio of charge produced to acetate consumed, was 40%. The average H_2O_2 production rate was 127 mg L^{-1} $hour^{-1}$ (12.2 mg cm^{-2} day^{-1} based on cathode area) and the maximum concentration achieved was 3075 mg L^{-1}. Cathodic coulombic efficiency decreased from nearly 100% over the first several hours to 66% by the end of the experiment, and the final overall coulombic efficiency, defined as

the product of anodic and cathodic coulombic efficiencies, was 27%. As shown in Figure 5.3D, catholyte pH increased from a starting value of 3.0 to 10.1 by the end of the batch experiment, with a sharp increase in pH near the end of the experiment as the pH deviated outside of the buffer range.

Electrode potentials and cell voltage over the course of the experiment are shown in Figure 5.4. Applied cell voltage varied between 0.7 and 1.0 V with an average of 0.78 ± 0.11 V. Many MFC and MPPC systems have reported high cathodic overpotentials on plain graphitic electrodes, and the relatively high applied voltage is not unusual at these current densities, despite the favorable thermodynamics of the reaction.[251] Average energy input per gram of H_2O_2 produced was 1.86 Wh g^{-1}. Although cell current was relatively stable at 0.91 mA cm^{-2} over the duration of the experiment (Fig. 2A), cell power input decreased over the course of the experiment from 20 mW to 15 mW (Fig S5A). We observed a 0.2V decrease in cathode overpotential (Fig. S5B) during operation concurrent with a decrease in average cathodic coulombic efficiency. Together, these observations suggest that some overreduction of H_2O_2 to water occurred as the peroxide concentration increased. Cathode activation during oxygen reduction may also contribute to reduced overpotential and decreased cathodic efficiency due to enhanced reduction of H_2O_2 to water compared to a pristine electrode.[243]

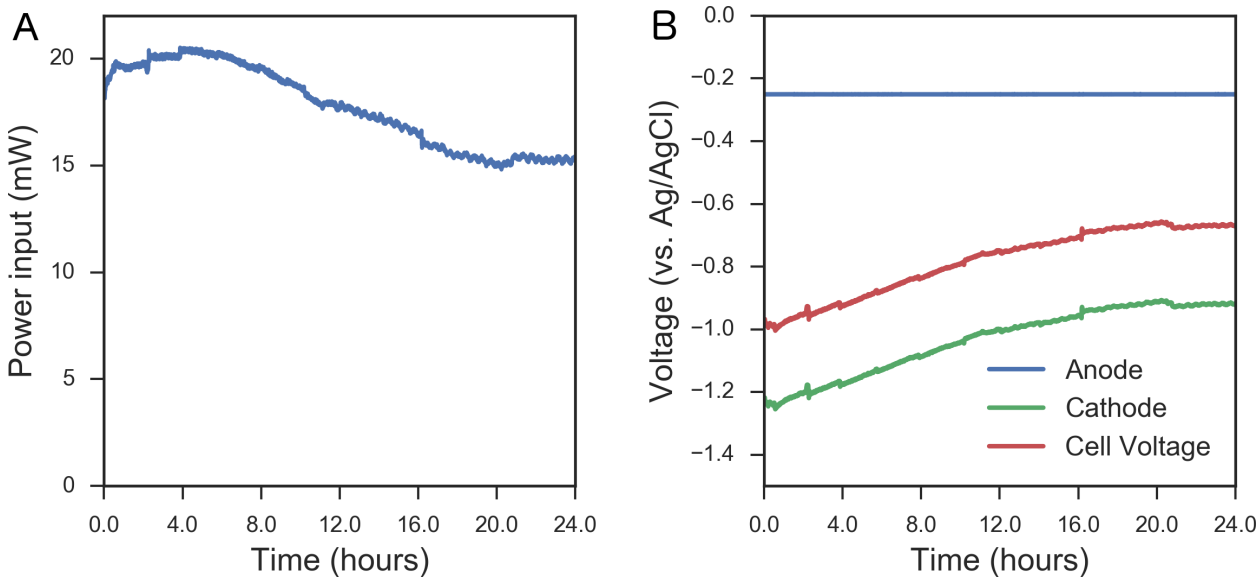

Figure 5.4 (A) Cell power input during the batch H2O2 production experiment and (B) electrode and cell voltage traces.

Although directly comparing performance from bioelectrochemical cells with different operating conditions is difficult, the MPPCs in this study performed favorably in terms of reported H_2O_2 titer and primary energy demand for peroxide compared to previous literature examples (Table S2). Compared to observations by Rozendal et al.[62], our phosphate buffered batch experiment produced three times higher current at a slightly lower applied cell voltage (0.47 vs. 0.53V in Rozendal et al.). One study achieved comparable concentrations of H_2O_2 (2.26 g L^{-1} in 21 hours) using real wastewater as feed, although they used an external energy input of 8.3 Wh g^{-1} H_2O_2, similar to that in the traditional anthraquinone process.[227] Ki et al. recently demonstrated H_2O_2 production from primary sludge at a slightly higher energy input (0.87 Wh g^{-1} H_2O_2) although they obtained a H_2O_2 concentration of 0.23 g L^{-1}.[252] These studies demonstrate the challenges of using less readily biodegradable carbon as feedstocks. A recently reported study achieved considerably higher volumetric H_2O_2 production rates, (18.6 vs. 4.6 g H_2O_2 L^{-1} day^{-1}) due in part to increased cathode surface area: volume ratio (2.72 cm^{-1} vs. 0.25 cm^{-1} in this

study) that reduced the total aerial H_2O_2 production rate needed to achieve similar concentrations.[56] Conventional industrial processes for electrochemically producing H_2O_2 require 4.4 – 8.9 Wh g^{-1},[253] which is significantly higher than the 0.5 – 2.8 Wh g^{-1} reported here. Despite this, significant improvements in energy intensity could be realized through cell design improvements such as decreasing electrode spacing and increasing anode surface area.

Because our primary objective was to assess feasibility and demonstrate proof-of-concept of coupling bioelectrochemical peroxide production to downstream green chemistry applications, we elected to use a synthetic acetate-based wastewater as feed in order to focus on coupling bioelectrochemical peroxide production to downstream applications; however, there are examples of BESs treating real wastewater for extended durations over a year[254] and at pilot scale.[26] The scalability of bioelectrochemical peroxide production via MPPCs is a future research need.

Table 5.1 Comparison of performance in reported literature MEC systems.

Ref.	Process type	Mem-brane	Catholyte HRT Hours	Anolyte	Catholyte	H_2O_2 Concentration g H_2O_2 L^{-1}	Current Density and Voltage mA & V	Energy Intensity Wh g^{-1} $H_2O_2^{-1}$
1	batch	CMI-7000 CEM	4	domestic and hospital wastewaters	50 mm NaCl	0.34	10 A m-2 at 0.6V	2.5
2	batch	CMI-7000 CEM	24	domestic wastewater and and acetate media	50 mm NaSO4	0.2	18.4 A m-2 at 0.04V produced	0.09 output
3	batch	Nafion PEM	12	glucose media w/ 50 mm PBS	50 mm $K3Fe(CN)_6$+PBS	0.08	0.2 A m-2 at 0.4V produced	0.06 output
4	Cont.	CMI-7000 CEM	0.023	12 mm acetate media w/ 50 mm PBS	50 mm NaSO4	8.50E-05	6.1 A m-2 at 0.6V	56
5	batch	Nafion PEM	47	Acetate media	50 mm NaCl	5	2.5 A m-2 at 3.8V	2.3
5	batch	Nafion PEM	21	domestic wastewater	50 mm NaCl	0.08	0.4 A m-2 at 0.9V	1.8
6	batch	Nafion PEM	21	6 mm acetate media/10 mm PBS	50 mm NaCl	9.7	1.7 A m-2 at 11.8V	3
7	batch	CMI-7000 CEM	8	12 mm acetate media/190 mm PBS	50 mm NaCl	1.3	5.3 Am 2 at 0.5 V	0.93
8	batch	CMI-7000 CEM	4-24	5 mm acetate media/50 mm PBS	DI Water	1.3	7.7Am 2 at1V	2.6
8	batch	CMI-7000 CEM	2-10	raw domestic wastewater	DI Water	0.15	0.56 Am 2 at 6.3 V	28
9	Cont.	AMI-7001	4	acetate media	200 mm NaCl	3.1	10.1 A m 2 at 0.31 V	1.1

5.3.2 Abiotic H2O2 buffer optimization

Designing a buffer compatible with both the bioelectrochemical and catalytic system components in this hybrid system presents unique challenges because the buffer species must be electrochemically inert in the cathode potential window, non-reactive with H_2O_2, and compatible with the downstream sulfoxidation process. Highly alkaline conditions reduce H_2O_2 stability[56] and can destroy the active phase of the oxidation catalyst or alter the reactivity in the downstream oxidation reactor. Excessive buffer concentration is also undesirable, however, since excess phosphate could precipitate out in the less-polar environment of the downstream oxidation reactor or otherwise inhibit the oxidation catalyst. Most buffered MPPC systems have used carbonate or phosphate buffers that have effective pH ranges between 5.8-10.8.[251] Despite this, cell overpotential can be reduced in acidic cathodic conditions[255] due to increased availability of protons at the cathode surface and improve oxygen reduction thermodynamics by 59 mV per catholyte pH unit decrease.

Table 5.2 Results of 24-hour batch assays performed with different buffer systems

	Name	Composition	Ionic Strength	Initial pH	Final pH	Final H_2O_2	Cathodic Coulombic Efficiency	Compatible with Nb-SiO2 catalyst?
			mM			g / L	%	
1	(Mono/di)-basic Sodium Phosphate	NaH_2PO_4 / $Na_2HPO_4 \cdot (H_2O)_7$	100	6	12.8	3.2	82%	Yes
2	(Mono/di)-basic Sodium Phosphate	NaH_2PO_4 / $Na_2HPO_4 \cdot (H_2O)_7$	150	6	11.3	2.9	76%	Yes
3	(Mono/di)-basic Sodium Phosphate	NaH_2PO_4 / $Na_2HPO_4 \cdot (H_2O)_7$	300	6	10.7	2.7	52%	No (precipitation)
4	(Mono/di)-basic Sodium Phosphate	NaH_2PO_4 / $Na_2HPO_4 \cdot (H_2O)_7$	500	6	6.9	3.4	76%	No (precipitation)
5	Citric Acid / Sodium Phosphate dibasic	$C_6H_8O_7$ / $Na_2HPO_4 \cdot (H_2O)_7$	120	3	10	3.4	90%	No (no sulfoxidation products observed)

To produce MPPC effluent compatible with downstream green chemistry applications, we optimized buffer conditions with the goal of producing modestly concentrated H_2O_2 solution at near neutral pH while maintaining solubility in the 25% ethanol solutions used for sulfoxidation. We screened pH 6 phosphate buffers at 100-500 mM total phosphate concentration in 24 hour abiotic batch tests along with a 200 mM pH 3 citric acid/phosphate buffer (Table S1). As expected, higher buffer concentration mitigated cathodic pH rise due to proton consumption. Although peroxide decomposition in alkaline conditions is expected, we did not notice a relationship between coulombic efficiency and final pH over the course of the 24-hour trials suggesting that homogenous decomposition of H_2O_2 at high pH was less important than overreduction at the electrode. The citric acid-phosphate buffer system was effective at this stage, however it was not selected as no product formation was observed in the sulfoxidation reactor. The 500 mM phosphate buffer maintained a near-neutral pH (6.9) even after 24 hours of H_2O_2 formation. We selected the 200 mM phosphate catholyte previously screened in the biotic MPPC for downstream usage due to precipitation observed in the 300 and 500 mM phosphate catholytes when diluted with ethanol. Implementation of bioelectrochemical peroxide production for chemical synthesis may require further optimization of redox-stable, low cost buffers compatible with particular catalytic systems. Nonetheless, our efforts demonstrate the feasibility of combining electrochemical H_2O_2 production with downstream catalytic processes in aqueous solvents.

5.3.3. Bioelectrochemical H2O2 generation in a continuously fed MPPC cathode

Batch bioelectrochemical H_2O_2 production has been reported; however, few reports of continuous distributed bioelectrochemical H_2O_2 generation exist in the literature, despite its potential advantages in terms of avoiding decomposition and improved process control. Thus, we

explored H_2O_2 production in a continuously fed cathode chamber. We operated the cathode in batch mode for 24 hours to build up an initial H_2O_2 residual and then switched operation to continuous flow and monitored effluent H_2O_2 concentration for several days until H_2O_2 concentrations reached steady state.

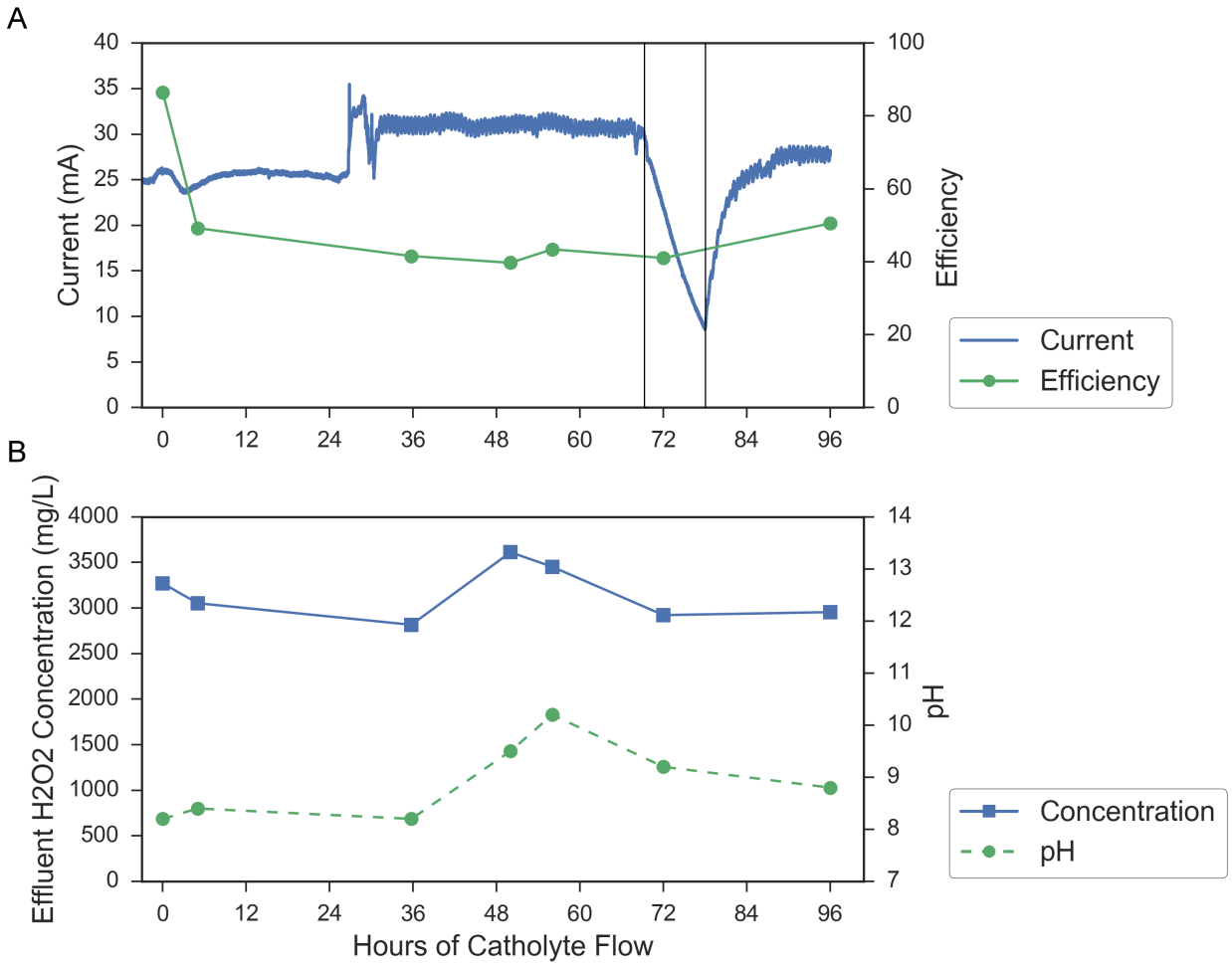

Figure 5.5 Continuous H_2O_2 Production test at a 1.67 day HRT. (A) Current and effluent H_2O_2 concentration shown over a 5 day period. A decrease in current was observed during an anodic acetate consumption test. (B) The decrease in current led to a decrease in effluent H_2O_2 concentration and pH that recovered after restoring anode feed.

Results from a four-day experiment conducted at a cathodic flow rate of 60 mL day^{-1} (39 hour HRT) are shown in Figure 5.5. Cell current was stable for the first three days and averaged 1.1 ± 0.2 mA cm^{-2} over the duration of the experiment. Effluent H_2O_2 concentration averaged 3150 ± 280 mg L^{-1} over the first three days with a maximum concentration of 3610 mg H_2O_2 L^{-1}. Effluent pH varied between 7.9 and 10.1. Current and H_2O_2 concentration partially recovered after a feed disruption on day four indicated by the vertical lines in Figure 5.5, although pH continued to lag somewhat. Effluent pH and H_2O_2 concentration appeared to be correlated (Pearson correlation coefficient R = 0.61) but this relationship was not significant at this sample size. Cathodic coulombic efficiency was calculated by averaging the initial and final H_2O_2 concentrations during each sampling period and then estimating the total H_2O_2 produced based on the flow rate and average concentration. Compared with the 24-hour batch experiments, cathodic coulombic efficiency was lower in continuous cathodic flow operation at this HRT and averaged 44 ± 5%.

We compared effluent concentration and pH over a range of cathode HRTs from 12 to 40 hours to demonstrate how titer and pH can be tuned for different uses by adjusting HRT (Table S3). H_2O_2 concentration and pH increased with HRT from a steady-state concentration 1400 mg L^{-1} at a 12-hour HRT to 3150 mg L^{-1} at 40 hours. Cathodic coulombic efficiency decreased with increasing retention time from 75% at 12 hours to 44% at 40 hours, which was expected given the tendency for H_2O_2 to degrade in MPPC conditions.[56] We did not evaluate HRTs beyond 40 hours but it is likely that efficiency would decrease at longer residence times due to decomposition. Few studies have investigated continuous bioelectrochemical H_2O_2 production. However, Li et al. operated BESs for 10 minute increments with and without catholyte flow and identified a tradeoff between current and coulombic efficiency they ascribed to H_2O_2

overoxidation.[247] Young et al. operated a system similar to ours with a four hour HRT which achieved lower energy intensity, in part due to lower HRT and resulting higher coulombic efficiency.[56] Further optimization of cell design to minimize electrode spacing, cathode surface area to volume ratio and increase anode surface area could increase titer and reduce the need for long HRTs used in this study.

5.3.4 Conversion, yield and kinetics of H2O2 assisted sulfoxidation on niobium catalyst

We have previously described a highly dispersed niobium(V)-silica (Nb(V)-SiO_2) catalyst for alkene epoxidation[256, 257] and sulfoxidation with H_2O_2 oxidant in acetonitrile.[238] An equivalent catalyst material was prepared here, and its high optical edge energy of 3.6 eV from diffuse reflectance UV-visible spectroscopy (Figure 5.11) indicates the high dispersion of the niobia phase that is correlated with high activity[256, 257, 258, 259]. We hypothesized that a similar selective sulfoxidation reaction could be catalyzed with dilute MPPC-derived H_2O_2 in an aqueous buffer, rather than in an organic solvent. There are few reports of sulfoxidation by solid oxide catalysts in protic solvents[260] and to the best of our knowledge, no previous work using these materials in high ionic strength aqueous solutions such as the MPPC catholyte used in this study. Therefore, we evaluated the kinetic parameters of this reaction across a range of temperatures and catalyst loadings using mock MPPC catholyte and then compared these results with MPPC derived H_2O_2 for a subset of these conditions.

Concentration time course plots for the disappearance of 4-hydroxythioanisole catalyzed by Nb(V)-SiO_2 using mock MPPC catholyte is shown in Figure 5.6A, and the relation between rate constant and catalyst loading is shown in 5.6B. 4-hydroxythioanisole disappearance kinetics fit well to an assumed pseudo-second order model (i.e. pseudo-first order with respect to each the sulfide and H_2O_2), as has been shown for related systems,[238, 261-263] and the resulting rate

constants increase linearly with increasing catalyst loading, also as expected. This is a generally useful finding, given the lack of studies of such catalysts in high ionic strength aqueous solutions.

The uncatalyzed system, corresponding to the 0 mol% data in Figure 5.6, revealed some activity due to H_2O_2 alone. The rate constant for the uncatalyzed reaction was subtracted from the overall rate observed in the catalyzed system to determine an apparent rate constant for just the Nb(V)-SiO_2-catalyzed reaction. In our previous work with a similar Nb(V)-SiO_2 catalyst in acetonitrile (rather than in an aqueous solution, as in this study),[238] complete oxidation of thioanisole to methyl phenyl sulfone was observed for a system with a 3:1 initial H_2O_2:thioanisole ratio. Significant accumulation of the sulfoxide intermediate was observed at low conversion, and the selectivity (k_1/k_2) was 2.4. Given the much lower excess H_2O_2 (1.1:1 starting ratio) in this study, over-oxidation to 4-(methylsulfonyl)-phenol sulfone was unexpected and not observed.

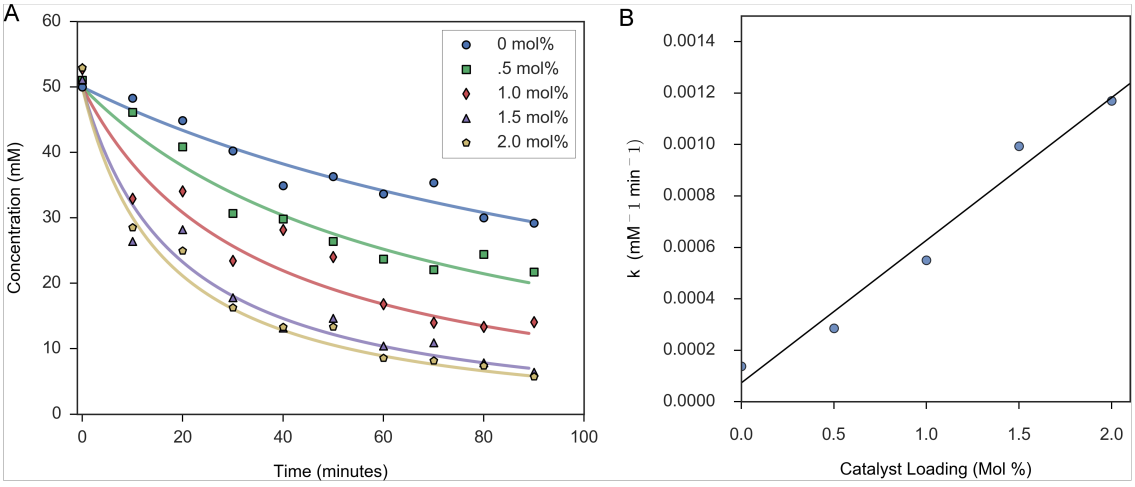

Figure 5.6 (A) Reactant concentration time course data at different catalyst loadings as measured by GC-FID. Solid lines are best fits to an overall 2nd order reaction rate that is first order w.r.t. both 4-(methylmercapto)phenol and H_2O_2 concentration. (B) Second order rate constant fits as a function of catalyst loading.

Arrhenius plots were developed for reactions carried out at 25 °C, 35 °C and 45 °C and are shown in Figure 5.11 for both the 0.5 mol% Nb(V)-SiO_2-catalyzed and uncatalyzed systems. As seen in Figure 5.7, initial reaction rates were highest at 55 °C but the Nb(V)-SiO_2 appeared to lose activity over time at this temperature; these data were not included in the activation energy calculations. A yellow tint was observed at the end of the 55 °C experiment, suggesting gradual leaching of the niobia active phase by chloride at this temperature, as has been observed previously for related catalysts.[264] We have previously assessed leaching and recyclability of the Nb(V)-SiO_2 catalyst used in this study for epoxidation of cyclohexene at 65 °C.[256] Some leaching was observed, but recalcining spent catalysts at 550 °C in air between batch experiments recovered the selectivity and most of the reactivity of the initial material. Activity tests by others on similar Nb-SiO_2 catalysts demonstrated gradual activity losses over five catalytic cycles, but found that these materials were relatively robust in the presence of aqueous H_2O_2.[265] Additional studies of long-term catalyst stability and the capacity of the anion exchange membrane to shield the catholyte from contamination with wastewater anions is necessary to ensure performance in real wastewater operating conditions.

The apparent activation energy of the uncatalyzed and Nb-catalyzed system is 52.0 kJ mol^{-1} and 44.8 kJ mol^{-1}, respectively. The apparent catalyzed barrier here is 5-10 kJ mol^{-1} higher than other activation energies previously reported in organic solvents,[238] suggesting a slight enthalpic penalty incurred by competitive binding of anions and water to Nb active sites that may occur under aqueous ionic conditions.

5.3.5 Selective sulfoxidation of 4-hydroxythioanisole using MPPC derived H2O2

Finally, we combined separate knowledge of the MPPC and the sulfoxidation catalyst in order to demonstrate a hybrid MPPC-catalytic process for sulfoxidation using bioelectrochemically derived H_2O_2 as a novel green oxidant. We demonstrated both a batch process as well as the continuous two-stage reactor shown in Figure 5.1. In the batch process, MPPC catholyte generated over a 24-hour period was neutralized to pH 7 with HCl. The MPPC effluent was diluted to 83 mM H_2O_2 with unused phosphate buffer and mixed 3:1 with an ethanol containing the catalyst and 4-hydroxythioanisole. Batch sulfoxidation experiments were performed as above and product distribution was quantified with GC-FID. Figure 5.7A shows a typical time course of 4-hydroxythioanisole sulfoxidation using MPPC derived H_2O_2 at 35 °C along with results from experiments using mock MPPC catholyte and commercial H_2O_2 performed between 25-55 °C. The system using MPPC derived H_2O_2 reached 82% yield after 2 hours with a selectivity of 99%. Compared to the mock MPPC effluent, the 4-hydroxythioanisole sulfoxidation rate constant differed by only 18% in actual MPPC catholyte (k_{app} = 4.8x10^{-4} mM^{-1} min^{-1} vs. 5.9x10^{-4} mM^{-1} min^{-1}), possibly due to higher phosphate and chloride concentrations. We analyzed catholyte samples for acetate leaching through the anion exchange membrane after 24 hours of H_2O_2 generation but observed low acetate concentration (<5 mg L^{-1}) in all samples. We did not directly quantify catholyte phosphate concentration after H_2O_2 production here, though previous studies have found no significant change in catholyte conductivity during MFC operation.[266]

In the continuous reactor, H_2O_2 generated in the MPPC was directly used downstream in a CSTR-style sulfoxidation reactor fed with 4-hydroxythioanisole. After allowing the MPPC effluent catholyte H_2O_2 concentration to reach steady state, we collected and neutralized 100 mL

of catholyte, and charged the sulfoxidation reactor with the same initial peroxide and reactant concentrations used in the batch experiments. Using the rate constant determined from the batch system, we selected a two hour HRT with a 50% 4-hydroxythioanisole solution: 50% diluted MPPC catholyte flow split as feed to the CSTR and monitored product concentrations over time for six hours. 5-minute composite effluent samples were collected for each time point and remaining H_2O_2 was neutralized with sodium bisulfite. The rate constant was determined from the 4-hydroxythioanisole reactor effluent time series data by numerically integrating the CSTR design equation as described in the Methods section. Figure 5.7B shows effluent reactant concentrations along with a best-fit line to an overall second order rate law. After curve fitting, the rate constant was found to be $k = 5.8 \times 10^{-4} \pm 0.1 \times 10^{-4}$ mM^{-1} min^{-1} within a 95% confidence interval. This value agrees with the rate constant obtained from the stand-alone batch reactor under similar conditions. Conversion reached approximate steady state after two hours and then declined slightly, possibly due to catalyst loss through the effluent filter. The agreement between continuous and batch performance demonstrates the ability to design and operate a tandem MPPC-catalysis system.

The tandem MPPC-catalysis system described here provides the first demonstration of usage of MPPC derived H_2O_2 for industrially relevant chemical transformations beyond disinfection or nonselective advanced oxidation processes. It provides proof-of-concept of a route of valorization of wastewater organics via onsite H_2O_2 generation coupled to direct use for bio-based chemical production. Direct on-site generation of H_2O_2 reduces the need for transport and storage of concentrated H_2O_2 solutions and has already been demonstrated (for abiotic H_2O_2 production) for paper bleaching[253] and electro-Fenton based oxidation of organics.[267]

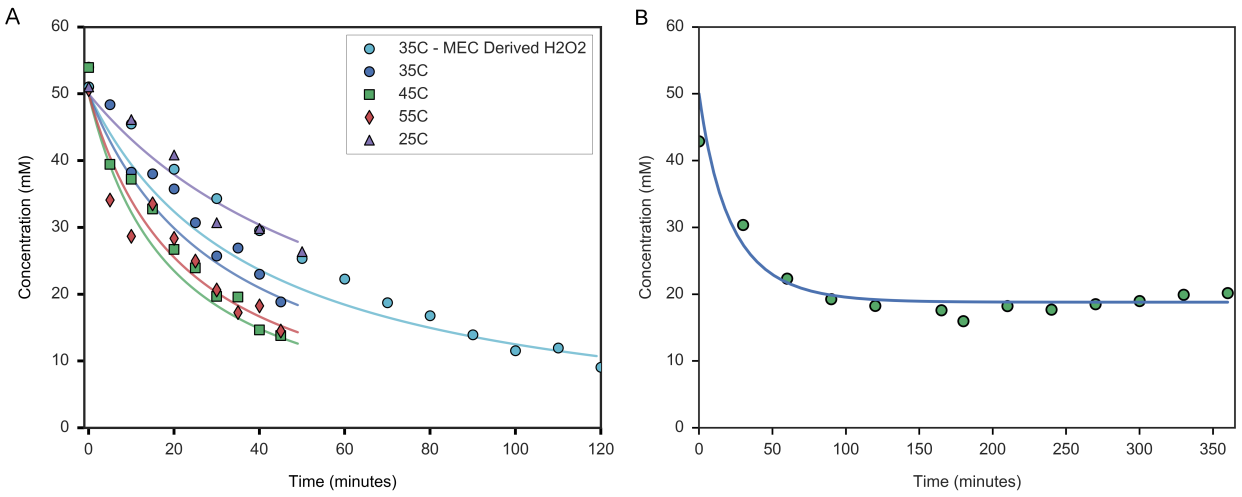

Figure 5.7 (A) Typical time course of sulfide consumption using MEC derived H_2O_2 at 35°C along with results from synthetic H_2O_2 experiments taken between 25-55°C. **(B)** Performance of a 20 mL sulfoxidation CSTR operated at 35°C and a two hour HRT. The solid line represents the best fit line from a second order overall reaction rate model fit to the CSTR data.

MPPC-based resource recovery from real wastewater may benefit from a "closed loop" system in which produced H_2O_2 can be used on-site, further reducing waste treatment and oxidant costs. BESs excel in low-temperature and lower organic loading scenarios[52] where mature technologies like anaerobic digestion do not function well. Additionally, recent studies have demonstrated successful BES treatment of hypersaline produced water that poses problems for traditional biological treatment processes.[268] However, it should be noted that the complex nature of real wastewater provides challenges for BESs such as competing electron-accepting reactions like sulfate reduction and methanogenesis which can reduce coulombic efficiency.[254] Treatment of complex organics would likely require longer HRTs and anode biofilm adaptation to these conditions[252], or a multi-stage treatment process in which complex organics are partially fermented prior to resource recovery in a BES.[269]

5.4 CONCLUSION

In this work we demonstrate a new pathway for resource recovery, as well as shown the flexibility of this solid oxide catalyst for applications in high ionic strength aqueous media. This differs substantially from previous demonstrations for this catalyst in organic solvents (e.g. acetonitrile).[238, 265] Substitution of "greener" solvents has benefits in terms of cost as well as environmental and human health.[270] This overall scheme could be tuned to other reactions such as dihydroxylation of alkenes and/or epoxidation,[239] as well as other H_2O_2-dependent catalytic processes. Building on the success of this proof-of-concept system for green chemistry using MPPC derived H_2O_2, performance could likely be improved by more tightly coupling the catalyst and MPPC design, including electrode construction, buffer choice, and catalyst architecture. Furthermore, catalytic systems that are unaffected or enhanced by the alkalinity inherent to bioelectrochemically derived H_2O_2 are a promising application.

5.5 METHODS

5.5.1 MPPC Reactor Configuration

Bioelectrochemical cells were constructed from laser-cut acrylic with an anode chamber volume of 200 mL equipped with an Ag/AgCl reference electrode (BASI, West Lafayette, IN). During MPPC operation, a 25 cm^2 piece of AvCarb carbon felt (Fuel Cell Earth, Woburn, MA) with a copper mesh current collector was used as the anode. The carbon felt was pretreated for 24 hours in 1N nitric acid, followed by 24 hours in acetone and 24 hours in ethanol.[271] During abiotic cathode optimization, the copper-backed carbon felt was replaced by a 10 cm^2 Pt mesh (Sigma) that was used as a counter anode. The cathode chamber had a 100 mL chamber with an exposed cathode surface area of 25 cm^2. A gas diffusion cathode was constructed as previously described[245] using a carbon cloth electrode (GDL-CT, Fuel Cells Etc., College Station, TX)

coated with PTFE on one side and 5 mg cm^{-2} carbon black (Vulcan XC-72, Fuel Cell Earth, Woburn, MA) dispersed in Nafion on the other side. The chambers were separated by a 33 cm^2 anion exchange membrane (Ultrex AMI-7001, Membranes Intl., Ringwood, NJ). The electrode spacing was 3.4 cm. Both chambers contained inlet and outlet ports to operate in continuous mode. During continuous H_2O_2 production and usage, cathode effluent was stored in a holding tank and neutralized and diluted before being used for sulfoxidation. A schematic of the combined MPPC and sulfoxidation reactor setup is shown in Figure 5.1, and photographs of the reactor setups are available in Figure 5.8.

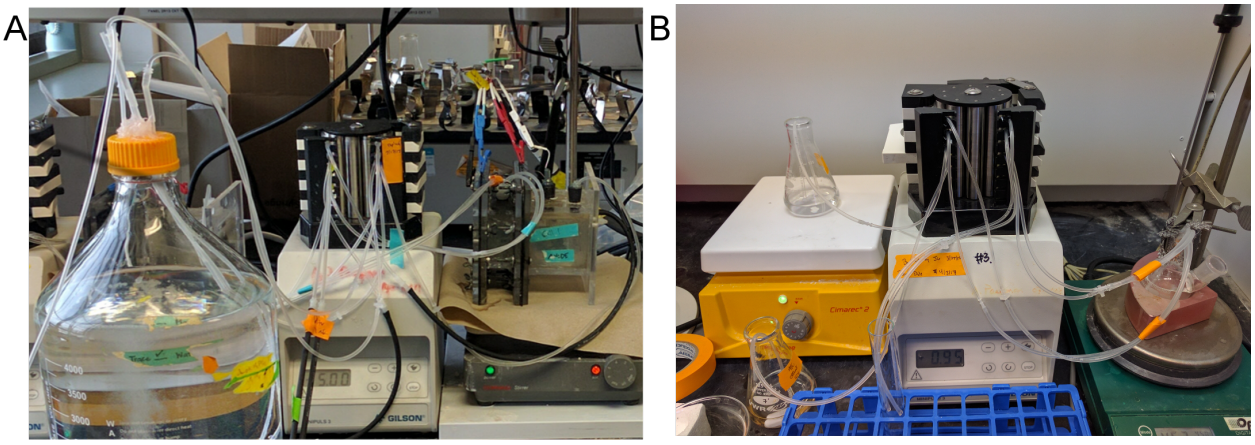

Figure 5.8 Reactor Setups. (A) MEC reactor with continuous cathode chamber on the left and bioanode on the right. (B) Sulfoxidation reactor using MEC derived H_2O_2.

5.5.2 MPPC Operating Conditions

The anodic biofilms used in this study were inoculated with 80 vol% synthetic acetate media, 10 vol% primary effluent from the O'Brien Water Reclamation Plant (Skokie, IL), and 10 vol% effluent from a previously operated MPPC. The anode potential was set at -0.3V vs. Ag/AgCl using a VMP-3 potentiostat (Bio-logic USA, Knoxville, TN) and cell current, working and counter electrode potentials were monitored every 30 seconds. The anode was fed with a synthetic wastewater medium containing 25 mM CH_3COONa in a 100 mM phosphate buffer.[272]

Each liter of anode medium also contained 0.3 mL of a trace metal solution.[273] MPPC anodic hydraulic residence time (HRT) was initially 12 hours and then decreased to 5 hours as acetate consumption rates increased to avoid substrate limitation.[274] The medium was continuously sparged with N_2 gas to minimize oxygen introduction into the system. The reactor was operated continuously for over six months, during which time cathode HRT and buffer concentration were varied to study the effect of these variables on efficiency and effluent H_2O_2 concentration. All MPPC experiments were performed at room temperature (22 ±1°C) with continuous mixing.

To identify buffer systems compatible with downstream catalytic processes, H_2O_2 production was tested in abiotic cells equipped with Pt anodes. H_2O_2 production rates, cathodic coulombic efficiency and catholyte pH increase were evaluated in 24 hour batch assays with several phosphate and citric acid buffer compositions, listed in Table S1. Cathode potential was fixed at -0.6V vs. Ag/AgCl and samples were collected for pH and H_2O_2 measurement. During continuous MPPC operation, catholyte chambers were triple rinsed with deionized water and then allowed to accumulate H_2O_2 for 24 hours before starting continuous flow. Effluent H_2O_2 concentrations were monitored until they reached steady state, and effluent was collected for sulfoxidation.

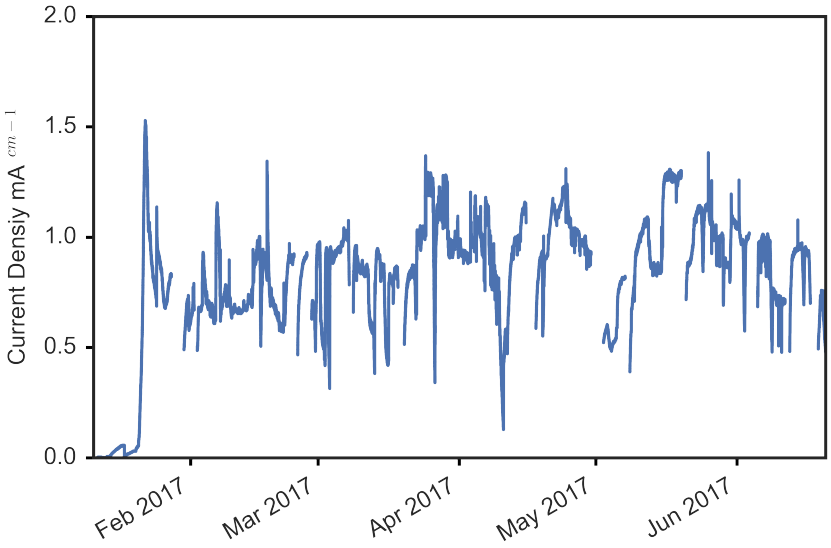

Figure 5.9 Longterm cell current performance of MEC operated at fixed anode potential of -0.3V (vs. Ag/AgCl) for 5 months. Bioelectrochemical current began after 5 days and rapidly entered an exponential growth phase. Current peaked 10 days after inoculation at 1.4 mA/cm^2 but declined to an average of 0.8 ± 0.3 mA/cm^2 for the duration of the experiment.

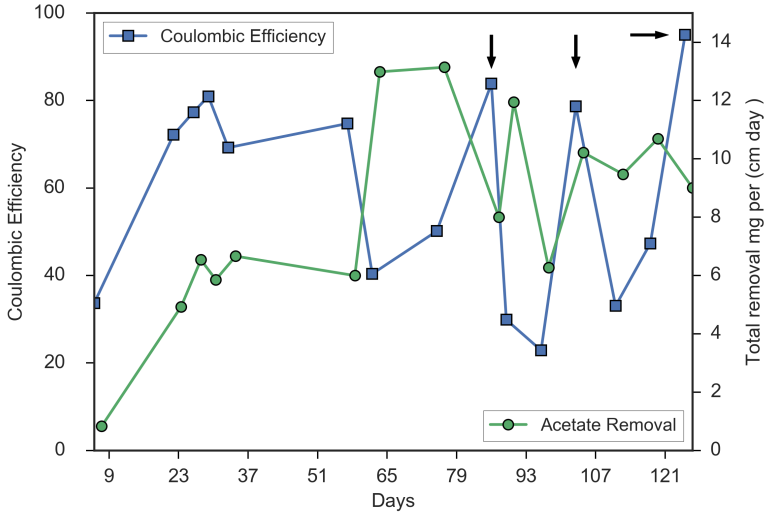

Figure 5.10 Biofilm areal acetate removal rate (mg cm^{-2} day^{-1}) and anodic coulombic efficiency for 5 months of reactor operation. Arrows indicate tests performed after removal of attached growth.

Table 5.3 H_2O_2 Production vs. Cathode HRT. Results from MEC cathode HRT experiments performed with, 12, 25, and 40 hour HRTs. For each experiment, the cathode chamber was triple rinsed and then operated in batch mode for 24 hours, then switched to continuous flow at the specified HRT for 4 days. For the 12 and 25 hour HRT experiments, only data from the last two days was averaged once steady state was reached.

HRT	Flowrate	Current Density*	Effluent pH	Effluent H2O2	Cathodic Coulombic Efficiency
hours^{-1}	**ml day^{-1}**	**mA cm^{-2}**		**mg L^{-1}**	**%**
12.2	197	0.99 ± 0.03	7.1 ± .1	1360 ± 200	75 ± 6**
25.2	95	0.93 ± .04	8.7 ± .2	2490 ± 200	67 ± 4**
39.8	60	1.10 ± 0.2	8.9 ± .7	3150 ±280	44 ± 5

* Average ± Standard Deviation

** Average over final two 48 hours (pseudo-steady state)

5.5.3 Analytical Methods

A Trace 1310 gas chromatograph (GC) (Thermo Scientific, USA) equipped with a diphenyl dimethyl polysiloxane column and an FID detector was used to measure anolyte acetate concentrations. A 0.1 μL sample was injected in the GC with a 50:1 split ratio under a constant flow rate of 1.0 ml min^{-1}. The GC oven was held at 70 °C for one minute, then increased at 10 °C per minute up to 180 °C, and then held at a constant temperature of 180°C for five minutes. 24 hour anodic acetate removal rates were calculated by measuring the difference between acetate medium and effluent concentrations at a given HRT. H_2O_2 concentration was measured in the catholyte effluent using the ammonium metavanadate method.[275] A vanadate colorimetric reagent was prepared by slowly adding 9 M sulfuric acid to a 12.4 mM ammonium metavanadate solution under magnetic stirring at 50°C until complete dissolution. Triplicate 0.75 mL catholyte samples were diluted to 20-200 mg L^{-1} H_2O_2 in deionized water and added to an equal volume of the vanadate solution. Absorbance was measured at 450 nm on a BioSpectrometer UV-Vis spectrophotometer (Eppendorf). Anodic and cathodic coulombic efficiencies were calculated

based on the total current and measured changes in acetate and H_2O_2 concentrations using equations 3 and 4.[244]

Anode coulombic efficiency (%): $$\frac{\int Idt}{\mathrm{nF}\,\Delta[H_3COO^-]V_{anode}} \quad \text{(Eq. 3)}$$

Cathode coulombic efficiency (%): $$\frac{\mathrm{nF}\int[H_2O_2]Qdt}{\int Idt} \quad \text{(Eq. 4)}$$

where I is the total current, n is the electron equivalents per mole of H_2O_2 or acetate (2 and 8, respectively), $[CH_3COO^-]$ and $[H_2O_2]$ are the concentrations of acetate and H_2O_2, respectively, in mol L^{-1}, V is the anode volume and Q is the volumetric flow rate of catholyte. Overall coulombic efficiency was calculated by multiplying the anodic and cathodic coulombic efficiencies together. Aerial current and power density were calculated by normalizing current and power to the anode and cathode surface area (25 cm^2). Batch cathodic coulombic calculations were performed using the volume of catholyte remaining in the reactor while sampling to account for changes in catholyte volume during the batch assays.

5.5.4 Catalyst Synthesis and Characterization

The niobia-silica catalyst was synthesized by grafting a niobium(V)–calixarene coordination complex to partially-dehydroxylated silica gel (Alfa Aesar). 4.65 mmol $NbCl_5$ (Strem Chemicals) and 4.65 mmol *p*-tert-butylcalix[4]arene (Sigma) were dissolved in 150 mL anhydrous, degasified toluene inside of an Ar glovebox. After sealing and transferring the flask to Schlenk line under N_2 purge, a reflux condenser was attached, and the solution was refluxed for 24 h. Then, 15.0 g silica gel, which was previously partially dehydroxylated at 300 °C for 10 h under dynamic vacuum (<25 mTorr), was added to the flask under N_2 purge, and the suspension was refluxed for an additional 24 h. Solids were vacuum-filtered in air, then washed with 5×100 mL toluene and dried under dynamic vacuum overnight. Dried, as-made solids were

sieved to <125 μm, and the sieved material was calcined in air immediately before use as catalyst at 550°C for 6 h using a ramp rate of 10 °C min^{-1}.

Catalyst metal content was quantified using inductively coupled plasma atomic emission spectroscopy (ICP-AES; Thermo iCAP 7600). Prepared Nb(V)-SiO_2 catalysts were digested in 48 wt % HF, diluted with 0.9 wt % HNO_3 and calibrated against a standard curve of a commercial Nb solution (Fluka Analytical) in 0.9 wt % HNO_3. Diffuse reflectance UV-visible (DRUV-vis) spectra were collected from 800–200 nm at ambient conditions using a Shimadzu UV-3600 with a Harrick Praying Mantis diffuse reflectance accessory, using polytetrafluoroethylene (PTFE) powder as a baseline perfect reflector and a 20:1 catalyst diluent.

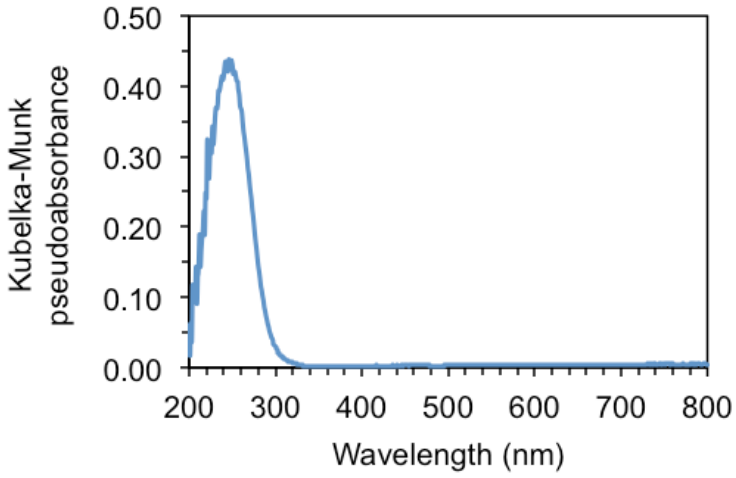

Figure 5.11 Diffuse reflectance UV-visible spectrum of Nb-SiO_2 catalyst.

5.5.5 4-hydroxythioanisole Oxidation

4-hydroxythioanisole oxidation kinetics were first investigated across a range of temperatures, catalyst loadings and aqueous buffer conditions using commercial H_2O_2 in a "mock MPPC catholyte" as follows: 1 mmol 4-hydroxythioanisole (Sigma) and 0.1 mmol phenol (internal standard, Sigma) were dissolved in a 2:1 deionized water: ethanol solution. Mock

MPPC catholyte was produced by pH adjusting a 200 mM pH 6 phosphate buffer (21 g L^{-1} KH_2PO_4 (Sigma) and 6.6 g L^{-1} $Na_2HPO_4 \cdot (H_2O)_7$ (Sigma)) to pH 13 with NaOH and then neutralized with HCl to pH 7 to mimic the neutralization procedure used for the MPPC catholyte and achieve a similar chloride concentration in the mock and actual MPPC catholyte. 31.1 mg Nb(V)-SiO_2 (0.16 mmol Nb g^{-1}) was then added to a solution containing 15 mL of the 4-hydroxythioanisole solution and 5 mL of mock MPPC effluent. Finally, 110 μL of H_2O_2 (30 wt % aqueous, 1.07 mmol) was added to start the reaction. Experiments were run between 25°C and 75°C on a hotplate and magnetically stirred at 500 rpm. At the start of reaction, the 4-hydroxythioanisole concentration was 50 mM and the molar ratios inside of the reactor were 100:107:0.5 for 4-hydroxythioanisole:H_2O_2:Nb. Reaction aliquots (approx. 200 μL) were quenched with approximately 5 mg $NaHSO_3$ (Sigma) at desired time points before GC analysis. Product identities were initially determined using GC-MS (Shimadzu QP2010, Zebron ZB-624 capillary column) and quantified using GC-FID (Shimadzu 2010, TR-1 capillary column) via calibrated standards. The reaction was assumed to be first order in both H_2O_2 and 4-hydroxythioanisole, giving second order overall, and rate constants were found by a satisfactory linear regression of

$$\ln \frac{M - X_A}{M(1 - X_A)} = C_{A0}(M - 1)kt \qquad \text{(Eq. 5)}$$

where $M = \frac{C_{B0}}{C_{A0}} = 1.1$, k is the apparent rate constant, and A and B as 4-hydroxythioanisole and H_2O_2, respectively. Apparent activation barriers were calculated from the slope of ln(k) vs. 1000/T (K) conforming to an Arrhenius model.

4-hydroxythioanisole oxidation kinetics were quantified with MPPC-derived H_2O_2 using a similar procedure. The reactor sampling and GC quantification were as described above, but the mock MPPC catholyte and 30 wt% H_2O_2 were replaced with H_2O_2-laden MPPC effluent

adjusted to pH 7 with HCl. The volumes of MPPC effluent and deionized water added to the 4-hydroxythioanisole solution were adjusted to give initial molar ratios of 100:110:0.5 for 4-hydroxythioanisole:H_2O_2:Nb in the reactor.

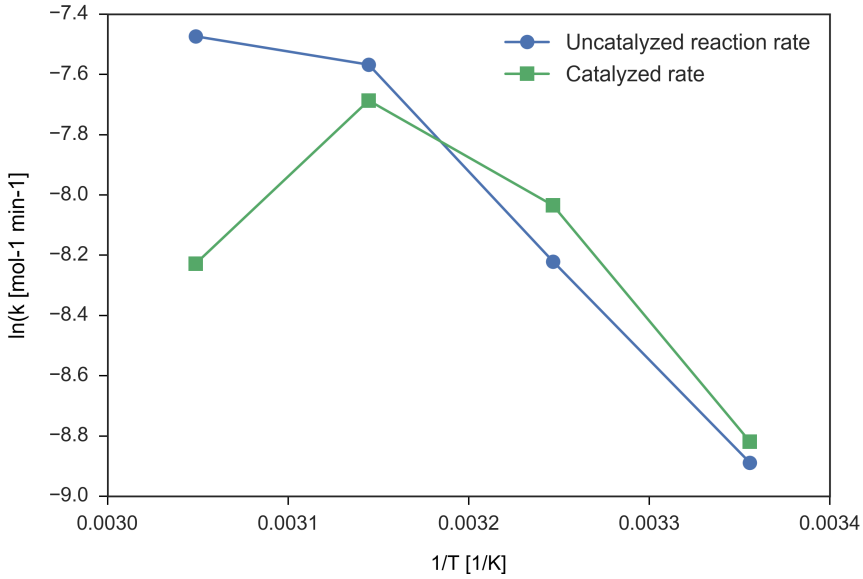

Figure 5.12 Arrhenius plot of (blue) uncatalyzed and (green) catalyzed (Total – uncatalyzed) ln(rate constants) vs. 1/T. Catalyzed reactivity decreased above 45C due to catalyst deactivation via possible niobium leaching from surface.

In addition to batch experiments, continuous stirred tank reactor (CSTR) sulfoxidation experiments were performed in a 50 mL stirred round bottom flask controlled at 35 ºC with a 20 mL holdup volume. Separate MPPC effluent and 4-hydroxythioanisole solutions were pumped into the reactor at 5 mL hr^{-1} each using a Minipuls 3 multichannel pump (Gilson, Inc., Middleton WI) to give a two-hour HRT. H_2O_2 was generated in an MPPC operated at sufficiently long HRTs (on average 1.7 days) in order to reach 110 mM H_2O_2 (3.7 g H_2O_2 L^{-1}). MPPC effluent was acidified with HCl to pH 7 prior to usage. The sulfide reactant solution initially contained 100 mM 4-hydroxythioanisole and 40 mM phenol dissolved in a 50:50 solution of water and ethanol. The reactor was initially filled with 10 mLs of 4-hydroxythioanisole and catholyte

solutions each and contained 0.5 mol% Nb(V)-SiO_2 (relative to 4-hydroxythioanisole). Cotton dead-end filters were used to minimize catalyst loss through the effluent ports on the CSTR. Effluent sulfide product distribution (unreacted sulfide [4-hydroxythioanisole], product sulfoxide [4-(methylsulfinyl)-phenol], and sulfone [4-(methylsulfonyl)-phenol] concentrations) was measured as in the batch experiments, and the reactor was operated until the product distribution reached steady state. The apparent rate constant was determined from the effluent concentration from the CSTR using the Levenberg-Marquardt nonlinear least squares regression algorithm and the Runge-Kutta method for numerically integrating differential equations in MATLAB R2016b (The Mathworks, Natick MA). The *nlparci* function was used to obtain a 95% confidence interval. The effluent concentration was fit to the following differential equation, derived from a mass balance:

$$\frac{dC_A}{dt} = \frac{C_{A0} - C_A}{\tau} - kC_A(C_{B0} - C_{A0} + C_A) \qquad \text{(Eq. 6)}$$

where A and B are 4-hydroxythioanisole and H_2O_2, respectively.

CHAPTER SIX

A Genome-Centric View of Anode Respiring Biofilms Enriched With Different Carbon Sources

6.1 INTRODUCTION

Microbial communities are critical to global biogeochemical cycles as well as environmental biotechnology processes. Until recently, our ability to understand and engineer microbial communities for specific functions is limited by our ability to isolate and culture microbial species. Recently, metagenomics, or direct sequencing of environmental DNA has allowed us to better understand the physiology of the vast majority of uncultivable microbes. Many early metagenomic studies used gene-centric approaches that aggregated community genomic potential. More recent studies have attempted to resolve metagenomic data into "bins", ideally representing individual populations within a community. Genome-centric metagenomics provides a more biologically meaningful view of communities as discrete populations. By resolving metabolic functions into separate species, microbial interaction networks and metabolic networks describing energy and material flux can be developed for prediction of community metabolic behavior.

Microbial electrochemical cells (MECs) are environmental biotechnology processes that convert waste carbon into electrical current by harnessing a microbial metabolism process known as extracellular electron transfer (EET). During EET, microbes use insoluble compounds like Fe(III) as terminal electron acceptors, allowing them to oxidize un-fermentable carbon sources under anaerobic conditions. In addition to their resource recovery applications, MECs represent an excellent opportunity for developing community metabolic models for complex consortia. Currently, quantitative community metabolic models exist only for mock communities and highly simplified two-member interactions.[276-278] MECs facilitate *in situ* monitoring of EET metabolism in the form of produced current. While full quantitative models of metabolic networks in MECs are not currently feasible, qualitative metabolic networks can be derived

using genome-centric metagenomics, and measurements of EET efficiency for various carbon feedstocks can be related to abundance of metabolic pathways.[25, 279]

Deployment of MECs for resource recovery applications will require engineering a microbial consortium capable of efficiently converting diverse carbon sources into electricity. The communities and metabolic functions in MECs are similar to those found in anaerobic digesters used to generate methane from waste carbon.[280] In addition to cellulose hydrolysis and degradation, a critical step is fermentation of simple sugars and short chain fatty acids into acetate. Many canonical exoelectrogens including *Geobacter spp*. require acetate or hydrogen as electron donors, and thus require flanking fermentative and acetogenic organisms to support their growth on more complex organic feedstocks. Achieving robust and resilient electricity generation from other substratres requires production of exogenous acetate through flanking community metabolism, or non-canonical exoelectrogens using other EET mechanisms. There are several relevant metabolic pathways that can generate acetate in anaerobic conditions including glycolysis, propanoate metabolism via the methmalonyl-coA pathway and butanoate metabolism via the butyryl-CoA:acetate-CoA transferase mechanism.

The impact of carbon feedstock on MFC anode community composition and electrochemical performance has been studied, but there is a lack of metagenomic studies of long-term enrichment with different carbon feedstock. Significant work on performance and 16S based community analyses has been performed.[280] More recently, metagenome assembled genomic (MAG) and metatranscriptomic analyses have revealed, at the genome scale, the major components of metabolic networks in acetate-fed MFCs and the (short-term) regulatory responses of these networks to perturbations (availability of a closed circuit for EET).[25] MECs operated on mixed fatty-acids (butyrate, propionate, and acetate) have shown preferential use of

acetate and butyrate as electron donors[281] as well as internal storage and secondary syntrophic metabolism in MECs fed with complex organics.[282] There is a need to understand longterm community adaptation to carbon source in MEC metabolic networks at a mechanistic level. This has important practical implications for enrichment of MECs for treating real waste streams and relevance to fundamental microbial ecology questions such as resilience & resistance, intermediate disturbance theory and diversity function relationships in MECs.[283] To address this need, we designed and operated replicate multi-anode MEC systems using different feedstocks. The multi-anode approach facilitated routine community biomass sampling and metagenomic sequencing. Finally, spike-in metabolic assays were used to test for volatile-fatty acid production and community wide EET efficiency.

6.2 RESULTS

6.2.1 Assembly & Differential Coverage Based Binning

The influence of carbon availability of community structure and function in MECs were evaluated with parallel enrichment systems. Three lab-scale two-chamber MEC systems operated with acetate (Cell 1), a 1:1 mix of butyrate and propionate (Cell 2), and glucose (Cell 3) were operated continuously for 3 months. Biomass samples for metagenomic sequencing were collected on days 36 and 72 of operation. Raw metagenomic reads in each individual sample and the co-assembly were *de novo* assembled using IDBA[284] with a kmer range of 21-101 and increment of 10, minimal contig length of 500bp and scaffolding enabled.[284] The quality for each assembled scaffold was checked by QUAST.[114] Raw sequence quality and assembly statistics are shown in Table 6.1 and 6.2, respectively. IDBA was selected for assembly after comparison of N50 and total assembly length with metaSpades and megahit, two recently benchmarked assemblers.[112, 115, 117] Average GC content for the 2nd acetate cell timepoint (C1B) was

significantly higher than other samples, (Table 6.1) and this sample differed in composition as well (Section 6.2.2)

Table 6.1 Raw Metagenomic Read and QC Summary

Sample	Feedstock	Day of Operation	# Raw Reads	# Clean Reads	% passing QC	%GC
C1A_S1	Acetate	36.00	5.00E+07	4.95E+07	99.1	61
C1B_S2	Acetate	72.00	4.29E+07	4.23E+07	98.7	66
C2A_S3	VFA	36.00	5.91E+07	5.81E+07	98.2	59
C2B_S4	VFA	72.00	4.57E+07	4.50E+07	98.5	62
C3A_S5	Glucose	36.00	6.06E+07	5.99E+07	99.0	60
C3B_S6	Glucose	72.00	5.12E+07	5.07E+07	98.9	61

Table 6.2 Assembly Statistics

Sample	# Contigs	Total Length (bp)	Average Length (bp)	Largest Contig (bp)	N50
C1A_S1	67224	2.78E+08	4141	227528	5709
C1B_S2	43097	1.52E+08	3531	130373	3861
C2A_S3	78728	3.04E+08	3858	210262	4719
C2B_S4	47939	1.67E+08	3483	229561	3596
C3A_S5	90791	3.05E+08	3356	243550	3524
C3B_S6	43508	1.51E+08	3472	128226	3856
Co-assembly	179916	7.97E+08	4432	322784	6505
* All statistics count only contigs > 1000 bp					

Population genome binning was performed on the co-assembled reads by aligning reads from each sample to the co-assembled contigs, and clustering groups on contigs with similar coverage profiles and genomic signatures (tetramer & GC content) using maxbin. Genome bin completeness and contamination were calculated in checkM by comparison of predicted coding sequences to known essential single-copy genes. While this approach may overestimate quality estimates for novel genomic portions, it captures the essential metabolic functions evaluated in this study. Genome binning resulted in 73 medium or high quality bins. (Figure 6.1) “High quality” genomes have >90% completeness and < 5% contamination and “medium quality have

50%<completeness<90% and <10% contamination, defined by the Genome Standards Consortium.[126] Phylogenetic analysis was performed via phylophlan, a multi-protein based alignment and tree construction approach incorporating >3000 genomes. Predicted ORFs were translated and aligned against a reference database and phylogenetic ranks were assigned to the highest-resolution consensus rank.[285] At a phylum level, the majority of assembled genomes (44) were identified as *Proteobacteria*, with 14 *Bacteroidetes* genomes, 7 *Firmicutes*, 5 *Acidobacteria* and 3 singleton phyla detected. No archaeal phyla were detected, despite being common members of anaerobic and fermentative communities. Bin refinement via ensemble bin merging (Metawrap) & manual bin curation (refineM) will be performed prior to more in-depth analysis.

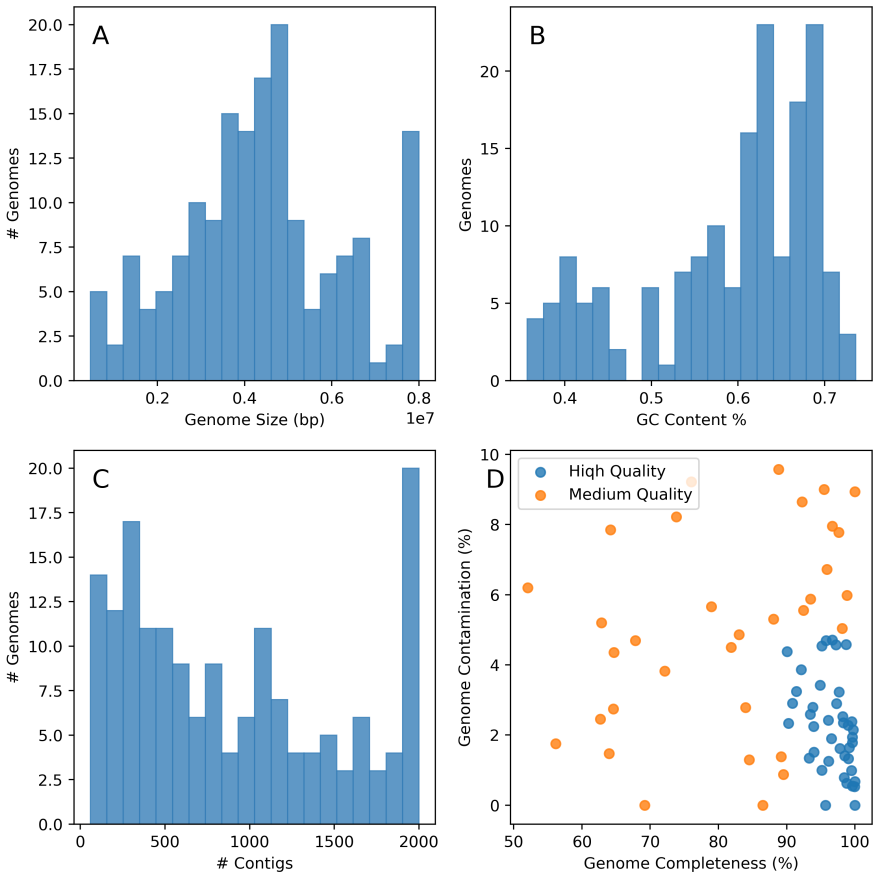

Figure 6.1 Metagenome-assembled genome (MAG) basic statistics and quality. (A) – (C) show histograms of genome size, GC content (out of 1.0) and number of contigs per bin. (D) shows

estimates of completeness and contamination based on inclusion of single copies of essential single copy genes. “High quality” genomes have >90% completeness and < 5% contamination and “medium quality have 50%<completeness<90% and <10% contamination, defined from MIMAG.[126]

6.3.2 Functional analysis of metagenome assembled genomes

MAGs were annotated using prokka. Enzyme classification (EC) annotated genes for each bin were compared to manually curated pathways for microbial beta-oxidation (butyrate metabolism), and propionate metabolism. Putative exoelectrogens were identified based on taxonomic affiliation to reported anode-respiring bacteria or the presence of multiple omC-like genes from COG annotated coding sequences. Due to partially incomplete genomes, genomes were thresholded as containing pathways if they contained genes coding for at least 75% of the overall pathway. Butyrate oxidizers were discarded if they lacked genes for either the butyrate kinase or butyrate acetyl-transferase pathways responsible for conversion of butyrate to butanoyl-CoA. Once butanoyl-CoA is produced, it can undergo beta oxidation, a 4 step process resulting in acetyl-CoA formation, where it enters the TCA cycle.

Putative functional genome categories and per sample abundances are shown in Figure 6.2 and 6.3. Figure 6.2A shows exoelectrogen abundance. Exoelectrogens were identified from the *Geobacteraceae* family, as well as a number of *Desulfovibrio* taxa, a sulfate-reducing *Deltaproteobacteria*. *D. desulfuricans* and *D. vulgaris* have previously been identified as anode respiring bacteria.[286] Desulfovibrio and Geobacter possess a number of outer membrane cytochrome (omc_) and periplasmic cytochrome genes necessary for extracellular electron transfer. Genomes taxonomically identified as *Rhodopseudomonas*, *Ochrobactrum* and *Achromobacter* were included as potential exoelectrogens on the basis of previous literature, but

their EET mechanism(s) are poorly characterized. Ongoing work in the Wells Lab is focused on better identifying EET pathways in bins assigned to these phylogenies. Geobacter abundance was highest in the more mature glucose and VFA fed biofilms, with the highest coverage being found in the glucose-fed systems. This makes sense given the relatively fast conversion of glucose to acetate under anaerobic conditions due to acetate kinase (*ack*), a ubiquitous enzyme involved in glycolysis. there was almost no coverage of Geobacter genomes in the 2nd timepoint of the acetate sample. This is extremely interesting as it suggests either a novel exoelectrogen responsible for current production or sequencing failure. *Achromobacter*, an *Alphaproteobacteria* found in high abundance in this sample has been reported as both an exoelectrogen as well as a laboratory contaminant[287, 288]

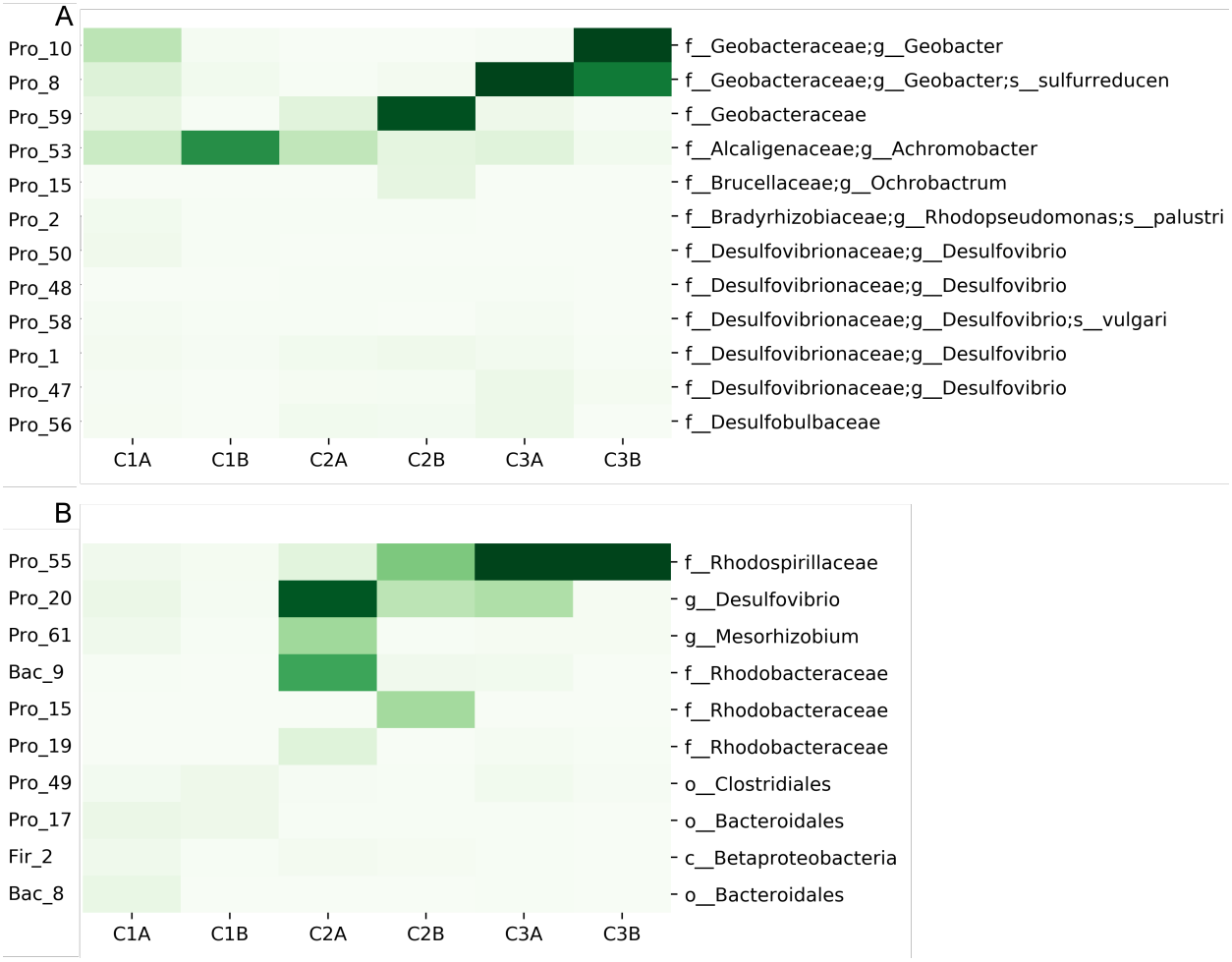

Figure 6.2 Genome bin coverage by sample for putative exoelectrogens (A) and butyrate oxidizers (B). Phylum level bin names are shown on the left and the lowest taxonomic level provided by phylophlan is shown on the right. Average coverage was calculated by weighting coverage across each contig included in the bin by the contig length.

Putative beta-oxidizers bins encoding genes for butyrate conversion to butyryl-CoA, butyrate kinase (*buk*) or butyryl- CoA:acetate CoA-transferase (*but*) as well as 75% of the genes responsible for beta-oxidation pathway to acetyl-CoA are shown in Figure 6.2B. Relatively few butyrate oxidizers were identified in the acetate cell samples, and the highest abundance of beta oxidizers was identified in VFA-fed cells as expected. Figure 6.3 shows abundance of several propionate oxidizers. Anaerobic propionate oxidation is relatively complicated and can be

directed to synthesis of amino acids, as well as via propionyl-CoA formation and conversion to succinyl-CoA or acetyl-CoA. Genomes encoding at least 75% of any of four degradation or bidirectional pathways identified in KEGG are included in Figure 6.3. In general, early timepoints possessed more propionate oxidizers than later timepoints, suggesting that propionate oxidation was not selected for in any of the MEC communities. Further classification of the specific pathways involved in propionate oxidation can help elucidate the reasoning for this. In addition, 20 recovered genomes have not currently been classified in any of the 3 major fucntions guilds described here. Their role fermentative organisms consuming glucose will be investigated.

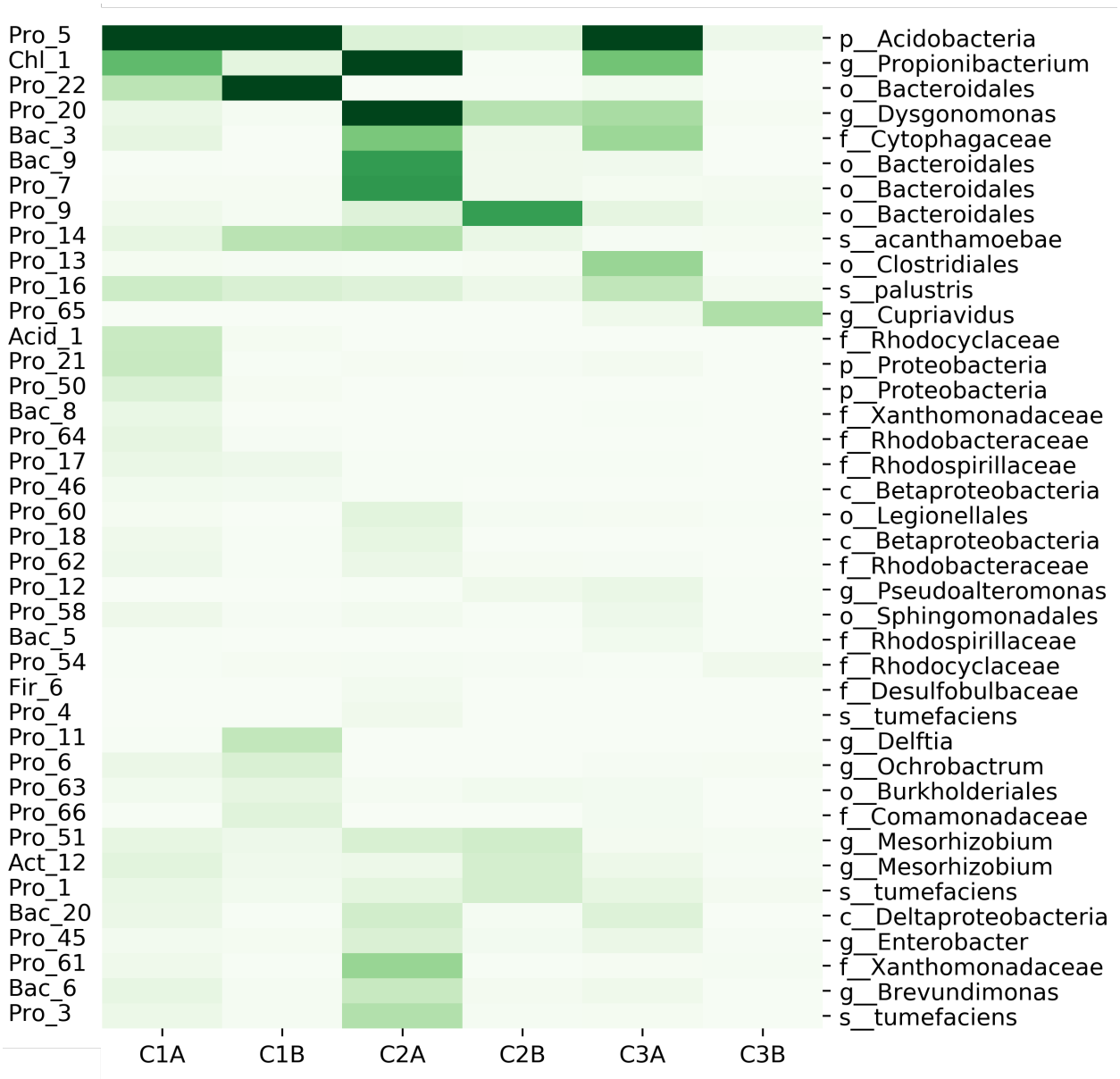

Figure 6.3 Genome bin coverage by sample for putative propionate oxidizers calculated similarly to Figure 6.2

6.3 METHODS

6.3.1 MEC Reactor Configuration and Operation

Triplicate bioelectrochemical cells (Cells 1, 2 and 3) were designed in Autodesk Inventor (Autodesk, San Rafael, CA) and 3D printed from PMMA on a FormLabs V2 printer. Each call had an anode chamber volume of 95 mL and was equipped with an Ag/AgCl reference electrode (BASI, West Lafayette, IN) and two carbon fiber brush anodes. The carbon brushes was pretreated for 24 hours in 1N nitric acid, followed by 24 hours in acetone and 24 hours in ethanol prior to usage in the MFCs to remove surface contamination.[271] Titanium wire was used to connect both anodes in a cell into a single electrode stack. This cell configuration was selected to facilitate routine subpassaging of and DNA extraction from biofilms to analyze microbial community structure and metabolic capabilities. The cathode chamber had a 30 mL chamber with an exposed cathode surface area of 17.5 cm^2. A gas diffusion cathode was constructed as previously described[245] using a carbon cloth electrode (GDL-CT, Fuel Cells Etc., College Station, TX) coated with PTFE on one side and 5 mg cm^{-2} carbon black (Vulcan XC-72, Fuel Cell Earth, Woburn, MA) dispersed in Nafion on the other side. The chambers were separated by a 37.6 cm^2 anion exchange membrane (Ultrex AMI-7001, Membranes Intl., Ringwood, NJ). The average electrode spacing was 1.57 cm. A schematic of the MEC is shown in Figure 6.1.

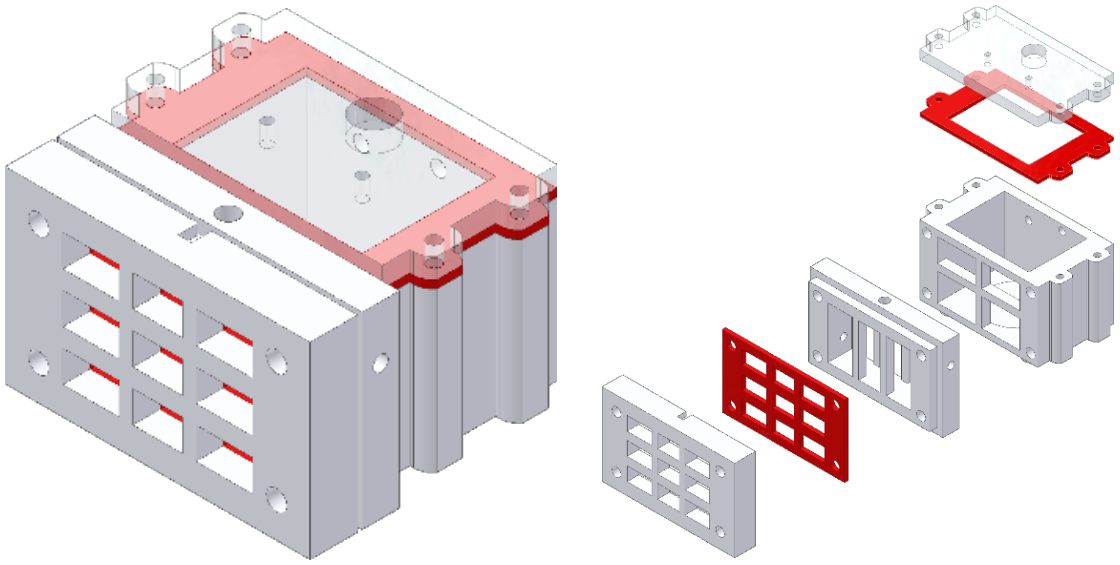

Figure 6.4 Two chamber 3D-printed mini-MEC schematic. The larger anode chamber is equipped with a reference electrode port and 2 small ports for anodes. Biomass can be harvested for DNA extraction from a single mature anode without disturbing the biofilm on the other anode.

Cells were inoculated from existing bioelectrochemical reactors and operated as previously described in chapter 5 except the electron donor in the synthetic wastewater media was varied between the cells. Cells 1, 2 and 3 were operated with acetate, a 1:1 mixture of butyrate and propionate, and glucose, respectively. Cell feed concentrations were operated on equal electron-equivalent basis. Anolyte VFA concentrations were measured using a Trace 1310 gas chromatograph (GC) (Thermo Scientific, USA) equipped with a diphenyl dimethyl polysiloxane column and an FID detector as previously described was used to measure anolyte acetate concentrations. A 0.1 µL sample was injected in the GC with a 50:1 split ratio under a constant flow rate of 1.0 ml min^{-1}. The GC oven was held at 70 °C for one minute, then increased at 10 °C per minute up to 180 °C, and then held at a constant temperature of 180°C for five minutes. 24 hour anodic acetate removal rates were calculated by measuring the difference between acetate medium and effluent concentrations at a given HRT. Batch assays were

performed by allowing cell current to reach 0 and then spiking known quantities of VFAs, and subsequently measuring exogenous VFA concentrations.

6.3.2 DNA extraction and amplicon sequencing

BES subpassaging was performed approximately every two weeks of reactor operation by removing each anode chamber top and replacing the older anode with a fresh anode. Each removed anode was vortexed for 1 minute and then centrifuged at 4000G in a 50mL falcon tube for 20 minutes to pellet biomass. Replicate DNA extractions were performed using a Fast DNA SPIN Kit for soil (MP Biomedicals, OH). Duplicate DNA extracts were pooled, quantified and preserved for sequencing library preparation. V4-V5 16S rDNA amplicon sequencing library preparation and sequencing of biweekly samples was performed as previously described in chapter 3 via PCR amplifiation, barcoding and 2x250bp paired end sequencing on an Illumina Miseq.

6.3.3 Metagenomic sequencing and assembly

Duplicate DNA extracts from two time points approximately 1 month apart (days 36 and 72 of operation) were selected for whole genome shotgun (WGS) sequencing. Library preparation and sequencing were performed by the NUSeq Core facility at Northwestern University (Evanston, IL) using a Nextera XT library preparation kit and 2x150bp paired end sequencing on an Illumina HiSeq 4000 sequencer (Illumina, San Diego CA) with an average insert size of 650 bp. 92 Gbp of raw sequencing data was produced. Quality control was performed in cutadapt[289] by removing read ends with Q<20, trimming sequencing adapters and then removing any resulting reads less than 20bp long. 98.7% of reads passed QC and were used for assembly. The quality of metagenomic sequencing data for each sample was checked before and after QC by FastQC.[290]

Sequence coverage for each assembled scaffold was calculated by mapping raw reads from the sample to assembled contigs using BWA with default parameters.[291] Open reading frame (ORF) calling and annotation were performed for each sample using prokka.[292] Genome Representative metabolic pathways were constructed by matching prokka-generated enzyme classifier numbers to metabolic pathways in KEGG and MetaCyc.[293, 294] When partial pathways were identified in genome bins, pathway presence or absence was determined using a 75% or greater criteria as well as by manual curation. Circular "ideogram" visualizations were generated in Circos.[295]

CHAPTER SEVEN

Future Work & Conclusions

The work in this thesis focuses on characterizing the mechanisms influencing microbial ecology as opposed to merely quantifying diversity. The next logical steps in this progression towards microbial resource management or community design are improved environmental and physiological monitoring of microbial communities. In this chapter, I present ongoing work focused on integrating community analysis with novel long-term hydrological monitoring. Finally, I discuss future directions to incorporate metatranscriptomic sequencing to improve resolution of functional populations in anode-respiring biofilms or other mixed communities. While the shotgun sequencing work presented in Chapter 6 represents a significant improvement in our ability to identify functional responses to carbon availability in MEC communities, future work should incorporate more direct functional analysis of populations.

7.1 Future Work & Outlook

7.1.1 Microbial community assembly in a native tallgrass prairies

In Chapter 4, trends in soil microbiome biogeography in an intensively managed (predominantly agricultural) landscape were described and related to broad mechanisms such as ease of dispersal. As a follow-up to this work, we are currently investigating the impact of soil moisture on community diversity in the Indian Boundary Prairie, a native tallgrass prairie using a novel soil moisture sensing platform developed by Argonne. The soil microbiome is critical for nutrient cycling, carbon sequestration, and agricultural productivity yet the dominant drivers of community composition are poorly understood. Soil microbial communities display predictable biogeographical patterns despite their high diversity and functional redundancy.[296, 297] At continental scales, pH is the most apparent driver of microbial community structure in near surface soils.[176, 298] While some microbial surveys have measured soil moisture during sampling, wetting and drying can occur rapidly in response to precipitation,[299] and to our knowledge, no

studies have characterized the effects of longitudinal hydrological conditions on soil microbial community structure in native prairies.

Due to the difficulty of measuring soil moisture, many soil microbiology studies do not explicitly consider variation in moisture content due to the hydrological characteristics of a site, and this property has been ignored in favor of more easily measured variables such as pH, carbon content, redox status, and nutrient variability. Many of these factors covary with soil depth, [300] and community variation along individual soil depth profiles can exceed that of surface variation across diverse biomes.[301] In addition to empirical correlations observed for many soil parameters with depth, soil hydrology influences these parameters directly due to leaching of acids and cations in low-lying soil due to rain and subsurface flow.[302] In this ongoing work, we used the 16S methods described in Chapters 3 & 4 on samples taken from 120 cm deep soil cores at 10 locations across an undisturbed Midwestern tall grass prairie. We combined depth-resolved 16S rRNA amplicon sequencing with long-term monitoring of soil moisture across a hydrological and elevation gradient (Figure 6.1).

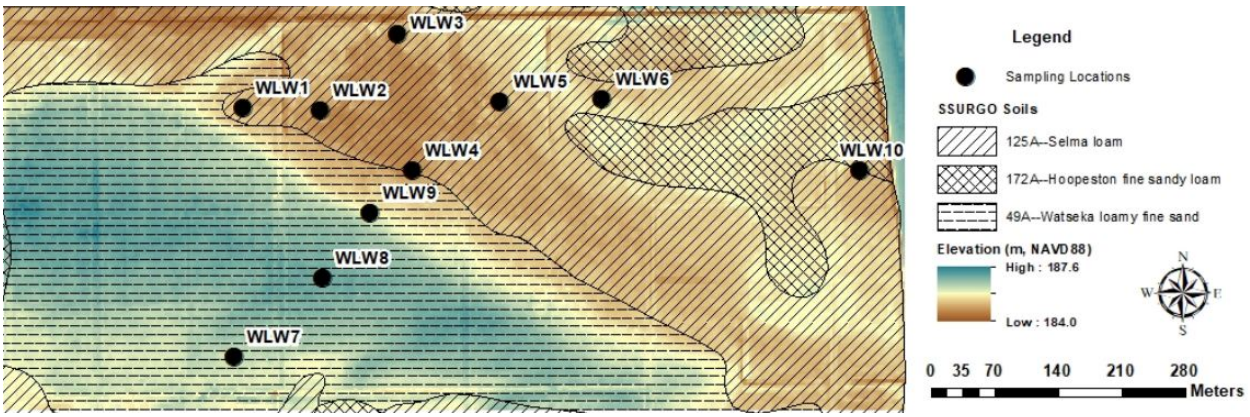

Figure 7.1 Elevation map of the native tall grass Indian Boundary Prairie. Soil sample locations and water level sensors are shown by the black dots, with soil type indicated by fill pattern. Higher elevations are shown in green and lower elevations are shown in brown. (Unpublished)

We hypothesized that relatively modest elevation and soil type differences can lead to compounding effects in flooding frequency and duration, leading to predictable differences in community structure between upland and lowland community structure. We demonstrated these effects at the community scale as well as individual OTU level and predicted metagenomic potential for denitrification. Moreover, these community differences persist even when accounting for factors like depth and pH. Our results suggest a stronger link between geographic and hydrological site characteristics and microbial community structure at field scales.

Using longitudinal inundation data collected from piezometers installed at the sampling points used for microbial analysis, we quantified the fraction of time that each 10 cm depth section was inundated with groundwater in the year following sampling. These patterns reflect both the elevation gradients, subsurface flow paths and soil texture differences within the site (Figure 7.1). The low lying, clayey soils characterized as Selma loam were frequently fully inundated with water, whereas the elevated sand ridge samples were significantly drier (ANOVA $p < .05$). Samples from the Hoopeston fine sandy loam soil series were intermediate in terms of both elevation and flooding duration (Figure & 7.1 & 7.2). Although this data is 'forward looking' and does not represent the exact conditions experienced by the soil microbial communities prior to sampling, clear differences in hydrological regime could be observed.

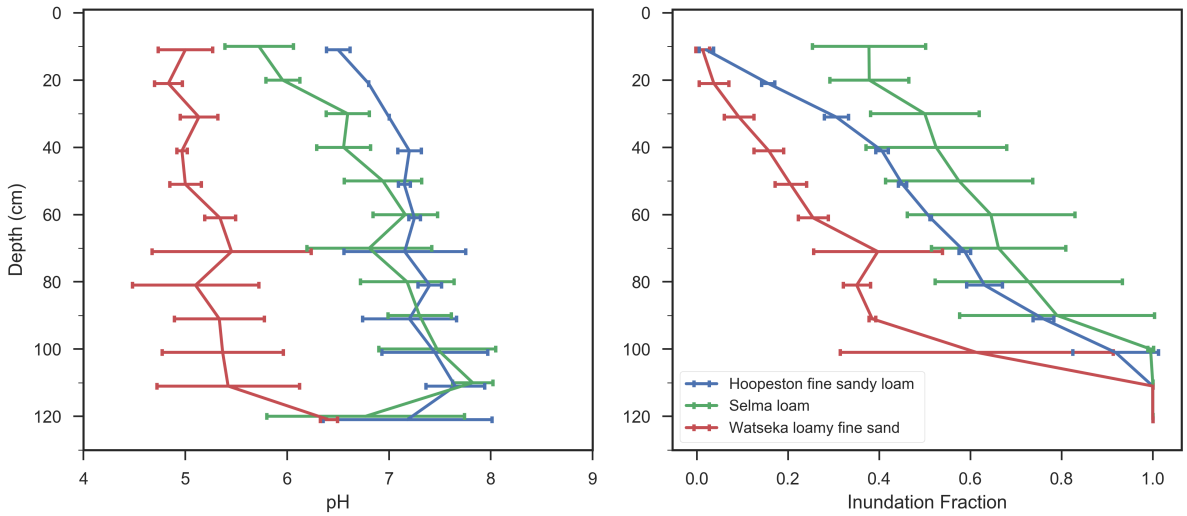

Figure 7.2 pH and annualized inundation fraction as a function of depth for soil samples collected in the IBP. Deeper soils are more acid & flooded more frequently, although the trajectory of these patterns depends on local factors like soil type and elevation.

Microbial community composition in the Indian Boundary Prairie (IBP) varied consistently with soil classification and edaphic factors. Samples obtained from communities in the three different soil types varied significantly from one another (PERMANOVA, Pseudo-F = 21.9, $P < 0.001$). We used Constrained Analysis of Principal Coordinates or "distance based redundancy analysis" (RDA) to constrain the observed variation in community structure with measured environmental and hydrological gradients. The ordination revealed that samples collected from the sand ridge (wells 7-9) and classified as "Watseka loamy fine sand" were distinct from those found in the rest of the prairie (Figure 7.3). Soil characterization revealed that the sand ridge samples had consistently higher pH, and lower clay and cation content compared to those in the low-lying wells. These factors were highly correlated with Axis 1, which explained 27% of the total community variation. Communities found in the loam and sandy loam soils, (circles and squares in Figure 7.3, respectively) were not easily distinguished on the basis

of soil type. These communities were most clearly differentiated by differences in depth and measured annual inundation.

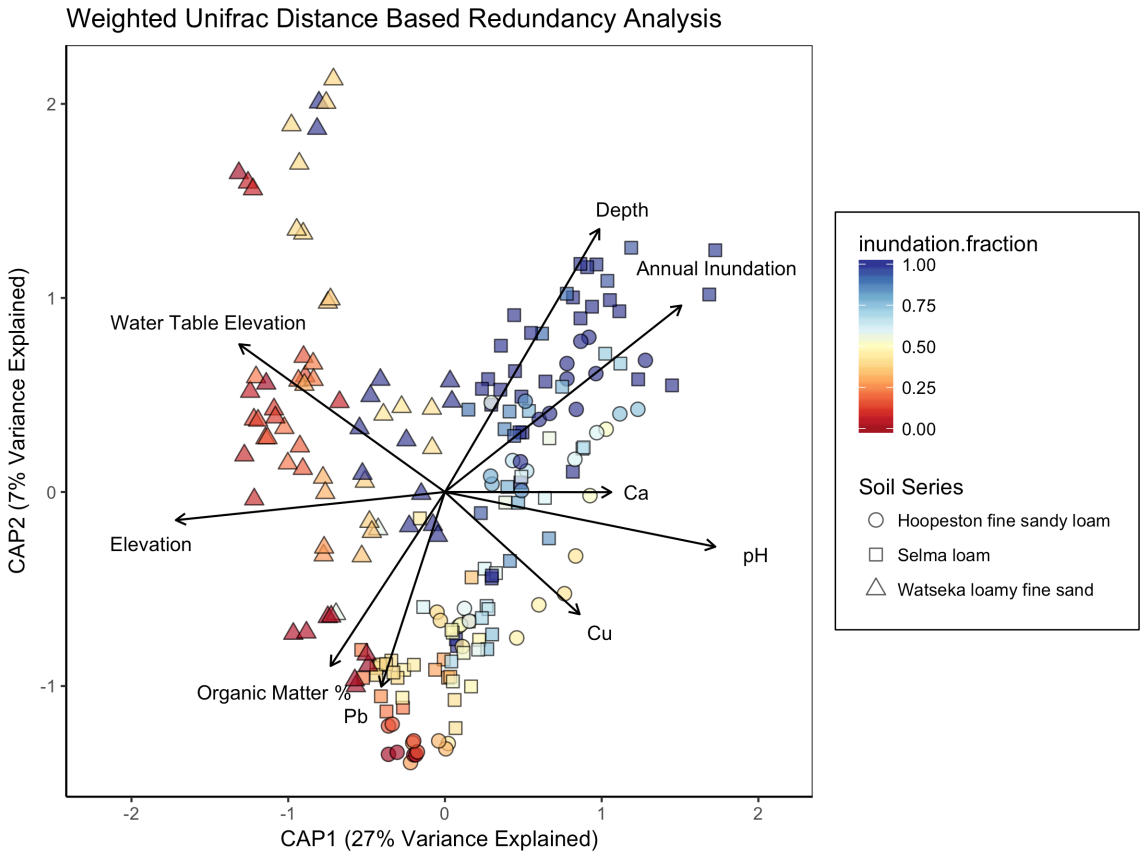

Figure 7.3 Weighted Unifrac distance based RDA ordination of IBP samples. Annualized inundation fraction is shown by color, and sample soil type is indicated by the marker. Samples cluster by inundation fraction as well as source material. Arrows length and direction indicates the strength of correlation with environmental variables and constrained axes. (Unpublished)

Additional work remains to categorize the community-wide differences seen in Figure 7.3 at higher phylogenetic resolution, and to explore the impact on predicted functional potential, using functional prediction tools like Tax4Fun.[111] In addition, we are working on comparing environmental predictors of beta diversity using depth-corrected, depth-only, and depth-naïve models to explore the extent to which classical depth-based approaches capture the soil diversity patterns explained by soil type and hydrological characterization used here. Beyond these field-

scale studies, sensor networks could be used to improve monitoring of mesocosm-based experiments to capture specific and environmentally relevant soil community responses such as N_2O production in flooded soils.

7.1.2 Realizing Metabolic Potential with Transcriptomics

An important question for scale up of MECs is their ability to resist and recover from dynamic conditions or pulse disturbances. To date, there are few studies of MEC community responses to dynamic conditions, and even fewer metatranscriptomic studies of microbial fuel cell communities.[71] In this work, the authors identified specific c-type cytochromes associated with bacterial families that were differentially expressed under open circuit (no extracellular electron transfer) and closed circuit conditions. Using this approach, the authors were able to address specific populations responsible for EET. In addition to potential short-term responses to operational perturbations addressed in this study, there is a need to understand community responses to other types of disturbances, such as changes in carbon availability.

The work presented in chapter 6 focuses on community responses to long-term press disturbances in the form of differences in electron donor. In addition to changes in anode community functional or genomic potential described in Chapter 6, metatranscriptomic responses to addition of electron donors could be used to profile ecological niches of the biofilm community via metatranscriptomic sequencing. The metagenomic approach used here can only address genome potential, and does not directly capture activity. By directly measuring transcriptomic responses to spike-in assays of different electron donors, it should be easier to assign functions to genome bins. A schematic of proposed methods is shown in Figure 7.4. Metatranscriptomic sequencing of anaerobic communities requires careful sampling to minimize community response to oxygen perturbations.[71] In addition, recovery of mRNA from mixed

communities for sequencing can be difficult due to contamination with inhibitory compounds and the high ratio of rRNA to mRNA that must be filtered from sequencing results.[303]

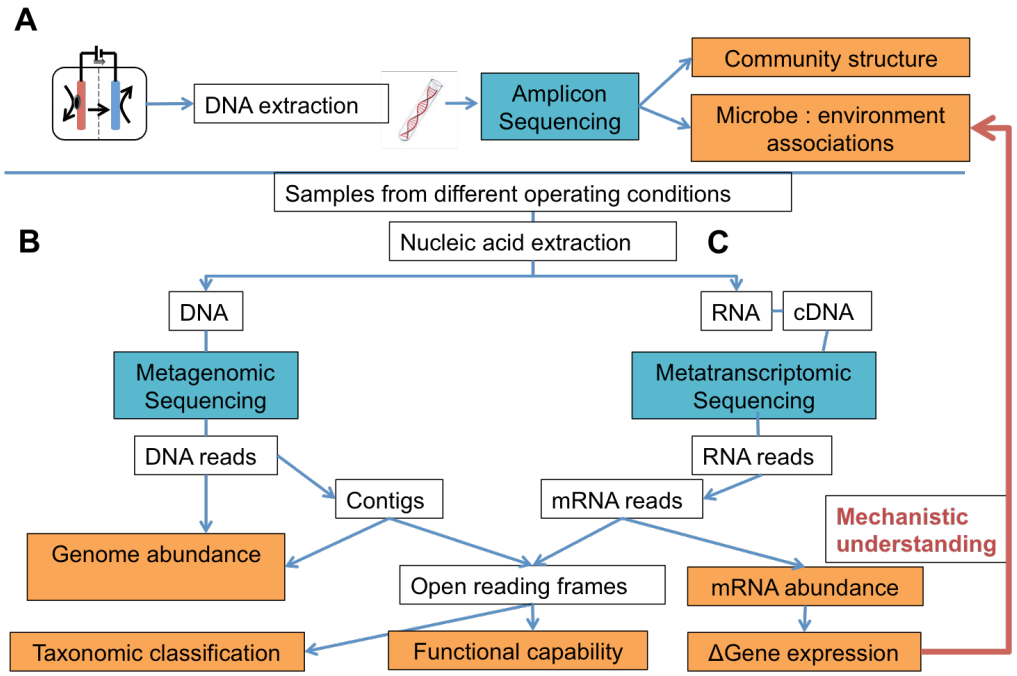

Figure 7.4 Schematic showing metagenomic and transcription based (metatranscriptomic) analysis of MEC communities. After combined nucleic acid extraction and separation, assembled genome bins (B) can be used as reference databases to map transcriptomic reads (C), enabling microbes with functional pathways to be distinguished from populations with incomplete pathways.

To further improve metatranscriptomic analysis, high quality near-complete genome bins should be assembled when possible. Incorporating long-read sequencing or linked-read sequencing for *de novo* genome assembly[304] with metatranscriptomic sequencing could improve functional classification of exoelectrogens and flanking community members in MECs. These sequencing approaches improve assembly of complex samples by producing short reads bound to a particular ~50kb stretch or generating longer individual reads that can be used as scaffolds. Increasing genome quality will make mapping reads and making functional predictions more straight-forward, which is often difficult for complex carbon metabolism in diverse

environments. In addition to improving detection of genomes with functional carbon metabolism pathways, electron-donor perturbation metatranscriptomics could be useful to identify novel instances of direct interspecies electron transfer (DIET). Anaerobic communities are known to harbor commensal relationships where metabolic functions are split between acetate producing and consuming microbes.[305] Detecting these relationships is difficult but preliminary DIET interactions may be identified through complementary transcriptional activity in discrete populations.

7.2 CONCLUSIONS

This thesis combines modern advances in metagenomics with bioprocess design for resource recovery. Significant improvements in sequencing capabilities and cost have made it feasible to incorporate routine sequencing into process design and optimization cycles. Due to the relative underutilization of metagenomics in engineered systems to date, the bioinformatics analysis work presented here advances our knowledge of community structure in the systems studied. The metagenomic and bioinformatics tools and methods I developed and made available have broad application for microbial communities monitoring in bioprocesses as well as in human or environmental associated microbial communities.

Despite these technical advances, microbial resource management still requires fundamental understanding of microbe physiology and ecological theory, as well as effective reactor design and operation. The work presented in chapters 3 & 4 improves our understanding of community assembly mechanisms in natural and engineered environments and demonstrates several new approaches to analyzing 16S marker gene datasets. The process developed in Chapter 5 demonstrates a novel bioelectrochemical resource recovery platform for chemical

synthesis from wastewater. Finally, the ongoing metagenomic analysis described in Chapter 6 adds to our knowledge of anode-respiring biofilm communities and microbial electrochemical cells in general. Continued integration of community ‘omics approaches, hypothesis driven and statistically sound NGS experimental design, and a broad view of resource management will enable new bioprocesses to protect human and environmental health.

REFERENCES

1. Bar-On, Y. M.; Phillips, R.; Milo, R., The biomass distribution on Earth. *Proc. Natl. Acad. Sci.* **2018**, 201711842.
2. Ciais, P.; Sabine, C.; Bala, G.; Bopp, L.; Brovkin, V.; Canadell, J.; Chhabra, A.; DeFries, R.; Galloway, J.; Heimann, M., Carbon and other biogeochemical cycles. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press: 2014; pp 465-570.
3. Falkowski, P. G.; Fenchel, T.; Delong, E. F., The microbial engines that drive Earth's biogeochemical cycles. *science* **2008,** *320*, (5879), 1034-1039.
4. Verstraete, W.; Wittebolle, L.; Heylen, K.; Vanparys, B.; de Vos, P.; van de Wiele, T.; Boon, N., Microbial Resource Management: The Road To Go for Environmental Biotechnology. *Eng. Life Sci.* **2007,** *7*, (2), 117-126.
5. Ardern, E.; Lockett, W. T., Experiments on the oxidation of sewage without the aid of filters. *Journal of the society of chemical industry* **1914,** *33*, (10), 523-539.
6. Goldstein, R.; Smith, W., *Water & sustainability (volume 4): US electricity consumption for water supply & treatment-the next half century*. Electric Power Research Institute: 2002.
7. Shade, A.; Peter, H.; Allison, S. D.; Baho, D.; Berga, M.; Bürgmann, H.; Huber, D. H.; Langenheder, S.; Lennon, J. T.; Martiny, J. B., Fundamentals of microbial community resistance and resilience. *Frontiers in microbiology* **2012,** *3*, 417.
8. Vanwonterghem, I.; Jensen, P. D.; Dennis, P. G.; Hugenholtz, P.; Rabaey, K.; Tyson, G. W., Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J* **2014,** *8*, (10), 2015-2028.
9. Kim, T.-S.; Jeong, J.-Y.; Wells, G. F.; Park, H.-D., General and rare bacterial taxa demonstrating different temporal dynamic patterns in an activated sludge bioreactor. *Appl. Microbiol. Biotechnol.* **2012,** *97*, (4), 1755-1765.
10. Stenuit, B.; Agathos, S. N., Deciphering microbial community robustness through synthetic ecology and molecular systems synecology. *Curr. Opin. Biotechnol.* **2015,** *33*, 305-317.
11. Cole, J. R.; Chai, B.; Marsh, T. L.; Farris, R. J.; Wang, Q.; Kulam, S.; Chandra, S.; McGarrell, D. M.; Schmidt, T. M.; Garrity, G. M., The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic acids research* **2003,** *31*, (1), 442-443.
12. Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S. M.; Betley, J.; Fraser, L.; Bauer, M., Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **2012,** *6*, (8), 1621-1624.
13. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I., QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* **2010,** *7*, (5), 335-336.

14. Butler, J. E.; Young, N. D.; Lovley, D. R., Evolution of electron transfer out of the cell: comparative genomics of six Geobacter genomes. *BMC genomics* **2010,** *11*, (1), 40.
15. van Kessel, M. A.; Speth, D. R.; Albertsen, M.; Nielsen, P. H.; den Camp, H. J. O.; Kartal, B.; Jetten, M. S.; Lücker, S., Complete nitrification by a single microorganism. *Nature* **2015**.
16. Speth, D. R.; Guerrero-Cruz, S.; Dutilh, B. E.; Jetten, M. S., Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nature communications* **2016,** *7*, 11172.
17. Schmidt, I.; Sliekers, O.; Schmid, M.; Bock, E.; Fuerst, J.; Kuenen, J. G.; Jetten, M. S.; Strous, M., New concepts of microbial treatment processes for the nitrogen removal in wastewater. *FEMS microbiology reviews* **2003,** *27*, (4), 481-492.
18. Nielsen, P. H.; Saunders, A. M.; Hansen, A. A.; Larsen, P.; Nielsen, J. L., Microbial communities involved in enhanced biological phosphorus removal from wastewater—a model system in environmental biotechnology. *Curr. Opin. Biotechnol.* **2012,** *23*, (3), 452-459.
19. Daigger, G. T., Evolving urban water and residuals management paradigms: Water reclamation and reuse, decentralization, and resource recovery. In *Water Environment Research*, Water Environment Federation: 2009; Vol. 81, pp 809-823.
20. Mo, W.; Zhang, Q., Energy–nutrients–water nexus: integrated resource recovery in municipal wastewater treatment plants. *Journal of environmental management* **2013,** *127*, 255-267.
21. Guest, J. S.; Skerlos, S. J.; Barnard, J. L.; Beck, M. B.; Daigger, G. T.; Hilger, H.; Jackson, S. J.; Karvazy, K.; Kelly, L.; Macpherson, L., A new planning and design paradigm to achieve sustainable resource recovery from wastewater. In ACS Publications: 2009.
22. Logan, B. E.; Rabaey, K., Taking the "Waste" Out of "Wastewate" for Human Water Security and Ecosystem Sustainability. *Science* **2012,** *337*, (6095), 681-686.
23. Reguera, G.; McCarthy, K. D.; Mehta, T.; Nicoll, J. S.; Tuominen, M. T.; Lovley, D. R., Extracellular electron transfer via microbial nanowires. *Nature* **2005,** *435*, (7045), 1098.
24. Rabaey, K.; Rozendal, R. A., Microbial electrosynthesis—revisiting the electrical route for microbial production. *Nature Reviews Microbiology* **2010,** *8*, (10), 706.
25. Ishii, S. i.; Suzuki, S.; Tenney, A.; Norden-Krichmar, T. M.; Nealson, K. H.; Bretschger, O., Microbial metabolic networks in a complex electrogenic biofilm recovered from a stimulus-induced metatranscriptomics approach. *Scientific reports* **2015,** *5*.
26. Jiang, D.; Curtis, M.; Troop, E.; Scheible, K.; McGrath, J.; Hu, B.; Suib, S.; Raymond, D.; Li, B., A pilot-scale study on utilizing multi-anode/cathode microbial fuel cells (MAC MFCs) to enhance the power production in wastewater treatment. *Int. J. Hydrogen Energy* **2011,** *36*, (1), 876-884.
27. Foley, J. M.; Rozendal, R. A.; Hertle, C. K.; Lant, P. A.; Rabaey, K., Life Cycle Assessment of High-Rate Anaerobic Treatment, Microbial Fuel Cells, and Microbial Electrolysis Cells. *Environ. Sci. Technol.* **2010,** *44*, (9), 3629-3637.
28. Cusick, R. D.; Bryan, B.; Parker, D. S.; Merrill, M. D.; Mehanna, M.; Kiely, P. D.; Liu, G.; Logan, B. E., Performance of a pilot-scale continuous flow microbial electrolysis cell fed winery wastewater. *Appl. Microbiol. Biotechnol.* **2011,** *89*, (6), 2053-2063.
29. O'Malley, M. A., The nineteenth century roots of'everything is everywhere'. *Nature Reviews Microbiology* **2007,** *5*, (8), 647.
30. Nocker, A.; Burr, M.; Camper, A. K., Genotypic microbial community profiling: a critical technical review. *Microbial ecology* **2007,** *54*, (2), 276-289.

31. Hubbell, S. P., *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press: 2001; Vol. 32.
32. Lozupone, C. A.; Knight, R., Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci.* **2007,** *104*, (27), 11436-11440.
33. Torsvik, V.; Øvreås, L., Microbial diversity and function in soil: from genes to ecosystems Vigdis Torsvik* and Lise Øvreås. In 2002; pp 1-6.
34. Curtis, T. P.; Sloan, W. T., Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Current opinion in microbiology* **2004,** *7*, (3), 221-226.
35. Ofiţeru, I. D.; Lunn, M.; Curtis, T. P.; Wells, G. F.; Criddle, C. S.; Francis, C. A.; Sloan, W. T., Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, (35), 15345-15350.
36. Woodcock, S.; Van Der Gast, C. J.; Bell, T.; Lunn, M.; Curtis, T. P.; Head, I. M.; Sloan, W. T., Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* **2007,** *62*, (2), 171-180.
37. Leibold, M. A.; Holyoak, M.; Mouquet, N.; Amarasekare, P.; Chase, J. M.; Hoopes, M. F.; Holt, R. D.; Shurin, J. B.; Law, R.; Tilman, D.; Loreau, M.; Gonzalez, A., The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **2004,** *7*, (7), 601-613.
38. Wells, G. F.; Park, H.-D.; Eggleston, B.; Francis, C. A.; Criddle, C. S., Fine-scale bacterial community dynamics and the taxa-time relationship within a full-scale activated sludge bioreactor. In *Water Research*, Elsevier Ltd: 2011; Vol. 45, pp 5476-5488.
39. Ofiţeru, I. D.; Lunn, M.; Curtis, T. P.; Wells, G. F.; Criddle, C. S.; Francis, C. A.; Sloan, W. T., Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci.* **2010,** *107*, (35), 15345-15350.
40. Werner, J. J.; Knights, D.; Garcia, M. L.; Scalfone, N. B.; Smith, S.; Yarasheski, K.; Cummings, T. A.; Beers, A. R.; Knight, R.; Angenent, L. T., Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108*, (10), 4158-4163.
41. Vuono, D. C.; Benecke, J.; Henkel, J.; Navidi, W. C.; Cath, T. Y.; Munakata-Marr, J.; Spear, J. R.; Drewes, J. E., Disturbance and temporal partitioning of the activated sludge metacommunity. *ISME J.* **2014,** *9*, (2), 425-435.
42. Daims, H.; Taylor, M. W.; Wagner, M., Wastewater treatment: a model system for microbial ecology. *Trends Biotechnol.* **2006,** *24*, (11), 483-489.
43. Ju, F.; Guo, F.; Ye, L.; Xia, Y.; Zhang, T., Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environ. Microbiol. Rep.* **2013,** *6*, (1), 80-89.
44. Gilbert, J. A.; Field, D.; Swift, P.; Newbold, L.; Oliver, A.; Smyth, T.; Somerfield, P. J.; Huse, S.; Joint, I., The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* **2009,** *11*, (12), 3132-3139.
45. Crump, B. C.; Hobbie, J. E., Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnology and Oceanography* **2005,** *50*, (6), 1718-1729.
46. Griffin, J. S.; Wells, G. F., Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly. *ISME J.* **2017,** *11*, (2), 500.
47. Eiler, A.; Heinrich, F.; Bertilsson, S., Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* **2011,** *6*, (2), 330-342.

48. Kent, A. D.; Yannarell, A. C.; Rusak, J. A.; Triplett, E. W.; McMahon, K. D., Synchrony in aquatic microbial community dynamics. *ISME J.* **2007,** *1*, (1), 38-47.
49. Shade, A., Diversity is the question, not the answer. *ISME J.* **2017,** *11*, (1), 1.
50. Ishii, S. i.; Suzuki, S.; Norden-Krichmar, T. M.; Nealson, K. H.; Sekiguchi, Y.; Gorby, Y. A.; Bretschger, O., Functionally stable and phylogenetically diverse microbial enrichments from microbial fuel cells during wastewater treatment. *PloS one* **2012,** *7*, (2), e30495.
51. Ishii, S. i.; Suzuki, S.; Norden-Krichmar, T. M.; Tenney, A.; Chain, P. S.; Scholz, M. B.; Nealson, K. H.; Bretschger, O., A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. *Nature communications* **2013,** *4*, 1601.
52. Pham, T.; Rabaey, K.; Aelterman, P.; Clauwaert, P.; De Schamphelaire, L.; Boon, N.; Verstraete, W., Microbial fuel cells in relation to conventional anaerobic digestion technology. *Eng. Life Sci.* **2006,** *6*, (3), 285-292.
53. Fu, L.; You, S.-J.; Yang, F.-l.; Gao, M.-m.; Fang, X.-h.; Zhang, G.-q., Synthesis of hydrogen peroxide in microbial fuel cell. In *J. Chem. Technol. Biotechnol.*, 2010; Vol. 85, pp 715-719.
54. Borole, A. P.; Reguera, G.; Ringeisen, B.; Wang, Z.-W.; Feng, Y.; Kim, B. H., Electroactive biofilms: Current status and future research needs. In *Energy Environ. Sci.*, 2011; Vol. 4, pp 4813-22.
55. Logan, B. E.; Hamelers, B.; Rozendal, R.; Schröder, U.; Keller, J.; Freguia, S.; Aelterman, P.; Verstraete, W.; Rabaey, K., Microbial Fuel Cells: Methodology and Technology †. In *Environ. Sci. Technol.*, 2006; Vol. 40, pp 5181-5192.
56. Young, M. N.; Links, M. J.; Popat, S. C.; Rittmann, B. E.; Torres, C. s. I., Tailoring Microbial Electrochemical Cells for Production of Hydrogen Peroxide at High Concentrations and Efficiencies. *ChemSusChem* **2016,** *9*, (23), 3345-3352.
57. Schröder, U., Anodic electron transfer mechanisms in microbial fuel cells and their energy efficiency. In *Phys. Chem. Chem. Phys.*, 2007; Vol. 9, pp 2619-2629.
58. Fan, Y.; Sharbrough, E.; Liu, H., Quantification of the Internal Resistance Distribution of Microbial Fuel Cells. In *Environ. Sci. Technol.*, 2008; Vol. 42, pp 8101-8107.
59. Xie, X.; Hu, L.; Pasta, M.; Wells, G. F.; Kong, D.; Criddle, C. S.; Cui, Y., Three-Dimensional Carbon Nanotube−Textile Anode for High-Performance Microbial Fuel Cells. In *Nano Lett.*, 2011; Vol. 11, pp 291-296.
60. Fan, Y.; Sharbrough, E.; Liu, H., Quantification of the internal resistance distribution of microbial fuel cells. *Environmental science & technology* **2008,** *42*, (21), 8101-8107.
61. Zhuang, L.; Zheng, Y.; Zhou, S.; Yuan, Y.; Yuan, H.; Chen, Y., Scalable microbial fuel cell (MFC) stack for continuous real wastewater treatment. *Bioresource technology* **2012,** *106*, 82-88.
62. Rozendal, R. A.; Leone, E.; Keller, J.; Rabaey, K., Efficient hydrogen peroxide generation from organic matter in a bioelectrochemical system. *Electrochem. Commun.* **2009,** *11*, (9), 1752-1755.
63. Modin, O.; Fukushi, K., Development and testing of bioelectrochemical reactors converting wastewater organics into hydrogen peroxide. In *Water Science & Technology*, 2012; Vol. 66, pp 831-6.
64. Rozendal, R. A.; Leone, E.; Keller, J.; Rabaey, K., Efficient hydrogen peroxide generation from organic matter in a bioelectrochemical system. In *Electrochemistry Communications*, Elsevier B.V.: 2009; Vol. 11, pp 1752-1755.

65. Campos-Martin, J. M.; Blanco-Brieva, G.; Fierro, J. L. G., Hydrogen Peroxide Synthesis: An Outlook beyond the Anthraquinone Process. In *Angew. Chem. Int. Ed.*, WILEY‐VCH Verlag: 2006; Vol. 45, pp 6962-6984.
66. Malvankar, N. S.; Tuominen, M. T.; Lovley, D. R., Biofilm conductivity is a decisive variable for high-current-density Geobacter sulfurreducens microbial fuel cells. *Energy & Environmental Science* **2012,** *5*, (2), 5790-5797.
67. Ishii, S. a. i.; Suzuki, S.; Norden-Krichmar, T. M.; Phan, T.; Wanger, G.; Nealson, K. H.; Sekiguchi, Y.; Gorby, Y. A.; Bretschger, O., Microbial population and functional dynamics associated with surface potential and carbon metabolism. In Nature Publishing Group: 2013; Vol. 8, pp 963-978.
68. Aelterman, P.; Freguia, S.; Keller, J.; Verstraete, W.; Rabaey, K., The anode potential regulates bacterial activity in microbial fuel cells. In *Appl Microbiol Biotechnol*, 2008; Vol. 78, pp 409-418.
69. Methe, B. A.; Nelson, K. E.; Lovley, D. R.; Fraser, C. M., Genome of Geobacter sulfurreducens: Metal Reduction in Subsurface Environments. In 2016; pp 1-4.
70. Kiely, P. D.; Regan, J. M.; Logan, B. E., The electric picnic: synergistic requirements for exoelectrogenic microbial communities. In *Current Opinion in Biotechnology*, Elsevier Ltd: 2011; Vol. 22, pp 378-385.
71. Ishii, S. i.; Suzuki, S.; Norden-Krichmar, T. M.; Tenney, A.; Chain, P. S. G.; Scholz, M. B.; Nealson, K. H.; Bretschger, O., A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. In *Nature Communications*, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 2013; Vol. 4, p 1601.
72. Butler, J. E.; Young, N. D.; Lovley, D. R., Evolution of electron transfer out of the cell: comparative genomics of six Geobacter genomes. In 2010; pp 1-12.
73. Lovell, C. R.; Leaphart, A. B., Community‐Level Analysis: Key Genes of CO2‐Reductive Acetogenesis. *Environ. Microbiol.* 2005, pp 454-469.
74. Friedrich, M. W., Methyl‐Coenzyme M Reductase Genes: Unique Functional Markers for Methanogenic and Anaerobic Methane‐Oxidizing Archaea. *Environ. Microbiol.* 2005, pp 428-442.
75. Griffin, J. S.; Lu, N.; Sangwan, N.; Li, A.; Dsouza, M.; Stumpf, A. J.; Sevilla, T.; Culotti, A.; Keefer, L. L.; Kelly, J. J., Microbial diversity in an intensively managed landscape is structured by landscape connectivity. *FEMS Microbiol. Ecol.* **2017,** *93*, (10), fix120.
76. Griffin, J.; Taw, E.; Gosavi, A.; Thornburg, N. E.; Pramanda, I.; Lee, H.-S.; Gray, K. A.; Notestein, J. M.; Wells, G., Hybrid Approach for Selective Sulfoxidation via Bioelectrochemically Derived Hydrogen Peroxide over a Niobium (V)–Silica Catalyst. *ACS Sustainable Chemistry & Engineering* **2018,** *6*, (6), 7880-7889.
77. Collins, F. S.; Morgan, M.; Patrinos, A., The Human Genome Project: lessons from large-scale biology. *Science* **2003,** *300*, (5617), 286-290.
78. Metzker, M. L., Sequencing technologies—the next generation. *Nature reviews genetics* **2010,** *11*, (1), 31.
79. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F. O., Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research* **2013,** *41*, (1), e1-e1.
80. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M., Comparison of next-generation sequencing systems. *BioMed Research International* **2012,** *2012*.

81. Scholz, M. B.; Lo, C.-C.; Chain, P. S., Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.* **2012,** *23*, (1), 9-15.
82. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X., The potential and challenges of nanopore sequencing. In *Nanoscience And Technology: A Collection of Reviews from Nature Journals*, World Scientific: 2010; pp 261-268.
83. Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W., Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods* **2010,** *7*, (6), 461.
84. Mitra, R. D.; Shendure, J.; Olejnik, J.; Church, G. M., Fluorescent in situ sequencing on polymerase colonies. *Anal. Biochem.* **2003,** *320*, (1), 55-65.
85. Metzker, M. L., Emerging technologies in DNA sequencing. *Genome research* **2005,** *15*, (12), 1767-1776.
86. Caporaso, J. G.; Knight, R., Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. **2011**, 1-7.
87. Woese, C. R.; Fox, G. E., Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci.* **1977,** *74*, (11), 5088-5090.
88. Han, X. Y.; Pham, A. S.; Tarrand, J. J.; Sood, P. K.; Luthra, R., Rapid and accurate identification of mycobacteria by sequencing hypervariable regions of the 16S ribosomal RNA gene. *American journal of clinical pathology* **2002,** *118*, (5), 796-801.
89. Schloss, P. D.; Westcott, S. L.; Ryabin, T.; Hall, J. R.; Hartmann, M.; Hollister, E. B.; Lesniewski, R. A.; Oakley, B. B.; Parks, D. H.; Robinson, C. J., Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* **2009,** *75*, (23), 7537-7541.
90. Edgar, R. C., UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Meth* **2013,** *10*, (10), 996-998.
91. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F., VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **2016,** *4*, e2584.
92. Callahan, B. J.; McMurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J. A.; Holmes, S. P., DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods* **2016,** *13*, (7), 581.
93. Mahé, F.; Rognes, T.; Quince, C.; de Vargas, C.; Dunthorn, M., Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* **2014,** *2*, e593.
94. Callahan, B. J.; McMurdie, P. J.; Holmes, S. P., Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017,** *11*, (12), 2639.
95. Amir, A.; McDonald, D.; Navas-Molina, J. A.; Kopylova, E.; Morton, J. T.; Xu, Z. Z.; Kightley, E. P.; Thompson, L. R.; Hyde, E. R.; Gonzalez, A., Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems* **2017,** *2*, (2), e00191-16.
96. Faith, D. P., Conservation evaluation and phylogenetic diversity. In *Biological conservation*, Elsevier: 1992; Vol. 61, pp 1-10.
97. Hill, M. O., Diversity and evenness: a unifying notation and its consequences. In *Ecology*, Eco Soc America: 1973; Vol. 54, pp 427-432.
98. Attrill, M. J., A testable linear model for diversity trends in estuaries. *Journal of Animal Ecology* **2002,** *71*, (2), 262-269.

99. Lozupone, C.; Knight, R., UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **2005,** *71*, (12), 8228-8235.
100. Chen, J.; Bittinger, K.; Charlson, E. S.; Hoffmann, C.; Lewis, J.; Wu, G. D.; Collman, R. G.; Bushman, F. D.; Li, H., Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **2012,** *28*, (16), 2106-2113.
101. Legendre, P.; FORTIN, M. J., Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular ecology resources* **2010,** *10*, (5), 831-844.
102. Liu, Q., Variation partitioning by partial redundancy analysis (RDA). *Environmetrics: The official journal of the International Environmetrics Society* **1997,** *8*, (2), 75-85.
103. Legendre, P.; Legendre, L., Numerical ecology: second English edition. *Developments in environmental modelling* **1998,** *20*.
104. Buttigieg, P. L.; Ramette, A., A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol. Ecol.* **2014,** *90*, (3), 543-550.
105. Oksanen, J.; Blanchet, F. G.; Kindt, R.; Legendre, P.; Minchin, P. R.; O'hara, R.; Simpson, G. L.; Solymos, P.; Stevens, M. H. H.; Wagner, H., Package 'vegan'. *Community ecology package, version* **2013,** *2*, (9).
106. McMurdie, P. J.; Holmes, S., phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **2013,** *8*, (4), e61217.
107. Turnbaugh, P. J.; Ley, R. E.; Hamady, M.; Fraser-Liggett, C. M.; Knight, R.; Gordon, J. I., The human microbiome project. *Nature* **2007,** *449*, (7164), 804.
108. Gilbert, J. A.; Jansson, J. K.; Knight, R., The Earth Microbiome project: successes and aspirations. *BMC biology* **2014,** *12*, (1), 69.
109. Edgar, R. C.; Haas, B. J.; Clemente, J. C.; Quince, C.; Knight, R., UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011,** *27*, (16), 2194-2200.
110. Acinas, S. G.; Marcelino, L. A.; Klepac-Ceraj, V.; Polz, M. F., Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J. Bacteriol.* **2004,** *186*, (9), 2629-2635.
111. Aßhauer, K. P.; Wemheuer, B.; Daniel, R.; Meinicke, P., Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* **2015,** *31*, (17), 2882-2884.
112. Sczyrba, A.; Hofmann, P.; Belmann, P.; Koslicki, D.; Janssen, S.; Dröge, J.; Gregor, I.; Majda, S.; Fiedler, J.; Dahms, E., Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nat. Methods* **2017,** *14*, (11), 1063.
113. Compeau, P. E.; Pevzner, P. A.; Tesler, G., How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.* **2011,** *29*, (11), 987.
114. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G., QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013,** *29*, (8), 1072-1075.
115. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P. A., metaSPAdes: a new versatile metagenomic assembler. *Genome research* **2017**, gr. 213959.116.
116. Peng, Y.; Leung, H. C.; Yiu, S.-M.; Chin, F. Y., IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012,** *28*, (11), 1420-1428.
117. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W., MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015,** *31*, (10), 1674-1676.

118. Vollmers, J.; Wiegand, S.; Kaster, A.-K., Comparing and evaluating metagenome assembly tools from a microbiologist's perspective-Not only size matters! *PloS one* **2017,** *12*, (1), e0169662.
119. Langmead, B.; Salzberg, S. L., Fast gapped-read alignment with Bowtie 2. *Nature methods* **2012,** *9*, (4), 357.
120. Li, H.; Durbin, R., Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **2009,** *25*, (14), 1754-1760.
121. Quinlan, A. R.; Hall, I. M., BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010,** *26*, (6), 841-842.
122. 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* **2014,** *2*, (1), 26.
123. Kang, D. D.; Froula, J.; Egan, R.; Wang, Z., MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015,** *3*, e1165.
124. Alneberg, J.; Bjarnason, B. S.; de Bruijn, I.; Schirmer, M.; Quick, J.; Ijaz, U. Z.; Loman, N. J.; Andersson, A. F.; Quince, C., CONCOCT: clustering contigs on coverage and composition. *arXiv preprint arXiv:1312.4038* **2013**.
125. Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **2015**, gr. 186072.114.
126. Bowers, R. M.; Kyrpides, N. C.; Stepanauskas, R.; Harmon-Smith, M.; Doud, D.; Reddy, T.; Schulz, F.; Jarett, J.; Rivers, A. R.; Eloe-Fadrosh, E. A., Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **2017,** *35*, (8), 725.
127. Uritskiy, G. V.; DiRuggiero, J.; Taylor, J., MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *bioRxiv* **2018**, 277442.
128. Sieber, C. M.; Probst, A. J.; Sharrar, A.; Thomas, B. C.; Hess, M.; Tringe, S. G.; Banfield, J. F., Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature microbiology* **2018**, 1.
129. Song, W.-Z.; Thomas, T., Binning_refiner: improving genome bins through the combination of different binning programs. *Bioinformatics* **2017,** *33*, (12), 1873-1875.
130. Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J., Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **2010,** *11*, (1), 119.
131. Finn, R. D.; Clements, J.; Eddy, S. R., HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **2011,** *39*, (suppl_2), W29-W37.
132. Leprevost, F. d. V.; Barbosa, V. C.; Francisco, E. L.; Perez-Riverol, Y.; Carvalho, P. C., On best practices in the development of bioinformatics software. *Frontiers in genetics* **2014,** *5*, 199.
133. da Veiga Leprevost, F.; Grüning, B. A.; Alves Aflitos, S.; Röst, H. L.; Uszkoreit, J.; Barsnes, H.; Vaudel, M.; Moreno, P.; Gatto, L.; Weber, J., BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics* **2017,** *33*, (16), 2580-2582.
134. Belmann, P.; Dröge, J.; Bremges, A.; McHardy, A. C.; Sczyrba, A.; Barton, M. D., Bioboxes: standardised containers for interchangeable bioinformatics software. *Gigascience* **2015,** *4*, (1), 47.

135. Di Tommaso, P.; Palumbo, E.; Chatzou, M.; Prieto, P.; Heuer, M. L.; Notredame, C., The impact of Docker containers on the performance of genomic pipelines. *PeerJ* **2015,** *3*, e1273.
136. Kurtzer, G. M.; Sochat, V.; Bauer, M. W., Singularity: Scientific containers for mobility of compute. *PloS one* **2017,** *12*, (5), e0177459.
137. Arkin, A. P.; Stevens, R. L.; Cottingham, R. W.; Maslov, S.; Henry, C. S.; Dehal, P.; Ware, D.; Perez, F.; Harris, N. L.; Canon, S., The DOE systems biology knowledgebase (KBase). *bioRxiv* **2016**, 096354.
138. Seviour, R.; Halkjær, P.; Nielsen, R., *Microbial ecology of activated sludge*. IWA publishing: 2010; Vol. 9, p 9781780401645.
139. Wells, G. F.; Park, H.-D.; Eggleston, B.; Francis, C. A.; Criddle, C. S., Fine-scale bacterial community dynamics and the taxa-time relationship within a full-scale activated sludge bioreactor. *Water Res.* **2011,** *45*, (17), 5476-5488.
140. Strous, M.; Fuerst, J. A.; Kramer, E. H.; Logemann, S.; Muyzer, G.; van de Pas-Schoonen, K. T.; Webb, R.; Kuenen, J. G.; Jetten, M. S., Missing lithotroph identified as new planctomycete. *Nature* **1999,** *400*, (6743), 446-449.
141. Shade, A.; Caporaso, J. G.; Handelsman, J.; Knight, R.; Fierer, N., A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J* **2013,** *7*, (8), 1493-1506.
142. Bjørnstad, O. N.; Ims, R. A.; Lambin, X., Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends in Ecology & Evolution* **1999,** *14*, (11), 427-432.
143. Liebhold, A.; Koenig, W. D.; Bjørnstad, O. N., Spatial synchrony in population dynamics. *Annual Review of Ecology, Evolution, and Systematics* **2004**, 467-490.
144. Kruskal, W., Relative importance by averaging over orderings. *The American Statistician* **1987,** *41*, (1), 6-10.
145. Wagner, M.; Loy, A.; Nogueira, R.; Purkhold, U.; Lee, N.; Daims, H., Microbial community composition and function in wastewater treatment plants. **2002**, 1-16.
146. Martins, A. M. P.; Pagilla, K.; Heijnen, J. J.; van Loosdrecht, M. C. M., Filamentous bulking sludge—a critical review. *Water Res.* **2004,** *38*, (4), 793-817.
147. Guo, F.; Zhang, T., Profiling bulking and foaming bacteria in activated sludge by high throughput sequencing. *Water Res.* **2012,** *46*, (8), 2772-2782.
148. Moran, P., The statistical analysis of the Canadian lynx cycle. *Australian Journal of Zoology* **1953,** *1*, (3), 291-298.
149. Ydenberg, R. C.; Dill, L. M., The economics of fleeing from predators. *Advances in the Study of Behavior* **1986,** *16*, (C), 229-249.
150. Andersson, A. F.; Riemann, L.; Bertilsson, S., Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* **2009,** *4*, (2), 171-181.
151. Lee, S.-H.; Kang, H.-J.; Park, H.-D., Influence of influent wastewater communities on temporal variation of activated sludge communities. In *Water Res.*, Elsevier Ltd: 2015; Vol. 73, pp 132-144.
152. Saunders, A. M.; Albertsen, M.; Vollertsen, J.; Nielsen, P. H., The activated sludge ecosystem contains a core community of abundant organisms. *ISME J* **2015,** *10*, (1), 11-20.
153. Wells, G. F.; Wu, C. H.; Piceno, Y. M.; Eggleston, B.; Brodie, E. L.; DeSantis, T. Z.; Andersen, G. L.; Hazen, T. C.; Francis, C. A.; Criddle, C. S., Microbial biogeography across a full-scale wastewater treatment plant transect: evidence for immigration between coupled processes. *Appl. Microbiol. Biotechnol.* **2014,** *98*, (10), 4723-4736.

154. Ju, F.; Xia, Y.; Guo, F.; Wang, Z.; Zhang, T., Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environ. Microbiol.* **2014,** *16*, (8), 2421-2432.
155. Poretsky, R.; Rodriguez-R, L. M.; Luo, C.; Tsementzi, D.; Konstantinidis, K. T., Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE* **2014,** *9*, (4), e93827.
156. Zaneveld, J. R.; Lozupone, C.; Gordon, J. I.; Knight, R., Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Research* **2010**, gkq066.
157. Langille, M. G. I.; Zaneveld, J.; Caporaso, J. G.; McDonald, D.; Knights, D.; Reyes, J. A.; Clemente, J. C.; Burkepile, D. E.; Thurber, R. L. V.; Knight, R.; Beiko, R. G.; Huttenhower, C., Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **2013,** *31*, (9), 814-821.
158. Wittebolle, L.; Marzorati, M.; Clement, L.; Balloi, A.; Daffonchio, D.; Heylen, K.; De Vos, P.; Verstraete, W.; Boon, N., Initial community evenness favours functionality under selective stress. *Nature* **2009,** *458*, (7238), 623-626.
159. Martins, A. M. P.; Pagilla, K.; Heijnen, J. J.; van Loosdrecht, M. C. M., Filamentous bulking sludge—a critical review. In *Water Res.*, Vol. 38, pp 793-817.
160. De los Reyes, F. L., Foaming. In *Microbial ecology of activated sludge. IWA Publishing, London*, 2010; pp 215-259.
161. Randall, C. W.; Buth, D., Nitrite build-up in activated sludge resulting from temperature effects. *Water Pollution Control Federation Journal* **1984**, 1039-1044.
162. Fuhrman, J. A.; Hewson, I.; Schwalbach, M. S.; Steele, J. A.; Brown, M. V.; Naeem, S., Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl. Acad. Sci.* **2006,** *103*, (35), 13104-13109.
163. Shade, A.; Kent, A. D.; Jones, S. E.; Newton, R. J.; Triplett, E. W.; McMahon, K. D., Interannual dynamics and phenology of bacterial communities in a eutrophic lake. . *Limnology and Oceanography* **2007,** *52*, 483-494.
164. Ju, F.; Zhang, T., Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* **2014,** *9*, (3), 683-695.
165. Fuhrman, J. A.; Cram, J. A.; Needham, D. M., Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology* **2015,** *13*, (3), 133-146.
166. Eaton, A. D., Clesceri, L. S., Greenberg, A. E., Franson, M. A. H., *Standard methods for the examination of water and wastewater*. 1998.
167. Rognes, T. Vsearch. https://github.com/torognes/vsearch. https://github.com/torognes/vsearch
168. DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L., Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* **2006,** *72*, (7), 5069-5072.
169. Caporaso, J. G.; Bittinger, K.; Bushman, F. D.; DeSantis, T. Z.; Andersen, G. L.; Knight, R., PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010,** *26*, (2), 266-267.
170. Price, M. N.; Dehal, P. S.; Arkin, A. P., FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010,** *5*, (3), e9490.

171. Edgar, R. C., Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010,** *26*, (19), 2460-2461.
172. Lindeman, R. H.; Merenda, P. F.; Gold, R. Z., *Introduction to bivariate and multivariate analysis.* Scott, Foresman: 1980.
173. Grömping, U., Relative importance for linear regression in R: the package relaimpo. *Journal of statistical software* **2006,** *17*, (1), 1-27.
174. Aitchison, J., The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)* **1982**, 139-177.
175. Zeglin, L. H., Stream microbial diversity in response to environmental changes: review and synthesis of existing research. *Front. Microbiol.* **2015,** *6*, (20131760), 11512-11515.
176. Fierer, N.; Jackson, R. B., The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103*, (3), 626-631.
177. Hanson, C. A.; Fuhrman, J. A.; Horner-Devine, M. C.; Martiny, J. B., Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* **2012,** *10*, (7), 497-506.
178. Battin, T. J.; Besemer, K.; Bengtsson, M. M.; Romani, A. M.; Packmann, A. I., The ecology and biogeochemistry of stream biofilms. *Nature Reviews Microbiology* **2016,** *14*, (4), 251-263.
179. Hullar, M. A. J.; Kaplan, L. A.; Stahl, D. A., Recurring Seasonal Dynamics of Microbial Communities in Stream Habitats. *Applied and Environmental Microbiology* **2006,** *72*, (1), 713-722.
180. Crump, B. C.; Amaral-Zettler, L. A.; Kling, G. W., Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J* **2012,** *6*, (9), 1629-1639.
181. Besemer, K.; Singer, G.; Quince, C.; Bertuzzo, E.; Sloan, W.; Battin, T. J., Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proc. R. Soc. B* **2013,** *280*, (1771), 20131760.
182. Drummond, J. D.; Davies-Colley, R. J.; Stott, R.; Sukias, J. P.; Nagels, J. W.; Sharp, A.; Packman, A. I., Microbial Transport, Retention, and Inactivation in Streams: A Combined Experimental and Stochastic Modeling Approach. *Environ. Sci. Technol.* **2015,** *49*, (13), 7825-7833.
183. Basu, N. B.; Thompson, S. E.; Rao, P. S. C., Hydrologic and biogeochemical functioning of intensively managed catchments: A synthesis of top-down analyses. *Water Resour. Res.* **2011,** *47*, (10).
184. Mulholland, P. J.; Helton, A. M.; Poole, G. C.; Hall, R. O.; Hamilton, S. K.; Peterson, B. J.; Tank, J. L.; Ashkenas, L. R.; Cooper, L. W.; Dahm, C. N., Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature* **2008,** *452*, (7184), 202-205.
185. Li, H.; Sivapalan, M.; Tian, F.; Liu, D., Water and nutrient balances in a large tile-drained agricultural catchment: a distributed modeling study. *Hydrology and Earth System Sciences* **2010,** *14*, (11), 2259-2275.
186. Jangid, K.; Williams, M. A.; Franzluebbers, A. J.; Schmidt, T. M.; Coleman, D. C.; Whitman, W. B., Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. *Soil Biology and Biochemistry* **2011,** *43*, (10), 2184-2193.

187. Allison, S. D.; Martiny, J. B., Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* **2008,** *105*, (Supplement 1), 11512-11519.
188. Sutton, S. D.; Findlay, R. H., Sedimentary microbial community dynamics in a regulated stream: East Fork of the Little Miami River, Ohio. *Environ. Microbiol.* **2003,** *5*, (4), 256-266.
189. Crump, B. C.; Peterson, B. J.; Raymond, P. A.; Amon, R. M.; Rinehart, A.; McClelland, J. W.; Holmes, R. M., Circumpolar synchrony in big river bacterioplankton. *Proc. Natl. Acad. Sci.* **2009,** *106*, (50), 21208-21212.
190. Ma, B.; Wang, H.; Dsouza, M.; Lou, J.; He, Y.; Dai, Z.; Brookes, P. C.; Xu, J.; Gilbert, J. A., Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **2016**.
191. Fierer, N.; Leff, J. W.; Adams, B. J.; Nielsen, U. N.; Bates, S. T.; Lauber, C. L.; Owens, S.; Gilbert, J. A.; Wall, D. H.; Caporaso, J. G., Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci.* **2012,** *109*, (52), 21390-21395.
192. Fierer, N.; Ladau, J.; Clemente, J. C.; Leff, J. W.; Owens, S. M.; Pollard, K. S.; Knight, R.; Gilbert, J. A.; McCulley, R. L., Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* **2013,** *342*, (6158), 621-624.
193. Whitaker, R. J., Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science* **2003,** *301*, (5635), 976-978.
194. Martiny, J. B. H.; Bohannan, B. J. M.; Brown, J. H.; Colwell, R. K.; Fuhrman, J. A.; Green, J. L.; Horner-Devine, M. C.; Kane, M.; Krumins, J. A.; Kuske, C. R.; Morin, P. J.; Naeem, S.; Øvreås, L.; Reysenbach, A.-L.; Smith, V. H.; Staley, J. T., Microbial biogeography: putting microorganisms on the map. *Nat Rev Micro* **2006,** *4*, (2), 102-112.
195. Tischendorf, L.; Fahrig, L., On the usage and measurement of landscape connectivity. *Oikos* **2000,** *90*, (1), 7-19.
196. Nemergut, D. R.; Costello, E. K.; Hamady, M.; Lozupone, C.; Jiang, L.; Schmidt, S. K.; Fierer, N.; Townsend, A. R.; Cleveland, C. C.; Stanish, L., Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* **2011,** *13*, (1), 135-144.
197. Wang, J.; Shen, J.; Wu, Y.; Tu, C.; Soininen, J.; Stegen, J. C.; He, J.; Liu, X.; Zhang, L.; Zhang, E., Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* **2013,** *7*, (7), 1310-1321.
198. Connell, J. H., Diversity in Tropical Rain Forests and Coral Reefs. *Science* **1978,** *199*, (4335), 1302-1310.
199. Schilling, K. E.; Helmers, M., Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds: Exploratory hydrograph analysis. *Hydrological Processes* **2008,** *22*, (23), 4497-4506.
200. Crump, B. C.; Armbrust, E. V.; Baross, J. A., Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Applied and environmental microbiology* **1999,** *65*, (7), 3192-3204.
201. Feris, K. P.; Ramsey, P. W.; Frazar, C.; Rillig, M. C.; Gannon, J. E.; Holben, W. E., Structure and Seasonal Dynamics of Hyporheic Zone Microbial Communities in Free-Stone Rivers of the Western United States. *Microb Ecol* **2003,** *46*, (2), 200-215.
202. Lindström, E. S.; Bergström, A.-K., Community composition of bacterioplankton and cell transport in lakes in two different drainage areas. *Aquatic Sciences* **2005,** *67*, (2), 210-219.

203. Kobayashi, T.; Ryder, D. S.; Gordon, G.; Shannon, I.; Ingleton, T.; Carpenter, M.; Jacobs, S. J., Short-term response of nutrients, carbon and planktonic microbial communities to floodplain wetland inundation. *Aquat Ecol* **2008,** *43*, (4), 843-858.
204. Horner-Devine, M. C.; Silver, J. M.; Leibold, M. A.; Bohannan, B. J.; Colwell, R. K.; Fuhrman, J. A.; Green, J. L.; Kuske, C. R.; Martiny, J. B.; Muyzer, G., A Comparison of taxon co‐occurrence patterns for macro‐and microorganisms. *Ecology* **2007,** *88*, (6), 1345-1353.
205. Barberán, A.; Bates, S. T.; Casamayor, E. O.; Fierer, N., Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **2012,** *6*, (2), 343-351.
206. Dumbrell, A. J.; Nelson, M.; Helgason, T.; Dytham, C.; Fitter, A. H., Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J.* **2010,** *4*, (3), 337-345.
207. Widder, S.; Besemer, K.; Singer, G. A.; Ceola, S.; Bertuzzo, E.; Quince, C.; Sloan, W. T.; Rinaldo, A.; Battin, T. J., Fluvial network organization imprints on microbial co-occurrence networks. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111*, (35), 12799-12804.
208. Rhoads, B. L.; Lewis, Q. W.; Andresen, W., Historical changes in channel network extent and channel planform in an intensively managed landscape: Natural versus human-induced effects. *Geomorphology* **2016,** *252*, 17-31.
209. Grimley, D. A.; Anders, A. M.; Bettis, E. A.; Bates, B. L.; Wang, J. J.; Butler, S. K.; Huot, S., Using magnetic fly ash to identify post-settlement alluvium and its record of atmospheric pollution, central USA. *Anthropocene* **2017,** *17*, 84-98.
210. Keefer, L.; Bauer, E.; Markus, M. *Hydrologic and Nutrient Monitoring of the Lake Decatur Watershed*; 2010; p 82 p. + appendices.
211. NASS, U. United States Department of Agriculture-National Agricultural Statistics Service. (April 25, 2017),
212. Collman, R. D., Cochran, C.C., and Werner, S.E. *Soil survey of McLean County, Illinois*; United States Department of Agriculture: Washington, D.C., 2004; p 421.
213. Endres, T. J. *Soil survey of Champaign County, Illinois*; United States Department of Agriculture: Washington, D.C., 2001.
214. Soller, D. R., S.D. Price, J.P. Kempton, and R.C. Berg *Three-dimensional geologic maps of Quaternary sediments in east-central Illinois*; United States Geological Survey: Reston, Virginia, 1999.
215. Stumpf, A. J., and W.S. Dey *Understanding the Mahomet aquifer: Geological, geophysical, and hydrogeological studies in Champaign County and adjacent areas*; Illinois State Geological Survey: Champaign, 2012.
216. US EPA, M., Methods for chemical analysis of water and wastes. *Cincinnati, Ohio, USA* **1979**.
217. Parada, A. E.; Needham, D. M.; Fuhrman, J. A., Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **2016,** *18*, (5), 1403-1414.
218. Navas-Molina, J. A.; Peralta-Sánchez, J. M.; González, A.; McMurdie, P. J.; Vázquez-Baeza, Y.; Xu, Z.; Ursell, L. K.; Lauber, C.; Zhou, H.; Song, S. J., Advancing our understanding of the human microbiome using QIIME. *Methods in enzymology* **2013,** *531*, 371.
219. Oksanen, J.; Kindt, R.; Legendre, P.; O'Hara, B.; Stevens, M. H. H.; Oksanen, M. J.; Suggests, M. *The vegan package*, 2007.
220. Weiss, S.; Van Treuren, W.; Lozupone, C.; Faust, K.; Friedman, J.; Deng, Y.; Xia, L. C.; Xu, Z. Z.; Ursell, L.; Alm, E. J.; Birmingham, A.; Cram, J. A.; Fuhrman, J. A.; Raes, J.; Sun, F.;

Zhou, J.; Knight, R., Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J* **2016,** *10*, (7), 1669-1681.
221. Williams, R. J.; Howe, A.; Hofmockel, K. S., Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in microbiology* **2014,** *5*, 358.
222. Schwager, E., Bielski, C., and Weingart, G. *ccrepe: ccrepe_and_nc.score*, R package version 1.4.0; 2014.
223. Strimmer, K., A unified approach to false discovery rate estimation. *BMC Bioinformatics* **2008,** *9*, (1), 303.
224. Csardi, G.; Nepusz, T., The igraph software package for complex network research. *InterJournal, Complex Systems* **2006,** *1695*, (5), 1-9.
225. Knights, D.; Kuczynski, J.; Charlson, E. S.; Zaneveld, J.; Mozer, M. C.; Collman, R. G.; Bushman, F. D.; Knight, R.; Kelley, S. T., Bayesian community-wide culture-independent microbial source tracking. *Nat Meth* **2011,** *8*, (9), 761-763.
226. Zhang, Y.; Angelidaki, I., Microbial electrolysis cells turning to be versatile technology: Recent advances and future challenges. *Water Res.* **2014,** *56*, (C), 11-25.
227. Modin, O.; Fukushi, K., Production of high concentrations of H 2O 2in a bioelectrochemical reactor fed with real municipal wastewater. *Environ. Technol.* **2013,** *34*, (19), 2737-2742.
228. Fu, L.; You, S.-J.; Yang, F.-l.; Gao, M.-m.; Fang, X.-h.; Zhang, G.-q., Synthesis of hydrogen peroxide in microbial fuel cell. *J. Chem. Technol. Biotechnol.* **2010,** *85*, (5), 715-719.
229. Wang, Y.; Feng, C.; Li, Y.; Gao, J.; Yu, C.-P., Enhancement of emerging contaminants removal using Fenton reaction driven by H 2 O 2-producing microbial fuel cells. *Chem. Eng. J.* **2017,** *307*, 679-686.
230. Noyori, R.; Aoki, M.; Sato, K., Green oxidation with aqueous hydrogen peroxide. *Chem. Commun.* **2003**, (16), 1977-1986.
231. Sheldon, R., Synthetic and mechanistic aspects of metal-catalysed epoxidations with hydroperoxides. *J. Mol. Catal.* **1980,** *7*, (1), 107-126.
232. Thomas, J. M.; Raja, R.; Lewis, D. W., Single‐Site Heterogeneous Catalysts. *Angew. Chem. Int. Ed.* **2005,** *44*, (40), 6456-6482.
233. Adam, W.; Mitchell, C. M.; Saha-Möller, C. R., Chemoselective methyltrioxorhenium (VII)-catalyzed sulfoxidations with hydrogen peroxide. *Tetrahedron* **1994,** *50*, (46), 13121-13124.
234. Cojocariu, A. M.; Mutin, P. H.; Dumitriu, E.; Aboulaich, A.; Vioux, A.; Fajula, F.; Hulea, V., Non-hydrolytic SiO 2–TiO 2 mesoporous xerogels—Efficient catalysts for the mild oxidation of sulfur organic compounds with hydrogen peroxide. *Catal. Today* **2010,** *157*, (1), 270-274.
235. Maggi, R.; Chitsaz, S.; Loebbecke, S.; Piscopo, C. G.; Sartori, G.; Schwarzer, M., Highly chemoselective metal-free oxidation of sulfides with diluted H 2 O 2 in a continuous flow reactor. *Green Chem.* **2011,** *13*, (5), 1121-1123.
236. Zhang, H.; Chen, C.; Liu, R.; Xu, Q.; Zhao, W., Transition-metal-free highly efficient aerobic oxidation of sulfides to sulfoxides under mild conditions. *Molecules* **2009,** *15*, (1), 83-92.
237. Sloboda‐Rozner, D.; Witte, P.; Alsters, P. L.; Neumann, R., Aqueous Biphasic Oxidation: A Water‐Soluble Polyoxometalate Catalyst for Selective Oxidation of Various Functional Groups with Hydrogen Peroxide. *Adv. Synth. Catal.* **2004,** *346*, (2‐3), 339-345.

238. Thornburg, N. E.; Notestein, J. M., Rate and Selectivity Control in Thioether and Alkene Oxidation with H2O2 over Phosphonate‐Modified Niobium (V)–Silica Catalysts. *ChemCatChem* **2017,** *9*, (19), 3714-3724.
239. Contreras, C. A.; Ignacio-de Leon, P. A. A.; Notestein, J. M., Synthesis of a family of peracid-silica materials and their use as alkene epoxidation reagents. *Microporous Mesoporous Mater.* **2016,** *225*, 289-295.
240. Zakzeski, J.; Bruijnincx, P. C.; Jongerius, A. L.; Weckhuysen, B. M., The catalytic valorization of lignin for the production of renewable chemicals. *Chem. Rev.* **2010,** *110*, (6), 3552-3599.
241. Gralnick, J. A.; Newman, D. K., Extracellular respiration. *Mol. Microbiol.* **2007,** *65*, (1), 1-11.
242. Song, C.; Zhang, J., Electrocatalytic Oxygen Reduction Reaction. In *PEM Fuel Cell Electrocatalysts and Catalyst Layers: Fundamentals and Applications*, Zhang, J., Ed. Springer London: London, 2008; pp 31-134.
243. Paliteiro, C.; Hamnett, A.; Goodenough, J. B., The electroreduction of oxygen on pyrolytic graphite. *J. Electroanal. Chem. Interfacial. Electrochem.* **1987,** *233*, (1-2), 147-159.
244. Logan, B. E., *Microbial fuel cells*. John Wiley & Sons: 2008.
245. Cheng, S.; Liu, H.; Logan, B. E., Increased performance of single-chamber microbial fuel cells using an improved cathode structure. *Electrochem. Commun.* **2006,** *8*, (3), 489-494.
246. Chen, J.-y.; Li, N.; Zhao, L., Three-dimensional electrode microbial fuel cell for hydrogen peroxide synthesis coupled to wastewater treatment. *J. Power Sources* **2014,** *254*, (c), 316-322.
247. Li, N.; An, J.; Zhou, L.; Li, T.; Li, J.; Feng, C.; Wang, X., A novel carbon black graphite hybrid air-cathode for efficient hydrogen peroxide production in bioelectrochemical systems. *J. Power Sources* **2016,** *306*, 495-502.
248. Wei, J.; Liang, P.; Huang, X., Recent progress in electrodes for microbial fuel cells. *Bioresour. Technol.* **2011,** *102*, (20), 9335-9344.
249. Nicoll, W.; Smith, A., Stability of dilute alkaline solutions of hydrogen peroxide. *Ind. Eng. Chem.* **1955,** *47*, (12), 2548-2554.
250. McTavish, D.; Buckley, M. M.-T.; Heel, R. C., Omeprazole. *Drugs* **1991,** *42*, (1), 138-170.
251. Ki, D.; Popat, S. C.; Torres, C. s. I., Reduced overpotentials in microbial electrolysis cells through improved design, operation, and electrochemical characterization. *Chem. Eng. J.* **2016,** *287*, (C), 181-188.
252. Ki, D.; Popat, S. C.; Rittmann, B. E.; Torres, C. I., H2O2 production in microbial electrochemical cells fed with primary sludge. *Environ. Sci. Technol.* **2017**.
253. Foller, P.; Bombard, R., Processes for the production of mixtures of caustic soda and hydrogen peroxide via the reduction of oxygen. *J. Appl. Electrochem.* **1995,** *25*, (7), 613-627.
254. Zhang, F.; Ge, Z.; Grimaud, J.; Hurst, J.; He, Z., Long-term performance of liter-scale microbial fuel cells treating primary effluent installed in a municipal wastewater treatment facility. *Environ. Sci. Technol.* **2013,** *47*, (9), 4941-4948.
255. Zhuang, L.; Zhou, S.; Li, Y.; Yuan, Y., Enhanced performance of air-cathode two-chamber microbial fuel cells with high-pH anode and low-pH cathode. *Bioresour. Technol.* **2010,** *101*, (10), 3514-3519.

256. Thornburg, N. E.; Thompson, A. B.; Notestein, J. M., Periodic Trends in Highly Dispersed Groups IV and V Supported Metal Oxide Catalysts for Alkene Epoxidation with H2O2. *ACS Catal.* **2015,** *5*, (9), 5077-5088.
257. Thornburg, N. E.; Nauert, S. L.; Thompson, A. B.; Notestein, J. M., Synthesis− Structure–Function Relationships of Silica-Supported Niobium (V) Catalysts for Alkene Epoxidation with H2O2. *ACS Catal.* **2016,** *6*, (9), 6124-6134.
258. Tiozzo, C.; Bisio, C.; Carniato, F.; Gallo, A.; Scott, S. L.; Psaro, R.; Guidotti, M., Niobium–silica catalysts for the selective epoxidation of cyclic alkenes: the generation of the active site by grafting niobocene dichloride. *PCCP* **2013,** *15*, (32), 13354-13362.
259. Carniti, P.; Gervasini, A.; Marzo, M., Dispersed NbO x catalytic phases in silica matrixes: influence of niobium concentration and preparative route. *J. Phys. Chem. C* **2008,** *112*, (36), 14064-14074.
260. Hulea, V.; Moreau, P.; Di Renzo, F., Thioether oxidation by hydrogen peroxide using titanium-containing zeolites as catalysts. *J. Mol. Catal. A: Chem.* **1996,** *111*, (3), 325-332.
261. Kholdeeva, O. A.; Kovaleva, L. A.; Maksimovskaya, R. I.; Maksimov, G. M., Kinetics and mechanism of thioether oxidation with H 2 O 2 in the presence of Ti (IV)-substituted heteropolytungstates. *J. Mol. Catal. A: Chem.* **2000,** *158*, (1), 223-229.
262. Iwamoto, M.; Tanaka, Y.; Hirosumi, J.; Kita, N., Asymmetric oxidation of sulfide to sulfoxides on Ti-containing mesoporous silica catalysts with hydrogen peroxide in the presence of optically active tartaric acid. *Chem. Lett.* **2001,** *30*, (3), 226-227.
263. Moreau, P.; Hulea, V.; Gomez, S.; Brunel, D.; Di Renzo, F., Oxidation of sulfoxides to sulfones by hydrogen peroxide over Ti-containing zeolites. *Appl. Catal., A* **1997,** *155*, (2), 253-263.
264. Mania, P.; Verel, R.; Jenny, F.; Hammond, C.; Hermans, I., Thermal Restructuring of Silica‐Grafted TiClx Species and Consequences for Epoxidation Catalysis. *Chem. Eur. J.* **2013,** *19*, (30), 9849-9858.
265. Tiozzo, C.; Palumbo, C.; Psaro, R.; Bisio, C.; Carniato, F.; Gervasini, A.; Carniti, P.; Guidotti, M., The stability of niobium-silica catalysts in repeated liquid-phase epoxidation tests: A comparative evaluation of in-framework and grafted mixed oxides. *Inorg. Chim. Acta* **2015,** *431*, 190-196.
266. Ahn, Y.; Logan, B. E., Saline catholytes as alternatives to phosphate buffers in microbial fuel cells. *Bioresour. Technol.* **2013,** *132*, 436-439.
267. Isarain-Chávez, E.; De La Rosa, C.; Martínez-Huitle, C. A.; Peralta-Hernández, J. M., On-site hydrogen peroxide production at pilot flow plant: application to electro-Fenton process. *Int. J. Electrochem. Sci* **2013,** *8*, 3084-3094.
268. Monzon, O.; Yang, Y.; Kim, J.; Heldenbrand, A.; Li, Q.; Alvarez, P. J., Microbial fuel cell fed by Barnett Shale produced water: Power production by hypersaline autochthonous bacteria and coupling to a desalination unit. *Biochem. Eng. J* **2017,** *117*, 87-91.
269. Li, W.-W.; Yu, H.-Q.; He, Z., Towards sustainable wastewater treatment by using microbial fuel cells-centered technologies. *Energy Environ. Sci.* **2014,** *7*, (3), 911-924.
270. Sherman, J.; Chin, B.; Huibers, P.; Garcia-Valls, R.; Hatton, T. A., Solvent replacement for green processing. *Environ. Health Perspect.* **1998,** *106*, (Suppl 1), 253.
271. Lee, H.-S.; Rittmann, B. E., Significance of biological hydrogen oxidation in a continuous single-chamber microbial electrolysis cell. *Environ. Sci. Technol.* **2009,** *44*, (3), 948-954.

272. Torres, C. s. I.; Kato Marcus, A.; Rittmann, B. E., Proton transport inside the biofilm limits electrical current generation by anode-respiring bacteria. *Biotechnol. Bioeng.* **2008,** *100*, (5), 872-881.
273. Smolders, G.; Van der Meij, J.; Van Loosdrecht, M.; Heijnen, J., Model of the anaerobic metabolism of the biological phosphorus removal process: stoichiometry and pH influence. *Biotechnol. Bioeng.* **1994,** *43*, (6), 461-470.
274. Torres, C. I.; Marcus, A. K.; Lee, H.-S.; Parameswaran, P.; Krajmalnik-Brown, R.; Rittmann, B. E., A kinetic perspective on extracellular electron transfer by anode-respiring bacteria. *FEMS Microbiol. Rev.* **2009,** *34*, (1), 3-17.
275. Nogueira, R.; Oliveira, M.; Paterlini, W., Simple and fast spectrophotometric determination of H2O2 in photo-Fenton reactions using metavanadate. *Talanta* **2005,** *66*, (1), 86-91.
276. Zhuang, K.; Izallalen, M.; Mouser, P.; Richter, H.; Risso, C.; Mahadevan, R.; Lovley, D. R., Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. *The ISME journal* **2011,** *5*, (2), 305.
277. Khandelwal, R. A.; Olivier, B. G.; Röling, W. F.; Teusink, B.; Bruggeman, F. J., Community flux balance analysis for microbial consortia at balanced growth. *PloS one* **2013,** *8*, (5), e64567.
278. Song, H.-S.; Cannon, W. R.; Beliaev, A. S.; Konopka, A., Mathematical modeling of microbial community dynamics: a methodological review. *Processes* **2014,** *2*, (4), 711-752.
279. Liu, Z.; Liu, J.; Zhang, S.; Su, Z., Study of operational performance and electrical response on mediator-less microbial fuel cells fed with carbon-and protein-rich substrates. *Biochem. Eng. J.* **2009,** *45*, (3), 185-191.
280. Pant, D.; Van Bogaert, G.; Diels, L.; Vanbroekhoven, K., A review of the substrates used in microbial fuel cells (MFCs) for sustainable energy production. *Bioresour. Technol.* **2010,** *101*, (6), 1533-1543.
281. Freguia, S.; Teh, E. H.; Boon, N.; Leung, K. M.; Keller, J.; Rabaey, K., Microbial fuel cells operating on mixed fatty acids. *Bioresour. Technol.* **2010,** *101*, (4), 1233-1238.
282. Freguia, S.; Rabaey, K.; Yuan, Z.; Keller, J., Electron and carbon balances in microbial fuel cells reveal temporary bacterial storage behavior during electricity generation. *Environ. Sci. Technol.* **2007,** *41*, (8), 2915-2921.
283. Stratford, J. P.; Beecroft, N. J.; Slade, R. C.; Grüning, A.; Avignone-Rossa, C., Anodic microbial community diversity as a predictor of the power output of microbial fuel cells. *Bioresour. Technol.* **2014,** *156*, 84-91.
284. Peng, Y.; Leung, H. C.; Yiu, S.-M.; Chin, F. Y., Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics* **2011,** *27*, (13), i94-i101.
285. Segata, N.; Börnigen, D.; Morgan, X. C.; Huttenhower, C., PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* **2013,** *4*, 2304.
286. Logan, B. E., Exoelectrogenic bacteria that power microbial fuel cells. *Nature Reviews Microbiology* **2009,** *7*, (5), 375.
287. Gray, J. S.; Birmingham, J. M.; Fenton, J. I., Got black swimming dots in your cell culture? Identification of Achromobacter as a novel cell culture contaminant. *Biologicals* **2010,** *38*, (2), 273-277.

288. Yu, J.; Cho, S.; Kim, S.; Cho, H.; Lee, T., Comparison of exoelectrogenic bacteria detected using two different methods: U-tube microbial fuel cell and plating method. *Microbes Environ.* **2012,** *27*, (1), 49-53.
289. Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **2011,** *17*, (1), pp. 10-12.
290. Andrews, S., FastQC: a quality control tool for high throughput sequence data. **2010**.
291. Li, H., Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* **2013**.
292. Seemann, T., Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **2014,** *30*, (14), 2068-2069.
293. Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000,** *28*, (1), 27-30.
294. Caspi, R.; Foerster, H.; Fulcher, C. A.; Kaipa, P.; Krummenacker, M.; Latendresse, M.; Paley, S.; Rhee, S. Y.; Shearer, A. G.; Tissier, C., The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2007,** *36*, (suppl_1), D623-D631.
295. Krzywinski, M. I.; Schein, J. E.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S. J.; Marra, M. A., Circos: an information aesthetic for comparative genomics. *Genome Res.* **2009**.
296. Dunbar, J.; Barns, S. M.; Ticknor, L. O.; Kuske, C. R., Empirical and theoretical bacterial diversity in four Arizona soils. *Applied and environmental microbiology* **2002,** *68*, (6), 3035-3045.
297. Staley, J. T.; Reysenbach, A.-L., *Biodiversity of microbial life*. Wiley: 2002.
298. Griffiths, R. I.; Thomson, B. C.; James, P.; Bell, T.; Bailey, M.; Whiteley, A. S., The bacterial biogeography of British soils. *Environ. Microbiol.* **2011,** *13*, (6), 1642-1654.
299. Evans, S. E.; Wallenstein, M. D., Soil microbial community response to drying and rewetting stress: does historical precipitation regime matter? *Biogeochemistry* **2012,** *109*, (1-3), 101-116.
300. Fierer, N.; Schimel, J. P.; Holden, P. A., Variations in microbial community composition through two soil depth profiles. *Soil Biology and Biochemistry* **2003,** *35*, (1), 167-176.
301. 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.* **2012,** *50*, 58-65.
302. Gambrell, R., Trace and toxic metals in wetlands—a review. *Journal of environmental Quality* **1994,** *23*, (5), 883-891.
303. Stewart, F. J.; Ottesen, E. A.; DeLong, E. F., Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME journal* **2010,** *4*, (7), 896.
304. Koren, S.; Phillippy, A. M., One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* **2015,** *23*, 110-120.
305. Rotaru, A.-E.; Shrestha, P. M.; Liu, F.; Shrestha, M.; Shrestha, D.; Embree, M.; Zengler, K.; Wardman, C.; Nevin, K. P.; Lovley, D. R., A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to Methanosaeta for the reduction of carbon dioxide to methane. *Energy & Environmental Science* **2014,** *7*, (1), 408-415.