Impact of Environmental Disturbances on Aquatic Microbial Community Structure and Function

BY

ADIT CHAUDHARY
B.E., M.Sc., Birla Institute of Technology and Science, Pilani, India, 2012

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biological Sciences
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Defense Committee:

Rachel Poretsky, Chair and Advisor
Stefan Green
Eric Stabb
Karl J. Rockne, Civil and Materials Engineering
John J. Kelly, Loyola University Chicago

ACKNOWLEDGMENTS

There are so many people who were instrumental in my success during graduate school. First, I would like to thank my advisor and mentor Dr. Rachel Poretsky, who provided me constant support, guidance and encouragement. Her enthusiasm, curiosity and love for science is truly infectious. I am very grateful for all the opportunities that she provided throughout my PhD. Most importantly, I am thankful to her for creating a wonderful lab atmosphere that encourages freedom and creativity which thoroughly helped me become an independent scientist.

I would like to thank my committee members: Dr. Stefan Green, Dr. John Kelly, Dr.Karl Rockne, Dr. Donald Morrison and Dr. Eric Stabb for their valuable suggestions and guidance. I greatly appreciate Dr. Green's help with sequencing problems and providing valuable suggestions for troubleshooting. I am grateful for my current and previous lab mates: Imrose, Dr. Binh Chu, Anirban, Dr. Melissa Pierce, Emily, Dolores and Sarah for their immense help, support and friendships. I am also thankful to the friends I made in the Ecology and Evolution program for their collegiality and support in graduate school.

A special thanks to my family for always encouraging me to be my best self. To my mother, who made me the person I am and who is my constant source of support. To my sister, for the much-needed laughs and to my father for teaching me the value of hard work and perseverance. I would not be here without their unconditional love and belief in my capabilities.

Last but definitely not the least, I would like to thank my fiancée Isha for being on this journey with me. For her love, support and delightful companionship all through these years. Thank you!

CONTRIBUTION OF AUTHORS

Chapter 1 is an introduction to the dissertation topic and includes some background information related to it. I am the sole author for this chapter.

Chapter 2 is a published research article in *mSphere* journal. I was the primary author and contributed to the study's data analysis, drafting the manuscript and partially contributed to the sample collection and laboratory work. Imrose Kauser and Anirban Ray contributed to sample collection and laboratory work. Rachel Poretsky was the principal investigator and contributed to the study's conception and sampling design, sample collection and drafting the manuscript.

Chapter 3 is currently in review for publication in *Limnology and Oceanography* journal. I was the primary author and contributed to the study's sampling and laboratory work, data analysis and drafting the manuscript. Sarah Turner contributed to the sample collection and field work. Rachel Poretsky was the principal investigator and contributed to the study's conception, data analysis and drafting the manuscript.

Chapter 4 is an unpublished work that is an extension of some of the preliminary work done in Chapter 3. I am the primary author and contributed to the study's conception, sampling and laboratory work, data analysis and drafting the manuscript. Personnel from the Poretsky Lab contributed to the sampling effort. Rachel Poretsky is the principal investigator and contributed to the study's conception and drafting the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

List of Supplementary Tables

LIST OF FIGURES

List of Supplementary Figures

LIST OF ABBREVIATIONS

C	Carbon
DOM	Dissolved Organic Matter
DOC	Dissolved Organic Carbon
t-DOM	Terrestrial-derived Dissolved Organic Matter
WGS	Whole-Genome Shotgun
WWTP	Wastewater Treatment Plant
OTU	Operational Taxonomic Unit
MGD	Million Gallons per Day
DAPI	4',6-diamidino-2-phenylindole
CAWS	Chicago Area Waterway System
GO	Gene Ontology
BLAST	Basic Local Alignment Search Tool
blastn	Nucleotide BLAST
blastp	Protein BLAST
MAG	Metagenome-Assembled Genome
RPKM	Reads Per Kilobase of transcript, per Million mapped reads
HMW	High Molecular Weight
MiGA	Microbial Genomes Atlas
RNA	Ribonucleic Acid
DNA	Deoxyribonucleic Acid
QIIME	Quantitative Insights Into Microbial Ecology

SIMPER	Similarity Percentages analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
ANOSIM	Analysis of Similarities
BIX	Biological Index
HIX	Humification Index
CDOM	Chromophoric Dissolved Organic Matter
FDOM	Fluoroscent Dissolved Organic Matter
FT-ICR-MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

SUMMARY

Despite their observed importance in mediating ecological and human health, our ability to predict aquatic microbial community response to major environmental disturbances is still limited. In this dissertation, the research focuses on understanding the various ways environmental perturbations can impact microbial community structure and function in regionally important aquatic ecosystems – the Chicago Area Waterways and Lake Michigan. I have aimed to address the following broad questions:

1. **What are the short-term implications of stormflow as a perturbation event on the microbial community structure and functional potential in a highly urbanized section of the Chicago Area Waterways?**
2. **How do the bacterial community diversity and dissolved organic matter (DOM) metabolism compare between the nearshore and offshore regions of Lake Michigan in light of the recent ecological changes caused by the invasive dreissenid mussels?**
3. **What is the relative potential of bacterial communities in nearshore and offshore Lake Michigan in the post-mussel period to utilize terrestrial-derived DOM (t-DOM)?**

Chapter 2 of the dissertation addresses the first question. Urban streams are susceptible to various anthropogenic stressors on their ecology and environment, and rain-induced storm flow events represent one such source of complex physical, chemical and biological perturbations. In this study, we focused on the short-term impacts of storm flow events on the microbial community dynamics of North Shore Channel, a highly urbanized section of the Chicago Area

Waterways. Using a combination of 16S rRNA gene sequencing and shotgun metagenomics, we investigated the stream microbial community composition and functional potential during dry and wet weather conditions between 2013-2015. The results demonstrated general trends present in the stream under storm flow versus base flow conditions and also highlighted the influence of increased wastewater treatment plant (WWTP) effluent flow following rain in shifting the stream microbial community from abundant freshwater taxa to those more associated with urban/anthropogenic settings. Shifts in the taxonomic composition were also linked to changes in the functional gene content. Overall, results from this study highlighted the significant changes in an urban stream microbial community during rain-induced storm flow conditions, with potential environmental and public health implications.

Chapter 3 primarily focuses on addressing the third question, but also provides some preliminary data for the second question. Lake Michigan is one of the largest lakes in the world, and over the last two decades has witnessed significant ecological changes due to proliferation of invasive dreissenid mussels into deeper regions of the lake. Carbon fixed by phytoplankton production constitutes a major source of labile DOM for bacterioplankton in Lake Michigan. However, the recent expansion of invasive dreissenid mussels into offshore lake waters has caused dramatic declines in phytoplankton production, negatively impacting the labile DOM pool available for bacteria. In addition, coastal waters in the southeastern part of the lake receive terrestrial-derived DOM (t-DOM) and nutrients from large tributaries such as the Kalamazoo River. How this spatial variation in the DOM pool impacts the bacterial community composition and function in Lake Michigan, and the relative importance of t-DOM in coastal and offshore bacterial community metabolism in the post-mussel period are poorly understood. In this project, we performed a preliminary investigation of Lake Michigan bacterial community structure and

activity across a nearshore-to-offshore transect beginning near the mouth of the Kalamazoo River. In addition, using water from the nearshore and offshore locations we evaluated short-term bacterioplankton response to a pulse of t-DOM (leaf litter leachate). The bacterial community composition and activity for the natural and t-DOM enriched samples was characterized using combined metagenomics and metatranscriptomics. The results from this study showed a significantly higher number of transcripts for *Synechococcus* in the offshore as compared to nearshore, but despite this and certain other differences for DOM-related transporter gene transcripts, the nearshore and offshore bacterial communities showed a similar capacity to utilize t-DOM. These findings have important implications in explaining bacterial community dynamics related to carbon metabolism in southern Lake Michigan in the post-mussel period.

Chapter 4 addresses the second question. By performing further microbial sampling of the nearshore-to-offshore Kalamazoo River transect, this project aimed at extending the preliminary work done in Chapter 3 to identify the spatiotemporal variation in the *in situ* dissolved organic matter (DOM) metabolism and the associated microorganisms in southern Lake Michigan. In addition to the samples collected across the nearshore-to-offshore transect in September 2015 for Chapter 3, we collected samples across the transect in summer 2017. These samples were processed to obtain deeply sequenced metagenomes, which were combined with the 2015 metagenomes to more reliably test the broad questions related to carbon metabolism in nearshore and offshore Lake Michigan. Results demonstrated broadly similar communities in the two regions both in terms of taxonomic composition and functional gene content. However, there were differences in the relative abundance of specific bacterial groups and functional processes across the transect, such as a higher abundance of genes encoding for aromatic compound metabolism in nearshore versus offshore. In addition, the chemical composition of the bulk

DOM pool in Lake Michigan was characterized using spectrofluorometric methods and showed correlation with some of the microbial community trends, with the nearshore waters comprising a significantly higher fraction of aromatic and terrestrial-derived humic DOM in comparison to offshore. Overall, the results from this study highlighted the specific differences in microbial community composition and potential carbon metabolism between the oligotrophic offshore and the more productive nearshore waters of southern Lake Michigan.

I. INTRODUCTION

Freshwater lakes and rivers are critical water sources. Despite covering a small fraction of Earth's surface, they contribute significantly to regional and global carbon budgets. It is estimated that of all the carbon delivered annually to inland waters from terrestrial sources, about 40% is released to the atmosphere and about 12% is stored in the sediments of these water bodies before the remaining carbon is delivered to the oceans (1). Freshwater ecosystems are also home to a rich biodiversity adapted to the unique habitats within these systems. Microorganisms such as heterotrophic bacteria are an important component of this biodiversity and perform critical ecological functions. Their role in nutrient cycling and assimilating constituents of the dissolved organic matter (DOM) pool is fundamental to the biogeochemical flux and food web dynamics in these ecosystems (2,3). However, aquatic bacterial communities are sensitive to changes in nutrient regimes and substrate availability arising from fluxes in terrestrial inputs and other perturbations (4). Moreover, anthropogenically influenced terrestrial loadings to aquatic systems are important for the aquatic microbiome not only from an ecological perspective, but also in a public health context (5). Despite the importance of freshwater microbial communities, our understanding of the impact of environmental disturbances on their diversity and activity remains limited.

In ecology and particularly in microbial ecology, a disturbance can be defined as any event that causes a change in the physical or chemical characteristics of the direct environment of a community (6). The response of a microbial community to a disturbance event can be assessed in terms of its stability. Community stability can be defined as the degree to which a community is resistant or resilient (recovery of a community to its original state after a

disturbance) to a disturbance event (6). Disturbances can either be discreet, short-term events (pulse) or more long-term and continuous in nature (press). For microbial communities in aquatic ecosystems, examples of pulse disturbances include rainfall-associated flooding and algal blooms. The proliferation of an invasive species in an aquatic body and the associated changes in the food web structure and nutrient regime is an example of a press disturbance. A microbial community's response to a disturbance can be evaluated by assessing changes in the population structure or community functional diversity, or both. Modern omics technologies such as metagenomics and metatranscriptomics provide an opportunity to measure microbial community composition, functional gene content and expression at a high resolution and thus enable a robust investigation of community responses to disturbance events.

A. Anthropogenic and Natural Sources of Disturbance to Aquatic Microbial Communities

Aquatic ecosystems in urban landscapes are susceptible to various types of anthropogenically driven disturbances. Microbes and chemical pollutants from the wastewater infrastructure in cities reach surface waters of stream, rivers, lakes and estuaries in large quantities everyday through various dissemination routes (7). These routes include non-point discharge from leaky and failing sewage infrastructure, stormwater runoff, and treated effluent from wastewater treatment facilities. Studies using cultivation and/or molecular approaches have documented the prevalence of fecal indicator bacteria, potential pathogens and antibiotic resistant bacteria/genes in surface waters exposed to inputs from the urban wastewater infrastructure (5, 8, 9). However, efforts so far to establish broad patterns of pollution in the urban aquatic microbiome have been challenging due to important effects of local factors such as

weather and impervious land cover (7). Also, while most work in these urban aquatic ecosystems has focused on tracking pollutant microorganisms/genes in the context of public health issues, the impact of wastewater discharge on the resident microbial community in aquatic systems from an ecological standpoint remains less explored (10). The use of whole-genome shotgun (WGS) metagenomics-based approaches to evaluate the impact of wastewater discharge on microbial community structure and functional diversity in urban aquatic ecosystems holds promise from both a public health and ecological perspective.

For large freshwater ecosystems such as the Great Lakes, the sources of ecological disturbance can be more varied and at different scales in comparison to urban aquatic systems. Although anthropogenic activities in the catchment area can directly affect the microbial ecology of large lakes (11), it is the indirect effects of human activity on the internal ecological processes and food web dynamics of these lakes that often have a larger impact on the overall microbial community dynamics. For example, the increased nutrient levels in Lake Erie due to loadings from tributaries with intensively farmed landscapes has caused frequent occurrence of cyanobacterial harmful algal blooms. These blooms have acted as a significant biological disturbance to the lake's bacterial community, affecting its composition and diversity (12). On the other hand, the impact of invasive dreissenid mussels on the ecology and microbial food web of Lake Michigan has been quite different – these mussels are prolific filter feeders and their expansion into deeper waters of the lake over the last 2 decades has decimated the annual spring diatom bloom (13, 14). This loss of primary productivity can have implications for the rest of the food web, including bacterioplankton production as now there would be less availability of labile phytoplankton-derived dissolved organic matter (DOM) for bacterial consumption (15). Large

lakes in other parts of the world are also facing similar stresses related to habitat alteration, invasive species and climate change (16, 17).

B. Using Omics Approaches to Investigate Disturbance Scenarios in Aquatic Ecosystems

Advances in DNA sequencing technology as well as computational tools in the last two decades have enabled microbial ecologists to investigate questions about microbial community dynamics in various habitats without being limited by the need to isolate members of a microbial community and study them in laboratory settings. For instance, the use of next-generation amplicon sequencing of the phylogenetic marker gene encoding for 16S ribosomal RNA in prokaryotes has enabled evaluating the microbial ecology of urban rivers and estuaries and the source tracking of potential pathogens and fecal indicator bacteria in these ecosystems (4, 5, 18). Direct retrieval and sequencing of the whole community genomic DNA (shotgun metagenomics) has provided researchers studying aquatic microbial communities information about not only the community composition (who is there?) but also about their genetic content (what are they potentially doing?) (19, 20). Additionally, development of methods to assess gene expression in environmental microbial communities (metatranscriptomics, (21)) provides a powerful approach to investigate community metabolic response to disturbance events, such as in short-term manipulation experiments (22). These omics approaches when combined with the advanced bioinformatics tools such as isolation of population genomes from metagenomes (23) and their metabolic modeling (24) can help us arrive at some level of mechanistic understanding about the disturbance effects on microbial community dynamics and potential ecosystem-level consequences.

C. References

1. Cole JJ, Prairie YT, Caraco NF, Mcdowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ, Melack J. 2007. Plumbing the Global Carbon Cycle : Integrating Inland Waters into the Terrestrial Carbon Budget 171–184.

2. Cole J, Findlay S, Pace M. 1988. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Mar Ecol Prog Ser 43:1–10.

3. Pernthaler J, Amann R. 2005. Fate of Heterotrophic Microbes in Pelagic Habitats: Focus on Populations. Microbiol Mol Biol Rev 69:440–461.

4. Newton RJ, McLellan SL. 2015. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. Front Microbiol 6:1–13.

5. Newton RJ, Bootsma MJ, Morrison HG, Sogin ML, McLellan SL. 2013. A Microbial Signature Approach to Identify Fecal Pollution in the Waters Off an Urbanized Coast of Lake Michigan. Microb Ecol 65:1011–1023.

6. Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, Matulich KL, Schmidt TM, Handelsman J. 2012. Fundamentals of microbial community resistance and resilience. Front Microbiol 3:1–19.

7. Newton RJ, McClary JS. 2019. The flux and impact of wastewater infrastructure microorganisms on human and ecosystem health. Curr Opin Biotechnol 57:145–150.

8. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, Michael I, Fatta-Kassinos D. 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Environ 447:345–360.

9. Zhang S, Pang S, Wang PF, Wang C, Han N, Liu B, Han B, Li Y, Anim-Larbi K. 2016. Antibiotic concentration and antibiotic-resistant bacteria in two shallow urban lakes after stormwater event. Environ Sci Pollut Res 23:9984–9992.

10. Drury B, Rosi-Marshall E, Kelly JJ. 2013. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol 79:1897–1905.

11. Chu BTT, Petrovich ML, Chaudhary A, Wright D, Murphy B, Wells G, Poretsky R. 2018. Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. Appl Environ Microbiol 84:1–15.

12. Berry MA, Davis TW, Cory RM, Duhaime MB, Johengen TH, Kling GW, Marino JA, Den Uyl PA, Gossiaux D, Dick GJ, Denef VJ. 2017. Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities. Environ

Microbiol 19:1149–1162.

13. Fahnenstiel G, Pothoven S, Vanderploeg H, Klarer D, Nalepa T, Scavia D. 2010. Recent changes in primary production and phytoplankton in the offshore region of southeastern Lake Michigan. J Great Lakes Res 36:20–29.

14. Vanderploeg H a., Bunnell DB, Carrick HJ, Höök TO. 2015. Complex interactions in Lake Michigan's rapidly changing ecosystem. J Great Lakes Res 41:1–6.

15. Biddanda B a., Cotner JB. 2002. Love handles in aquatic ecosystems: The role of dissolved organic carbon drawdown, resuspended sediments, and terrigenous inputs in the carbon balance of Lake Michigan. Ecosystems 5:431–445.

16. Moore M V., Hampton SE, Izmest'Eva LR, Silow EA, Peshkova E V., Pavlov BK. 2009. Climate change and the world's "sacred sea"-lake Baikal, Siberia. Bioscience 59:405–417.

17. McKenna JE. 2019. The Laurentian Great Lakes: A case study in ecological disturbance and climate change. Fish Manag Ecol 26:486–499.

18. Fisher JC, Newton RJ, Dila DK, McLellan SL. 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elem Sci Anthr 3:000064.

19. Zhang S-Y, Tsementzi D, Hatt JK, Bivins A, Khelurkar N, Brown J, Tripathi SN, Konstantinidis KT. 2018. Intensive allochthonous inputs along the Ganges River and their effect on microbial community composition and dynamics. Environ Microbiol 1–40.

20. Meziti A, Tsementzi D, Ar. Kormas K, Karayanni H, Konstantinidis KT. 2016. Anthropogenic effects on bacterial diversity and function along a river-to-estuary gradient in Northwest Greece revealed by metagenomics. Environ Microbiol 18:4640–4652.

21. Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, Pickering M, Pate WM, Moran MA, Hollibaugh JT. 2005. Analysis of microbial gene transcripts in environmental samples. Appl Environ Microbiol 71:4121–4126.

22. Poretsky RS, Sun S, Mou X, Moran MA. 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ Microbiol 12:616–627.

23. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.

24. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia JM, Chia JM, Colasanti R, Conrad N, Davis JJ, Davison BH, Dejongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W,

Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States department of energy systems biology knowledgebase. Nat Biotechnol 36:566–569.

.

II. Taxon-Driven Functional Shifts Associated with Storm Flow in an Urban Stream Microbial Community[1]

A. Abstract

Urban streams are susceptible to stormwater and sewage inputs that can impact their ecological health and water quality. Microbial communities in streams play important functional roles, and their composition and metabolic potential can help assess ecological state and water quality. Although these environments are highly heterogenous, little is known about the influence of isolated perturbations, such as those resulting from rain events on urban stream microbiota. Here, we examined the microbial community composition and diversity in an urban stream during dry and wet weather conditions with both 16S rRNA gene sequencing across multiple years and shotgun metagenomics to more deeply analyze a single storm flow event. Metagenomics was used to assess population-level dynamics as well as shifts in the microbial community taxonomic profile and functional potential before and after a substantial rainfall. The results demonstrated general trends present in the stream under storm flow versus base flow conditions and also highlighted the influence of increased effluent flow following rain in shifting the stream microbial community from abundant freshwater taxa to those more associated with urban/anthropogenic settings. Shifts in the taxonomic composition were also linked to changes in functional gene content, particularly for transmembrane transport and organic substance biosynthesis. We also observed an increase in relative abundance of genes encoding degradation

---

of organic pollutants and antibiotic resistance after rain. Overall, this study highlighted some differences in the microbial community of an urban stream under storm flow conditions and showed the impact of a storm flow event on the microbiome from an environmental and public health perspective.

B. Introduction

Streams and rivers are important freshwater resources, used for recreation, agriculture, domestic water sources, and industrial purposes. By storing, processing, and transporting terrestrially derived nutrients and organic matter, rivers play an important ecological role in linking biogeochemical cycles between terrestrial and aquatic ecosystems (1). Over the last century, many streams and rivers have witnessed rapid urbanization and anthropogenic development of their drainage basins, which has exposed them to frequent external inputs in the form of wastewater treatment plant (WWTP) effluent, industrial discharge, and sewer/stormwater overflows. These inputs often impact stream hydrological, physicochemical, and biological characteristics (2). For streams and rivers that serve as wastewater and/or stormwater outfall sites, rain-induced storm flow events are especially influential, as they often lead to an increased influx of WWTP effluent and unregulated waste via combined sewer overflows (CSOs) (3, 4). These perturbations bring in nutrients, a variety of microorganisms, including pathogens, and chemical pollutants such as steroid hormones that impact water quality, biodiversity, and ecosystem health (2, 3, 5, 6).

Because urban aquatic streams are typically highly variable systems that are regularly subject to anthropogenic inputs, it is unclear how much isolated perturbations such as rainfall and associated increases in storm flow might influence the water column microbial community,

even in the short-term. Studies investigating urban river microbiota using genetic markers for fecal bacteria or 16S rRNA gene-based microbial community surveys have shown the presence of human fecal contamination, “urban signature” bacteria, and changes in community composition in streams and rivers impacted by WWTP effluent, stormwater, and CSOs (7–11). Moreover, others have documented the possible influx of antibiotic-resistant bacteria and pathogens from WWTP effluent (12, 13) and stormwater events (6, 14) into urban environments, further signifying the importance of evaluating the persistence of these organisms and their impact on the riverine microbiome from a public health perspective. While these studies provide valuable information about the effects of storm flow events on urban stream microbial content, they are limited to specific taxonomic and pollutant marker genes. Recent whole-genome shotgun (WGS) metagenomics-based approaches have explored community composition and functional dynamics in urban-impacted streams (15, 16), although a direct effect of storm flow on microbial dynamics remains less explored. A robust evaluation of the impacts of such isolated and short-term perturbations is critical for making predictions about the public health and possible longer-term ecological implications.

In this study, we used both 16S rRNA gene amplicon and shotgun metagenomics to analyze the water column microbial community during base flow and storm flow conditions in the North Shore Channel (NSC) stream, a section of the highly urbanized Chicago Area Waterway System (CAWS) (see Supplementary Figure S2.1). We focused on a site downstream of a WWTP and numerous CSO outflow points using 16S rRNA gene amplicon sequencing of samples from both base flow and storm flow over the course of multiple seasons and years. Additionally, samples obtained immediately before and shortly (<24 h) after a single rain event at the same site provided an opportunity for a deep analysis of short-term variability in the

taxonomic and functional composition of the water column microbiome using WGS metagenomics. Coupled with the 16S rRNA data from multiple samples, we were able to link some of these changes in the stream microbial taxonomic and functional profiles to storm flow conditions. Although our deep metagenomics-based analysis is centered around a single event, our findings provide a window into the variability and short-term changes in an urban freshwater system and set the groundwork for making predictions about possible ecosystem-level and public-health-related impacts of rainfall events on these systems. Overall, our results show that rain-associated WWTP effluent flow and perhaps CSOs impact the stream microbiome composition and functional potential, with the introduction of exogenous organisms to the system being a significant driver of the observed change.

C. Results and Discussion

Impact of rainfall on NSC microbial community composition. Rainfall can impact urban waterways by increasing effluent flow from WWTPs or causing combined sewer overflow events (CSOs) at outflow points along streams (4). The NSC site that we investigated has a WWTP (O'Brien Water Reclamation Plant) and several CSO outfall sites within a few kilometers upstream (Supplementary Figure S2.1) and often experiences increased flow from both following rainfall, including the two rain events reported in this study (Supplementary Figure S2.2). Sequences from 16S rRNA gene amplicons at five distinct times between 2013 and 2015 representing both summer and fall and stream base flow (dry weather; three samples) and storm flow (<24 h after rain; two samples) (with additional details in Supplementary Table S2.1) revealed both a temporal and rainfall-associated clustering of the samples at the operational taxonomic unit (OTU) level (principal-coordinate analysis [PCoA], Bray-Curtis metric) (Figure

2.1A). In particular, the separate clustering of storm flow and base flow samples along the principal axis 2 highlights the strong influence of rain on the microbial community composition, regardless of time/year sampled. Such changes might result from either a direct influx of allochthonous microbes or a shift in the resident microbial community in response to altered chemical conditions following rain, although none of the measured physicochemical parameters showed a statistically significant difference between storm flow and base flow conditions ($P > 0.05$, Welch's *t* test [Supplementary Table S2.1]). In addition to shifts in community composition, microbial diversity based on OTU richness and Good's coverage was slightly higher in the storm flow samples than the base flow samples (Supplementary Table S2.2), although the differences were not significant ($P > 0.05$, Welch's *t* test).

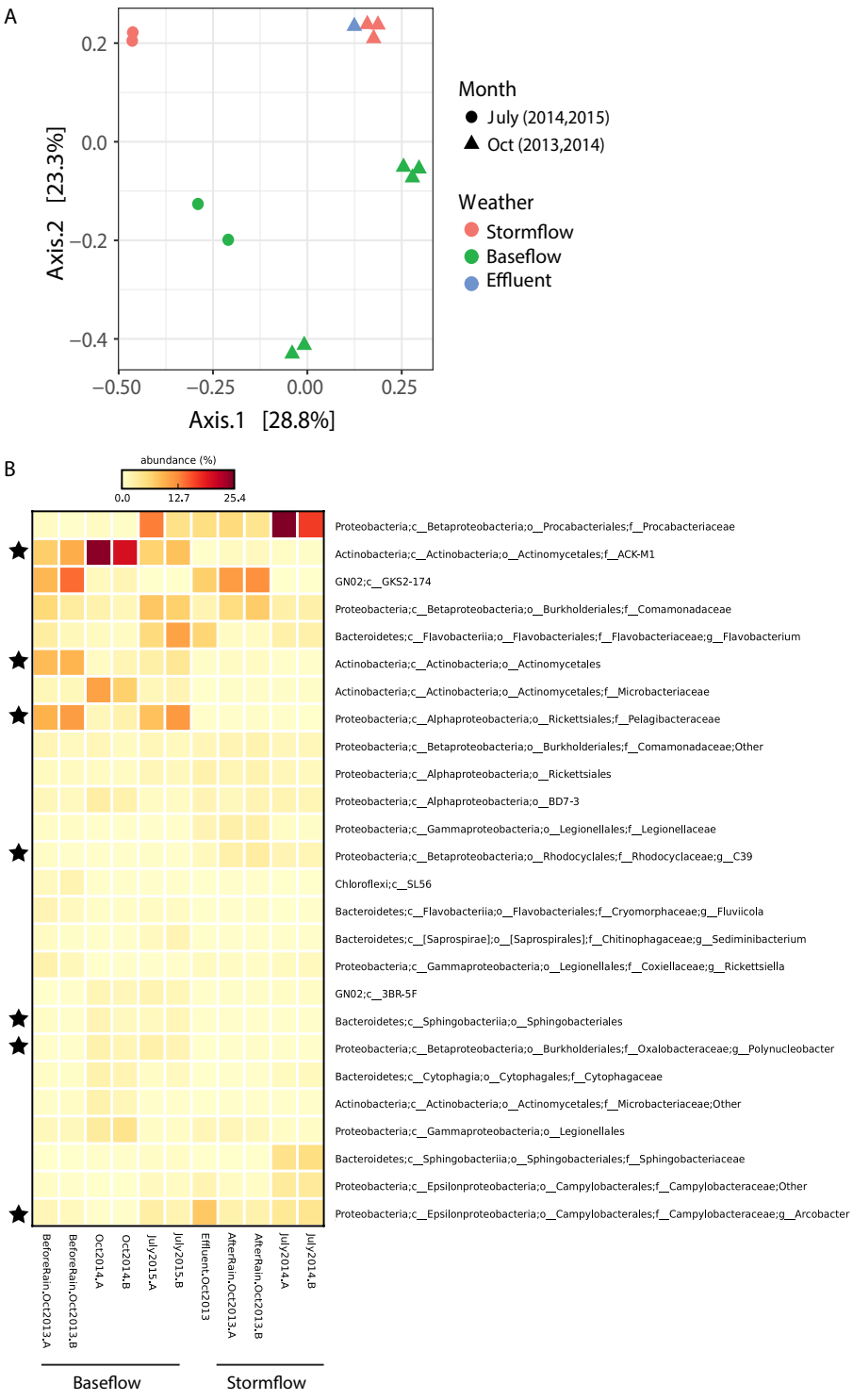
To analyze shifts in the microbial community across all storm flow versus base flow samples, OTUs were clustered at various hierarchical taxonomic levels. There was a difference in genus-based community compositions between the storm flow and base flow samples as per analysis of similarity (ANOSIM; Bray-Curtis metric, $R^2 = 0.5$, $P = 0.1$). Genus-level comparisons of microbial community composition revealed a significantly lower abundance of unknown genera within groups *Pelagibacteraceae*, ACK-M1, and *Actinomycetales* and a significantly higher abundance of *Arcobacter* and genus C39 within the family *Rhodocyclaceae* during storm flow compared to base flow ($P < 0.05$, Welch's *t* test) (Figure 2.1B). The ACK-M1 family of *Actinobacteria* and *Pelagibacteraceae* include common freshwater organisms that do not favor nutrient-rich conditions (17, 18), while genera within *Rhodocyclaceae* are *Betaproteobacteria*, known to take advantage of nutrient/substrate-rich conditions, likely due to higher growth rates (17). *Rhodocyclaceae* has previously been associated with urban streams and was reported to be abundant in impacted Milwaukee waterways (19). Similarly,

Arcobacter has often been associated with sewage and WWTP effluent (8, 9, 20). The increase in the relative abundance of these organisms in the NSC following rainfall could be due to point source inputs from the increased effluent flow and/or CSOs and was analyzed in more detail with shotgun metagenomics (described below).

Overall, the rain-associated changes in the microbial community composition appeared to be directly related to increased effluent; the after-rain community OTUs were more similar to those in the WWTP effluent than to those in the before-rain community (Figure 2.1A). This could be linked to a few taxa, such as unknown genera within families *Procabacteriaceae* and *Legionellacaea* as well as the genus *Arcobacter*, which were abundant in the effluent and increased in the stream after rain (Figure 2.1B).

Metagenomics-based microbial community composition before and after rain in North Shore Channel. The overall trends from the 16S rRNA gene-based analysis across seasons and years warranted a whole-community metagenomic analysis of more temporally resolved samples clustered around a large rainfall event. Here, we report our observations of a single, isolated event, acknowledging that this might not be representative of every rainfall event in this dynamic urban system. Instead, our results allow us to make predictions and better understand how urban microbial communities might be influenced by system-wide perturbations. Metagenomes with 4.06 to 16.21 million reads per library were obtained (Supplementary Table S2.3) from the same NSC site discussed above (Supplementary Figure S2.1) before and <24 h after a heavy rainfall that followed a dry period in October 2013 (Supplementary Figure S2.2). These were used to comprehensively identify short-term changes in the microbial taxonomic profile after the rain. The rain resulted in increased WWTP effluent flow into the stream for ~24 h following precipitation, from <200 million gal per day (MGD) to >300 MGD, and several

Figure 2.1. (A) Principal-coordinate analysis (PCoA; Bray-Curtis metric) of OTU-based microbial community diversity for North Shore Channel (NSC) water and WWTP effluent. Samples were obtained during either base flow or storm flow conditions between 2013 and 2015 in the summer (July) and fall (October). Each NSC time point is represented on the PCoA by biological duplicates, except for October 2013 storm flow and base flow samples, which also have sequencing duplicates for one of their biosamples. (B) Heat map representing the relative abundance (percentage of total 16S rRNA gene sequences) of dominant bacterial taxa classified until the lowest possible level (up to genus) for the NSC and effluent samples. Taxa highlighted with a star represent bacterial groups with significantly different relative abundance ($P < 0.05$, Welch's t test) between the storm flow and base flow samples of NSC. Two biological replicates marked as A and B represent each NSC time point, and the average value of these replicates per time point was used in Welch's t test between the two groups (storm flow and base flow).


A
0.2
0.0
−0.2
−0.4
Axis.2 [23.3%]
−0.50
−0.25
0.00
0.25
Axis.1 [28.8%]
Month
July (2014,2015)
Oct (2013,2014)
Weather
Stormflow
Baseflow
Effluent
B
abundance (%)
0.0
12.7
25.4
Proteobacteria;c__Betaproteobacteria;o__Procabacteriales;f__Procabacteriaceae
Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__ACK-M1
GN02;c__GKS2-174
Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae
Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium
Actinobacteria;c__Actinobacteria;o__Actinomycetales
Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae
Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Pelagibacteraceae
Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;Other
Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales
Proteobacteria;c__Alphaproteobacteria;o__BD7-3
Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Legionellaceae
Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__C39
Chloroflexi;c__SL56
Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Cryomorphaceae;g__Fluviicola
Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__Chitinophagaceae;g__Sediminibacterium
Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__Rickettsiella
GN02;c__3BR-5F
Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales
Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Polynucleobacter
Bacteroidetes;c__Cytophagia;o__Cytophagales;f__Cytophagaceae
Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;Other
Proteobacteria;c__Gammaproteobacteria;o__Legionellales
Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__Sphingobacteriaceae
Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;Other
Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Arcobacter
BeforeRain.Oct2013.A
BeforeRain.Oct2013.B
Oct2014.A
Oct2014.B
July2015.A
July2015.B
Effluent.Oct2013
AfterRain.Oct2013.A
AfterRain.Oct2013.B
July2014.A
July2014.B
Baseflow
Stormflow

CSO events at at least three outfall locations upstream of our sampled site within 10 h of rain (http://www.mwrd.org/irj/portal/anonymous/overview) (Supplementary Figure S2.2). Community coverage estimates using read redundancy (21) showed that the before-rain metagenomes captured between 50 and 60% of the community and the after-rain libraries captured approximately 40% (Supplementary Figure S2.3), indicating only a nominal increase in diversity after rainfall; as described above, a small increase in community OTU richness after rain was also observed with the 16S rRNA gene amplicon data (Supplementary Table S2.2). Furthermore, the concentrations of microbial cells in the before- and after-rain samples were determined by DAPI (4',6-diamidino-2-phenylindole) counts and found to be similar: 1.39×10^6 and 1.25×10^6 cells/ml, respectively. Previous studies have reported conflicting responses of microbial community diversity to urban inputs, with some showing an increase (19) and others a decrease (15, 22) relative to less-impacted conditions/systems. This may be due to different base conditions (operationally defined here as dry weather for at least 72 h); the NSC is characterized by significant urban effluent flow even in the absence of rain. While Lake Michigan provides the primary freshwater input, about 70% of the annual flow through the CAWS is contributed by the treated effluent discharge from WWTPs in the city (23) during both base flow and storm flow conditions. Our results do not show a strong pattern of change in microbial community diversity/richness during storm flow in NSC, perhaps because of the variable nature of urban stream microbial communities or due to the small size of this study. However, we hypothesize based on our results that individual rain events might not significantly impact microbial diversity in this system.

Despite overall similarities in microbial diversity and cell counts, numerous taxonomic differences were seen following rain, indicating that these changes likely reflect actual changes

in microbial populations. The microbial communities pre- and post- rainfall determined both from 16S rRNA genes and by assigning taxa to assembled metagenomic contigs showed overall concordance; however, we focused on the assembled contigs for a high-resolution, population-level characterization of the community and to evaluate possible links between taxonomic and functional changes in the microbiome (24). About ~67% of the large (>500-bp) contigs used by MyTaxa were classifiable at the phylum level, ~35% at the genus level, and 24% at the species level. At the phylum level (*Proteobacteria* divided into subphyla), several individual taxa showed significantly different relative abundances after rain with large effect sizes (Figure 2.2A). *Actinobacteria* and *Bacteroidetes* significantly decreased in relative abundance after rain, whereas *Gammaproteobacteria*, *Betaproteobacteria*, and *Chlamydia* significantly increased ($P < 0.05$, *t* test, false-discovery rate corrected) (Figure 2.2A). Similarity percentage (SIMPER) analysis (25) revealed that *Actinobacteria*, *Gammaproteobacteria*, and unclassified *Proteobacteria* contributed the most (35, 14, and 21%, respectively) to the differences in community compositions between the before- and after-rain samples at the phylum level. At the genus level, the decrease in relative abundance of innominate (unclassified at genus level) *Actinobacteria*,"*Candidatus Pelagibacter*," and *Streptomyces* as well as the increase in relative abundance of *Legionella* and *Rickettsia*-affiliated sequences after rain contributed to the major change (>50%) in community composition (Figure 2.2B). *Francisella*, *Nitrospira*, *Chlamydia*, and *Pseudomonas* were other major genera that increased significantly ($P < 0.05$, *t* test, FDR corrected) in relative abundance in the after-rain microbiome. As was observed with 16S rRNA amplicons in all samples (described above), the urban signature bacterium *Arcobacter* increased by >50% in relative abundance following rain, although the increase was not statistically significant (Figure 2.2B). *Legionella*, *Pseudomonas*, and *Arcobacter* have all been

previously associated with effluent contamination of urban waterways (20), supporting the significant role of increased effluent flow on the NSC microbiome. Increases in the relative abundance of other taxa such as *Francisella*, *Rickettsia*, and *Chlamydia* that comprise pathogenic species (26, 27) and are usually not abundant in aquatic environments could be a result of microbial influx from the effluent and/or the CSOs upstream. The decrease in the freshwater groups of *Actinobacteria* and *Pelagibacteria* after rain likely reflects a dilution effect on base flow NSC waters from the increased effluent and CSO flow. Several species, including *Francisella tularensis*, "*Candidatus Nitrospira defluvii*," *Legionella longbeachae*, and *Enterococcus faecalis*, were rare (<0.1% of the total sequences characterized by MyTaxa) in the before-rain microbiome but increased in relative abundance after rain to >0.1% (Supplementary Table S2.3). Most of these species are not common freshwater bacteria and are indicative of contamination.

Population-level changes in response to rainfall in the North Shore Channel. We followed population-level trends for abundant organisms that exhibited large changes in their relative abundance after rain. Organisms most similar to *Legionella pneumophila* increased 10-fold in relative abundance after rain and also comprised the largest fraction of characterized species (11%) in the after-rain microbiome. Reads were recruited to the longest contig assigned to *L. pneumophila* in the rain-associated samples, with roughly equal similarities (about 90 to 100% nucleotide identity) from each sample, suggesting the presence of the same population both before and after rain that increased substantially after rain (Supplementary Figure S2.4). This was supported by similarities in the average amino acid identity (AAI) of predicted protein-coding genes from *L. pneumophila* before and after rainfall contigs (60% and 63%, respectively) to the genome sequences of the environmental isolate *L. pneumophila strain LPE509* and the

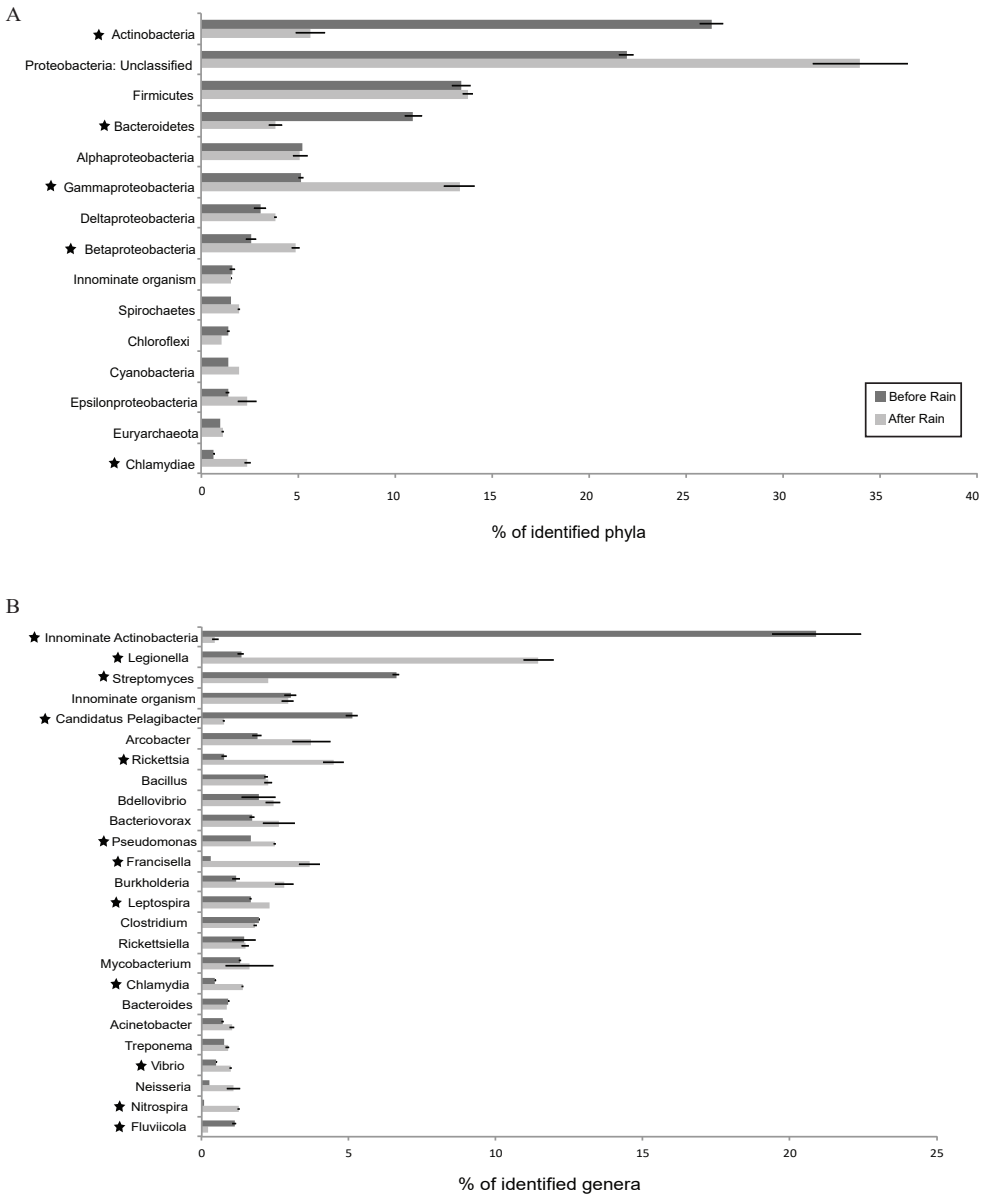


Figure 2.2. Rank abundance plots for (A) phylum (Proteobacteria subdivided into classes)- and (B) genus-level classifications of metagenomic contigs from October 2013 before- and after-rain samples. The relative abundances of different taxa are averages of biological replicates for each sample (n = 2). Based on taxon mean relative abundance across the samples, only the top 15 phyla and top 25 genera are shown. Phyla and genera highlighted with a star represent taxa with significant difference in relative abundance between the before- and after-rain microbiota ($P < 0.05$, *t* test, false-discovery rate corrected). "Innominate organism" comprises contigs classified as organisms that either belonged to no known phylum/genus or a candidate phylum/genus.

clinical isolate *L. pneumophila subsp. pneumophila strain Philadelphia 1*. The AAI between genes attributed to *L. pneumophila* in the before- and after-rain metagenomes was 83%. Although genome pairs for the same species typically exhibit higher AAIs (~90%) (28, 29), 83% still signifies close genetic relatedness and not necessarily distinct populations. Overall, these results indicate that the before- and after-rain *Legionella* isolates are members of the same species, but different from any currently known, sequenced members of *Legionella*. The discordance between our *Legionella*-like organisms and well-characterized *L. pneumophila* strains also makes it unclear if the corresponding populations are pathogenic, although a few predicted genes (1 and 3 for the before- and after-rain metagenomes, respectively) had high identity matches (>90%) to known *L. pneumophila* virulence genes in the Virulence Factor Database (http://www.mgc.ac.cn/VFs/). Organisms within *Legionella* have been associated with artificial aquatic environments, such as water distribution systems and cooling towers in buildings (30, 31), as well as WWTP effluent (20): thus their dramatic post-rain surge is not surprising.

Another notable increase in relative abundance after rain (~16-fold) was attributed to *Francisella tularensis*, an organism with known soil- and waterborne pathogenic subspecies (27, 32). Using a similar approach to that described above, AAIs between genes attributed to *F. tularensis* in before- and after-rain samples and a reference genome of pathogenic subspecies *F. tularensis subsp. tularensis SCHU S4* were 47% and 54%, respectively. Similar AAI values were observed between the metagenomic sequences and genomes of low-virulence subspecies of this organism. The AAI between the before- and after-rain *F. tularensis* genes was 68%. Thus, sequences classified as *F. tularensis* in our samples likely share the same taxonomic order *Thiotrichales*, but are different species from the known *F. tularensis* and might represent

different populations within the same genus in the before- and after-rain samples, although the low number of sequences in the before-rain data set could bias AAI calculation.

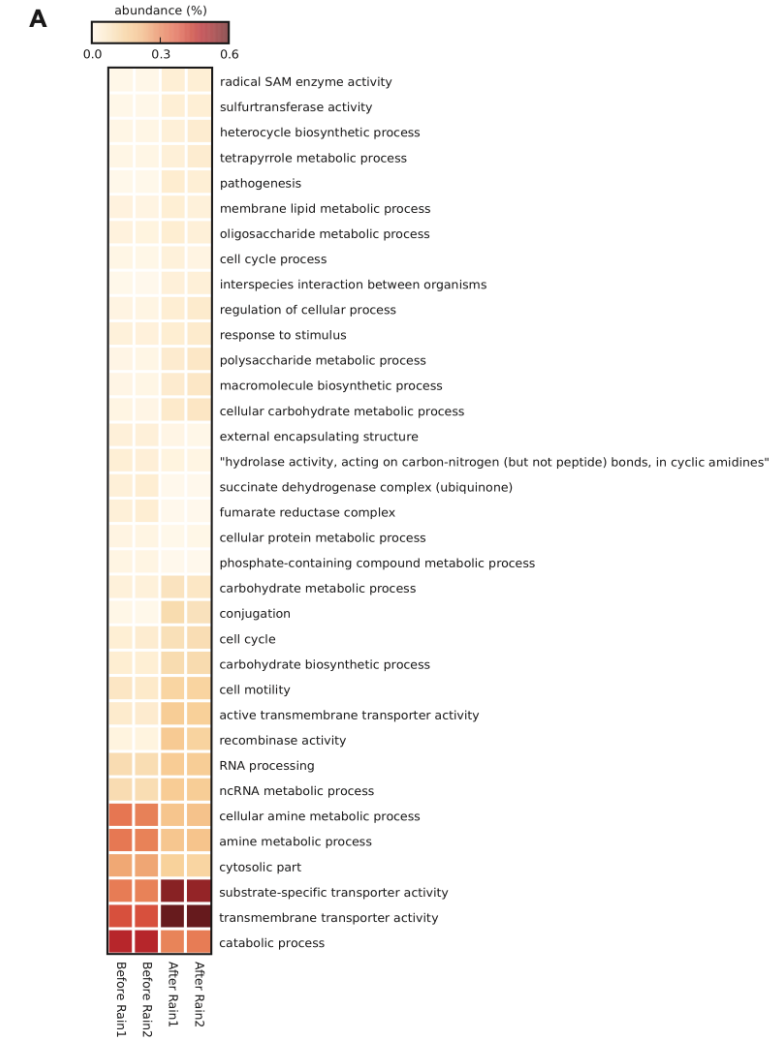
We also evaluated the population dynamics for species that dramatically dropped in relative abundance after the rain. *Actinobacterium* SCGC AAA027-L06 is a member of the ubiquitous freshwater *Actinobacteria* lineage acI-B (33), and the relative abundance of contigs affiliated with this organism decreased dramatically (43-fold) after rain. Read recruitment indicated similarity between the before- and after-rain populations, with reads from each sample sharing ~90 to 100% nucleotide identity to the largest contig of this organism, although fewer reads mapped to the contig from the after-rain samples (Supplementary Figure S2.5). As with the *L. pneumophila* population, the 84% AAI between the before- and after-rain sequences indicates close genetic relatedness between the two populations. Furthermore, the AAIs with respect to the *Actinobacterium* SCGC AAA027-L06 draft genome were similar for the sequences from the before- and after-rain microbial communities (81% and 83%, respectively), indicating close genetic relatedness to this organism. Members of the acI-B lineage have been detected in diverse freshwater habitats (19, 34–36) and tend to prefer oligotrophic environments due to their small cell size and oligotrophic life strategies (18, 37). Their decrease in relative abundance after rain likely reflects the reduced influence of freshwater flow from Lake Michigan due to increased wastewater flow.

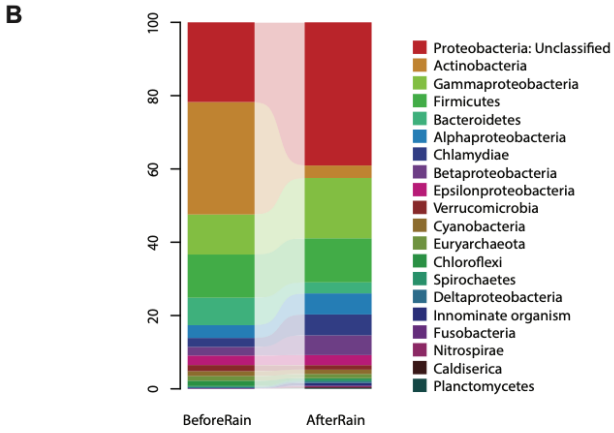
Overall functional gene content in before- and after-rain microbial communities. Functional gene profiles revealed taxon-driven shifts in the microbial community functional potential after rain. Although many abundant Gene Ontology (GO) terms related to housekeeping functions, such as nucleic acid and small molecule binding, did not significantly change in relative abundance after rain (data not shown), we observed an increase of >50% of

functions within the broad terms of transporter activity and carbohydrate metabolism after rain (Figure 2.3A). Little is known about the selective increase in transporter genes under various environmental conditions, although transporters are the primary microbial mechanism for the uptake and subsequent assimilation of nutrients and organic matter. Transporter gene expression has been shown to change in response to organic carbon inputs (38) and a phytoplankton bloom (39) in marine systems. In freshwater systems, transporters are important for cyanobacterial phosphorus acquisition (40). More recently, amino acid and amine transporter genes were among those found to be associated with various environmental conditions in *Polynucleobacter* populations in the CAWS (41). Here, we identified transporter genes that were more abundant change in relative abundance after rain (data not shown), we observed an increase of >50% of functions within the broad terms of transporter activity and carbohydrate metabolism after rain (Figure 2.3A). Little is known about the selective increase in transporter genes under various environmental conditions, although transporters are the primary microbial mechanism for the uptake and subsequent assimilation of nutrients and organic matter. Transporter gene expression has been shown to change in response to organic carbon inputs (38) and a phytoplankton bloom (39) in marine systems. In freshwater systems, transporters are important for cyanobacterial phosphorus acquisition (40). More recently, amino acid and amine transporter genes were among those found to be associated with various environmental conditions in *Polynucleobacter* populations in the CAWS (41). Here, we identified transporter genes that were more abundant following the observed rain event and were primarily related to transmembrane and substrate-specific transporter activity (Figure 2.3A).

Within the broad GO term of transporter activity, genes related to substrate-specific transmembrane transporter activity, specifically organic acid and ion transmembrane transporter

Figure 2.3. (A) Heat map showing relative abundance (percentage of total predicted genes) at level 3 of Gene Ontology (GO) terms for the before- and after-rain microbiomes. GOs that had a higher relative abundance (>50%) in one of the two groups (before versus after rain) compared to the other are shown. GOs that had less than 100 gene counts (in situ abundance) across all the samples have been excluded from the plot. Samples numbered 1 and 2 for each time point represent biological replicates. (B) Taxonomic composition at the phylum level of genes from the rain event microbial communities classified within the GO term "transmembrane transporter activity." Relative abundances are a fraction of total sequences identified at the phylum level.


A
abundance (%)
0.0
0.3
0.6
radical SAM enzyme activity
sulfurtransferase activity
heterocycle biosynthetic process
tetrapyrrole metabolic process
pathogenesis
membrane lipid metabolic process
oligosaccharide metabolic process
cell cycle process
interspecies interaction between organisms
regulation of cellular process
response to stimulus
polysaccharide metabolic process
macromolecule biosynthetic process
cellular carbohydrate metabolic process
external encapsulating structure
"hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines"
succinate dehydrogenase complex (ubiquinone)
fumarate reductase complex
cellular protein metabolic process
phosphate-containing compound metabolic process
carbohydrate metabolic process
conjugation
cell cycle
carbohydrate biosynthetic process
cell motility
active transmembrane transporter activity
recombinase activity
RNA processing
ncRNA metabolic process
cellular amine metabolic process
amine metabolic process
cytosolic part
substrate-specific transporter activity
transmembrane transporter activity
catabolic process
Before Rain1
Before Rain2
After Rain1
After Rain2


B
100
80
60
40
20
0
BeforeRain
AfterRain
Proteobacteria: Unclassified
Actinobacteria
Gammaproteobacteria
Firmicutes
Bacteroidetes
Alphaproteobacteria
Chlamydiae
Betaproteobacteria
Epsilonproteobacteria
Verrucomicrobia
Cyanobacteria
Euryarchaeota
Chloroflexi
Spirochaetes
Deltaproteobacteria
Innominate organism
Fusobacteria
Nitrospirae
Caldiserica
Planctomycetes

activity, doubled in relative abundance after rain from an average of 0.06% to an average of 0.12% (Supplementary Figure S2.6). Genes encoding all transmembrane transporters were primarily attributed to *Actinobacteria* (31% of the identified sequences at phylum level) and unclassified *Proteobacteria* (22%) before rain, whereas unclassified *Proteobacteria* (39%) and *Gammaproteobacteria* (16%) were the major groups encoding transporters after rain (Figure 2.3B). *Gammaproteobacteria* harboring transporter genes increased by 51% after rain, while *Actinobacteria* encoding these genes exhibited more than 9-fold decrease, mirroring the shifts observed for the overall taxonomic profiles for these groups (Figures 2.2 and 2.3B). Genera contributing to the increase in gammaproteobacterial sequences included *Legionella*, *Francisella*, and *Pseudomonas*, exhibiting a pattern similar to the shifts in their relative abundance in the overall microbial community. Furthermore, as with the overall microbial community, *Actinobacterium* SCGC AAA027-L06 (unclassified at genus level) contributed the largest fraction of sequences containing transmembrane transporter activity genes within *Actinobacteria* in the before-rain community. Interestingly, based on the functional gene content of organisms with dominant shifts in their relative abundance, those organisms that increased after rain had a higher proportion of their genes affiliated to transporter functions compared to those that dropped in abundance after rain. For instance, 3.7% and 6.8% of the *L. pneumophila* and *F. tularensis* genes, respectively, were associated with transmembrane transport, whereas *Actinobacterium* SCGC AAA027-L06 and the genus *Pelagibacter* had ≤2%. Thus, the increase in transporter functions following the rain appears to be directly associated with an increase in the relative proportion of a subset of the organisms that harbor these functions rather than an increase in the distribution of these genes across the community. Organisms with transmembrane

transporter genes, especially for organic substrates like organic acids, may be more suited to take advantage of the heterogeneous environment resulting from storm flow conditions.

Additional GOs showing differential abundances included genes related to photosynthesis, biosynthesis of organic compounds such as amines, vitamins, and pigments, as well as the activity of enzyme groups oxidoreductase (acting on the CH-NH_2 group of donors) and ligase (forming phosphoric ester bonds) that were twice as abundant in the before-rain microbiome (Supplementary Figure S2.6). Genes related to multiorganism processes such as pathogenesis and conjugation were >50% more abundant after rain, while the before-rain microbiome had >50% more functions related to the catabolic process, amine metabolic process, and phosphate-containing compound metabolic process (Figure 2.3A). Should the trend of increased pathogenesis and conjugation genes commonly occur with rainfall and persist in the system, it could pose a public health threat, particularly if it promotes the spread of pathogenicity genes throughout the community. Thus, this could be an important group of genes to investigate in future studies.

Further evidence that changes in community composition drove the overall changes in the metabolic capacity came from genes that decreased in relative abundance after rain, such as those encoding biosynthesis of organic substances, which mirrored the overall shifts in taxa (Figure 2.2); *Actinobacteria* (39% of the identified sequences at phylum level) and unclassified *Proteobacteria* (31%) were the major taxa encoding organic substance biosynthesis before rain and unclassified *Proteobacteria* (45%) and *Gammaproteobacteria* (13%) after rain. The short-term nature and lack of gene expression data make it difficult to know about the viability and activity of these organisms, but taxon-driven shifts in community functional potential were recently observed in another river in response to sewage and terrestrial-derived organisms (15).

Biodegradation and antibiotic resistance gene abundance before and after rain. In addition to the GO-based functional analysis, we examined how rainfall impacted biodegradation and antibiotic resistance gene content. Predicted open reading frames (ORFs) from both the before- and after-rain metagenomes were searched against a compiled database of protein sequences of microbial enzymes involved in the degradation of 12 different compounds associated with wastewater contamination, stormwater runoff, and WWTP effluent input (Figure 2.4A). We detected biodegradation genes (BDGs) in both the before- and after-rain samples for 8 out of the 12 contaminants tested, but observed a significant increase ($P < 0.05$, *t* test) in the relative abundance of genes involved in the degradation of nicotine, phenol, 1,4-dichlorobenzene, and pentachlorophenol and a decrease ($P < 0.05$) in cholesterol-degrading genes after rain (Figure 2.4A). Additionally, the total relative abundance of all BDGs was significantly higher in the after-rain sample ($P < 0.05$, *t* test). BDGs before rain were primarily affiliated with unclassified *Proteobacteria* and *Actinobacteria* (35% and 30% of the identified sequences at phylum level, respectively), with the profile shifting to unclassified *Proteobacteria* and *Betaproteobacteria* (49% and 19%, respectively) as the dominant members of the community after rain, similar to the overall taxonomic shifts described above. These results reflect the increase in effluent flow from the WWTP as well as the suspected presence of these compounds in untreated wastewater and CSOs (3, 42–47)(Figure 2.4A).

Changes in the relative abundance of antibiotic resistance genes (ARGs) after rain were evaluated using the Comprehensive Antibiotic Resistance Gene Database (CARD). As only a few ORFs (~10 per library) could be classified as ARGs from both the time points, we queried the unassembled paired-end reads against CARD. This resulted in several hits for various ARG categories at both time points (0.04% and 0.07% of the total number of reads for before- and

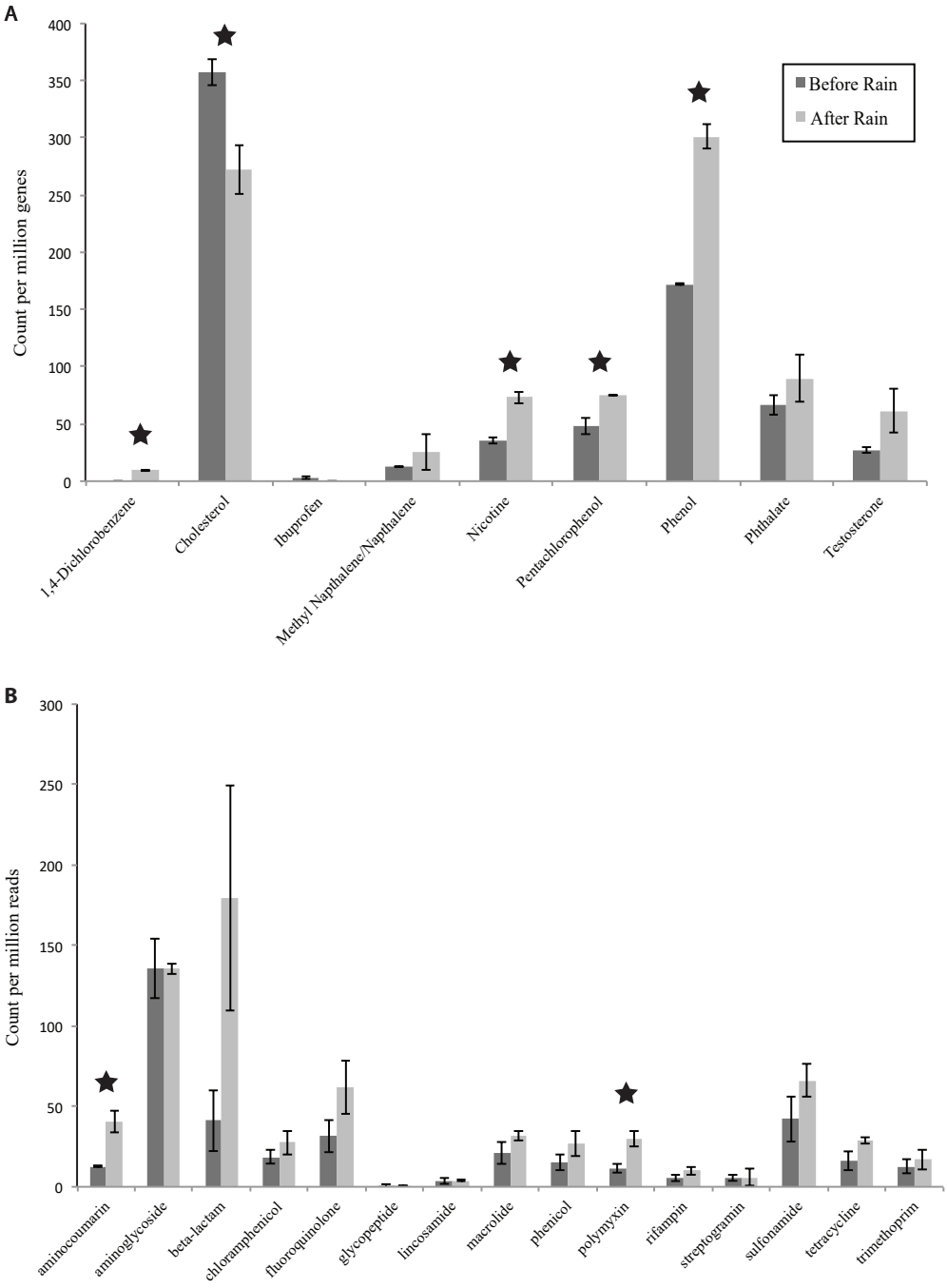


Figure 2.4. Relative abundance of (A) biodegradation genes (BDGs) and (B) antibiotic resistance genes (ARGs) in the before- and after-rain microbial communities. Relative abundance of BDGs refers to gene count (in situ abundance) per million genes per library averaged for each sample for their replicates (n = 2) (see Materials and Methods). For ARGs, relative abundance refers to read count per million reads per library averaged for each sample for their replicates. BDGs and ARGs with significant differences in relative abundances between the two time points ($P < 0.05$, *t* test) are highlighted with stars.

after-rain samples, respectively) and revealed notable increases in the relative abundance of several ARG classes after rain (Figure 2.4B), including significant increases in aminocoumarin and polymyxin resistance genes ($P < 0.05$, *t* test). As with the BDGs, the total relative abundance for all ARGs pooled for each time point was significantly higher in the after-rain sample ($P < 0.05$, *t* test). Increases in ARGs with urban-impacted storm flow were recently observed elsewhere as well (14), indicating that this could be a significant and underexplored effect of storm flow. Reads with high matches to ARGs were queried against metagenomic contigs, revealing that unclassified *Proteobacteria* and *Firmicutes* were the abundant ARG-carrying phyla (40% and 23% of the identified sequences at the phylum level, respectively) in the before-rain microbiome, whereas unclassified *Proteobacteria* (50%) and *Gammaproteobacteria* (24%) were the dominant groups after the rain. This further supports the importance of taxon-driven changes on gene content.

The results for both community composition and functional gene analysis provide evidence for the significant influence of storm flow-related input on the microbial community, particularly from increased WWTP effluent flow rates associated with heavy rain. Overall, this study revealed a shift in microbial community composition following rain from organisms frequently associated with freshwater systems toward organisms associated with urban-impacted waters (9, 19, 20), as well as a shift in functional gene content. The increased relative abundance (and possibly actual abundance) of BDGs and ARGs along with the increase in genes associated with conjugation and pathogenesis in the after rain microbiome highlight the environmental and public health implications of storm flow in urban waterways. The extent to which these changes in gene content are expressed metabolically and persist is unknown. Although the WGS metagenomic analysis of a single rainfall event limits the scope of interpretations that can be

drawn, our results provide substantial insights into microbial community dynamics in an urban stream during storm flow conditions, highlighting the need to investigate the urban stream microbiome with longer temporal scales and systematic sampling design to better predict the impact of rain-associated storm flow events.

D. Materials and Methods

Site description and sample collection. The North Shore Channel (NSC) is a 12.3-km-long man-made stream of the Chicago Area Waterway System that receives freshwater input from Lake Michigan and effluent input from the O'Brien Water Reclamation Plant, a WWTP that serves over 1.3 million people residing in a 365-km^2 area (http://www.mwrd.org/irj/portal/anonymous/waterreclamation). Our study site is approximately 1 km downstream of the WWTP outfall (Supplementary Figure S2.1). The NSC also has 48 CSOs along its course, six of which are located within about 1 km upstream of WWTP, and two of which are located within 1 km downstream of the WWTP. These release excess stormwater mixed with untreated sewage into the river when the transport and storage capacity of the city's sewage network is exceeded following high rainfall (http://www.mwrd.org/irj/portal/anonymous/overview)(Supplementary Figure S2.1). Water from the selected NSC site was sampled five times between 2013 and 2015 (0- to 1-m depth): three samplings represent stream water during base flow (dry weather) conditions, and the other two represent storm flow (<24 h after rainfall) conditions (details are in Supplementary Table S2.1). We also sampled the WWTP effluent in October 2013 during base flow conditions. Additional sample metadata and water chemistry are given in Supplementary Table S2.1.

Water was collected using a horizontal sampler (Wildco, Yulee, FL) and passed on-site in succession through ~1.6-µm-pore-size glass fiber filters to remove larger particles (Whatman, Pittsburgh, PA), and cells were collected on 0.22-µm-pore-size polycarbonate membrane filters (EMD Millipore, Billerica, MA). WWTP effluent was collected from the WWTP outlet where the released effluent mixes with stream water. About 10 liters of water was filtered in duplicate for each NSC sampled time point (for effluent, a single ~10-liter sample was obtained), and ~20 ml of the filtrate was transported back to the lab for chemical analysis. Water temperature, pH, conductivity, and total dissolved solids were measured on-site using a portable water quality meter (Hanna Instruments, Woonsocket, RI). Additional water chemistry analysis is described in Supplementary Table S2.1.

DNA extraction and sequencing. DNA was extracted from filters as described in reference 48. Briefly, filters were incubated in lysis buffer (50 mM Tris-HCl, 40 mM EDTA, 0.75 M sucrose) containing 1 mg/ml lysozyme and 200 µg/ml RNase at 37°C for 30 min. Subsequently, the samples were incubated with 1% SDS and 10 mg/ml proteinase K at 55°C and rotated overnight. From the lysate, DNA was extracted using phenol-chloroform, followed by ethanol precipitation and elution in Tris-EDTA (TE) buffer.

Whole-genome shotgun (WGS) metagenomic sequencing was done on the Illumina HiSeq (v1) with a paired-end format and a read length of 150 bp at the Michigan State University Research Technology Support Facility. We obtained 2.82 and 3.18 Gbp of paired-end read data for the before- and after-rain samples, respectively. Replicate filters were sequenced at the University of Illinois at the Chicago DNA Services Facility (DNAS) on a single lane of the Illumina HiSeq platform with paired-end format and read length of 100 bp, yielding 4.04 and 1.31 Gbp of paired-end read data for the before- and after-rain libraries, respectively.

For 16S rRNA gene amplicon sequencing, 10 to 30 ng of DNA from each biological replicate (filter) was amplified with the V1 to V3 primers 27F and 534R (49, 50). Amplicons were sequenced at the DNAS on the Illumina MiSeq platform with the paired-end format and read length of 300 bp. Between 28,933 and 160,811 sequences per sample were obtained, with an average of 61,337 sequences per sample.

16S rRNA gene-based analysis of microbial community diversity. Paired-end bar-coded reads of 16S rRNA gene amplicons were obtained for all the time points sampled and quality filtered using Trimmomatic (51), with a minimum average quality score of 20 across a 4-base sliding window and a minimum read length of 100 bp (including primer) post-trimming. Trimmed, paired-end reads were merged using Pear (52), but due to low yield of the merged reads, likely due to issues related to the MiSeq V2 kit chemistry, further analysis was only performed on the trimmed forward reads. Reads were analyzed using QIIME version 1.8.0 (53). Library statistics are summarized in Supplementary Table S2.2. Chimeric sequences were removed using *identify_chimeric_seqs.py* with the usearch61 denovo method and *filter_fasta.py*. Filtered sequences were clustered into operational taxonomic units (OTUs) at a 97% identity level using scripts *pick_otus.py* and *pick_rep_set.py* based on usearch61 denovo OTU picking. Representative OTUs were assigned taxonomy based on the Greengenes reference database (May 2013 version) using *assign_taxonomy.py* with uclust. OTUs occurring as singletons or with sequences from just one library were excluded from analyses. Determination of community taxonomic composition and alpha diversity was performed using *summarize_taxa.py* and *alpha_diversity.py*, respectively, with a random subsample of 17,384 sequences per sample to avoid bias arising from variation in sequencing depth. Good's coverage for each library was

estimated using *alpha_diversity.py* and OTUs that included singletons, subsampled to an even depth of 18,289 sequences per library, the smallest library size.

Metagenomic sequence assembly and phylogenetic classification. Raw metagenomic sequences were quality filtered using a Phred average per sliding window with a quality threshold (Q) of ≥ 20 and not allowing any N values. Quality-filtered coupled reads for each metagenomic library were assembled as described in reference 48. Coupled reads were first assembled into contigs with Velvet (54) and SOAPdenovo2 (55) separately and input to Newbler 2.0 to obtain longer contigs with better N_{50} values (56). Additional metagenomic library statistics are provided in Supplementary Table S2.4. Gene calling was done with MetaGeneMark (57). Due to uneven data yields from sequencing, we used assemblies from the first sequencing run for each sample as the representative sequences for annotations and mapped the coupled reads from both the replicate libraries to these contigs for each sample to calculate the contig coverage in each library. The predicted protein-coding genes for each data set were used for phylogenetic classification of the corresponding contigs using MyTaxa (28) with a database of all sequenced bacterial and archaeal genomes (http://enve-omics.ce.gatech.edu/data/mytaxa) using DIAMOND blastp in the sensitive mode (58). Reads were mapped to contigs using blastn with cutoffs of ≥50% alignment length, identity of ≥97%, and an E value of $\leq 10^{-10}$. Contig coverage (sum of lengths of reads mapping to contig/contig length) was used as a proxy for *in situ* abundance in each library and calculated using the *BlastTab.seqdepth_nomedian.pl* script from the Enveomics bioinformatics toolbox (59). The script *aai.rb* from the same toolbox was used to calculate average amino acid identity (AAI) between any two sets of protein-coding genes.

Analysis of functional gene content and antibiotic resistance genes. Predicted metagenomic genes were searched against the Swiss-Prot database (60) using blastp and cutoffs

of at least 40% sequence identity, 70% coverage of the query sequence, and an E value of $\leq 10^{-10}$. The Swiss-Prot match for the best hit for each query sequence was mapped to its corresponding Gene Ontology (GO) term (61), followed by binning the characterized genes at various depths (distance of a GO term from the parent node) of the GO database using the Semantics collection of scripts in the Enveomics toolbox (http://enveomics.blogspot.com/2012/11/semantics.html). To evaluate the functional profile at a specific depth, *in situ* abundance for these GO terms was calculated using gene coverage (described above), and relative abundance for each GO term was obtained as a fraction of the total abundance of genes with identified functions in that library. The taxonomic affiliation of genes classified within a specific GO term was evaluated using MyTaxa, as described above.

To specifically evaluate the presence and abundance of genes involved in biodegradation of select wastewater contaminants in the rain-associated metagenomes, we created a database of protein sequences of enzymes related to degradation of select contaminants that are commonly found in WWTP effluent and sewage: testosterone, ibuprofen, caffeine, nicotine, cholesterol, 1,4- dichlorobenzene, methylnaphthalene, pentachlorophenol, phenol, *N,N*-diethyl-3-toluamide, tetrachloroethylene, and phthalate (3, 42–47). The enzymes were selected based on their role in the degradation pathways for these compounds (62), as well as the sequence availability in NCBI. This database is available from the corresponding author upon request. The predicted ORFs were searched against this database using blastp, and the best hits were filtered at same thresholds used for Swiss-Prot (described above). Coverage estimates were used for calculation of the *in situ* abundance for each BDG class and normalized for each library by dividing the abundance of each BDG class by the total coverage of all predicted genes in that library and multiplying the result by 1 million to obtain gene count per million genes per library.

Antibiotic resistance genes in the rain-associated samples were identified by searching the predicted ORFs as well as paired-end metagenomic reads against the Comprehensive Antibiotic Resistance Gene Database (CARD) (63) using blastp and blastx and a threshold of at least 80% sequence identity and 80% coverage of the query sequence (64, 65). Filtered reads for each library were binned into broad antibiotic resistance categories using the Resistance Gene Categories index file provided on the CARD website (http://arpcard.mcmaster.ca/), and the read counts for each category were normalized for the library size as read count for ARG category per million reads per library.

Microbial abundance estimation using fluorescence microscopy. October 2013 NSC samples were fixed with paraformaldehyde (1% final concentration) in triplicate and stored in 4°C. Samples were then vortexed and collected on 25-mm black polycarbonate filters (0.2-μm-pore size) and stained with 5 μl of a 10-mg/ml DAPI (4',6-diamidino-2-phenylindole) working solution diluted in 10× phosphate-buffered saline (PBS). Microbial cells were enumerated (three slides from three replicate samples per time point) with an epifluorescence microscope (Zeiss Axio Scope.A1).

Statistical analyses. Analysis of similarity (ANOSIM) and similarity percentage (SIMPER) analysis on 16S rRNA gene and metagenomic community composition data sets, respectively, were performed using the R vegan package (66). The Statistical Analysis of Metagenomic Profiles (STAMP) software package was used for two-tailed Student's *t* tests or Welch's *t* tests to evaluate differentially abundant taxonomic groups among the 16S rRNA gene and metagenomic data sets (67) (multiple test correction, if applied, was done using Storey's false-discovery rate correction), and R was used for these tests to evaluate differentially abundant physicochemical parameters, ARGs, and BDGs. Principal-coordinate analysis (PCoA; Bray-

Curtis metric) of OTUs (with singletons removed and the table subsampled to an even depth per sample) was performed with the Phyloseq package in R (68).

Accession number(s). All of the sequence data in this study have been submitted to the Sequence Read Archive at NCBI under accession no. SRP080963.

E. Acknowledgments

This work was supported by the University of Illinois at Chicago.

We thank Markeia Scruggs and Neil Mohindra for assistance with sampling and the personnel of the University of Illinois at Chicago DNA Services Facility for facilitating sample sequencing. We also thank the anonymous reviewers whose suggestions improved the manuscript.

F. References

1. Cole JJ, Prairie YT, Caraco NF, Mcdowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ, Melack J. 2007. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. Ecosystems 10:172–185. https://doi.org/10.1007/s10021-006-9013-8.

2. Paul MJ, Meyer JL. 2001. Streams in the urban landscape. Annu Rev Ecol Syst 32:333–365. https://doi.org/10.1146/annurev.ecolsys.32.081501.114040.

3. Phillips PJ, Chalmers AT, Gray JL, Kolpin DW, Foreman WT, Wall GR. 2012. Combined sewer overflows: an environmental source of hormones and wastewater micropollutants. Environ Sci Technol 46:5336–5343. https://doi.org/10.1021/es3001294.

4. Weyrauch P, Matzinger A, Pawlowsky-Reusing E, Plume S, von Seggern D, Heinzmann B, Schroeder K, Rouault P. 2010. Contribution of combined sewer overflows to trace contaminant loads in urban streams. Water Res 44:4451–4462. https://doi.org/10.1016/j.watres.2010.06.011.

5. Walsh CJ, Roy AH, Feminella JW, Cottingham PD, Groffman PM, Morgan RP. 2005. The urban stream syndrome: current knowledge and the search for a cure. J North Am Benthol Soc 24:706–723. https://doi.org/10.1899/04-028.1.

6. Rechenburg A, Koch C, Classen T, Kistemann T. 2006. Impact of sewage treatment plants and combined sewer overflow basins on the microbiological quality of surface water. Water Sci Technol 54:95–99. https://doi.org/10.2166/wst.2006.454.

7. Sercu B, Van De Werfhorst LC, Murray J, Holden PA. 2009. Storm drains are sources of human fecal pollution during dry weather in three urban Southern California watersheds. Environ Sci Technol 43:293–298. https://doi.org/10.1021/es801505p.

8. Newton RJ, Bootsma MJ, Morrison HG, Sogin ML, McLellan SL. 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microb Ecol 65:1011–1023. https://doi.org/10.1007/s00248-013-0200-9.

9. Fisher JC, Newton RJ, Dila DK, McLellan SL. 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elementa 3:000064. https://doi.org/10.12952/journal.elementa.000064.

10. Drury B, Rosi-Marshall E, Kelly JJ. 2013. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol 79:1897–1905. https://doi.org/10.1128/AEM.03527-12.

11. Drury B, Scott J, Rosi-Marshall EJ, Kelly JJ. 2013. Triclosan exposure increases triclosan resistance and influences taxonomic composition of benthic bacterial communities. Environ Sci Technol 47:8923–8930. https://doi.org/10.1021/es401919k.

12. Czekalski N, Berthold T, Caucci S, Egli A, Bürgmann H. 2012. Increased levels of multiresistant bacteria and resistance genes after wastewater treatment and their dissemination into Lake Geneva, Switzerland. Front Microbiol 3:106. https://doi.org/10.3389/fmicb.2012.00106.

13. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, Michael I, Fatta-Kassinos D. 2013. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. Sci Total Environ 447:345–360. https://doi.org/10.1016/j.scitotenv.2013.01.032.

14. Zhang S, Pang S, Wang PF, Wang C, Han N, Liu B, Han B, Li Y, Anim-Larbi K. 2016. Antibiotic concentration and antibiotic-resistant bacteria in two shallow urban lakes after stormwater event. Environ Sci Pollut Res 23:9984–9992. https://doi.org/10.1007/s11356-016-6237-9.

15. Meziti A, Tsementzi D, Ar Kormas K, Karayanni H, Konstantinidis KT. 2016. Anthropogenic effects on bacterial diversity and function along a river-to-estuary gradient in Northwest Greece revealed by metagenomics. Environ Microbiol 18:4640–4652. https://doi.org/10.1111/1462-2920.13303.

16. Jeffries TC, Schmitz Fontes ML, Harrison DP, Van-Dongen-Vogels V, Eyre BD, Ralph PJ, Seymour JR. 2016. Bacterioplankton dynamics within a large anthropogenically impacted urban estuary. Front Microbiol 6:1438. https://doi.org/10.3389/fmicb.2015.01438.

17. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75:14–49. https://doi.org/10.1128/MMBR.00028-10.

18. Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. 2014. Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. Mol Ecol 23:6073–6090. https://doi.org/10.1111/mec.12985.

19. Newton RJ, McLellan SL. 2015. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. Front Microbiol 6:1028. https://doi.org/10.3389/fmicb.2015.01028.

20. McLellan SL, Fisher JC, Newton RJ. 2015. The microbiome of urban waters. Int Microbiol 18:141–149. https://doi.org/10.2436/20.1501.01.244.

21. Rodriguez-R LM, Konstantinidis KT. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics 30:629–635. https://doi.org/10.1093/bioinformatics/btt584.

22. Kirs M, Kisand V, Wong M, Caffaro-Filho RA, Moravcik P, Harwood VJ, Yoneyama B, Fujioka RS. 2017. Multiple lines of evidence to identify sewage as the cause of water quality impairment in an urbanized tropical watershed. Water Res 116:23–33. https://doi.org/10.1016/j.watres.2017.03.024.

23. Illinois Department of Natural Resources. 2011. Illinois Coastal Management Program issue paper: Chicago River and North Shore Channel corridors. Illinois Department of Natural Resources, Springfield, IL.

24. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9:e93827. https://doi.org/10.1371/journal.pone.0093827.

25. Clarke KR. 1993. Non-parametric multivariate analyses of changes in community structure. Aust Ecol 18:117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x.

26. Kingry LC, Petersen JM. 2014. Comparative review of Francisella tularensis and Francisella novicida. Front Cell Infect Microbiol 4:35. https://doi.org/10.3389/fcimb.2014.00035.

27. Stratton CW, Mitchell WM. 1996. The pathogenesis of Chlamydia species. Antimicrob Infect Dis Newsl 15:83–88. https://doi.org/10.1016/S1069-417X(01)80014-5.

28. Luo C, Rodriguez-R LM, Konstantinidis KT. 2014. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res 42:e73. https://doi.org/10.1093/nar/gku169.

29. Rodriguez-R LM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species. Microbe 9:111–118. https://doi.org/10.1128/microbe.9.111.1.

30. Lee HK, Shim JI, Kim HE, Yu JY, Kang YH. 2010. Distribution of Legionella species from environmental water sources of public facilities and genetic diversity of L. pneumophila serogroup 1 in South Korea. Appl Environ Microbiol 76:6547–6554. https://doi.org/10.1128/AEM.00422-10.

31. Bartram J, Chartier Y, Lee JV, Pond K, Surman-Lee S (ed). 2007. Legionella and the prevention of legionellosis. WHO, Geneva, Switzerland.

32. Petersen JM, Mead PS, Schriefer ME. 2009. Francisella tularensis: an arthropod-borne pathogen. Vet Res 40:7. https://doi.org/10.1051/vetres:2008045.

33. Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R, Grossart HP, Woyke T, Warnecke F. 2013. Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. ISME J 7:137–147. https://doi.org/10.1038/ismej.2012.86.

34. Ghai R, Rodriguez-Valera F, McMahon KD, Toyama D, Rinke R, Cristina Souza de Oliveira T, Wagner Garcia J, Pellon de Miranda F, Henrique-Silva F. 2011. Metagenomics of the water column in the pristine upper course of the Amazon River. PLoS One 6:e23785. https://doi.org/10.1371/journal.pone.0023785.

35. Warnecke F, Amann R, Pernthaler J. 2004. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environ Microbiol 6:242–253. https://doi.org/10.1111/j.1462-2920.2004.00561.x.

36. Satinsky BM, Fortunato CS, Doherty M, Smith CB, Sharma S, Ward ND, Krusche AV, Yager PL, Richey JE, Moran MA, Crump BC. 2015. Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. Microbiome 3:39. https://doi.org/10.1186/s40168-015-0099-0.

37. Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT, Mutschler J, Dwulit-Smith J, Chan LK, Martinez-Garcia M, Sczyrba A, Stepanauskas R, Grossart HP, Woyke T, Warnecke F, Malmstrom R, Bertilsson S, McMahon KD. 2014. Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. ISME J 8:2503–2516. https://doi.org/10.1038/ismej.2014.135.

38. Poretsky RS, Sun S, Mou X, Moran MA. 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ Microbiol 12:616–627. https://doi.org/10.1111/j.1462-2920.2009.02102.x.

39. Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. 2012. Bacterial community transcription patterns during a marine phytoplankton bloom. Environ Microbiol 14:228–239. https://doi.org/10.1111/j.1462-2920.2011.02602.x.

40. Pitt FD, Mazard S, Humphreys L, Scanlan DJ. 2010. Functional characterization of Synechocystis sp. strain PCC 6803 pst1 and pst2 gene clusters reveals a novel strategy for phosphate uptake in a freshwater cyanobacterium. J Bacteriol 192:3512–3523. https://doi.org/10.1128/JB.00258-10.

41. Sangwan N, Zarraonaindia I, Hampton-Marcell JT, Ssegane H, Eshoo TW, Rijal G, Negri MC, Gilbert JA. 2016. Differential functional constraints cause strain-level endemism in Polynucleobacter populations. mSystems 1:e00003-16. https://doi.org/10.1128/mSystems.00003-16.

42. Boyd GR, Palmeri JM, Zhang S, Grimm DA. 2004. Pharmaceuticals and personal care products (PPCPs) and endocrine disrupting chemicals (EDCs) in stormwater canals and Bayou St. John in New Orleans, Louisiana, USA. Sci Total Environ 333:137–148. https://doi.org/10.1016/j.scitotenv.2004.03.018.

43. Glassmeyer ST, Furlong ET, Kolpin DW, Cahill JD, Zaugg SD, Werner SL, Meyer MT, Kryak DD. 2005. Transport of chemical and microbial compounds from known

wastewater discharges: potential for use as indicators of human fecal contamination. Environ Sci Technol 39:5157–5169. https://doi.org/10.1021/es048120k.

44. Benotti MJ, Brownawell BJ. 2007. Distributions of pharmaceuticals in an urban estuary during both dry- and wet-weather conditions. Environ Sci Technol 41:5795–5802. https://doi.org/10.1021/es0629965.

45. Phillips P, Chalmers A. 2009. Wastewater effluent, combined sewer overflows, and other sources of organic compounds to Lake Champlain. J Am Water Resour Assoc 45:45–57. https://doi.org/10.1111/j.1752-1688.2008.00288.x.

46. Sauvé S, Aboulfadl K, Dorner S, Payment P, Deschamps G, Prévost M. 2012. Fecal coliforms, caffeine and carbamazepine in stormwater collection systems in a large urban area. Chemosphere 86:118–123. https://doi.org/10.1016/j.chemosphere.2011.09.033.

47. Fang H, Cai L, Yu Y, Zhang T. 2013. Metagenomic analysis reveals the prevalence of biodegradation genes for organic pollutants in activated sludge. Bioresour Technol 129:209–218. https://doi.org/10.1016/j.biortech.2012.11.054.

48. Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R, Konstantinidis KT. 2011. Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. Appl Environ Microbiol 77:6000–6011. https://doi.org/10.1128/AEM.00107-11

49. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol 74:2461–2470. https://doi.org/10.1128/AEM.02272-07.

50. Somenahally AC, Mosher JJ, Yuan T, Podar M, Phelps TJ, Brown SD, Yang ZK, Hazen TC, Arkin AP, Palumbo AV, Van Nostrand JD, Zhou J, Elias DA. 2013. Hexavalent chromium reduction under fermentative conditions with lactate stimulated native microbial communities. PLoS One 8:e83909. https://doi.org/10.1371/journal.pone.0083909.

51. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

52. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620. https://doi.org/10.1093/bioinformatics/btt593.

53. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR,

Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.

54. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.

55. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory- efficient short-read de novo assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18.

56. Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT. 2012. Individual genome assembly from complex community short-read metagenomic datasets. ISME J 6:898-901. https://doi.org/10.1038/ismej.2011.147.

57. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38:e132. https://doi.org/10.1093/nar/gkq275.

58. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

59. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J Prepr 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.

60. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187–D191. https://doi.org/10.1093/nar/gkj161.

61. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29. https://doi.org/10.1038/75556.

62. Gao J, Ellis LBM, Wackett LP. 2010. The University of Minnesota Biocatalysis/Biodegradation Database: improving public access. Nucleic Acids Res 38:D488–D491. https://doi.org/10.1093/nar/gkp771.

63. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD,

Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/AAC.00419-13.

64. Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 9:207–216. https://doi.org/10.1038/ismej.2014.106.

65. Port JA, Cullen AC, Wallace JC, Smith MN, Faustman EM. 2014. Metagenomic frameworks for monitoring antibiotic resistance in aquatic environments. Environ Health Perspect 122:222–228. https://doi.org/10.1289/ehp.1307009.

66. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2015. vegan: Community Ecology package. R package version 2.2-1. R Foundation for Statistical Computing, Vienna, Austria. http://CRANR-project.org/package=vegan.

67. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30: 3123–3124. https://doi.org/10.1093/bioinformatics/btu494.

68. McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.

III. Bacterioplankton Response to Allochthonous Dissolved Organic Matter Across a Coastal to Offshore Transect in Lake Michigan[2]

A. Abstract

Heterotrophic bacterioplankton play an important role in the lake food web and the global carbon cycle through assimilation of constituents of the dissolved organic matter (DOM) pool. Lake Michigan is one of the largest lakes in the world, and second largest of the Great Lakes by volume. Over the last two decades, the lake has witnessed significant ecological changes due to proliferation of the invasive quagga mussels into deeper regions of the lake. The impact of these changes on the labile dissolved organic matter pool available for bacterial consumption, and the relative importance of terrestrially derived DOM (t-DOM) for bacterial metabolism across Lake Michigan in the post-mussel period is poorly understood. Here, we investigated Lake Michigan bacterial community structure and activity across a coastal-to-offshore transect beginning near the mouth of Kalamazoo River, one of the largest tributaries to southern Lake Michigan. In addition, we evaluated short-term bacterioplankton response to a pulse of t-DOM (leaf litter leachate) in shipboard mesocosms set up using nearshore and offshore lake-water. The bacterial community composition and activity for the natural and t-DOM enriched samples was characterized using combined metagenomics and metatranscriptomics. Despite observing differences in the active community composition and DOM related transporter gene transcripts across the transect, the nearshore and offshore bacterial communities showed a similar responses to t-DOM, primarily in the form of increased transcriptional activity for aromatic compound

[2] Chaudhary A, Turner S, Poretsky R. Submitted to *Limnology and Oceanography*.

metabolism. The use of metagenome assembled genomes identified populations within the Bacteroidetes phylum that play an important role in t-DOM response.

B. Introduction

Heterotrophic bacterioplankton play a critical role in biogeochemical cycling and food web dynamics in freshwater lake ecosystems by assimilating components of the dissolved organic matter (DOM) pool. Bacterioplankton secondary production and respiration are responsible for most of the carbon flux through these systems (1) and contribute significantly to regional and global carbon budgets (2, 3). Among the freshwater lakes in the world, the Laurentian Great Lakes comprise the largest group, containing about 21% of the world's surface freshwater by volume; however, there have been surprisingly few research efforts to investigate microbial food web and bacterioplankton dynamics in this system. Recent efforts to characterize bacterioplankton in the Great Lakes and especially in oligotrophic Lake Michigan have provided valuable information about their spatiotemporal dynamics but have largely been restricted to marker gene surveys (4–6).

Primary production (PP) and bacterial secondary production (BP) are generally tightly coupled in nutrient-limited lakes like Lake Michigan. Until recently, phytoplankton-derived DOM was estimated to support roughly 90% of the annual BP in Lake Michigan euphotic zone, with the rest supported by terrigenous carbon (7). However, the annual spring diatom bloom in offshore waters of the lake has been decimated in the last 10-15 years due to filtering effects from invasive dreissenid mussels (8, 9). This has a direct impact on the annual DOM pool available for bacteria. Changes in PP in the spring season can impact not only spring BP, but also the subsequent summer; BP typically exceeds PP in the summer and is hypothesized to use the

accumulated DOM from the previous spring, as spring PP is typically higher than BP (7). In contrast, nearshore waters (typically < 30m water-column depth) receive terrestrially-derived DOM (t-DOM) and nutrients from sources such as rivers (10). The higher nutrient levels especially during spring runoff events trigger phytoplankton blooms in the nearshore (11), and the DOM resulting from such blooms together with t-DOM likely leads to overall higher levels of DOM available for bacterial assimilation and production in the nearshore. While recent studies have provided important information about bacterial community diversity, secondary production, and respiration rates across this coastal to offshore gradient in southern Lake Michigan (5, 6, 12, 13), our understanding of functional activity of specific bacterial groups across this gradient remains limited.

Despite the recent decline in phytoplankton productivity and the associated oligotrophication of offshore Lake Michigan waters, bacterial respiration rates and cell abundance have remained relatively stable in comparison to pre-dreissenid mussel periods (11, 12), resulting in offshore Lake Michigan becoming a net source of carbon as compared to a net sink prior to dreissenid mussel proliferation (12). By contrast, nearshore waters in southern half of the lake continue to remain a net sink of carbon. The relative potential of nearshore and offshore Lake Michigan bacterial communities to use t-DOM, and consequently the extent to which bacterial metabolism in the post-mussel period is supported by t-DOM in different regions of the lake, remains unclear (12).

To provide insights into the impact of potentially differing water chemistries and carbon availability in coastal and offshore Lake Michigan on bacterial community structure and DOM metabolism, we investigated the bacterial community composition (both for whole and transcriptionally active fractions) as well as community transcriptional activity across a

nearshore-to-offshore transect beginning near the mouth of Kalamazoo River, one of the largest tributaries to southern Lake Michigan. Additionally, to evaluate the potential of Lake Michigan bacterial communities from nearshore and offshore regions to metabolize t-DOM and identify the DOM compounds being assimilated, we conducted mesocosm experiments to test the short-term bacterial community transcriptional response to a t-DOM pulse. Using a combination of metagenomics and metatranscriptomics, we identified a higher relative abundance for transcripts affiliated with Cyanobacteria in the offshore waters as compared to nearshore, as well as community-wide differences across the transect in the taxonomic composition of gene transcripts associated with DOM transporter activity. However, despite these differences, the nearshore and offshore bacterial communities showed a similar capacity at the transcriptional level to metabolize terrestrially derived DOM. The use of metagenome assembled genomes (MAGs) provided further evidence of a population-specific response to t-DOM across the transect, with populations within the Bacteroidetes phylum playing an important role.

C. Materials and Methods

Sample collection and experimental design. Water samples were collected in September 2015 from Lake Michigan onboard the R/V Lake Guardian across a coastal-to-offshore transect beginning near the mouth of Kalamazoo River, a major tributary to southern Lake Michigan (7). Water was collected in 10-20 L polycarbonate carboys from near-surface (2 m depth) in a nearshore location (total depth – 18 m) along the transect, and from the near-surface and hypolimnion (60 m depth) from an offshore location (total depth – 110 m). Approximately 60-70 L of water was collected from each location/depth. Collected water was stored at 4°C in the dark for 12-18 hours prior to the setup of mesocosm incubation experiments.

Nutrient concentrations and other environmental parameters (Table 3.1) were measured by US EPA personnel according to standard EPA methods.

Water from the nearshore and offshore sites was used in shipboard mesocosm experiments in acid-washed polycarbonate cubitainers (10 L capacity), where each mesocosm contained 2.7 L of lake water. Three mesocosms for each site were enriched initially with 120 μM t-DOM and three mesocosms were left unamended (control). t-DOM was prepared from Eastern Cottonwood (*Populus deltoides)* leaf litter collected from the Marian Byrnes Park, Chicago, IL. To prepare the leachate, dry leaf litter was incubated in 10 L sterile deionized water in the dark for 7 days followed by removal of cell debris using combusted GF/F filters (14). The DOC concentration in the leachate was measured using high-temperature catalytic oxidation at Gray Research Group, Northwestern University. All mesocosms were incubated in the dark for 19h, and subsampled for ~1 L water at 2 h and 19 h. Subsampled water was filtered immediately through 1.6 μm pore-size glass fiber filters (TISCH Scientific, North Bend, OH) to remove larger particles and organisms, and free-living cells were collected on 0.2 μm pore-size polycarbonate membrane filters (EMD Millipore, Billerica, MA). Filters were stored immediately in liquid nitrogen and transported back to lab for storage at -80 °C until DNA/RNA extraction.

DNA, RNA isolation and next generation sequencing. For microbial DNA extraction from the 0.2 μm filters, each frozen filter was first fragmented into small pieces and roughly 1/3rd of the fragments were randomly picked for use in an organic extraction method as described previously (15). Briefly, filter fragments were incubated in lysis buffer (50 mM Tris-HCl, 40 mM EDTA, 0.75 M sucrose) containing 1.15 mg/ml lysozyme and 200 μg/ml RNase at 37 °C for 30 min, followed by incubation with 1% SDS and 10 mg/ml proteinase K at 55 °C for 2 h while rotating. DNA was extracted from the lysate using phenol:chloroform, and isolated using ethanol

precipitation followed by elution in Tris-EDTA (TE) buffer. Genomic DNA (gDNA) for the filters corresponding to the triplicate control mesocosms at the 2 h time-point were pooled equally and used for whole-genome shotgun (WGS) sequencing at the University of Illinois at Chicago Sequencing Core. Sequencing was performed on an Illumina NextSeq500 with paired-end format and read-length of 150 bp, yielding 62, 47, and 52 million reads for metagenome libraries for control mesocosms using lake-water from nearshore surface, offshore surface and offshore hypolimnion, respectively.

RNA was isolated from the remaining fragments for each filter using the same organic extraction method except that acid phenol:chloroform (pH 4.5) was used for the extraction step and RNase was not used in the lysis buffer. Isolated RNA was treated with DNase using the TURBO DNA-free DNase kit (Invitrogen, Carlsbad, CA) to digest residual genomic DNA. Purified RNA from the triplicate mesocosms for each treatment/time-point was pooled in equal amounts and concentrated using ethanol precipitation. Each pooled, concentrated RNA sample was then assessed for RNA concentration and integrity using the Qubit RNA quantitation kit and the Agilent 2200 TapeStation, respectively. The RNA integrity (RIN) number obtained from the TapeStation results for all the samples ranged from 4-6. Between 100-200 ng of total RNA from each sample was then used for ribosomal RNA (rRNA) depletion using the RiboZero kit (Illumina, San Diego, CA). The rRNA-depleted RNA was subsequently used for cDNA synthesis and library preparation using the SMARTer Stranded RNA-Seq kit (Takara Bio USA, Mountain View, CA), providing cDNA libraries with an average fragment length of ~300 bp. All the cDNA libraries were then sequenced on a NextSeq500 with paired-end format and read-length of 150 bp, yielding between 13-32 million reads per library. All of the sequence data in

this study have been submitted to the Sequence Read Archive at NCBI under accession number PRJNA693412.

Metagenome assembly, annotation and read recruitment. The three metagenome libraries were first quality filtered using a Phred average per sliding window with a quality threshold (Q) of ≥ 20 and not allowing any N values. The filtered, short-read libraries were then individually assembled to obtain contigs using MEGAHIT (16) with default settings. Assembly yielded 243,721, 133,679 and 226,984 contigs longer than 500 bp for the nearshore-surface, offshore-surface and offshore-hypolimnion metagenomes with an N50 value of 1419 bp, 1623 bp and 1612 bp, respectively. Contigs from all the three metagenomes were mined for protein-coding genes using MetaGeneMark (17). The predicted protein-coding genes and contigs were phylogenetically classified using MyTaxa (18), using its database of bacterial and archaeal genomes (http://enve-omics.ce.gatech.edu/data/mytaxa) and DIAMOND blastp in the sensitive mode (19). For functional annotation of the metagenome-derived genes, gene sequences were searched against the Swiss-Prot database (20) using blastp with following cutoffs: 30% sequence identity, 70% coverage of query sequence, and an E value of $\leq 10^{-10}$. Subsequently, the Swiss-Prot match for the best hit for each query sequence was mapped to its corresponding term in the SEED (21) and Gene Ontology (GO) (22) databases. The SEED-based annotations were used for overall functional comparison between the metagenomes/metatranscriptomes (Figures 3.2 and 3.3), whereas the GO-based annotations were specifically used for evaluating the DOM-associated transporter activity (Figure 3.4). To calculate contig or gene abundance in the metagenomes, short-reads for each metagenome were mapped to the corresponding contigs or genes using blastn with cutoffs: ≥ 75 bp sequence alignment length, ≥ 95% sequence identity, and an E value cutoff of $\leq 10^{-10}$.

Metagenome assembled genomes (MAGs) reconstruction. To reconstruct population genomes from the metagenomic datasets, we performed a combined assembly of pooled short-reads from the three metagenomes using MEGAHIT with default settings. Subsequently, contigs larger than 1000 bp from the combined assembly were used for genome binning using MaxBin 2.0 (23) with default settings. Contig bins obtained were assessed for quality (degree of genome completeness and contamination) using CheckM (24). We obtained 30 bins or MAGs with a genome completeness of at least 50% and with less than 10% contamination. For each of these MAGs, a reassembly was performed using the metagenome reads that aligned to a MAG's contigs at ≥ 98% nucleotide identity and ≥ 100 bp length (25). Aligned reads for each MAG were reassembled using metaSPAdes (26). From the reassembly, only contigs longer than 1000 bp were retained in the final MAG (previously binned contigs were discarded), and the genome quality for the reassembled MAGs was assessed using CheckM. The reassembly process reduced the genome contamination levels for most of the final MAGs, and we focused further analysis on 17 MAGs that had more than 50% completeness and less than 10% contamination based on CheckM (Supplementary Table S3.1). These MAGs were taxonomically annotated and mined for protein coding genes using the MiGA webserver (27). Subsequently, the predicted genes in each MAG were annotated using the SEED database as described above for the metagenomes.

Metatranscriptome processing and analysis. The metatranscriptome libraries were first quality filtered similarly to the metagenomes, followed by removal of adapter sequences and first three bases of the forward read (based on recommendations in cDNA library preparation kit) using Trimmomatic (28). This was followed by removal of ribosomal RNA encoding cDNA reads from the metatranscriptomes using SortMeRNA (29). The quality filtered, non-ribosomal cDNA reads were used for whole-community gene expression analysis for the different

mesocosm treatments/time-points by mapping them to genes from the corresponding metagenome. For example, all metatranscriptomes derived from mesocosms that had water from nearshore Lake Michigan were mapped to genes (functionally annotated with SEED/GO) from the nearshore surface metagenome. cDNA reads were mapped to the genes with the same similarity threshold as the metagenomic reads, as described above. Similarly, to evaluate the gene expression profile for genes corresponding to a MAG population across the mesocosm treatments and time-points, the metatranscriptome libraries were mapped to the genes for that MAG with the same similarity threshold as for the metagenomic reads.

Statistical analyses and data visualization. To account for differences in the water column depth and effects of thermal stratification across the nearshore-to-offshore transect, we focused our analysis to comparison of the near-surface bacterial communities between the two regions. The number of reads from a metagenomic (gDNA) or metatranscriptomic (cDNA) library mapping to a specific gene either from an assembled metagenome or a MAG was calculated using the *BlastTab.seqdepth_nomedian.pl* script from the Enveomics bioinformatics toolbox (30). The counts for gene abundance/expression were used for comparison of microbial community/individual MAG-based functional profile (and taxonomy profile in the case of whole community) between different metagenomes/metatranscriptomes using DESeq 2.0 package in R (31). The implementation first involved a normalization of the raw gDNA/cDNA read counts for each gene/functional category/taxon in a library to account variation in sequencing depth using the default method in DESeq 2.0, followed by differential abundance testing between any metagenomes/metatranscriptomes of interest with the Wald test. The package was also used to generate ordination plots and clustered heatmaps of the functional/taxonomic profiles for the metagenomes/metatranscriptomes, with the raw count data first normalized using the ‘rlog’

(regularized-logarithm transformation) function in DESeq 2.0 for these plots so that the data were approximately homoskedastic (Supplementary Figures S3.1, S3.3 and Figure 3.2). For Figure 3.1B, Figure 3.4 and Supplementary Figure S3.4, the gene abundance/expression counts were normalized based on the RPKM formula (Reads Per Kilobase of gene or transcript, per Million mapped reads) before generating the plots. Visualization of MAG-specific differential gene expression (genes annotated and binned into functional categories using SEED Subsystems) between the t-DOM and control mesocosms was performed using the R package OmicCircos (32).

D. Results

Lake Michigan water chemistry across the nearshore-offshore transect. Lake Michigan is dimictic and experiences thermal stratification in the summer season, and stratification was still evident from the temperature levels across the water-column during our sampling in the late summer/early fall season (Table 3.1). Across the nearshore-to-offshore transect (Figure 3.1A), nearshore surface waters had higher ammonia (22.7 μg N/L), nitrate (311.7 μg N/L) and chlorophyll-a levels (2.2 μg/L) as compared to offshore surface waters (9.6 μg N/L, 242.3 μg N/L and 1.5 μg/L, respectively) (Table 3.1). Conversely, offshore surface had higher soluble reactive phosphate levels (13.9 μg P/L) as compared to nearshore surface (4.2 μg P/L).

Comparison of metagenome and metatranscriptome-based community composition of nearshore and offshore Lake Michigan. We compared the taxonomic composition of the whole as well as the active fractions of the bacterial community from metagenome (MG) and metatranscriptome (MTs) datasets from nearshore and offshore Lake Michigan surface waters

Table 3.1. Water chemistry and environmental characteristics for sampled Lake Michigan sites

Site	Nearshore	Offshore	Offshore
GPS coordinates	42°41.4488 N 086°15.2129 W	42°42.6924 N 086°42.3408 W	42°42.6924 N 086°42.3408 W
Water column depth (m)	18	106	106
Water sample depth	Near surface (~2 m)	Near surface (~2 m)	Hypolimnion (60.2 m)
Sampling Date	9/13/15	9/13/15	9/13/15
Nutrient sampling station	Mid Epilimnion	Mid Epilimnion	Lower Hypolimnion
NH_4 µg N/L	22.7	9.6	20.3
[a]SRP µg P/L	4.2	13.9	4.3
NO_x µg N/L	311.7	242.3	303.2
TP µg P/L	2.3	6.3	4.6
TN µg N/L	463.3	366.6	426.9
[b]chl-a µg /L	2.2	1.5	0.5

[a]SRP: Soluble Reactive Phosphate

[b]chl-a: Chlorophyll-a

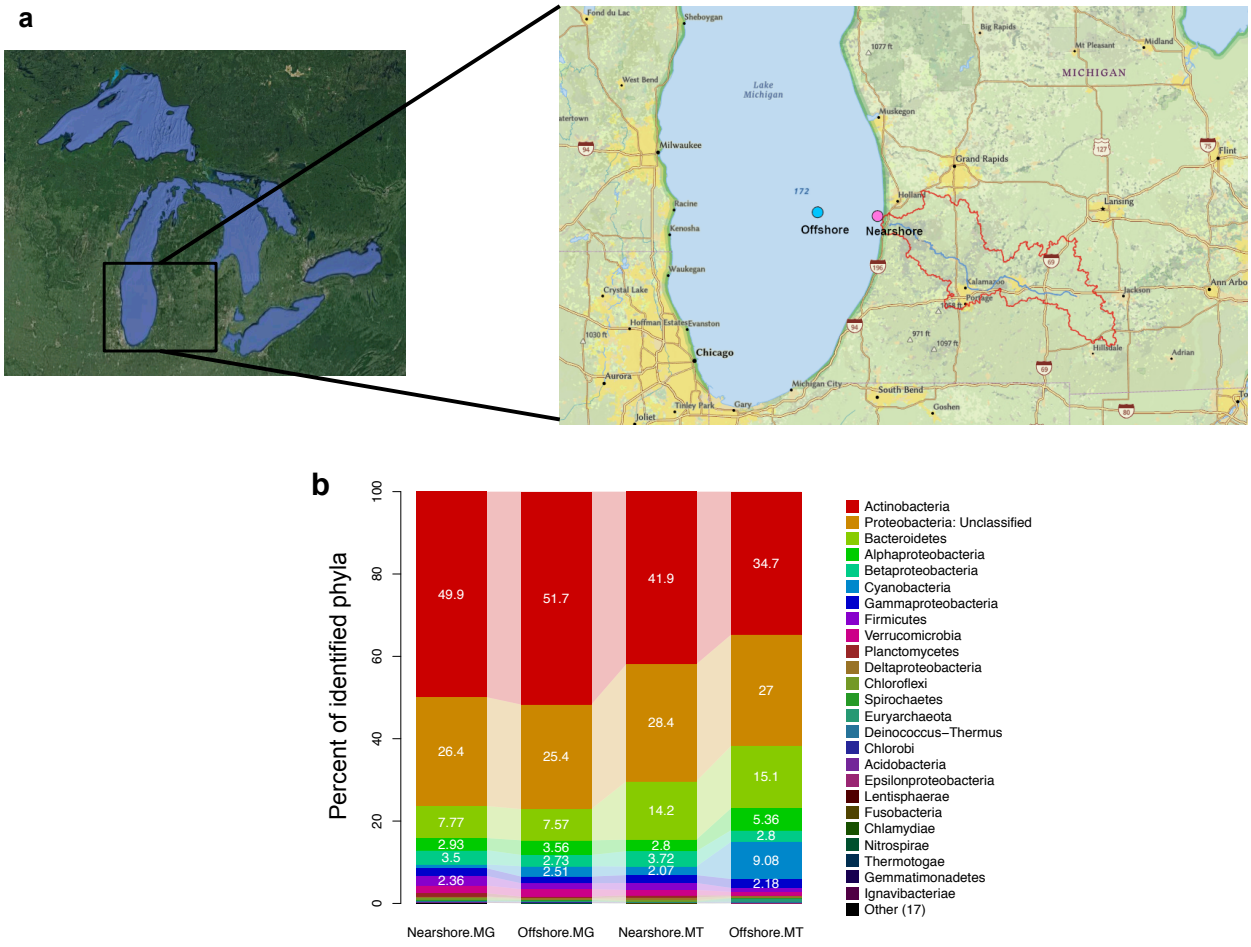


Figure 3.1. (a) Map showing the Great Lakes region (left panel) and southern Lake Michigan with the surrounding landscape (right panel). Nearshore and offshore sampling sites are highlighted in the right panel with pink and blue dots, respectively. The right panel also highlights the Kalamazoo River (in blue) and its watershed boundary (in red) **(b)** Taxonomic composition of the active (mRNA-based) and total (DNA-based) bacterial communities in the nearshore and offshore surface-waters of southern Lake Michigan. The taxonomic profile is shown at the phylum level, with Proteobacteria subdivided into classes. Relative abundance represents % of total phyla, and the "other" category represents organisms with low abundances.

(Figure 3.1B, Supplementary Figure S3.1). Clustering of taxonomic profiles at the phylum level revealed similarity between the nearshore MG, offshore MG and nearshore MT, whereas the offshore MT was relatively distinct from these samples (Supplementary Figure S3.1). Scatter

plots for phylum level comparison of taxonomic profile between the nearshore and offshore MGs/MTs showed a high similarity between the MGs ($R^2 = 0.99$) and the MTs ($R^2 = 0.96$) of the two regions. The only phylum differentially abundant between the nearshore and offshore MGs was 'Innominate organism,' which was present in low abundance in both sites (< 0.1% of the total phyla in the metagenome) but was ~6 times more abundant in the nearshore than offshore (DESeq2, $P < 0.05$) (Supplementary Figure S3.1). As the percentage of sequences from the MG and MT datasets annotated at more resolved taxonomic levels was low (~35%), we used 16S rRNA gene amplicon sequences for comparison of nearshore and offshore bacterial communities at genus level. Although the 16S rRNA gene amplicons showed a similar community profile between the nearshore and offshore at the phylum level ($R^2 = 0.86$), the regions were much less similar at the genus level ($R^2 = 0.64$). Several genera were differentially abundant between nearshore and offshore (t-test, $P < 0.01$): alfV-A and unclassified *Cryomorphaceae* were more abundant offshore, whereas unclassified bacV, unclassified *Comamonadaceae*, acIV-A, betIV-A and acSTL-A were more abundant nearshore (Supplementary Figure S3.2). Despite a general similarity in the community composition between the nearshore and offshore MGs and MTs, there were more phyla differentially abundant in the MTs between the nearshore and offshore bacterial communities as compared to the MGs, indicating differential activity of specific phyla in different regions of the lake. Among these differences, rare organisms (in both the MGs and MTs) Dictyoglomi and Gemmatimonadetes were attributed to more transcripts in the offshore community (DESeq2, $P < 0.05$) (Supplementary Figure S3.1). The abundant phylum Cyanobacteria had ~4-fold more transcripts in the offshore community, although the difference was not significant (DESeq2, $P = 0.06$) (Figure 3.1B). Nevertheless, the Cyanobacterial transcripts accounted for a large part of the difference between the nearshore and offshore active

taxa, and *Synechococcus* was the predominant genus contributing to Cyanobacterial transcripts both nearshore and offshore. Compared to their abundance in the total community (MG), Crenarchaeota and Thaumarchaeota were responsible for significantly fewer expressed genes (MT) nearshore (DESeq2, $P<0.05$) (Supplementary Figure S3.1). No phyla were significantly different in their expression versus abundance patterns in the offshore bacterial community.

Metabolic response of nearshore and offshore Lake Michigan bacterial communities to terrestrial dissolved organic matter pulse. In order to assess how different regions of the lake respond to pulses of terrestrial organic matter, we conducted mesocosm experiments with nearshore and offshore surface water. Given that these communities had similar taxonomic composition but differences in transcriptional active organisms, we expected to see a strong transcriptional response to t-DOM in our mesocosms. In general, the MTs for the nearshore mesocosms differed from the offshore mesocosm MTs (Figure 3.2A). The MTs were annotated based on SEED Subsystems and the major functional processes likely contributing to the difference between control nearshore and offshore MTs included photosynthesis (more transcripts in offshore MTs), membrane transport and regulation and cell signaling (more transcripts in nearshore MTs) (Supplementary Figure S3.3). In addition, the transcriptional activity of Lake Michigan nearshore and offshore bacterial communities changed with exposure to t-DOM in mesocosm-based incubations. The MTs were distinct depending on whether or not the sample received t-DOM (Figure 3.2A, Supplementary Figure S3.3). While the control MTs at 2h and 19h clustered together in both nearshore and offshore bacterial communities, indicating that the transcriptional profile changed little over time, the t-DOM addition shifted the transcriptional profiles in the treatment mesocosm over the 19h incubation (Figure 3.2A). Broad functional categories that exhibited variation between the control and t-DOM MTs included

phosphorus metabolism, regulation and cell signaling, motility and chemotaxis (fewer transcripts in t-DOM MTs) as well as stress response, and iron acquisition and metabolism (more transcripts in t-DOM MTs) (Supplementary Figure S3.3). Analyzing the functional profile at a more resolved level of SEED Subsystems provided information for the specific functional processes

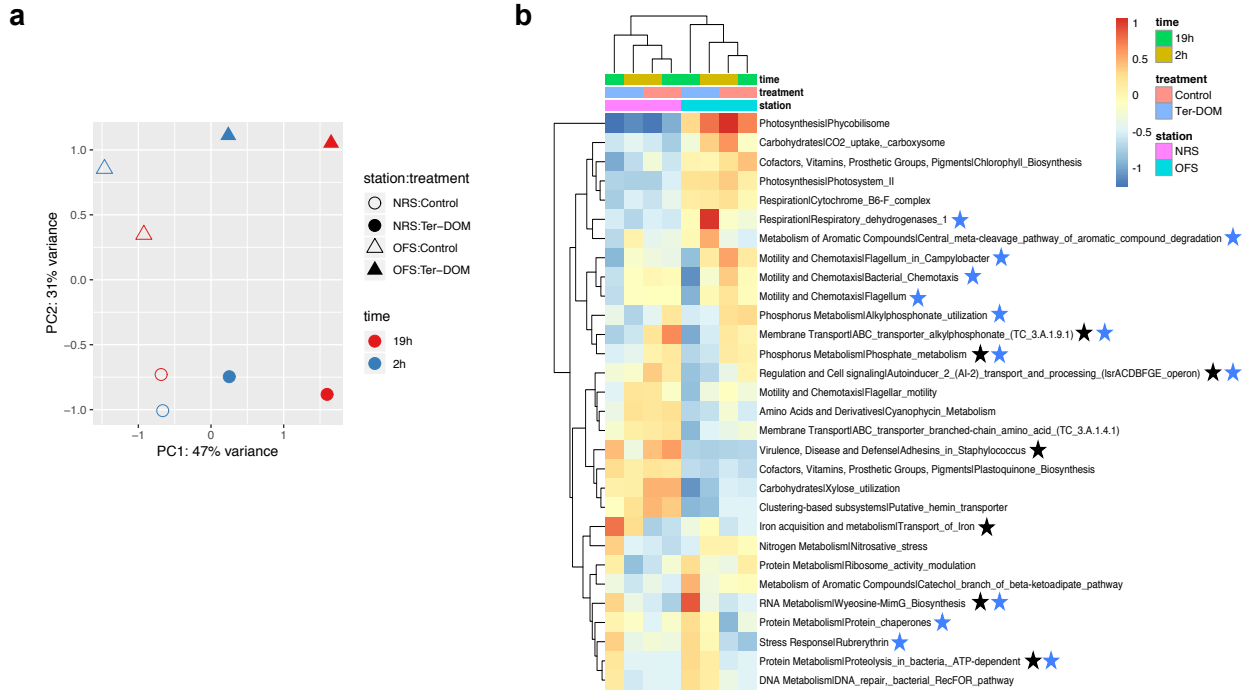


Figure 3.2. (a) Principal Components Analysis (PCA) plot of metatranscriptomes (functionally annotated with SEED database) for the different mesocosms and time-points. **(b)** Clustered heatmap of metatranscriptome-based active functional profile for the bacterial communities in the different mesocosms and time-points. Functional processes were annotated at level 3 of the SEED Subsystems database. Nearshore and offshore sampling sites across the Lake Michigan transect are labeled as NRS and OFS, respectively. Functional processes that are significantly different based on DESEq2 (Wald test, $p < 0.05$) between the control and treatment mesocosms at either time-point are highlighted with a black star for nearshore lake-water mesocosms, and with a blue star for offshore lake-water mesocosms.

contributing to the overall differences in t-DOM and control mesocosm MTs (Figure 3.2B). Notably, the SEED Subsystems category important for complex organic matter metabolism, 'central meta-cleavage pathway of aromatic compound degradation', had more transcripts in the t-DOM mesocosms compared to the control for the offshore (DESeq2, $P < 0.05$) and nearshore (DESeq2, $P = 0.07$) bacterial communities (Figure 3.2B). Transcripts related to RNA and protein metabolism – Wyeosine-MimG biosynthesis and ATP-dependent proteolysis in bacteria, respectively, were also significantly more abundant in the t-DOM mesocosms (DESeq2, $P < 0.05$) (Figure 3.2B). Conversely, transcripts related to phosphorus metabolism were significantly less abundant in the t-DOM mesocosms as compared to the control mesocosms. These included transcripts encoding an ABC transporter for alkylphosphonate; alkylphosphonate utilization; and phosphate metabolism (DESeq2, $P < 0.05$) (Figure 3.2B).

While these differences between the control and t-DOM mesocosm MTs were generally similar for the nearshore and offshore bacterial communities, there were some exceptions to this trend. Notably, iron acquisition and metabolism had more transcripts for t-DOM mesocosm MTs for only nearshore bacterial community (Supplementary Figure S3.3), whereas transcripts for alkylphosphonate utilization as well as for encoding rubrerythrin, a protein involved in oxidative stress response, were significantly more abundant only in the offshore t-DOM mesocosms (Figure 3.2B). Many functional processes had different temporal patterns in their transcriptional variation between the control and treatment mesocosms for the nearshore and offshore bacterial communities. For instance, differences in the transcript levels between the respective control and t-DOM mesocosm MTs for central meta-cleavage pathway of aromatic compound degradation and alkylphosphonate utilization were significant at only 2h, whereas transcripts encoding ABC transporter for alkylphosphonate were significantly different between nearshore control and t-

DOM MTs at only 19h. Similarly, transcripts encoding stress response and related processes were significantly abundant in the t-DOM MTs as compared to control MTs mostly at 19h for the nearshore and offshore bacterial communities. All other described differences between t-DOM and control mesocosms for both the offshore and nearshore bacterial communities were similar at both 2h and 19h.

Population specific bacterioplankton response to t-DOM based on metagenome-assembled genomes (MAGs). In addition to evaluating the changes in the Lake Michigan bacterial community activity in response to t-DOM enrichment mesocosms, we investigated population specific trends in metabolic activity by mapping transcripts to reconstructed population genomes or MAGs from the Lake Michigan MGs. The taxonomic profiles at the phylum level for MTs from t-DOM mesocosms were largely similar to their corresponding control MTs for the 2h time-point, and at 19h both nearshore and offshore t-DOM MTs exhibited a ~100% increase in the transcripts for the *Bacteroidetes* phylum as compared to their control MTs (Supplementary Figure S3.4). Because the taxa present were not vastly different between samples and had a similar response to t-DOM pulse in both the nearshore and offshore mesocosms, MAGs were used to evaluate how organisms might respond differently based on their location in the environment and to the t-DOM pulse. Three deeply sequenced MGs from nearshore and offshore lake water were co-assembled and the contigs were subsequently used for genome binning to generate MAG consensus populations that are present across the sampled transect. The MG co-assembly, genome binning, and refining process resulted in the generation of 17 MAGs of good quality (genome completeness ≥ 50%; contamination ≤ 10%) (Supplementary Table S3.1). Based on MAG quality, MAG putative taxonomy from the MiGA webserver (27) and the overall number of transcripts mapping to genes encoded on a MAG, we

narrowed our focus to 9 MAGs that had a higher number of mapped transcripts from the different MTs and were classified within the known, abundant freshwater taxonomic groups: bin004, bin008, bin020, bin035, bin040, bin093, bin100, bin104 and bin105 (these MAGs are highlighted in yellow in Supplementary Table S3.1). The differences in gene expression for specific functional processes between treatment and control bacterial community activity were tested using DESeq2 (here we describe relative expression of functional processes as more or less if there is a $\log_2$ fold-change of ≥ 0.5 and/or P value ≤ 0.05) (31). Certain MAGs, namely bin004 (*Limnohabitans*), bin008 (*Rhodoferax*), and bin040 (AcI-B1 Actinobacteria) showed differences in organic substrate metabolism between the treatment and control bacterial community activity at 2h for both the nearshore and offshore water-based mesocosms (Supplementary Table S3.2). We observed more transcripts related to tricarballylate utilization, salicylate and gentisate catabolism, and glycerol and glycerol-3-phosphate uptake and utilization in the 2h t-DOM MTs compared to control MTs for MAGs bin004, bin008 and bin040, respectively (Supplementary Table S3.2). These trends were generally similar for both the offshore and nearshore mesocosms, indicating that the same organisms responded similarly to t-DOM regardless of their location in the lake (Supplementary Table S3.2). After 19h, these and other MAGs (bin093, bin105) had more transcripts related to cellular stress response, such as proteolysis, DNA repair, and protein chaperone activity (Supplementary Table S3.3) and fewer transcripts mapped to these MAGs, suggesting that these organisms were more stressed and less active towards the end of the t-DOM incubation (Supplementary Table S3.1).

Conversely, MAGs bin020, bin035, bin100 and bin104 exhibited similar or higher transcript counts in t-DOM MTs as compared to control at all time-points. Although these MAGs also exhibited more stress response-related transcripts in treatment MTs at 19h, the response was

less intense (lower fold-change values) as compared to the MAGs described above. In addition, they had similar or more transcripts for certain processes related to carbon metabolism at both time-points, thereby displaying active organic matter processing throughout the incubation. Three of these four MAGs were classified as *Bacteroidetes*, and one (bin104) was classified as *Sphingomonas*, an *Alphaproteobacterium* (Supplementary Table S3.1). bin020, classified as *Flavobacteriia* within *Bacteroidetes* (Figure 3.3), was particularly active in the nearshore waters where the number of transcripts for metabolic processes related to TCA cycle, aromatic compound metabolism (protocatechuate branch of beta-ketoadipate pathway) and carbohydrate fermentation (actetyl-CoA fermentation to butyrate) increased after 19h with t-DOM (Figure 3.3, Supplementary Table S3.4).

Tracking the bacterial taxa exhibiting DOM-associated transporter gene expression in the mesocosms over time. In addition to evaluating the SEED Subsystems-based overall functional response to t-DOM pulse by Lake Michigan bacteria at the whole community as well as population level, we specifically evaluated DOM-related substrate transporter activity of different bacterial taxa across the nearshore-to-offshore transect and in response to the t-DOM pulse. We assessed the taxonomic profile of transporter genes expressed for different classes of DOM monomers by the bacteria. These included amino acids, carbohydrates, carboxylic acids, nucleic acids, polyamines, organic phosphonates and lipids. Of these, the first four classes listed above had either a similar or higher normalized expression of transporter genes in the t-DOM mesocosms relative to the controls at both the time-points. Transporter gene expression for these four classes also constituted the bulk (> 80%) of DOM related transporter genes expressed by bacteria in all mesocosms. Interestingly, the taxa associated with these transcripts were different nearshore vs. offshore (Figure 3.4). In nearshore control mesocosms, transcripts affiliated with

Bacteroidetes comprised the majority (~50%) of the transcripts for amino acids and carboxylic acids transporter activity, whereas the same was not the case for offshore control mesocosms

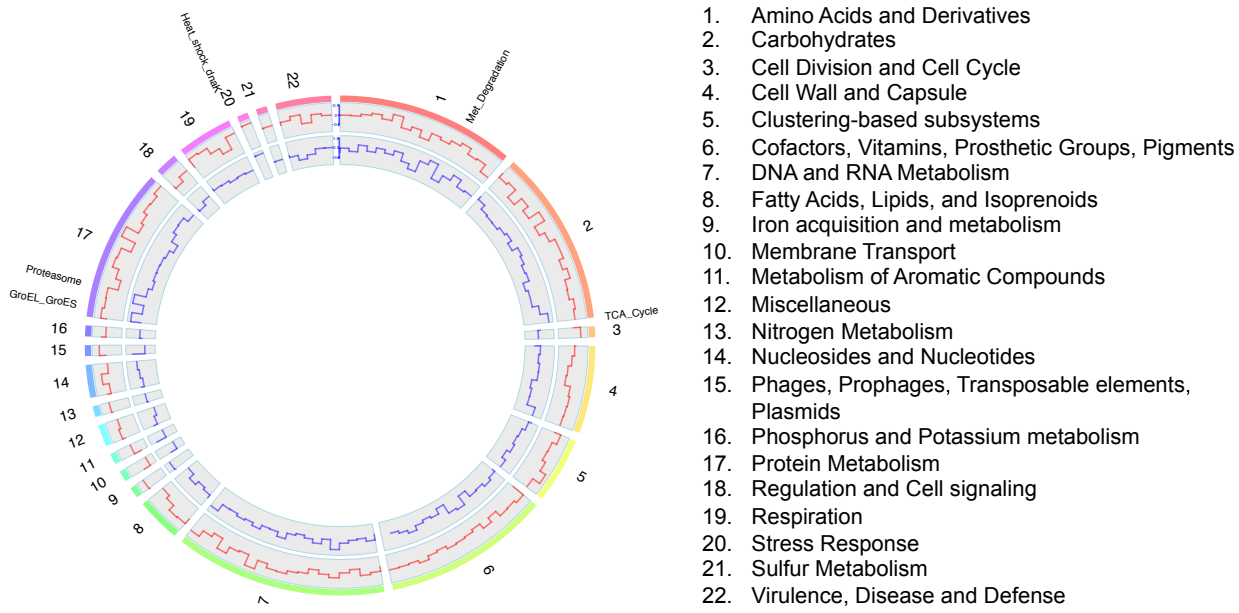


Figure 3.3: Differential gene expression for MAG bin020 (*Flavobacteriia*) between the nearshore t-DOM and control mesocosms at 2h and 19h. Following the circular tracks from outside to inside: track 1 shows the broad SEED Subsystems categories represented by arcs of different colors. Each of these arcs includes more resolved Subsystems functional processes within which the gene expression data are organized for this MAG. Only the specific functional processes that were significantly more or less abundant (DESEq2, Wald test, $P < 0.05$) between the t-DOM and control mesocosms at any time-point are labeled; track 2 uses a step plot to show $\log_2$ fold-change in expression for each functional process in the t-DOM treatment versus the control at 2h; track 3 uses a step plot to show $\log_2$ fold-change in expression for each functional process in the t-DOM treatment versus the control at 19h. Figure was generated with OmicCircos (32).

where *Actinobacteria* and *Proteobacteria* affiliated transcripts together constituted the bulk (~65-85%) of the amino acids and carboxylic acids-associated transporter gene expression. Furthermore, while *Actinobacteria* and *Firmicutes* affiliated transcripts comprised the majority transcripts in nearshore control mesocosms for nucleic acids and carbohydrates transporter activity, respectively, offshore control mesocosms had *Proteobacteria* and *Bacteroidetes* affiliated transcripts as the majority transcripts for nucleic acids and carbohydrates-associated transporter gene expression (Figure 3.4). In addition to these differences between unamended nearshore and offshore bacterial communities, communities in mesocosms that received a t-DOM pulse exhibited a change over time in the composition of the taxa involved in DOM-related transporter activity, primarily driven by an increase in the relative abundance of transcripts affiliated with *Bacteroidetes* after 19h in both the nearshore and offshore mesocosms (Figure 3.4). This was true for three of the four DOM classes; in the case of carboxylic acids, Bacteroidetes-affiliated transcripts only increased in relative abundance in the nearshore t-DOM mesocosms after 19h. In the offshore t-DOM mesocosms, Proteobacterial-affiliated transcripts increased after 19h (Figure 3.4).

E. Discussion

Using an integrated metagenomics and metatranscriptomics approach, this project aimed to investigate the active bacterial taxa in coastal and offshore regions of southern Lake Michigan, the importance of various DOM substrates for their metabolism, and specifically the potential role of terrestrial-derived DOM (t-DOM) in bacterial metabolism given the recent decline in the lake's phytoplankton abundance and production (8, 9). The significantly higher relative transcriptional activity of *Cyanobacteria*, which were predominantly *Synechococcus*, in the

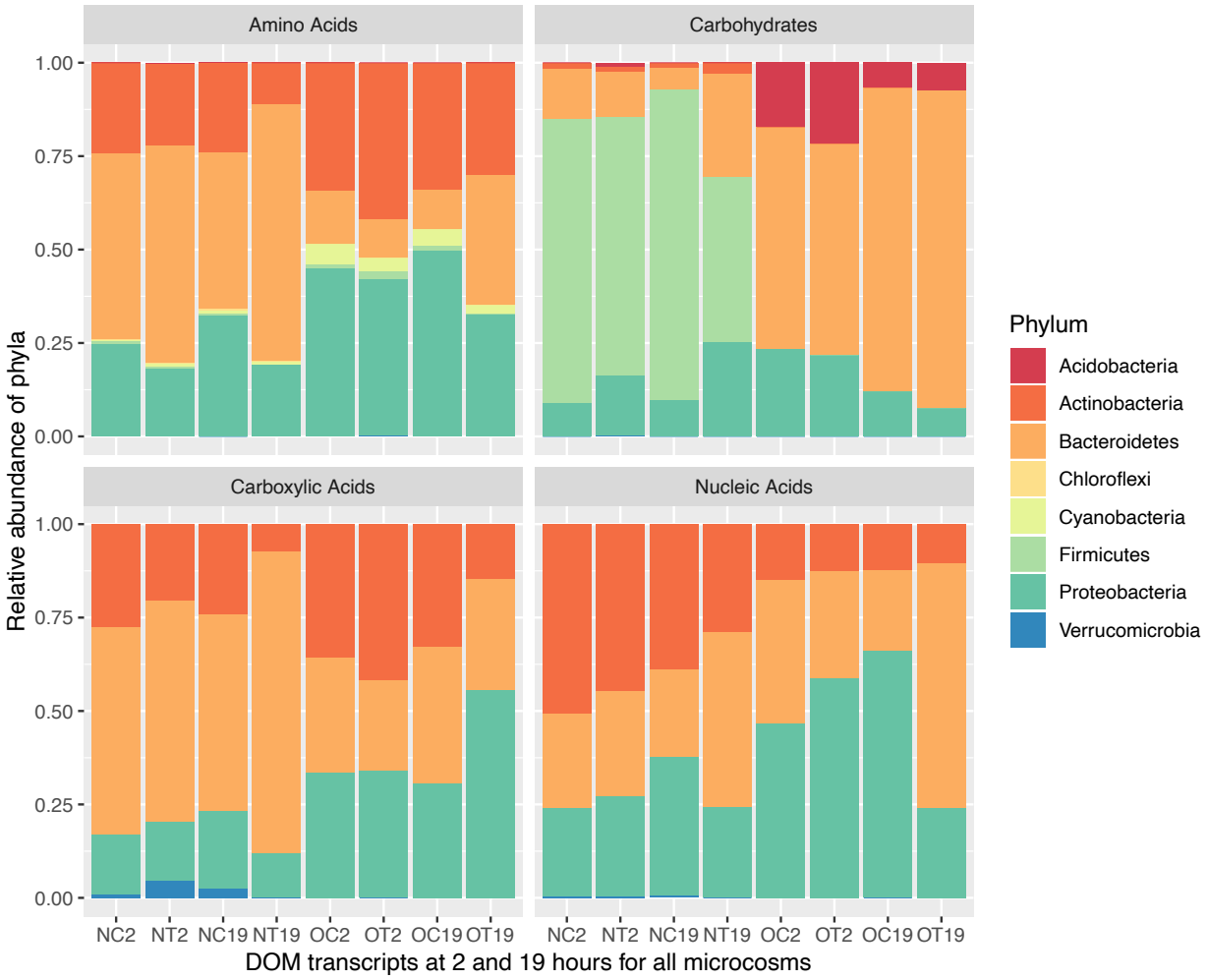


Figure 3.4: Taxonomic affiliation at the phylum level of active genes involved in transporter activity for four major classes of DOM compounds in all the mesocosm bacterial communities: amino acids, carboxylic acids, carbohydrates, and nucleic acids. The first two letters of the x-axis labels refer to the sampling site (N for nearshore, O for offshore) and treatment (C for control, T for t-DOM), followed by numbers 2 or 19 that refer to the sampling time-points – 2h and 19h, respectively. Relative abundance of each taxa is a fraction of the total transcript count mapping to transporter genes that were classified at the phylum level, with the transcript count normalized for gene length and metatranscriptome library size.

offshore surface waters provides evidence at the transcript level to support the recent findings of greater role of Cyanobacterial primary production in supporting the microbial food web in

offshore Lake Michigan as compared to in nearshore waters (11). The recent oligotrophication of offshore Lake Michigan as a result of invasive dreissenid mussels feeding on larger phytoplankton such as the diatoms has caused a decline in overall phytoplankton production (8, 9), resulting in a greater role of picophytoplankton primary production, as seen in other lakes (33). Although mussel feeding in nearshore waters would also similarly affect the phytoplankton community, the proximity of nearshore waters to terrestrial nutrient inputs could mitigate some of the negative impacts of mussel feeding on phytoplankton production and this, together with influx of terrestrially derived DOM, could possibly result in a more diverse pool of DOM sources for bacterial consumption in nearshore as comparison to offshore. The 137%, 29% and 44% more ammonium, nitrate and chlorophyll a in nearshore waters as compared to offshore at the time of sampling (Table 3.1) supports the hypothesis of terrestrial nutrient subsidies supporting nearshore primary production, although the SRP concentration in offshore was much higher (235%) as compared to in nearshore. The observed differences in Cyanobacterial transcripts between nearshore and offshore bacterial communities can likely be associated with the higher transcription of photosynthesis related processes observed in the offshore mesocosms as compared to nearshore (Figure 3.2B, Supplementary Figure S3.3).

In addition, we found differences in the taxonomic composition of transcripts associated with DOM-related transporter genes between the control nearshore and offshore bacterial communities (Figure 3.4). This is an interesting finding that suggests potentially different substrate preferences for organisms classified within the same phyla in different regions of the lake. This could be due to differential activity across the transect of phylogenetically related organisms that have different substrate preferences (34), and this in turn may reflect the local availability of these substrates and lake physicochemical conditions. Additionally, despite the

similarity in the phylum-level taxonomic profile for the total bacterial communities of control nearshore and offshore mesocosms, the differences in relative abundance of certain genera seen from the 16S rRNA gene amplicons (Supplementary Figure S3.2) may be associated with the observed taxonomic differences in the DOM-transporter gene transcripts. The significantly higher relative abundance of the alfV-A clade seen here in the offshore is consistent with the recent findings of high abundance for this group in offshore Lake Michigan waters in the summer (5). The low efficiency of annotating genes/transcripts at more resolved taxonomic levels (genus) limits the scope of our interpretations here. Lastly, the observed differences in the active bacterial community composition and transporter gene expression should be viewed in the context that they were evaluated using MTs from the control mesocosms (2h time-point), so potential variation in the transcriptional profile of the bacterial communities due to bottle effects cannot be ruled out (35). Nevertheless, these results established the basis for a differential transcriptional response to t-DOM amendments depending on the region of the lake. This is especially interesting, given the overall similarity in the total microbial communities at the two locations.

Surprisingly, the transcriptional activity of nearshore and offshore bacterial communities observed in response to a t-DOM pulse in the mesocosms followed a generally similar pattern over time (Figure 3.2). In terms of organic matter metabolism, the upregulation of the central-meta cleavage pathway of aromatic compound degradation both in offshore and nearshore mesocosms in response to the t-DOM pulse suggested the presence of such compounds in t-DOM and provided evidence for metabolic capability, at least at the transcription level, of the bacterial communities across different regions of Lake Michigan to utilize t-DOM. In addition to the transcriptional activity at the whole community level, we saw more transcripts for

aromatic compound degradation and C metabolism in response to t-DOM by MAGs, especially for populations associated with the phylum *Bacteroidetes*. Freshwater lineages of *Bacteroidetes* are known to be involved in complex DOM degradation derived either from humic-rich terrestrial sources or from protein-rich exudates of phytoplankton cells (36). *Flavobacteriia* have also been enriched in both DNA and RNA from freshwater-based microcosms amended with naphthalene, an aromatic hydrocarbon (37). Here, the observed relative increase in DOM transporter gene expression over time for the *Bacteroidetes* group from the t-DOM mesocosms (Figure 3.4) provides further evidence for the possible role of this group in breaking down complex components of t-DOM and assimilating the degradation products (38). Despite the generally similar response to t-DOM pulse observed in the nearshore and offshore lake water-based mesocosms, it is important to note that there were differences in the activity and response to t-DOM for certain organisms across the transect (Figures 3.3, 3.4). The predominantly higher transcriptional activity for the *Flavobacteriia*-affiliated organism represented by MAG bin _020 in nearshore mesocosms suggests its habitat preference for nearshore waters, and its ability to metabolize t-DOM may be linked to the potentially higher levels of terrestrial C subsidies in the nearshore. In the context of recent metagenomics and population genomics-based work that evaluated the carbon and nutrient metabolism of various possible C sources in freshwater bacteria (39), the short-term transcriptional response to the t-DOM pulse observed here seems specific to aromatic compounds.

It is unclear if the higher transcriptional activity observed for stress response-related processes for the bacterioplankton in t-DOM mesocosms at 19h is a result of bacterial competition in response to t-DOM enrichment or due to the chemical composition of the t-DOM treatment itself. From the transcriptional abundance patterns of the MAGs (Supplementary

Tables S3.1, S3.2 & S3.3), it seems that some organisms such as *Bacteroidetes* and *Alphaproteobacteria* (*Sphingomonas*) were more functionally active and less stressed in the t-DOM mesocosms than others. Thus, the relative increase in transcriptional activity of DOM-related metabolic processes such as transporter gene expression seen in *Bacteroidetes* could possibly be due the organisms' direct response to t-DOM enrichment or a result of other organisms being under more stress, or both. Despite the confounding effects of the stress response on limiting the interpretation of community metabolic response to the t-DOM pulse, the overall results from our incubation experiment suggest that bacterial community assemblages both nearshore and offshore in Lake Michigan have the capacity to metabolize terrestrially derived complex DOM. In addition, this ability is relatively independent of the distance of the assemblages from the shore with the possible exception of few organisms that may exhibit habitat preferences across the transect. Recent work in southern Lake Michigan reported that despite their relatively low abundance, various taxa within *Bacteroidetes* and *Proteobacteria* have a significantly higher cellular protein synthesis potential (PSP) than the more abundant freshwater groups like AcI and LD12 (6). Although their larger cell size may be a contributing factor to this, the higher PSPs could also be an indication of a more active role for these taxa in freshwater ecosystem processes than their relatively low abundance would suggest. The role of bacterial groups within the *Bacteroidetes* as the primary responders at the transcript level to the t-DOM pulse seen here provides further evidence of their importance to the ecosystem processes in freshwater lakes in assimilating complex carbon from diverse sources. Our findings are particularly significant in the context of recent ecological changes in the microbial food web in offshore Lake Michigan, where the decline in primary production due to invasive dreissenid mussels has likely reduced the available labile DOM pool released from autochthonous sources

i.e. phytoplankton cells. The ability to metabolize t-DOM has the potential to offset some of the negative effects of the declining autochthonous labile DOM on bacterial community function. The relatively stable bacterial respiration rates and cell abundance in the offshore region in the post-mussel invasion period in comparison to the pre-mussel period highlight the metabolic flexibility and resilience of Lake Michigan bacterioplankton (11, 12), and t-DOM may be playing an important role in supporting this stability. However, further evidence would be needed in order to conclusively validate the increased role of t-DOM in offshore bacterial metabolism, such as by assessing bacterioplankton transcription, production and growth-rates in response to diverse terrestrial DOM sources and comparing these with *in-situ* microbial community activity.

F. Acknowledgements

This work was supported by Illinois-Indiana Sea Grant Cooperative Science and Monitoring Initiative (CSMI 2015). We thank the crew of R/V *Lake Guardian* for assistance with water sampling. Rachel Macam assisted with nucleic acid isolation and library preparation for sequencing. We thank personnel of University of Illinois at Chicago Sequencing Core for facilitating sample sequencing, and Joel Hoffman (U.S. Environmental Protection Agency) for measuring water chemistry.

G. References

1. Cole J, Findlay S, Pace M. 1988. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Mar Ecol Prog Ser 43:1–10.

2. Cole JJ, Prairie YT, Caraco NF, Mcdowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ, Melack J. 2007. Plumbing the Global Carbon Cycle : Integrating Inland Waters into the Terrestrial Carbon Budget 171–184.

3. Downing J a., Cole JJ, Middelburg JJ, Striegl RG, Duarte CM, Kortelainen P, Prairie YT, Laube K a. 2008. Sediment organic carbon burial in agriculturally eutrophic impoundments over the last century. Global Biogeochem Cycles 22:1–10.

4. Paver SF, Newton RJ, Coleman ML. 2020. Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry. Environ Microbiol 22:433–446.

5. Fujimoto M, Cavaletto J, Liebig JR, McCarthy A, Vanderploeg H a., Denef VJ. 2016. Spatiotemporal distribution of bacterioplankton functional groups along a freshwater estuary to pelagic gradient in Lake Michigan. J Great Lakes Res 42:1036–1048.

6. Denef VJ, Fujimoto M, Berry M a., Schmidt ML. 2016. Seasonal succession leads to habitat-dependent differentiation in ribosomal RNA:DNA ratios among freshwater lake bacteria. Front Microbiol 7:1–13.

7. Biddanda B a., Cotner JB. 2002. Love handles in aquatic ecosystems: The role of dissolved organic carbon drawdown, resuspended sediments, and terrigenous inputs in the carbon balance of Lake Michigan. Ecosystems 5:431–445.

8. Vanderploeg H a., Bunnell DB, Carrick HJ, Höök TO. 2015. Complex interactions in Lake Michigan's rapidly changing ecosystem. J Great Lakes Res 41:1–6.

9. Fahnenstiel G, Pothoven S, Vanderploeg H, Klarer D, Nalepa T, Scavia D. 2010. Recent changes in primary production and phytoplankton in the offshore region of southeastern Lake Michigan. J Great Lakes Res 36:20–29.

10. Robertson DM, Saad D a. 2013. Reply to Discussion - "Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models." J Am Water Resour Assoc 49:725–734.

11. Carrick HJ, Butts E, Daniels D, Fehringer M, Frazier C, Fahnenstiel GL, Pothoven S, Vanderploeg H a. 2015. Variation in the abundance of pico, nano, and microplankton in Lake Michigan: Historic and basin-wide comparisons. J Great Lakes Res 41:66–74.

12. Weinke AD, Kendall ST, Kroll DJ, Strickler E a., Weinert ME, Holcomb TM, Defore A a., Dila DK, Snider MJ, Gereaux LC, Biddanda B a. 2014. Systematically variable

planktonic carbon metabolism along a land-to-lake gradient in a Great Lakes coastal zone. J Plankton Res 36:1528–1542.

13. Dila DK, Biddanda B a. 2015. From land to lake: Contrasting microbial processes across a Great Lakes gradient of organic carbon and inorganic nutrient inventories. J Great Lakes Res 41:75–85.

14. Poretsky RS, Sun S, Mou X, Moran MA. 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ Microbiol 12:616–627.

15. Chaudhary A, Kauser I, Ray A, Poretsky R. 2018. Taxon-Driven Functional Shifts Associated with Storm Flow in an Urban Stream Microbial Community. mSphere 3.

16. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676.

17. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38.

18. Luo C, Rodriguez-R LM, Konstantinidis KT. 2014. MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res 42.

19. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60.

20. Wu CH, Apweiler R, Bairoch A, Natale D a, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187-91.

21. Overbeek R, Begley T, Butler RM, Choudhuri J V., Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov D a., Rül;ckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33:5691–5702.

22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25:25–29.

23. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.

24. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

25. Johnston ER, Rodriguez-R LM, Luo C, Yuan MM, Wu L, He Z, Schuur E a. G, Luo Y, Tiedje JM, Zhou J, Konstantinidis KT. 2016. Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. Front Microbiol 7:1–16.

26. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. MetaSPAdes: A new versatile metagenomic assembler. Genome Res 27:824–834.

27. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282–W288.

28. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

29. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.

30. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection : a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J Prepr 4:e1900v1.

31. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21.

32. Hu Y, Yan C, Hsu CH, Chen QR, Niu K, Komatsoulis GA, Meerzaman D. 2014. Omiccircos: A simple-to-use R package for the circular visualization of multidimensional Omics data. Cancer Inform 13:13–20.

33. Callieri C, Stockner J. 2000. Picocyanobacteria success in oligotrophic lakes: Fact or fiction? J Limnol 59:72–76.

34. Jones SE, Newton RJ, McMahon KD. 2009. Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. Environ Microbiol 11:2463–2472.

35. Stewart FJ, Dalsgaard T, Young CR, Thamdrup B, Revsbech NP, Ulloa O, Canfield DE, DeLong EF. 2012. Experimental incubations elicit profound changes in community transcription in OMZ bacterioplankton. PLoS One 7.

36. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria.Microbiology and molecular biology reviews : MMBR.

37. Jurelevicius D, Alvarez VM, Marques JM, Lima LRFDS, Dias FDA, Seldin L. 2013. Bacterial community response to petroleum hydrocarbon amendments in freshwater, marine, and hypersaline water-containing microcosms. Appl Environ Microbiol 79:5927–5935.

38. Pernthaler J. 2017. Competition and niche separation of pelagic bacteria in freshwater habitats. Environ Microbiol 19:2133–2150.

39. Linz AM, He S, Stevens SLR, Anantharaman K, Rohwer RR, Malmstrom RR, Bertilsson S, McMahon KD. 2018. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. PeerJ 6:e6075.

IV. *In situ* Bacterial Community Dynamics Related to Dissolved Organic Matter Metabolism in Southern Lake Michigan

A. Introduction

Recent work has shown that despite the recent oligotrophication of offshore Lake Michigan and decline in phytoplankton productivity due to invasive dreissenid mussels, bacterial respiration rates have remained relatively stable in comparison to pre-dreissenid mussel periods, resulting in offshore Lake Michigan becoming a net source of carbon to the atmosphere as compared to a net sink prior to dreissenid mussel proliferation (1). In comparison, nearshore waters in southern half of the lake continue to remain a net sink of carbon due to both high phytoplankton and bacterial growth rates. Thus, investigating the bacterial functional activity in nearshore and offshore regions of the lake especially with respect to DOM uptake and metabolism will inform us of the relative importance of different classes of DOM compounds for bacterial metabolism and their potential source of origin i.e. autochthonous or allochthonous. Understanding this microbial-carbon link in Lake Michigan is not only important for the current monitoring efforts of the lake's microbial food web, but it can also help better predict the bacterial response in future disturbance scenarios or similar ecological disturbances in other large lakes.

In Chapter 3, based on the analysis of nearshore and offshore lake bacterial community dynamics in summer 2015, we observed that despite overall similarities in the bacterial community composition between the two regions, there were differences in the relative abundance of specific taxa and gene transcripts across the transect. However, the analysis of only a single time-point limits the scope of these results, and more sampling effort is required to arrive at robust predictions about *in situ* bacterial community structure and DOM-associated

metabolism in Lake Michigan. In addition, to arrive at a more mechanistic understanding of the link between bacterioplankton metabolic potential and DOM processing in the lake, it is important to investigate the microbial molecular data in the context of DOM characteristics in pelagic Lake Michigan. Work done by Zhou et al. (2) to evaluate the spectrofluorometric characteristics of the bulk DOM pool in all the Great Lakes provided valuable information about the DOM composition in offshore Lake Michigan. However, a more recent effort to characterize DOM in both the nearshore and offshore regions of the lake has been lacking.

In this study, we extended the preliminary work done in Chapter 3 by further sampling the nearshore-to-offshore Lake Michigan transect near the mouth of Kalamazoo River in summer 2017. The combined 2015 and 2017 samples were analyzed using a shotgun metagenomics-based approach enabling more reliable spatiotemporal evaluation of DOM metabolism and bacterial community structure across southern Lake Michigan. Additionally, we performed spectrofluorometric characterization of the bulk DOM pool as well as measurement of dissolved organic carbon (DOC) and nutrient concentrations across the transect in 2017-2018 to provide important water chemistry context to the molecular microbial datasets. Results from the bacterial community data supported the trends seen earlier in 2015 with broad similarities in community composition and functional diversity in nearshore and offshore Lake Michigan, however there were important differences related to the abundance of certain taxa and genes encoding for aromatic compound metabolism. The results from the bulk DOM characterization supported the microbial metabolic trends with a significantly higher aromaticity and humic content observed in nearshore DOM as compared to offshore. Lastly, tracking of specific bacterial populations across the transect using metagenome-assembled genomes (MAGs) only partially supported the trends

seen in the overall bacterial community diversity, such as the predominantly offshore presence of a MAG-based Cyanobacterial population.

B. Materials and Methods

Sample collection. Near-surface (0-2 m depth) Lake Michigan water samples were collected across a nearshore-to-offshore transect beginning at the mouth of Kalamazoo River, one of the largest tributaries to southern Lake Michigan (3). From two sampling events (September 2015 and July/August 2017), a total of six samples were collected from four sites across the transect – three samples from nearshore site NRS (~3.5 km from shore), and 1 each from offshore sites OFS-10, OFS-30 and OFS-40 with the numbers in the site names representing distance of that site (in km) from shore (samples in 2015 were only collected from sites NRS and OFS-40, Table 4.1). For samples from 2015, water sample collection, storage, filtration for free-living microorganisms and metagenome sequencing has been described in detail in Chapter 3. For samples from 2017, between 5-10L water was collected in 10 L polycarbonate cubitainers/carboys for each site. Collected water was filtered on-site through pre-combusted 1.6 μm pore-size glass fiber filters (Sterlitech, Kent, WA) to remove larger particles and organisms, and free-living cells were collected on 0.2 μm pore-size polycarbonate membrane filters (EMD Millipore, Billerica, MA). Filters were stored immediately either in liquid nitrogen or dry ice and transported back to lab for storage at -80 °C until DNA isolation. In addition, for samples from 2017, about 100 ml of the filtrate through the 0.2 μm pore-size filters was collected in pre-combusted scint vials for each site and stored at -20 °C until use for nutrient and dissolved organic carbon (DOC) measurements.

DNA isolation and metagenomic sequencing. For the 2017 samples, the 0.2 μm pore-size filters were broken into small fragments, and a portion of the fragments were picked at

Table 4.1. Samples collected in southern Lake Michigan and their water chemistry characteristics.

Sample Name	Sample Type	Sampling date	Distance from shore (km)	Water Temperature (°C)	Soluble Phosphate (PO_4^{3-}) μg P/L	NO_x^- [a] μg N/L	NH_4 μg N/L	DOC[b] (μM) ± SD[c] (μM)
NRS-A	Nearshore	9/13/15	3.5		4.2	312	22.7	
NRS-B	Nearshore	7/27/17	3.5	26.6	2.8	420		183±1.1
NRS-C	Nearshore	8/8/17	3.5	20.7	1.3	290		191±1.2
OFS-10	Offshore	8/2/17	10	22.1	2	258		185±0.9
OFS-30	Offshore	8/3/17	30	23.2	2.5	293		145±0.4
OFS-40	Offshore	9/13/15	40		13.9	242	9.6	

[a]NO_x^- : Nitrate + nitrite

[b]DOC: Dissolved organic carbon

[c]SD: Standard deviation

random for DNA isolation using an organic extraction method as described previously (4). Briefly, the filter fragments were first incubated in a lysis buffer that contained 1.15 mg/ml lysozyme and 200 μg/ml RNase at 37 °C for 30 min, and this was followed by incubation with

1% SDS and 10 mg/ml proteinase K at 55 °C for 2 h while rotating. From the lysate, DNA was extracted using phenol:chloroform and subsequently isolated using ethanol precipitation followed by elution in Tris-EDTA (TE) buffer. Complete description of DNA isolation, metagenomic sequencing and library yield for the nearshore and offshore samples collected in September 2015 is provided in the methods section of Chapter 3. The isolated DNA for all the samples was subsequently used for whole-genome shotgun sequencing on an Illumina NextSeq500 at the University of Illinois at Chicago Sequencing Core, in paired-end format and read-length of 150 bp (the 2015 and 2017 samples were sequenced in two separate sequencing runs). Overall, we obtained six deeply sequenced metagenomes with 26-62 million reads per library.

Analysis of metagenomic datasets. The raw metagenomic libraries were first quality filtered using a Phred average per sliding window with a quality threshold (Q) of ≥ 20 and not allowing any N values. The trimmed paired-end reads for each metagenome were then individually assembled into longer contiguous sequences or contigs using MEGAHIT (5) with default settings. For each metagenome, contigs ≥ 500 bp were mined for protein coding genes using MetaGeneMark (6). The predicted protein-coding genes were then used for phylogenetic classification of their corresponding contigs using MyTaxa (7), using its database of bacterial and archaeal genomes (http://enveomics.ce.gatech.edu/data/mytaxa) and DIAMOND blastp in the sensitive mode (8). The metagenome-derived gene sequences were functionally annotated by first searching them against the Swiss-Prot database (9) using blastp with following cutoffs: 30% sequence identity, 70% coverage of query sequence, and an E value of $\leq 10^{-10}$. Subsequently, the Swiss-Prot match for the best hit for each query sequence was mapped to its corresponding term in the SEED database (10). Contig or gene abundance in the metagenomes was calculated by

mapping the short-reads for each metagenome to the corresponding contigs or genes using blastn with cutoffs: ≥ 75 bp sequence alignment length, ≥ 95% sequence identity, and an E value cutoff of ≤ 10^{-10}.

Metagenome-assembled genomes (MAGs) reconstruction and analysis. To reconstruct MAGs from the metagenomic libraries, we first co-assembled all the six metagenomes using MEGAHIT with default settings. From the combined assembly, contigs ≥ 1000 bp were used for obtaining population genome bins or MAGs using MaxBin 2.0 (11) with default settings. The MAGs generated were subsequently checked for quality (genome completion and contamination) using CheckM (12). Based on CheckM results, MAGs with ≥ 50% genome completeness were selected for improving their genome quality using a read recruitment and reassembly process as described in Chapter 3 Materials and Methods. Overall, we obtained 26 MAGs with ≥ 50% genome completion and ≤ 10% genome contamination. These MAGs were then uploaded to the MiGA webserver (13) for taxonomic annotation and mining for protein coding genes, apart from providing other metrics about the MAG quality. Read recruitment plots for the selected MAGs were generated using the *BlastTab.recplot2.R* script in the Enveomics toolbox (14).

Nutrients, DOC measurements and DOM characterization. Nutrient levels and environmental parameters corresponding to 2015 samples were measured by US EPA personnel according to standard EPA methods. For the 2017 samples, part of the collected filtrate (check Materials and Methods: Sample collection) was used for measuring nutrients (soluble phosphate: PO_4^{3-}, nitrate + nitrite: NO_x^-) at Karl Rockne Lab, University of Illinois at Chicago using an autoanalyzer (AQ300, SEAL Analytical, Mequon, WI) (15). The remaining filtrate was used for measuring DOC concentration at Guo Lab, School of Freshwater Sciences, University of

Wisconsin-Milwaukee using the high temperature combustion method (16). In addition, we also collected more filtrate from the same sites (excluding OFS-40) and an additional offshore site (OFS-50, 50 km from the shore) in spring (March 31-April 12) and summer (August 11-September 4) 2018. These filtrate samples were used for characterizing the chromophoric DOM (CDOM) and fluorescent DOM (using fluorescence excitation-emission matrix spectra) at the Guo Lab using methods and techniques as described in (2).

Statistical Analyses. All statistical tests on the nutrient and DOC/DOM characteristics as well as the metagenomic data were performed in R (v.4.0.2, "Taking off Again"). Differences in the average concentrations of nutrients and DOC between nearshore and offshore samples were calculated using *t* test. *t* test was also used for calculating differences in chromophoric and fluorescent DOM metrics between nearshore and offshore for DOM samples collected in 2018. We evaluated the significance of environmental factors (sample type/sample distance from shore), nutrients and their interaction effects in shaping the microbial community composition in Lake Michigan using Permutational Multivariate Analysis of Variance (PERMANOVA). PERMANOVA was performed using the adonis function in the R package vegan (17). Differences in the relative abundance of specific phyla between nearshore and offshore were evaluated using *t* test. We used the DESeq2 package in R (18) to analyze the microbial community functional profile in nearshore and offshore Lake Michigan using clustered heatmaps as well as to identify differentially abundant functional processes between the two regions.

C. Results and Discussion

Nutrients and DOC/DOM characteristics across southern Lake Michigan. Average NO_x^- (nitrate + nitrite) levels were 29% higher in the nearshore samples as compared to offshore

(Table 4.1), although not statistically significant (*t* test, $P = 0.15$). There was no significant difference in average phosphate levels between nearshore and offshore. Ammonium was measured for only summer 2015 and was found to be 136.5% higher in the nearshore as compared to offshore (only site OFS-40 was sampled). DOC was measured for summer 2017, and average DOC was 13% higher in nearshore although not statistically significant (*t* test, $P = 0.39$). Similar trends for DOC were also observed for spring 2018 (Table 4.2).

Compared to the nutrient and DOC levels, the chromophoric and fluorescent DOM characteristics for 2018 across the transect displayed stronger patterns of differences between nearshore and offshore (Table 4.2). Optical properties for characterizing CDOM such as absorption coefficient (a_{254}), spectral slope ($S_{275\text{-}295}$) and slope ration (S_R) were all significantly different between the nearshore and offshore 2018 samples (*t* test, $P < 0.05$). Higher a_{254}, lower $S_{275\text{-}295}$ and lower S_R values for nearshore CDOM as compared to offshore suggest that a larger proportion of nearshore DOM in comparison to offshore comprises of high molecular weight (HMW) and aromatic component that is likely derived from terrigenous sources (2). From the fluorescence EEM spectra, we found the presence of signatures of humic-like DOM (peak A and C) and protein-like DOM (peak B) in all the samples (Figure 4.1). However, there were differences in the relative intensity of peaks B and C between the nearshore and offshore samples. Nearshore samples seemed to have a more pronounced peak C, which has been associated with terrestrial humic-like DOM (2, 19). Conversely, offshore samples had a relatively more pronounced peak B, which has been associated with autochthonous production (2, 19). The values of biological index (BIX) and humification index (HIX) derived from the EEM data complemented the trends from Figure 4.1: nearshore DOM had significantly lower mean BIX than offshore DOM (*t* test, $P = 0.01$) (Table 4.2), and with the exception of one

summer 2018 sample the mean HIX of all the other nearshore samples was significantly higher than the offshore (*t* test, $P = 0.03$). Overall, the trends seen here from the DOM characterization suggest a relatively higher presence of humic-like, HMW-DOM in nearshore Lake Michigan waters that is likely derived from terrestrial sources such as the inputs from Kalamazoo River, whereas DOM in the offshore waters comprises of a higher proportion of DOM produced from autochthonous sources and with relatively more protein-like components. As seen previously (20), nearshore waters also had slightly higher DOC levels than offshore, and the proximity to terrestrial humic-like DOM may be contributing to this increase in nearshore.

Taxonomic and functional diversity of bacterial communities in nearshore and offshore southern Lake Michigan. Bacterial community composition at the phylum level (Proteobacteria divided into subphyla) did not seem to differ significantly between nearshore and offshore Lake Michigan (Figure 4.2A), and the variation in the community composition between the offshore sites may be a contributing factor to this. Based on permutational multivariate analysis of variance (PERMANOVA), the variation in phylum-level community composition between the different samples was not significantly influenced by sample type i.e. nearshore or offshore (adonis, $R^2 = 0.11$, $P > 0.05$); distance of the sample site from the shore ($R^2 = 0.15$, $P > 0.05$); phosphate ($R^2 = 0.07$, $P > 0.05$) or NO_x^- levels ($R^2 = 0.04$, $P > 0.05$) or their interacting effects ($R^2 = 0.03$, $P > 0.05$). Comparison of relative abundance of specific phyla between nearshore and offshore samples similarly revealed insignificant differences for most of the taxa with the notable exception of *Betaproteobacteria* that were significantly higher in relative abundance in nearshore than offshore (*t* test, $P = 0.02$). In addition, although statistically not significant, *Alphaproteobacteria* and *Cyanobacteria* were 3.5 times and 2.5 times more abundant in the offshore bacterial community as compared to nearshore, respectively. Taken together,

these results seem to suggest broad similarities in nearshore and offshore bacterial community structure but with a few important exceptions that may be associated with some of the trends seen in water chemistry (nutrients, DOM composition) across the transect. For instance, Betaproteobacterial genus *Polynucleobacter* includes free-living freshwater tribes that can uptake carboxylic acid monomers released from the photodegradation of humic acids (21). The higher relative abundance of Cyanobacteria (likely *Synechococcus*) in offshore predicably correlates with the higher BIX values seen for the offshore DOM. We explore the presence/abundance of some of these specific bacterial groups across the transect in more detail with the use of MAGs in later sections (see below).

Analysis of bacterial community functional profiles highlighted overall similarity across Lake Michigan as seen for the taxonomic diversity, but also provided possible links between microbial C metabolism and the observed DOM characteristics (Figure 4.2B). Community functional diversity based on annotation with the SEED Subsystems database at the broad level highlighted relatively better clustering of the nearshore samples together in comparison to the offshore samples (Figure 4.2B). Comparison of the broad functional processes individually between the nearshore and offshore bacterial communities revealed a significantly higher relative abundance of aromatic compound metabolism and motility/chemotaxis in the nearshore as compared to offshore (DESeq, adjusted P value < 0.1), whereas processes related to photosynthesis and clustering-based subsystems were significantly higher in relative abundance in the offshore (DESeq, adjusted P value < 0.1). The higher abundance of genes associated with aromatic compound metabolism in nearshore correlates well with the higher degree of nearshore-DOM aromaticity that is likely arising from the terrestrial humic inputs. And the higher abundance of bacterial photosynthesis processes in offshore are likely associated with the higher

Table 4.2. DOC[a] levels and Chromophoric DOM/Fluorescent DOM characteristics in southern Lake Michigan in 2018.

Sample Station	NRS	OFS-10	OFS-30	OFS-50	NRS	NRS	NRS	NRS
Sample Type	Nearshore	Offshore	Offshore	Offshore	Nearshore	Nearshore	Nearshore	Nearshore
Season	spring	spring	spring	spring	summer	summer	summer	summer
DOC (μM)	162	128	147	93				
Absorption coefficient (m^{-1}) 254 nm	8.73	4.27	4.01	4.11	7.71	9.34	4.24	6.23
$S_{275-295}$[b]	0.022	0.03	0.031	0.029	0.023	0.023	0.031	0.022
S_R[c]	1.409	1.956	2.235	2.767	1.461	1.274	2.215	1.77
BIX[d]	0.69	0.9	0.91	0.97	0.74	0.74	0.87	0.8
HIX[e]	5.36	2.14	2	1.91	4.45	5.87	1.38	2.63

[a]DOC: Dissolved organic carbon

[b]$S_{275-295}$: Chromophoric DOM-based spectral slope

[c]S_R: Chromophoric DOM-based slope ratio

[d]BIX: Fluorescent DOM-based Biological index

[e]HIX: Fluorescent DOM-based Humification index

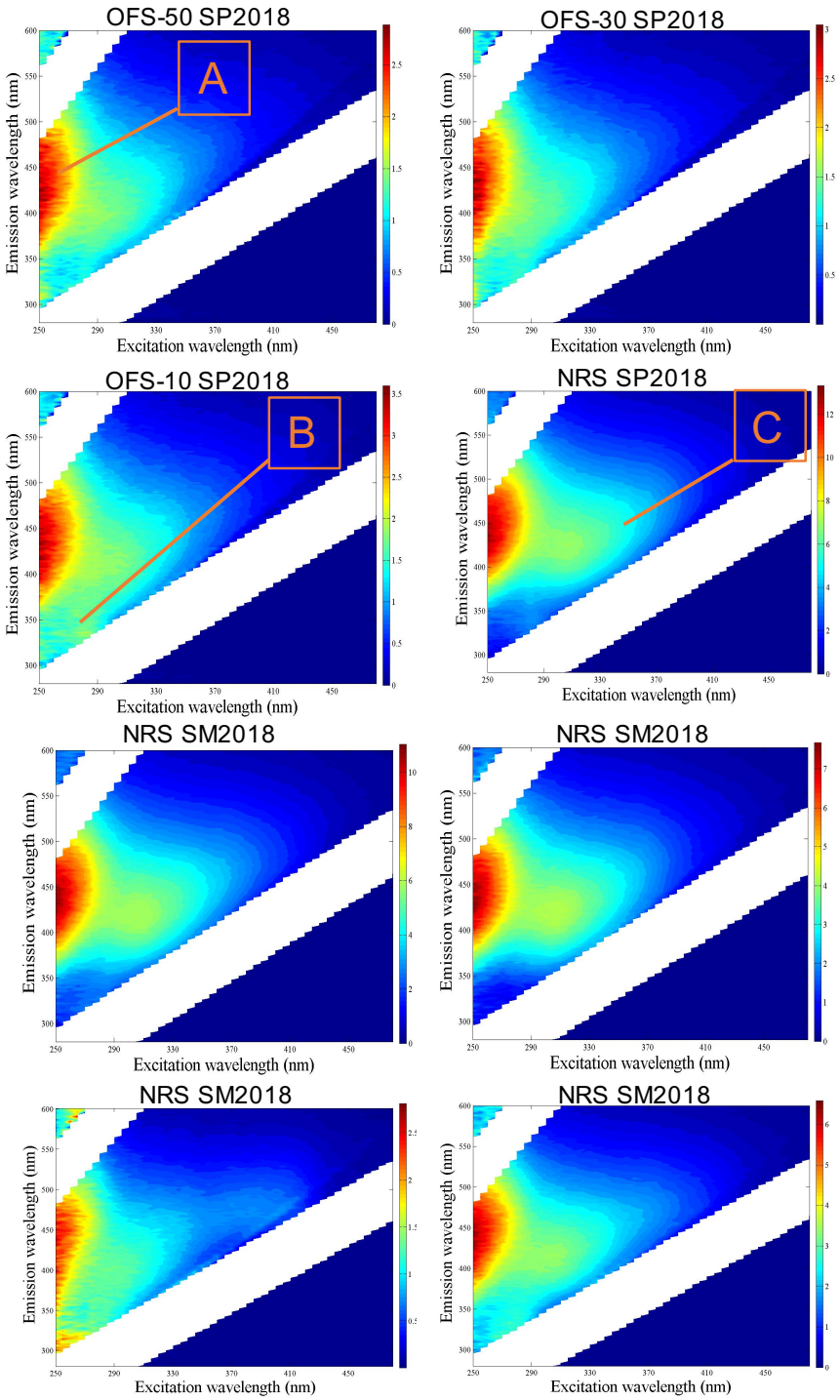


Figure 4.1. Fluorescence excitation-emission matrix spectra for spring and summer Lake Michigan DOM. Each spectra is labelled by its corresponding sample station (nearshore/offshore) and the season in which the DOM sample was obtained (spring 2018 – SP2018; summer 2018 – SM2018). Labelled peaks in the spectra show presence of humic-like DOM (peaks A and C) and protein-like DOM (peak B).

abundance of Cyanobacteria (*Synechococcus*) observed there.

Taken together, the results suggest broad similarities in the functional potential of major metabolic processes related to C, N and P between nearshore and offshore bacterial communities, but there exist certain differences related to C metabolism that could be tied with the local water chemistry. As seen earlier in Chapter 3, groups within *Bacteroidetes* are known to have the potential to metabolize complex aromatic compounds in freshwater systems. In addition, we observed a differential abundance of *Betaproteobacteria* across the transect and this phylum also includes groups that can utilize photodegradation products of humic DOM. Using genome-resolved metagenomics to isolate MAGs from within these groups from the Lake Michigan metagenomes provides us with the opportunity to test whether particular populations from these phyla have differential abundances in nearshore and offshore regions and if we can model their metabolism to evaluate the possible utilization of terrestrial DOM-derived substrates.

Tracking specific bacterial populations in Lake Michigan using metagenome-assembled genomes (MAGs). Tracking MAG-based consensus populations across nearshore and offshore Lake Michigan revealed variable trends in the presence and abundance of different populations representing major bacterial phyla. The combined assembly of the metagenomes followed by population binning and contig reassembly generated 26 MAGs of good quality (check Materials and Methods). We narrowed our focus on 6 of these MAGs that represented populations from the taxonomic groups that showed differential abundance across the transect and/or are known to include bacteria that can metabolize HMW-DOM in freshwater ecosystems (22) (Table 4.3). Four of these MAGs (LMS_bin181, LMS_bin079, LMS_bin056 and LMS_bin010) were classified by the MiGA webserver (13) within *Bacteroidetes*, one (LMS_bin035) within *Cyanobacteria* (*Synechococcales*) and one (LMS_bin009) within

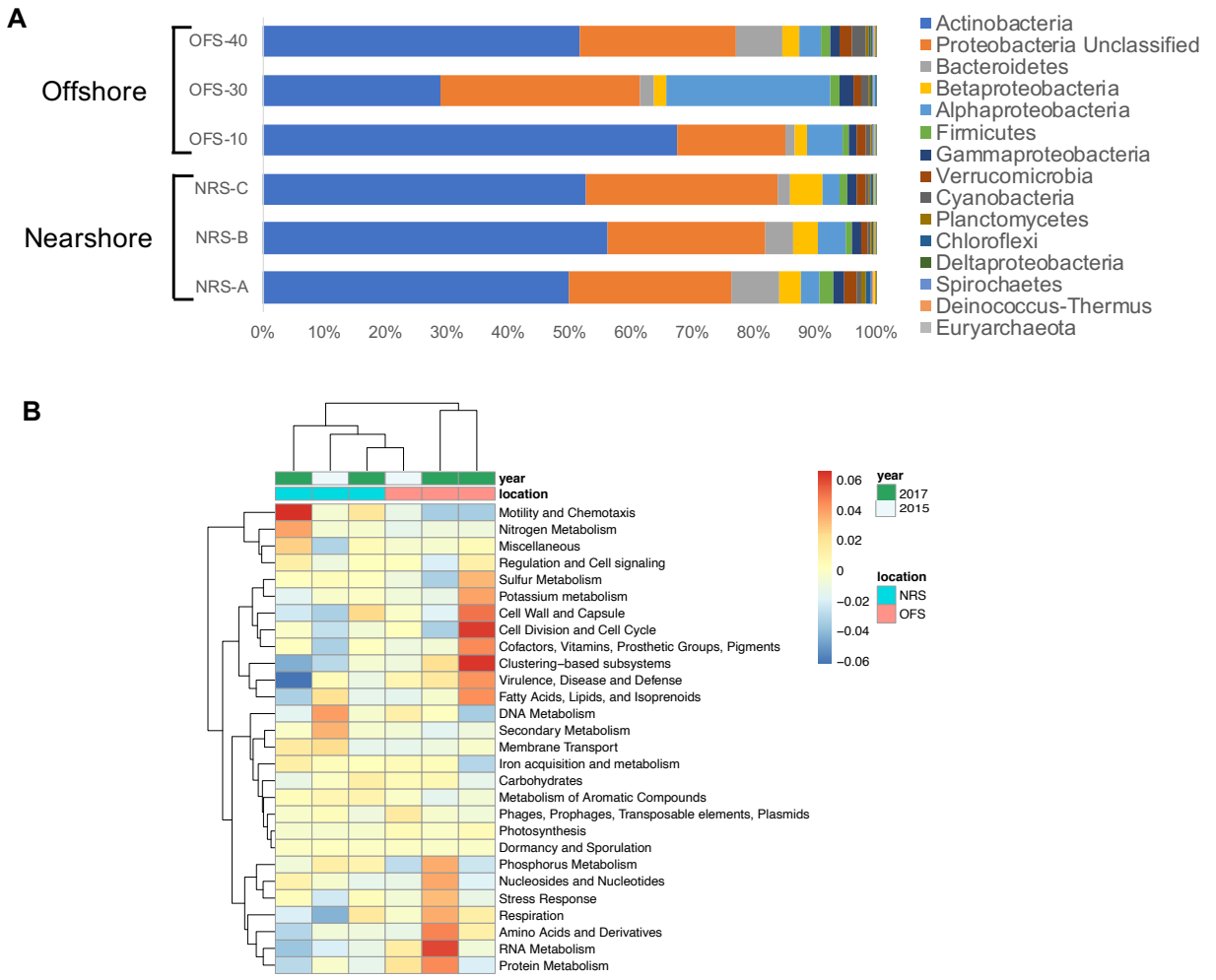


Figure 4.2. (A) Taxonomic composition of nearshore and offshore Lake Michigan microbial communities from 2015 and 2017 at the phylum level (Proteobacteria divided into subphyla). Nearshore site NRS was sampled once in 2015 (NRS-A) and twice in 2017 (NRS-B and -C). Offshore site OFS-40 was sampled in 2015, and sites OFS-10 and OFS-30 were sampled in 2017 (see Materials and Methods). **(B)** Microbial community functional diversity in nearshore (NRS) and offshore (OFS) Lake Michigan based on broad-level annotation with SEED Subsystems database. Gene abundance for individual SEED categories were normalized using the 'rlog' (regularized-logarithm transformation) function of DESeq2 such that the data are approximately homoskedastic.

Betaproteobacteria (*Polynucleobacter*). To assess the presence and abundance of these populations across the transect, we tracked genome coverage of each MAG for all the six

metagenomes using read recruitment plots (Figure 4.3). Contrary to our expectations, most of the MAG-based populations showed relatively consistent coverage and presence in both nearshore and offshore Lake Michigan (Supplementary Figures S4.1and S4.2). Only MAGs LMS_bin035 (*Synechococcales)* and LMS_bin181 (*Fluviicola*) differed in their presence or abundance across the transect, with the LMS_bin035 consistently present only in the offshore sites with moderate-to-high coverage (Figure 4.3). Conversely, LMS_bin181 exhibited sufficient coverage in only two nearshore samples and had incomplete coverage in the rest (Supplementary Figure S4.3). The differential presence of a Cyanobacterial population between nearshore and offshore provides further evidence for the trends already seen earlier in Chapter 3 as well as for the overall microbial community data in this study (Figure 4.2). With the oligotrophication of

Table 4.3. Summary statistics for the metagenome-assembled genomes (MAGs) used for population tracking across Lake Michigan. MAG completion and contamination were determined using CheckM (12), and overall quality and likely taxonomy were determined using MiGA webserver (13)**.**

bin ID	Likely taxonomic classification	MAG completeness (%)	MAG contamination (%)	MAG overall quality (out of 100)
LMS_bin009	*Polynucleobacter (Betaproteobacteria)*	82.6	2.1	37 (Intermediate)
LMS_bin010	*Chitinophagia (Bacteroidetes)*	90	4.3	67 (High)
LMS_bin035	*Synechococcaceae (Cyanobacteria)*	51.6	0.9	56 (High)
LMS_bin056	*Cyclobacteriaceae (Bacteroidetes)*	88.4	2.8	59 (High)
LMS_bin079	*Cytophagales (Bacteroidetes)*	76.2	8.3	53 (High)
LMS_bin181	*Fluviicola (Bacteroidetes)*	99.5	0.8	91 (Excellent)

offshore Lake Michigan and decline in microphytoplankton production, the higher transcriptional activity (Chapter 3, Figure 3.2) as well as abundance of Cyanobacterial populations in offshore (Figure 4.2, 4.3) strongly supports the hypothesis of a larger role of Cyanobacteria primary production in supporting the microbial food web in offshore Lake Michigan and maintaining stable microbial respiration rates in the post-mussel period (1, 23). On the contrary, the sporadic presence of *Fluviicola* population seen in the nearshore may be related to nearshore microphytoplankton bloom dynamics as *Fluviicola* are known to be primary responders to labile DOM released from phytoplankton in freshwater ecosystems (24). These phytoplankton blooms may be a product of high nutrient inputs from the Kalamazoo River during rain-associated high flow periods, as seen earlier in other coastal-to-offshore transects in the lake (23).

The *Bacteroidetes* and *Betaproteobacteria* MAG populations exhibited consistent presence across the transect, and given their family/genus level affiliations this was contrary to our initial hypothesis and some of the trends observed from the overall microbial community composition. *Chitinophagia* include organisms that can utilize HMW-DOM from various sources including terrestrial-derived humic content (22). Additionally, *Polynucleobacter* includes members that are ubiquitous in freshwater ecosystems, and can utilize diverse substrates derived from phytoplankton (25) as well as humic acids photodegradation products (21). The consistent distribution of these microorganisms across Lake Michigan perhaps reflects the availability of their preferred substrates in both nearshore and offshore, and thus a relative lack of dependence on terrestrial-derived humic content. Perhaps there exist other populations within these phyla that may have preference for terrestrial DOM and thus have higher abundance/activity in the nearshore.

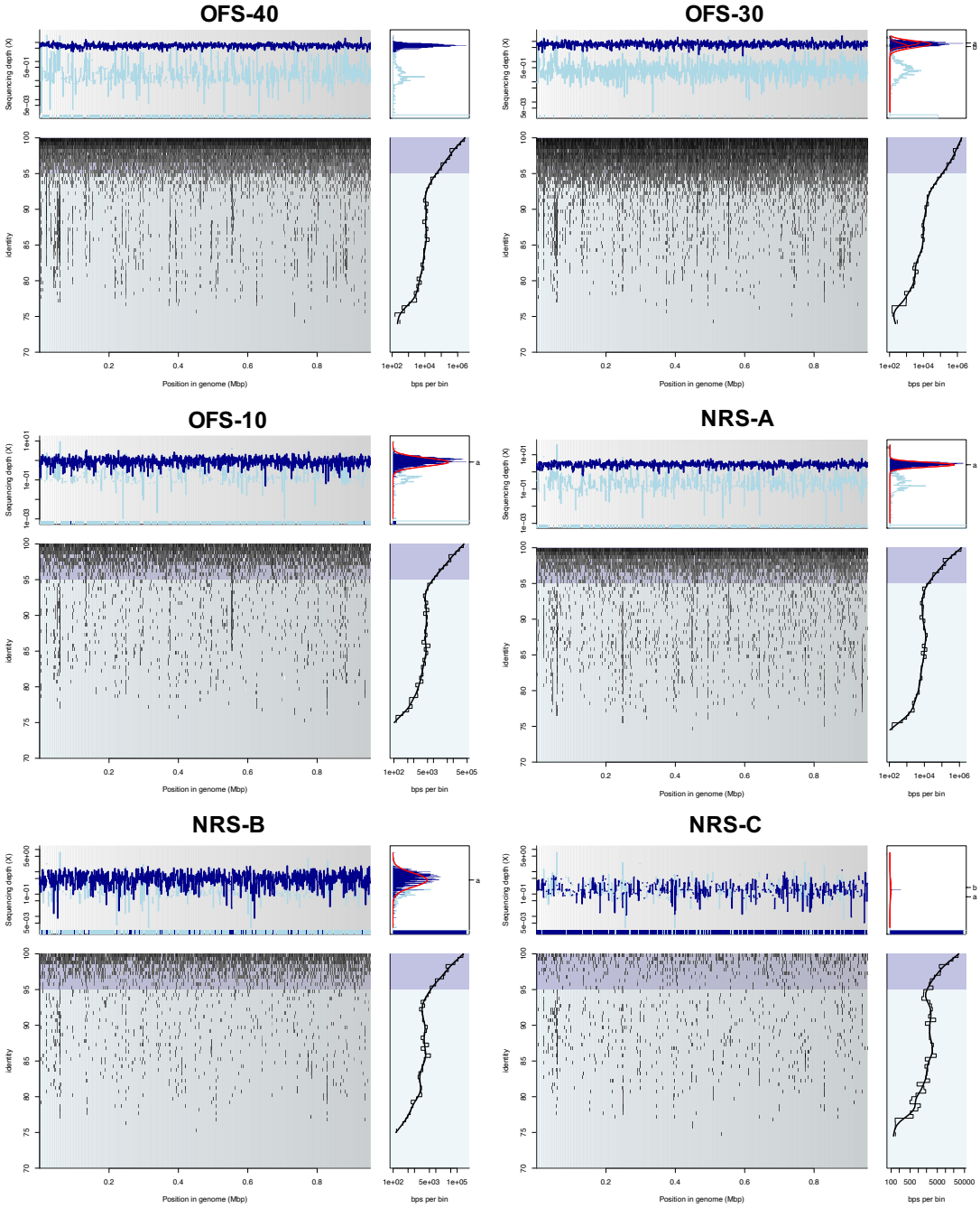


Figure 4.3. Read recruitment plots for MAG-based population LMS_bin035 (*Synechococcales*) in nearshore (NRS) and offshore (OFS) southern Lake Michigan. The coverage histogram (top left) in each plot shows coverage for the MAG in the corresponding Lake Michigan metagenome from reads that match at ≥ 95% nucleotide identity and ≥ 70 bp in length (dark blue) as well as reads that match at ≥ 70 bp in length and < 95% nucleotide identity (light blue). The recruitment plots (bottom left) show the individual reads mapping to the MAG at each position in the genome. The consistently high coverage of the MAG in offshore metagenomes at high identity (dark blue) in comparison to nearshore metagenomes can be seen.

The overall results of this study highlight that the microbial community composition and metabolic potential are not significantly different between nearshore and offshore Lake Michigan at least in the summer season. The same is true for certain water chemistry parameters. However, there are specific differences in the bulk DOM composition that also reflect in certain trends in the microbial community dynamics between nearshore and offshore. However, to more conclusively validate these differences such as with the use of genome-resolved metagenomics and metatranscriptomics, continued monitoring of the sampled sites across different seasons is necessary. In addition, the use of techniques such as fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) to characterize the DOM pool in Lake Michigan that provide a molecular resolution at a similar scale to the microbial omics data will be very valuable in mechanistically linking the carbon chemistry to microbial community function (26).

D. Acknowledgements

This work was supported by funds from Illinois-Indiana Sea Grant Cooperative Science and Monitoring Initiative (CSMI 2015), as well as Elmer Hadley Research Grants (2017, 2018) from the Department of Biological Sciences, University of Illinois at Chicago. We thank the crew of R/V *Lake Guardian* and members of the Poretsky Lab for assistance with water sampling. We thank the personnel of University of Illinois at Chicago Sequencing Core for facilitating sample sequencing. We are very grateful to Karl Rockne Lab (University of Illinois at Chicago) for measuring nutrients, and Laudong Guo Lab (School of Freshwater Sciences, University of Wisconsin-Milwaukee) for measuring DOC concentration and spectrofluorometric characterization of Lake Michigan DOM samples.

E. References

1. Weinke AD, Kendall ST, Kroll DJ, Strickler E a., Weinert ME, Holcomb TM, Defore A a., Dila DK, Snider MJ, Gereaux LC, Biddanda B a. 2014. Systematically variable planktonic carbon metabolism along a land-to-lake gradient in a Great Lakes coastal zone. J Plankton Res 36:1528–1542.

2. Zhou Z, Guo L, Minor EC. 2016. Characterization of bulk and chromophoric dissolved organic matter in the Laurentian Great Lakes during summer 2013. J Great Lakes Res 42:789–801.

3. Biddanda B a., Cotner JB. 2002. Love handles in aquatic ecosystems: The role of dissolved organic carbon drawdown, resuspended sediments, and terrigenous inputs in the carbon balance of Lake Michigan. Ecosystems 5:431–445.

4. Chaudhary A, Kauser I, Ray A, Poretsky R. 2018. Taxon-Driven Functional Shifts Associated with Storm Flow in 3.

5. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676.

6. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38.

7. Luo C, Rodriguez-R LM, Konstantinidis KT. 2014. MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res 42.

8. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60.

9. Wu CH, Apweiler R, Bairoch A, Natale D a, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187-91.

10. Overbeek R, Begley T, Butler RM, Choudhuri J V., Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov D a., Rül;ckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33:5691–5702.

11. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.

12. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

13. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282–W288.

14. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection : a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J Prepr 4:e1900v1.

15. Izadmehr M, Rockne K. 2018. "pocket Wetlands" for Nutrient Removal in Tile-Drained Agriculture. World Environ Water Resour Congr 2018 Watershed Manag Irrig Drainage, Water Resour Plan Manag - Sel Pap from World Environ Water Resour Congr 2018 404–414.

16. Guo L, Santschi PH, Warnken KW. 1995. Dynamics of dissolved organic carbon (DOC) in oceanic environments. Limnol Oceanogr 40:1392–1403.

17. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2015. vegan: Community Ecology Package. R package version 2.2-1. http://CRANR-project.org/package=vegan.

18. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21.

19. Coble PG. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Mar Chem 51:325–346.

20. Biddanda B a., Cotner JB. 2003. Enhancement of dissolved organic matter bioavailability by sunlight and its role in the carbon cycle of Lakes Superior and Michigan. J Great Lakes Res 29:228–241.

21. Hahn MW, Scheuerl T, Jezberová J, Koll U, Jezbera J, Šimek K, Vannini C, Petroni G, Wu QL. 2012. The passive yet successful way of planktonic life: Genomic and experimental analysis of the ecology of a free-living polynucleobacter population. PLoS One 7.

22. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria.Microbiology and molecular biology reviews : MMBR.

23. Carrick HJ, Butts E, Daniels D, Fehringer M, Frazier C, Fahnenstiel GL, Pothoven S,

Vanderploeg H a. 2015. Variation in the abundance of pico, nano, and microplankton in Lake Michigan: Historic and basin-wide comparisons. J Great Lakes Res 41:66–74.

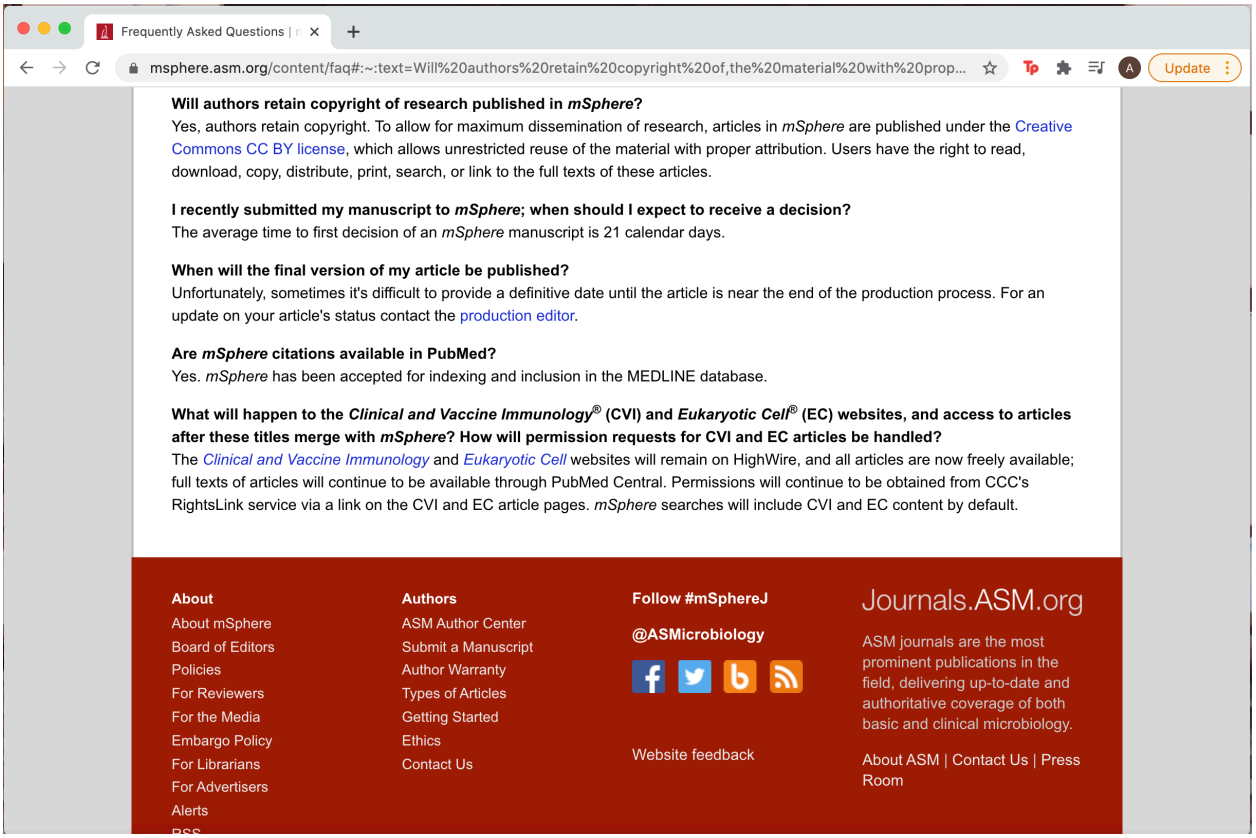
24. Eckert EM, Salcher MM, Posch T, Eugster B, Pernthaler J. 2012. Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom. Environ Microbiol 14:794–806.

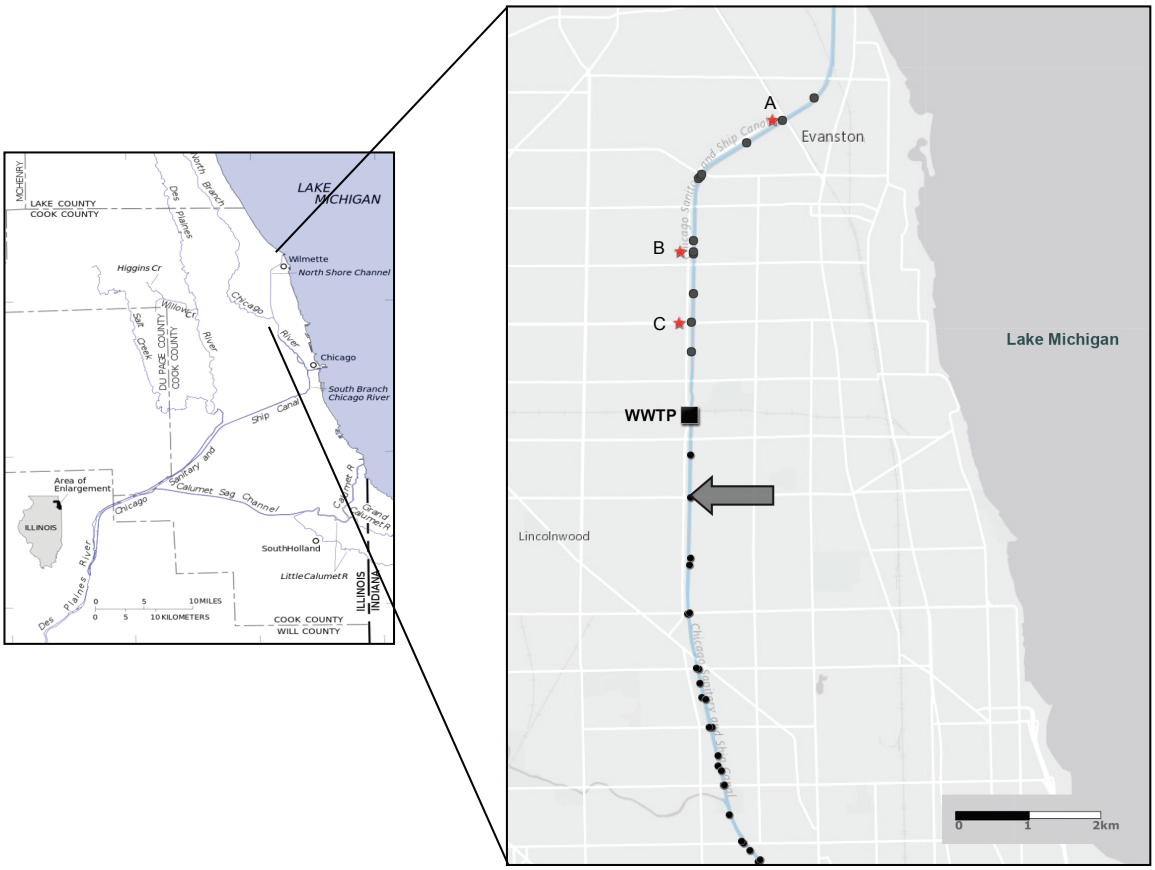
25. Paver SF, Youngblut ND, Whitaker RJ, Kent AD. 2015. Phytoplankton succession affects the composition of Polynucleobacter subtypes in humic lakes. Environ Microbiol 17:816–828.

26. Minor EC, Steinbring CJ, Longnecker K, Kujawinski EB. 2012. Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. Org Geochem 43:1–11.

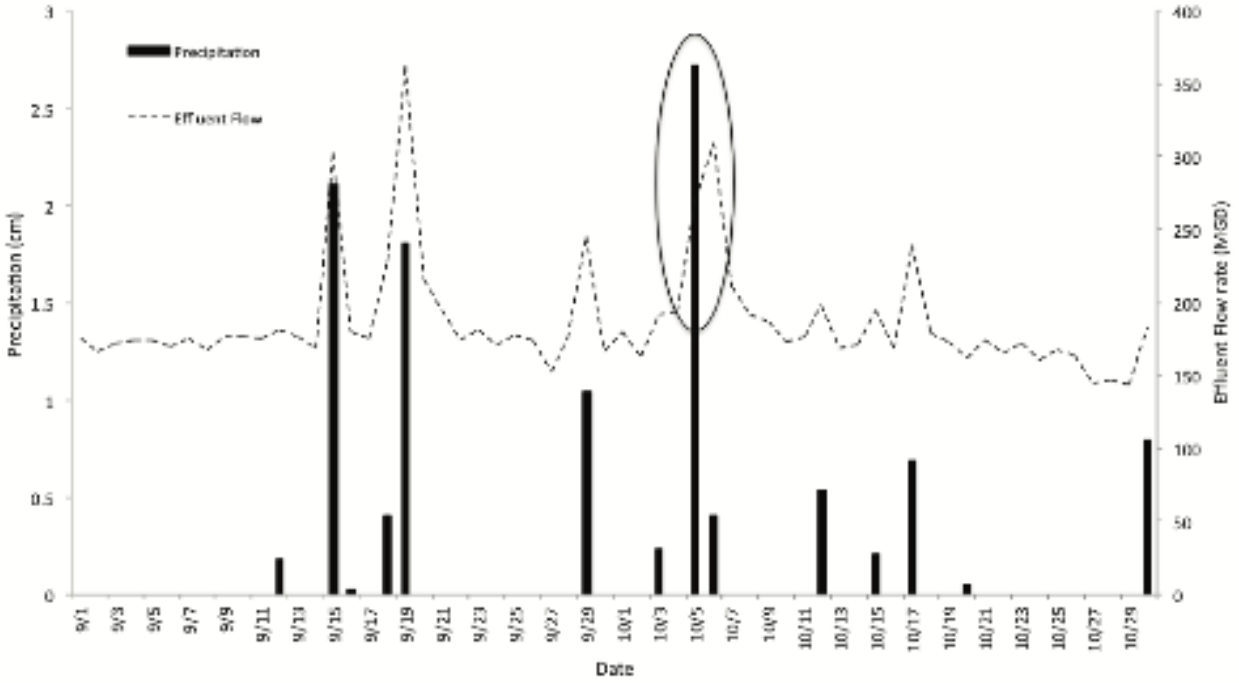
V. APPENDICES

A. Supplementary Materials – Chapter 2





Supplementary Figure S2.1. Map of the Chicago Area Waterway System (left panel) and the North Shore Channel (NSC) (right panel). Our study site at NSC is highlighted with an arrow. The point designated WWTP on the right panel represents the O'Brien Water Reclamation Plant. Black dots along the stream represent locations for monitored CSO outfalls. CSO outfalls marked with red stars (locations A, B, and C) recorded CSO events in the evening of 5 October 2013 with durations of 56, 50, and 5 min, respectively (http://www.mwrd.org/irj/portal/anonymous/overview).



Supplementary Figure S2.2 O'Brien Water Reclamation Plant effluent flow rate (million gallons per day [MGD]) and rain gauge data for the months of September and October 2013 (http://www.mwrd.org/irj/portal/anonymous/overview). The circled region of the plot corresponds to data around the rain event (5 October 2013), which is the focus of this study. No data were available for 17 September 2013 as the rain gauge was out of service.

Supplementary Table S2.1. Water chemistry and environmental characteristics for North Shore Channel sampled time points.

Sample ID	Sampling Date	Weather[a]	SRP[b,c] (mg P/L)	NH_3 /NH_4^{+c} (mg N/L)	NO_3^{-c} (mg N/L)	TDS[d] (ppm)	pH	C[d] (□S/cm)	T[d] (°C)
Before Rain Oct. 2013	10/05/2013	Baseflow	0.80 ± 0.09	0.41 ± 0.03	5.74 ± 0.56	286	7.6	566	22.7
After Rain Oct. 2013	10/06/2013	Stormflow	0.49 ± 0.02	0.26 ± 0.01	6.87 ± 0.06	308	7.5	604	18.5
July 2014	07/08/2014	Stormflow	0.56 ± 0.00	0.35 ± 0.00	5.11 ± 0.17	413	7.0	813	24
Oct. 2014	10/25/2014	Baseflow	1.82 ± 0.10	1.04 ± 0.00	9.62 ± 0.08	471	7.1	923	20.5
July 2015	07/23/2015	Baseflow	NA[e]	NA[e]	NA[e]	324	7.5	675	24.5
Effluent. Oct. 2013	10/05/2013 (WWTP Effluent)	Baseflow	1.5[f]	0.3[f]	10.93[f]	360	7.4	707	23.2

[a] Weather : 'Baseflow' condition represents no rainfall event (<2.5 mm precipitation) for at least 72 h prior to sample collection, and 'stormflow' represents sample collection <24 h after rainfall (>10 mm precipitation)
[b] SRP: Soluble Reactive Phosphate
[c] The concentrations for SRP, NH_3 and NO_3^- are mean values for duplicate samples with their standard errors. NH_3 or NH_4^+ was measured as per the methodology (Hach kits/AutoAnalyzer 3)
[d] TDS: Total Dissolved Solids; C: Conductivity; T: Temperature
[e] NA: Data not available
[f] MWRD data for the O'Brien Water Reclamation Plant effluent for the date sampled

October 2013 rain-associated filtrate samples were analyzed for nitrate, soluble reactive phosphate (SRP) and ammonium (NH_4^+) using an AutoAnalyzer 3 (Seal Analytical, Inc., Mequon, WI, USA) as described previously (1). Briefly, nitrate was measured using the cadmium reduction technique (2), SRP was measured using the antimonyl tartrate technique (3), and ammonium was measured using the phenol hypochlorite technique (4). Nitrate was calculated as the difference between nitrate+nitrite (NO_x^-) and nitrite (NO_2^-), which were measured with cadmium reduction and without cadmium reduction techniques, respectively. Samples from remaining time points were analyzed using Hach kits for nitrate (Nitrate TNTplus 835), ammonia (Ammonia TNTplus 830) and SRP (Phosphorus TNTplus 843) (Hach, Loveland, CO, USA) following the manufacturer's instructions.

References

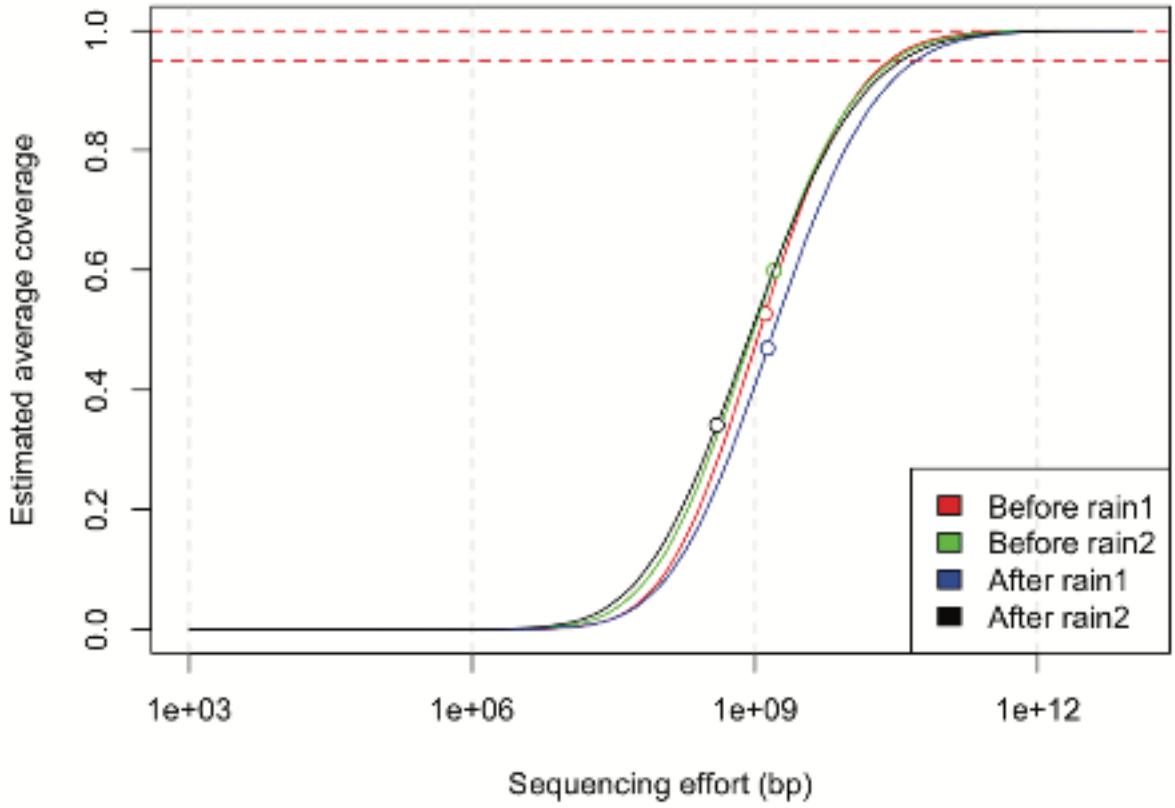
1. A. McCormick, T. J. Hoellein, S. A. Mason, J. Schluep, and J. J. Kelly, Environ Sci Technol 48:11863–11871, 2014, https://doi.org/10.1021/es503610r).
2. APHA, Standard Methods for the Examination of Water and Wastewater, 20th ed., 1998.
3. J. Murphy and J. P. Riley, Anal Chim Acta 27:31–36, 1962.
4. L. Solarzano, Limnol Oceanogr 14:799–801, 1969.

Supplementary Table S2.2. Sequencing statistics and diversity estimates for the 16S rRNA gene amplicon libraries used in the study.

16S rRNA gene library[a]	No of sequences after quality filtering (forward read only)	Median sequence length (bp)	Goods coverage estimate[b]	OTU richness (# OTUs)[b]
AfterRain.Oct2013.A	141,420	241	0.77	5,297
AfterRain.Oct2013.B1	19,313	206	0.80	5,018
AfterRain.Oct2013.B2	59,679	249	0.79	5,011
July2014.A	64,038	247	0.85	4,025
July2014.B	73,386	252	0.85	3,899
BeforeRain.Oct2013.A	34,150	289	0.87	3,695
BeforeRain.Oct2013.B1	24,075	197	0.85	3,988
BeforeRain.Oct2013.B2	28,635	290	0.88	3,503
Oct2014.A	38,619	289	0.89	3,211
Oct2014.B	35,751	293	0.90	3,124
July2015.A	54,255	256	0.89	3,277
July2015.B	66,487	253	0.88	3,428
Effluent.Oct2013	33,272	292	0.85	4,100

[a] Letters ‘A’ and ‘B’ following a sample ID represent biological replicate libraries for a sampled time point, and numbers ‘1’ and ‘2’ following these letters represent sequencing replicates

[b] Goods coverage estimate and OTU richness were calculated after subsampling each library to the smallest library size

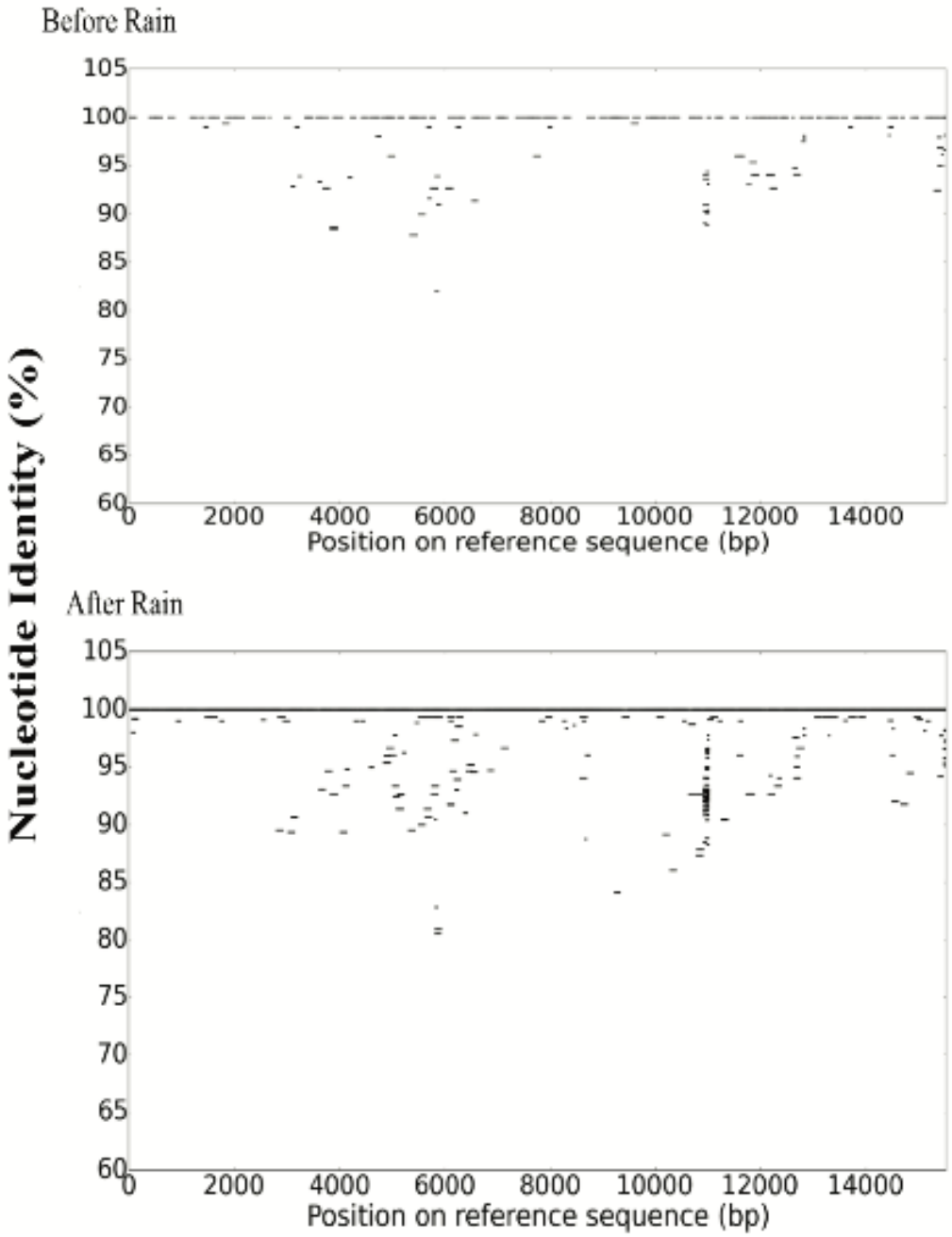


Supplementary Figure S2.3. Community coverage estimates based on metagenomic reads generated using Nonpareil for the before- and after-rain metagenomes. Sample numbers 1 and 2 for each time point represent biological replicate libraries.

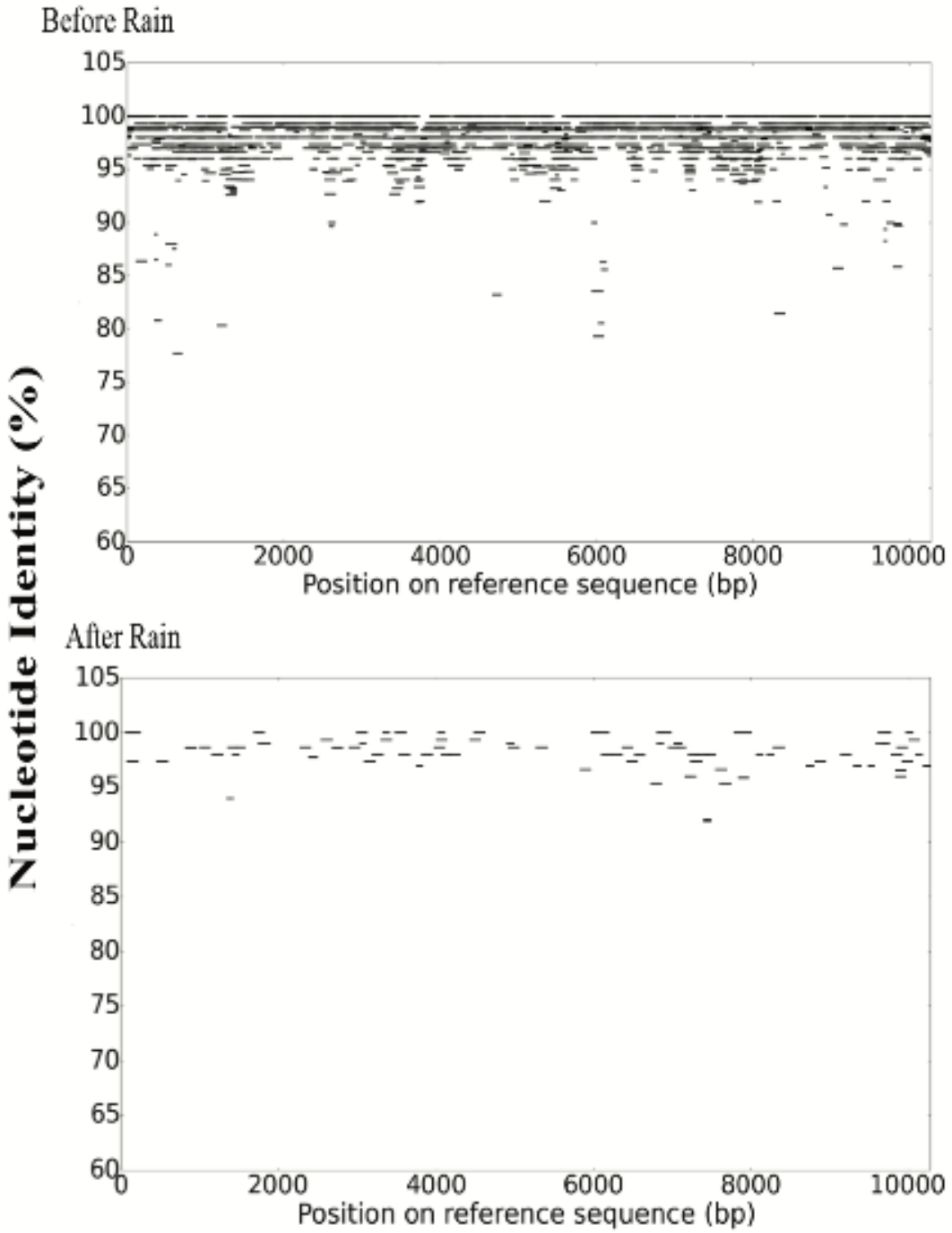
Supplementary Table S2.3: Rare species in before rain microbiome that were in the abundant fraction after rain.

Species	Relative abundance in after rain microbiome[a] (%)
Francisella tularensis	0.78
Candidatus Nitrospira defluvii	0.41
Simkania negevensis	0.32
Legionella longbeachae	0.29
Legionella drancourtii	0.25
Parachlamydia acanthamoebae	0.23
Chlamydia psittaci	0.16
Micavibrio aeruginosavorus	0.15
Chlamydia trachomatis	0.14
Arcobacter sp. L	0.14
Fluoribacter dumoffii	0.13
Neisseria meningitidis	0.11
Rickettsia endosymbiont of Ixodes scapularis	0.1
Enterococcus faecalis	0.1

[a] Abundances for each taxa are relative to the total number of sequences characterized by MyTaxa in the after rain library



Supplementary Figure S2.4. Reads from before-rain (top) and after-rain (bottom) data sets were mapped to the longest contig attributed to *Legionella pneumophila* from the after-rain metagenome. Reads for biological replicate libraries (n = 2) were pooled for both the before- and after-rain time points.



Supplementary Figure S2.5. Reads from before-rain (top) and after-rain (bottom) data sets were mapped to the longest contig attributed to *Actinobacterium* SCGC AAA027-L06 from the before-rain metagenome. Reads for biological replicate libraries (n = 2) were pooled for both the before- and after-rain time points.

Supplementary Figure S2.6. Heat map showing the relative abundance (percentage of total predicted genes) at the level 4 depth of Gene Ontology (GO) terms for the before- and after-rain microbiomes. GO terms that had a higher relative abundance (>100%) in one of the two groups (before versus after rain) compared to the other are shown, and terms that had less than a total of 75 gene counts across all the samples have been excluded from the plot. Samples numbered 1 and 2 for each time point represent biological replicates.

Supplementary Table S2.4: Sequencing statistics for the metagenomes used in the study.

Metagenome Library	No of paired-end reads after quality filtering (million)	Read Length (bp)	Estimated community coverage[a]	No of contigs (contig size > 500 bp)	Contig N50 (bp)	Assembly efficiency (for contigs > 500 bp)[b]
Before Rain1	8.74	150	52%	98708	1283	29-59%
Before Rain2	16.21	100	61%	68202	1403	23-45%
After Rain1	9.26	150	47%	97165	1224	24-49%
After Rain2	4.06	100	36%	5141	1068	6-13%

[a] Community coverage estimates were obtained from quality filtered reads using Nonpareil[c]
[b] Assembly efficiency is an estimate of the extent a metagenomic library is represented by its assembled contigs (size >500 bp). This was calculated using the following formula:

Assembly efficiency = $[\{\sum_i (contig_i\ length \times contig_i\ coverage)\} \div metagenome\ size] \times 100$

Where $contig_i$ coverage is the average number of reads that map to $contig_i$ from the corresponding metagenome (sum of lengths of reads that map to $contig_i$/$contig_i$ length) and metagenome size is the total number of base pairs from the library used in the assembly. Given the sequencing insert length for the paired end reads (600 bp for 2x150bp libraries and 500 bp for 2x100bp libraries) for these assembled libraries and the variable length of the contigs (> 500 bp), we estimate that some contigs would represent both the paired-end reads while some would only represent one, so the actual number of bases used in the assembly process should be somewhere between the following numbers – (total number of basepairs in the paired reads file/2) - (total number of basepairs in the paired end reads file). This leads to the estimated range for our calculated assembly efficiency.

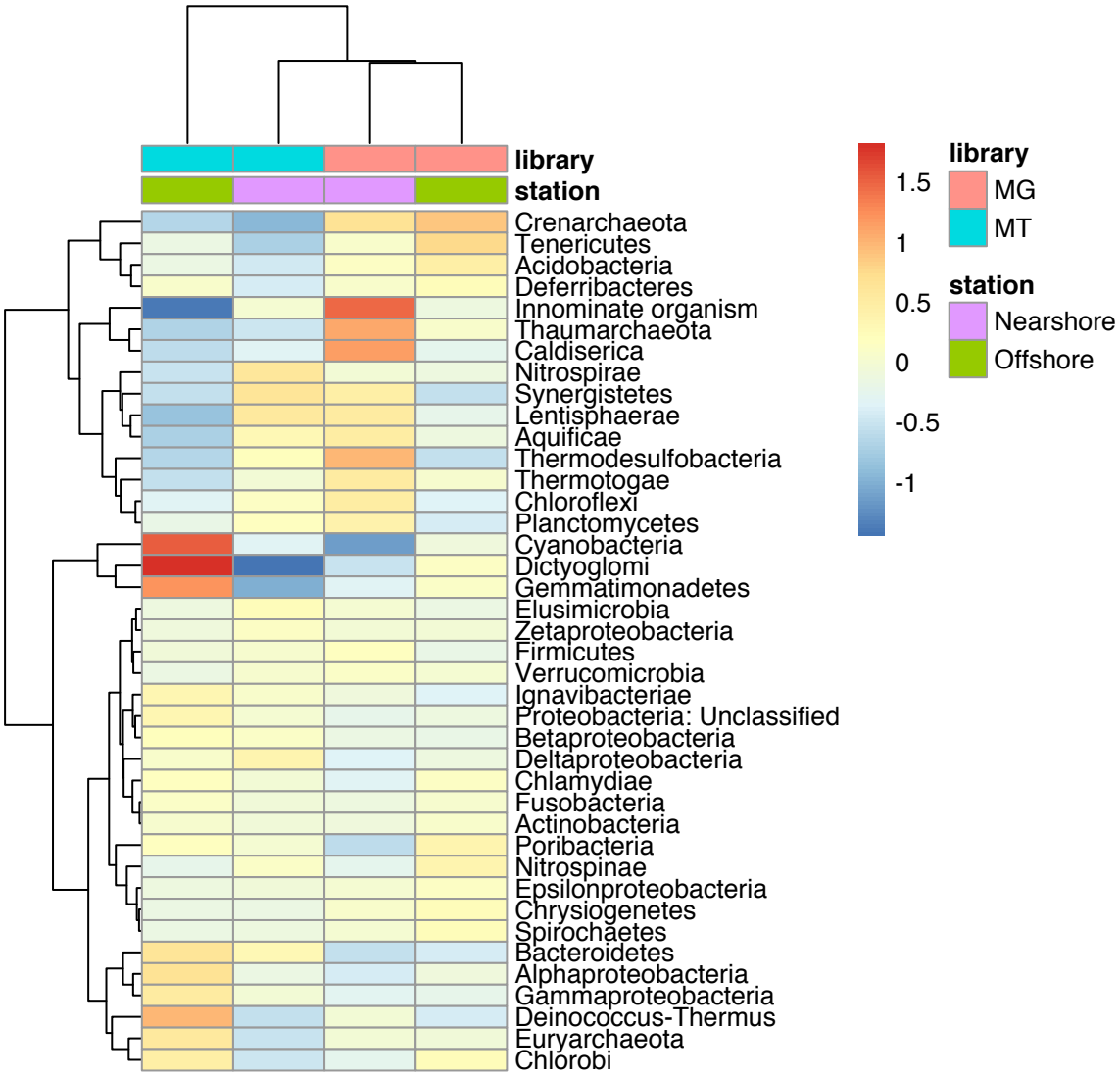
[c] L.M. Rodriguez-R and K. T. Konstantinidis, Bioinformatics 30:629–635, 2014, https://doi.org/10.1093/bioinformatics/btt584

B. Supplementary Materials – Chapter 3

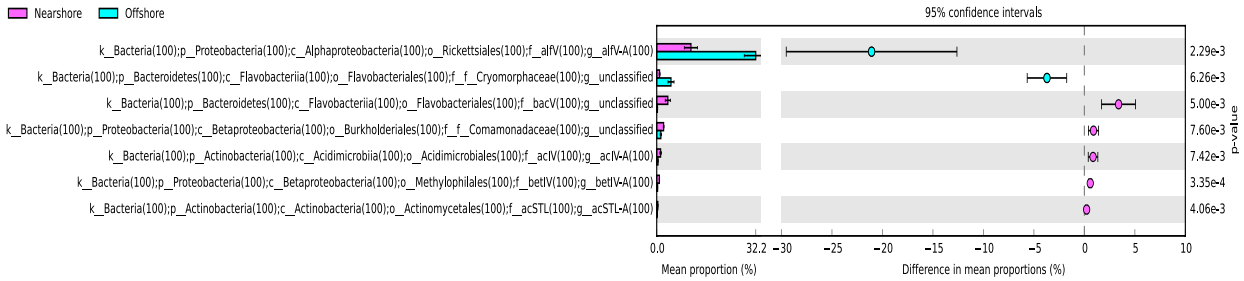
Supplementary Table S3.1. Summary statistics for the 17 good quality (completeness ≥ 50%, contamination ≤ 10%) metagenome assembled genomes (MAGs) using the metagenomes obtained from Lake Michigan. MAG completion, contamination, overall quality and likely taxonomy were determined using MiGA webserver.

bin ID	Likely taxonomic classification	MAG comlete ness (%)	MAG contamin ation (%)	MAG overall quality (out of 100)	Transcript read count (cDNA hits per library per bin)							
					Nears hore Contr ol 2h	Nears hore Contr ol 19h	Nears hore t-DOM 2h	Nears hore t-DOM 19h	Offsh ore Cont rol 2h	Offsh ore Cont rol 19h	Offsh ore t-DOM 2h	Offsh ore t-DOM 19h
bin_ 002	Clostridia (Firmicutes)	83.8	0	83.8 (Excelle nt)	20952	23917	27346	16495	457	371	792	201
bin_ 004	Limnohabitans (Betaproteobac teria)	57.7	7.2	41.5 (Interme diate)	80002	10978 9	62341	6070	8944 4	8213 6	7888 3	1564 6
bin_ 008	Rhodoferax (Betaproteobac teria)	69.4	3.6	51.4 (High)	79757	10675 9	68710	7569	4415 0	3617 8	3872 9	8428
bin_ 011	Candidatus Planktophila (acI-A Actinobacteria)	71.2	5.4	44.2 (Interme diate)	32612	31824	23206	9399	4013	3692	4390	1532
bin_ 018	Bacilli (Firmicutes)	73.9	2.7	60.4 (High)	30597	31068	35021	2757	323	336	210	69
bin_ 020	Flavobacteriia (Bacteroidetes)	93.7	2.7	80.2 (Excelle nt)	49334	51802	53016	45530	49	101	138	105
bin_ 035	Bacteroidetes	71.2	0.9	66.7 (High)	24933	15424	24578	13228	1949	1649	3759	870
bin_ 040	Candidatus Nanopelagicus abundans (acI-B1 Actinobacteria)	59.5	1.8	50.5 (High)	7537	7017	4652	1796	6863	5431	6697	2352
bin_ 047	Flavobacteriale s (Bacteroidetes)	55.9	2.7	42.4 (Interme diate)	11554	9357	12169	5506	7466 1	5107 7	1495 07	2425 8
bin_ 062	Chitinophagac eae (Bacteroidetes)	45.9	5.4	18.9 (Low)	11569	8041	7615	8341	2242	1152	2502	3220
bin_ 093	Acidimicrobial es (Actinobacteria)	79.3	4.5	56.8 (High)	4108	6924	3390	1231	5905	5334	5352	2570
bin_ 100	Chitinophagac eae (Bacteroidetes)	62.2	0	62.2 (High)	7907	6712	6083	7926	5614	4377	7904	9571
bin_ 104	Sphingomonas (Alphaproteob acteria)	72.1	0	72.1 (High)	5140	16399	6415	6932	1379 8	1459 8	2912 4	2499 1
bin_ 105	Rhodoluna (Actinobacteria)	65.8	1.8	56.8 (High)	7765	9771	6185	7709	1890 4	1159 3	2285 7	1964 8

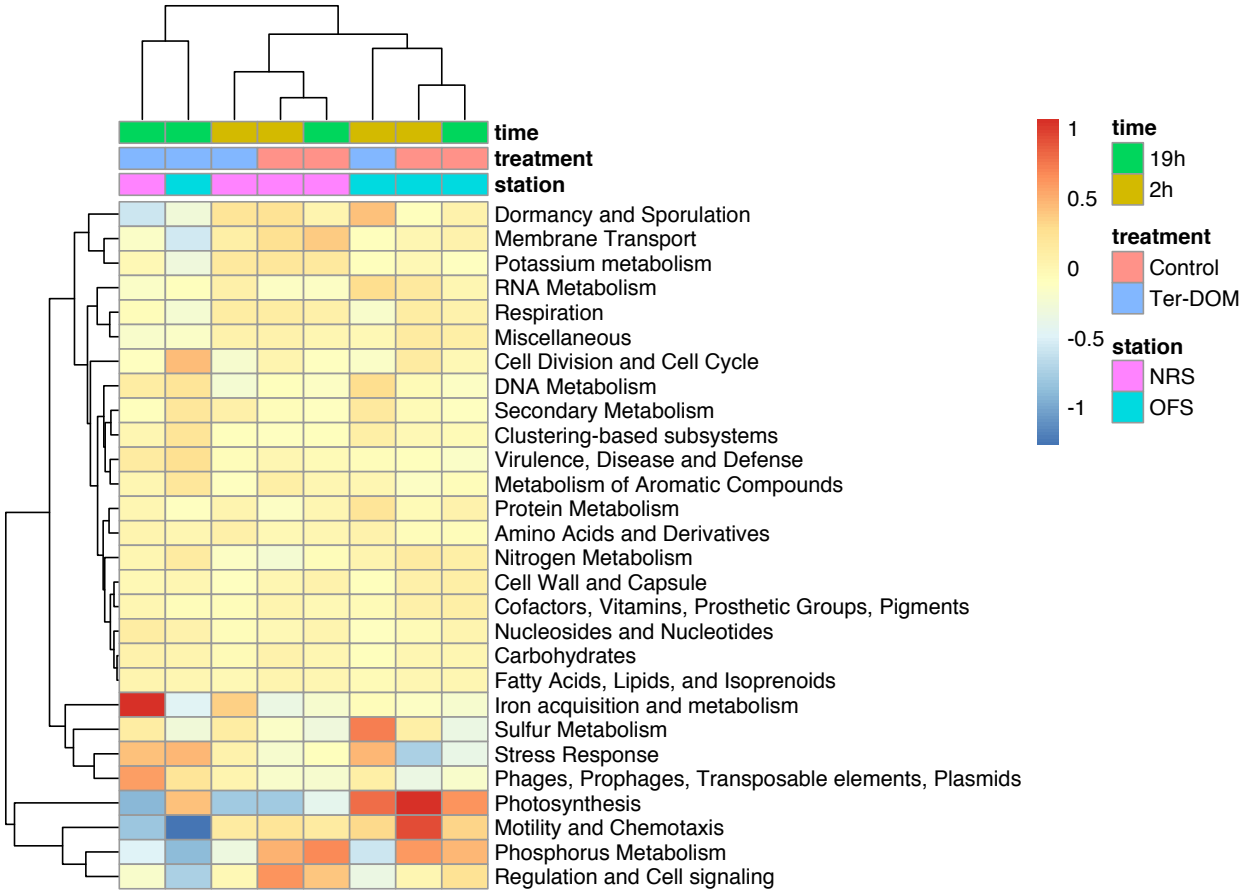
bin_119	Bacilli (Firmicutes)	58.6	0	58.6 (High)	907	1030	742	324	108	113	10	3
bin_130	Sinobacteraceae (Gammaproteobacteria)	71.2	7.2	35.2 (Intermediate)	2712	4172	1503	1122	3697	2540	3538	3045
bin_158	Rhodospirillaceae (Alphaproteobacteria)	55	2.7	41.5 (Intermediate)	23	10	23	9	24	24	9	1



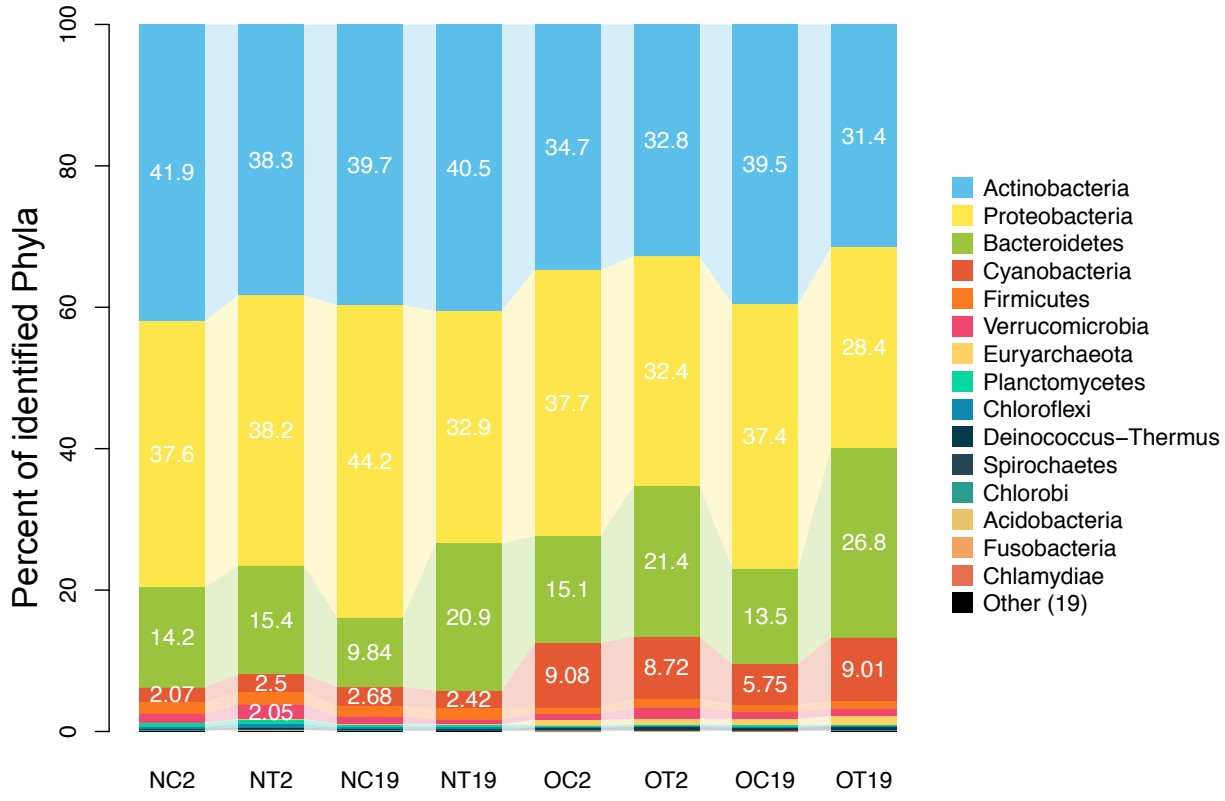
Supplementary Figure S3.1. Clustered heatmap of the taxonomic composition of the active (mRNA-based) and total (DNA-based) bacterial community in the nearshore and offshore surface-waters of southern Lake Michigan. The taxonomic profile is shown at the phylum level, with Proteobacteria subdivided into classes. Transcript abundances for individual phyla have been normalized using the 'rlog' (regularized-logarithm transformation) function of DESeq2 such that the data are approximately homoskedastic. All identified phyla are shown, regardless of abundance.



Supplementary Figure S3.2. Bar plots representing the relative abundance of significantly different genera (two-tailed t-test, p-value < 0.01) between nearshore and offshore bacterial communities based on 16S rRNA gene amplicon sequences. The relative abundance of each genus is an average for the nearshore/offshore triplicate control mesocosms (2h time-point) from which the sequences were obtained. Bar plots show only the abundant genera (abundance ≥ 1% of total sequences in a library) across the samples. Bioinformatics processing and taxonomic annotation of 16S amplicon datasets was performed using QIIME v1.8.0 and TaxAss, and the bar plots were generated using STAMP.



Supplementary Figure S3.3. Clustered heatmap of metatranscriptome-based functional profiles for the bacterial communities in the different mesocosms and time-points. Nearshore and offshore sampling sites across the Lake Michigan transect are labeled as NRS and OFS, respectively. Functional processes were annotated at the broadest level of the SEED Subsystems database. Transcript abundance for individual SEED categories were normalized using the 'rlog' (regularized-logarithm transformation) function of DESeq2 such that the data are approximately homoskedastic.



Supplementary Figure S3.4. Taxonomic composition of the active (mRNA-based) bacterial communities in all the mesocosms and time-points. The taxonomic profile is shown at the phylum level, and transcript counts were normalized based on the RPKM formula (Reads Per Kilobase of transcript, per Million mapped reads) before generating the plots. The first two letters of the x-axis labels refer to the sampling site (N for nearshore, O for offshore) and treatment (C for control, T for t-DOM), followed by numbers 2 or 19 that refer to the sampling time-points – 2h and 19h, respectively.

Supplementary Table S3.2. Differential expression of specific functional processes based on SEED annotations for MAGs bin004, bin008 and bin040 between the t-DOM and control mesocosms at 2h using DESeq2. The multiple sheets correspond to data for the different MAGs and different mesocosms being tested (nearshore/offshore lake-water). A positive value in the column 'log2FoldChange' for a specific functional process indicates a higher expression for that function in the t-DOM sample as compared to the control, and a negative value indicates the opposite. Only those functional processes that are different between the control and t-DOM mesocosms by a fold-change of 0.5 or more are shown. Functional processes that are significantly differentially expressed (Wald test, *P* value < 0.05) between the control and t-DOM mesocosm are highlighted in yellow.

MAG: bin_004
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
RNA Metabolism|Rrf2_family_transcriptional_regulators	118.191288	1.916752716	0.012031233
Iron acquisition and metabolism|Transport_of_Iron	103.8675864	1.775623056	0.019513232
Phosphorus Metabolism|Phosphate_metabolism	191.1629394	-1.734830322	0.021117742
Carbohydrates|Tricarballylate_Utilization	22.5129856	1.442962825	0.05983049
Stress Response|Hfl_operon	85.49200016	1.340095501	0.072806345
Amino Acids and Derivatives|Alanine_biosynthesis	499.3350527	1.288005988	0.073849492
Protein Metabolism|Protein_chaperones	260.359371	1.264409722	0.081057892
Protein Metabolism|GroEL_GroES	5802.349169	1.144728834	0.104565789
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	122.4896561	-1.18785102	0.105140501
Stress Response|Oxidative_stress	253.9357128	1.106948854	0.122171388
Nucleosides and Nucleotides|Pyrimidine_utilization	8.478524684	1.105956038	0.124751988
Secondary Metabolism|Auxin_biosynthesis	149.2834038	1.092214303	0.131258417
Amino Acids and Derivatives|Tryptophan_synthesis	16.7030206	1.12463796	0.141985432
Stress Response|Heat_shock_dnaK_gene_cluster_extended	2076.36831	0.99738843	0.152457845
Metabolism of Aromatic Compounds|p-Hydroxybenzoate_degradation	23.4813131	-0.97181358	0.205364842
Amino Acids and Derivatives|Aromatic_amino_acid_degradation	22.22374492	0.949974274	0.215832961
Amino Acids and Derivatives|Glycine_cleavage_system	80.4944242	0.850171522	0.242869021
RNA Metabolism|Polyadenylation_bacterial	16.66780869	0.876423149	0.253755251
RNA Metabolism|Queuosine-Archaeosine_Biosynthesis	72.89116697	0.831867772	0.254763671
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	78.73885902	-0.805526628	0.268245719
Amino Acids and Derivatives|Chorismate:_Intermediate_for_synthesis_of_PAPA_antibiotics	11.73311114	0.808530999	0.287588875

Membrane Transport|ABC_transporter_alkylphosphonate_(TC_3.A.1.9.1)	3.219374546	-0.61689911	0.301575881
Photosynthesis|Photosystem_II-type_photosynthetic_reaction_center	91.17117807	-0.730600135	0.310225648
Membrane Transport|Type_IV_pilus	13.78043214	0.772026889	0.313393204
Respiration|Biogenesis_of_c-type_cytochromes	52.53113811	0.71777459	0.330904915
Potassium metabolism|Potassium_homeostasis	44.25130919	-0.718227447	0.335300258
Regulation and Cell signaling|Murein_hydrolase_regulation_and_cell_death	10.93581291	-0.711494926	0.347938284
Protein Metabolism|Programmed_frameshift	39.38703547	-0.680557897	0.363436116
DNA Metabolism|RuvABC_plus_a_hypothetical	4.645456865	-0.613256382	0.36509535
DNA Metabolism|DNA_Repair_Base_Excision	42.25932119	0.659700506	0.375845772
Respiration|Terminal_cytochrome_C_oxidases	129.9621262	0.62408929	0.375968043
Protein Metabolism|Ribosomal_protein_S12p_Asp_methylthiotransferase	21.1019941	-0.665135006	0.385600742
Cofactors, Vitamins, Prosthetic Groups, Pigments|Chlorophyll_Biosynthesis	167.3043558	-0.58754168	0.399688947
Carbohydrates|L-rhamnose_utilization	75.0164572	0.583261968	0.418421508
Stress Response|Glutaredoxins	73.37155802	0.572184158	0.427617827
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	56.09760147	-0.562874897	0.441940507
Amino Acids and Derivatives|Cysteine_Biosynthesis	35.05848578	0.563802673	0.452300192
Respiration|Biogenesis_of_cytochrome_c_oxidases	14.36645891	0.556599392	0.468460846
Secondary Metabolism|Steroid_sulfates	15.83529855	-0.55164456	0.472740369
RNA Metabolism|tRNA_nucleotidyltransferase	7.095199684	-0.503290829	0.492491463
Miscellaneous|YbbK	40.98163192	0.507225656	0.494765191

MAG: bin_004
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	185.0120561	2.13808779	0.008277638
RNA Metabolism|Rrf2_family_transcriptional_regulators	238.6019217	2.133087289	0.011448239
Membrane Transport|ABC_transporter_alkylphosphonate_(TC_3.A.1.9.1)	25.16175499	-2.208655424	0.012811756
Phosphorus Metabolism|Phosphate_metabolism	299.8148627	-2.173114259	0.013344364
Amino Acids and Derivatives|Glycine_cleavage_system	128.7567327	1.719348356	0.014901973
RNA Metabolism|Transcription_factors_bacterial	132.1599957	-1.545428953	0.02469379
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	343.040767	-1.957277661	0.026645973

Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	121.114708	-1.412555685	0.032219992
Protein Metabolism|Protein_chaperones	759.1378372	1.885795729	0.032751226
Fatty Acids, Lipids, and Isoprenoids|Glycerolipid_and_Glycerophospholipid_Metabolism_in_Bacteria	46.26113317	-1.720993694	0.035965688
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	148.2126426	-1.438733423	0.03770294
Respiration|Respiratory_dehydrogenases_1	549.602559	1.806737444	0.039427666
DNA Metabolism|DNA_repair,_bacterial	244.4275673	1.458353597	0.064608603
Iron acquisition and metabolism|Transport_of_Iron	77.25027084	1.170451768	0.070870926
Virulence, Disease and Defense|Resistance_to_fluoroquinolones	83.1546606	-1.134787	0.072427412
Nucleosides and Nucleotides|Pyrimidine_utilization	14.03772316	1.576597403	0.073743895
Stress Response|Glutaredoxins	59.22549363	1.279054101	0.078357103
RNA Metabolism|Queuosine-Archaeosine_Biosynthesis	51.21885622	1.282868065	0.091826812
Regulation and Cell signaling|Murein_hydrolase_regulation_and_cell_death	14.33559289	-1.474942981	0.093761808
Carbohydrates|Pyruvate_Alanine_Serine_Interconversions	155.0500663	1.079126896	0.098657526
Virulence, Disease and Defense|MLST	46.36898529	-1.290089208	0.099096768
DNA Metabolism|DNA_repair,_bacterial_UvrD_and_related_helicases	10.49576303	-1.44783363	0.102002264
Clustering-based subsystems|Bacterial_Cell_Division	128.496498	-0.972514383	0.109240683
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	120.1954122	-0.943622565	0.112866414
Miscellaneous|YbbK	34.36502973	1.25911904	0.121462366
Protein Metabolism|Proteasome_bacterial	860.6682023	1.347310568	0.126361563
Respiration|Anaerobic_respiratory_reductases	115.2753769	0.887618406	0.127295183
Nucleosides and Nucleotides|Purine_conversions	115.770107	-0.879600403	0.131638656
Carbohydrates|Tricarballylate_Utilization	15.24462459	1.268731874	0.144307128
Amino Acids and Derivatives|Alanine_biosynthesis	572.4925769	1.198478734	0.17549461
Respiration|Respiratory_Complex_I	385.4346994	-1.117281216	0.195948931
Carbohydrates|Entner-Doudoroff_Pathway	43.40565778	-0.989618687	0.196098546
Photosynthesis|Photosystem_II-type_photosynthetic_reaction_center	87.0886571	0.706194005	0.214870083
Respiration|F0F1-type_ATP_synthase	662.9227543	-1.081201699	0.219499484
Amino Acids and Derivatives|Histidine_Biosynthesis	54.41157341	-0.869977962	0.220620274
Respiration|Formate_hydrogenase	71.95188463	0.755375844	0.220681702

Clustering-based subsystems|CBSS-262719.3.peg.410	68.26954305	-0.779489686	0.220765591
Stress Response|Heat_shock_dnaK_gene_cluster_extended	5267.957067	0.97652663	0.220909697
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	31.84337478	0.980759376	0.221666041
Stress Response|Oxidative_stress	217.68056	0.856503678	0.223689786
Cell Wall and Capsule|Sialic_Acid_Metabolism	62.33604532	0.786630154	0.234115055
Protein Metabolism|GroEL_GroES	5188.171403	0.962537135	0.234362135
Carbohydrates|Glycolysis_and_Gluconeogenesis	147.9832529	-0.714141635	0.234751523
Protein Metabolism|Ribosome_biogenesis_bacterial	10.53171374	-0.972486359	0.266171333
DNA Metabolism|DNA_Repair_Base_Excision	53.52480695	0.778835606	0.268599286
Cofactors, Vitamins, Prosthetic Groups, Pigments|Ubiquinone_Biosynthesis	9.870915606	0.94800861	0.277929909
Regulation and Cell signaling|Orphan_regulatory_proteins	35.3236976	-0.816282806	0.296096109
DNA Metabolism|DNA_repair,_UvrABC_system	31.66704262	-0.811764017	0.307452941
Respiration|Biogenesis_of_cytochrome_c_oxidases	18.2798535	0.860734033	0.307869151
Amino Acids and Derivatives|Chorismate_Synthesis	205.7656127	0.685429354	0.310379364
Cell Wall and Capsule|Lipid_A_modifications	19.30358003	0.843413051	0.315497157
Protein Metabolism|Peptidyl-prolyl_cis-trans_isomerase	15.61097436	0.852257802	0.317366658
RNA Metabolism|tRNA_processing	65.95674958	0.611981788	0.33037056
Cofactors, Vitamins, Prosthetic Groups, Pigments|Chlorophyll_Biosynthesis	225.6187174	0.654973536	0.344944511
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	23.80078666	0.744336411	0.364507189
Membrane Transport|Type_IV_pilus	19.08445444	0.740403255	0.376305196
Amino Acids and Derivatives|Arginine_and_Ornithine_Degradation	22.34223031	-0.679081521	0.410928597
Respiration|Biogenesis_of_c-type_cytochromes	43.43476576	0.596728921	0.416846437
Stress Response|Hfl_operon	48.95569892	0.573952751	0.419864386
Protein Metabolism|Universal_GTPases	582.00388	-0.69983067	0.423622181
Amino Acids and Derivatives|Phenylalanine_and_Tyrosine_Branches_from_Chorismate	25.19428433	-0.64468333	0.427977945
Virulence, Disease and Defense|Copper_homeostasis	9.030363953	0.679778832	0.434946909
RNA Metabolism|RNA_polymerase_bacterial	1365.271008	-0.681512779	0.436208954
Protein Metabolism|Ribosome_LSU_bacterial	3495.049877	-0.680289616	0.43871252
Carbohydrates|Maltose_and_Maltodextrin_Utilization	5.337758277	0.65319939	0.460251178
Amino Acids and Derivatives|Aromatic_amino_acid_degradation	24.38626202	0.570090339	0.482334489

Protein Metabolism|Ribosome_SSU_bacterial	1861.613995	-0.564086958	0.523240566
RNA Metabolism|Polyadenylation_bacterial	17.07637344	-0.532766739	0.525441494
Amino Acids and Derivatives|Methionine_Degradation	3.692605676	0.515350182	0.561545136

MAG: bin_008
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
RNA Metabolism|Rrf2_family_transcriptional_regulators	59.9266218	1.657511481	0.010896054
Iron acquisition and metabolism|Transport_of_Iron	97.38076042	1.466755737	0.023373152
Cell Wall and Capsule|mycolic_acid_synthesis	124.2135762	1.308447722	0.040588728
Protein Metabolism|GroEL_GroES	3043.400321	1.052537087	0.083428415
Cofactors, Vitamins, Prosthetic Groups, Pigments|Chlorophyll_Biosynthesis	47.96365812	-1.103097577	0.088246327
Sulfur Metabolism|Thioredoxin-disulfide_reductase	698.4358328	1.026348359	0.092395614
Stress Response|Hfl_operon	67.75285972	1.055342171	0.098817736
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	147.9158967	-1.002365108	0.107692202
Stress Response|Oxidative_stress	258.3776548	0.971184747	0.113649347
Potassium metabolism|Glutathione-regulated_potassium-efflux_system_and_associated_functions	20.90723559	1.01220403	0.119229058
Stress Response|Heat_shock_dnaK_gene_cluster_extended	1193.613086	0.903753828	0.132299561
Potassium metabolism|Potassium_homeostasis	63.61613396	-0.928880775	0.144719887
Protein Metabolism|Protein_chaperones	382.591231	0.859120759	0.154797122
RNA Metabolism|Transcription_initiation,_bacterial_sigma_factors	254.0173222	0.813511821	0.179157525
Carbohydrates|Glycerol_fermenation_to_1,3-propanediol	8.720665112	-0.798410868	0.191227964
RNA Metabolism|RNA_polymerase_bacterial	935.7944486	0.756822272	0.201479705
Nitrogen Metabolism|Dissimilatory_nitrite_reductase	3.913118961	0.62805336	0.210974532
Membrane Transport|Choline_Transport	10.1741093	-0.777119295	0.213833712
Secondary Metabolism|Steroid_sulfates	15.87608264	-0.803709452	0.213893711
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	109.1201173	-0.768322616	0.213991162
Metabolism of Aromatic Compounds|Salicylate_and_gentisate_catabolism	8.608861713	0.747407884	0.223351937
Clustering-based subsystems|Putative_hemin_transporter	205.7182539	-0.718885037	0.233877789
Amino Acids and Derivatives|Chorismate:_Intermediate_for_synthesis_of_PAPA_antibiotics,_PABA,_anthranilate,_3-hydroxyanthranilate_and_more.	25.15576475	0.751090987	0.248545369

Secondary Metabolism|Auxin_biosynthesis	117.8407824	0.697641139	0.255350842
DNA Metabolism|DNA_structural_proteins,_bacterial	5.031152949	-0.607038367	0.277531601
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	121.1948844	0.661372367	0.279193811
Amino Acids and Derivatives|Glycine_cleavage_system	207.9543219	0.645904725	0.281732465
Nucleosides and Nucleotides|Pyrimidine_utilization	7.490827725	0.648098327	0.284735307
Phosphorus Metabolism|Phosphate_metabolism	180.1152756	-0.639746857	0.288280023
Amino Acids and Derivatives|Glutamine,_Glutamate,_Aspartate_and_Asparagine_Biosynthesis	302.8754076	0.605066416	0.307432239
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_and_NADP_cofactor_biosynthesis_global	13.64001466	0.629726598	0.328998967
Respiration|Biogenesis_of_c-type_cytochromes	76.13811463	0.594900938	0.33804668
DNA Metabolism|DNA_repair,_bacterial_UvrD_and_related_helicases	12.74558747	-0.605173636	0.346391835
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	154.2886904	-0.562680563	0.350144329
Protein Metabolism|Selenocysteine_metabolism	25.37937154	0.588848202	0.364857566
Cell Wall and Capsule|Capsular_heptose_biosynthesis	17.10592003	-0.587078223	0.366743639
Respiration|Biogenesis_of_cytochrome_c_oxidases	45.28037654	0.559474351	0.37884499
Amino Acids and Derivatives|Alanine_biosynthesis	186.0408557	0.516353402	0.386660808
Amino Acids and Derivatives|Tryptophan_synthesis	15.09345885	0.559685611	0.388250825
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	40.36102699	-0.545177532	0.393670911
RNA Metabolism|Polyadenylation_bacterial	204.1530063	-0.506303281	0.3944534 5
DNA Metabolism|DNA_repair,_UvrABC_system	22.47248317	-0.548017242	0.399904677
Membrane Transport|Na(+)_H(+)_antiporter	23.25510697	0.544683055	0.402491323
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	38.01315562	-0.515851043	0.420490 06
Carbohydrates|Pyruvate_metabolism_II:_acetyl-CoA,_acetogenesis_from_pyruvate	18.11215062	0.523108027	0.421853051
Protein Metabolism|Ribosome_biogenesis_bacterial	30.2987211	-0.500195408	0.438741811

MAG: bin_008
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Protein Metabolism|Protein_chaperones	439.6829897	3.442735664	0.006526229
Stress Response|Heat_shock_dnaK_gene_cluster_extended	2206.949023	3.269308498	0.008837258
Respiration|Respiratory_dehydrogenases_1	164.6654699	3.322877275	0.008860022

Protein Metabolism|GroEL_GroES	1260.672803	2.973619962	0.01563203
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	253.7771914	2.19447031	0.06387272
Amino Acids and Derivatives|Glycine_cleavage_system	226.1412565	2.083583768	0.076812375
RNA Metabolism|Polyadenylation_bacterial	111.442797	-2.08147412	0.081207946
Sulfur Metabolism|Thioredoxin-disulfide_reductase	1876.819836	1.799957974	0.115142818
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	45.91398092	-1.856072541	0.127034842
DNA Metabolism|DNA_repair,_bacterial	98.99500431	1.784656679	0.128472364
Respiration|F0F1-type_ATP_synthase	304.5986236	-1.632741975	0.1521909
Protein Metabolism|Proteasome_bacterial	665.4959472	1.619067816	0.152721246
Protein Metabolism|Peptidyl-prolyl_cis-trans_isomerase	14.77020776	1.709449884	0.187354209
RNA Metabolism|Transcription_initiation,_bacterial_sigma_factors	209.5612441	1.495774155	0.187591187
Respiration|Respiratory_Complex_I	152.1287706	-1.502221866	0.18874834
Protein Metabolism|Translation_elongation_factor_G_family	118.3909735	-1.446672758	0.206758109
Protein Metabolism|Selenocysteine_metabolism	12.95536266	1.552406171	0.233279147
Stress Response|Glutaredoxins	60.82350017	1.364163006	0.241799783
Membrane Transport|Na(+)_H(+)_antiporter	19.44798324	1.446739324	0.249602446
Amino Acids and Derivatives|Chorismate_Synthesis	44.20856881	1.290486254	0.273931786
Motility and Chemotaxis|Bacterial_Chemotaxis	184.3877645	-1.213709242	0.2779588
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	140.8512796	-1.198527949	0.286413883
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_and_NADP_cofactor_biosynthesis_global	2.479550455	-1.391621738	0.287956015
DNA Metabolism|DNA_repair,_bacterial_UvrD_and_related_helicases	2.479550455	-1.391621738	0.287956015
DNA Metabolism|DNA_structural_proteins,_bacterial	2.479550455	-1.391621738	0.287956015
Phosphorus Metabolism|Phosphate_metabolism	140.0247628	-1.189860813	0.289749284
Amino Acids and Derivatives|Lysine_Biosynthesis_DAP_Pathway	48.05244895	-1.191785151	0.308710429
RNA Metabolism|RNA_polymerase_bacterial	124.8211346	1.138686357	0.310525661
RNA Metabolism|Transcription_factors_bacterial	68.97923536	-1.14764064	0.317298507
Protein Metabolism|Ribosome_LSU_bacterial	2631.95296	-1.069878613	0.325872006
Regulation and Cell signaling|Murein_hydrolase_regulation_and_cell_death	2.066292046	-1.239233551	0.334271891
Protein Metabolism|Universal_GTPases	468.1814537	-1.037553841	0.343060752
Clustering-based subsystems|NusA-TFII_Cluster	6.995514464	-1.270559487	0.347959037

Carbohydrates|Pyruvate_Alanine_Serine_Interconversions	46.68811927	1.070983408	0.357424777
Virulence, Disease and Defense|MLST	37.97243566	-1.075120044	0.362600974
Protein Metabolism|Ribosome_SSU_bacterial	667.4414805	-0.987259776	0.364437443
Nucleosides and Nucleotides|De_Novo_Purine_Biosynthesis	39.81715926	-1.064357295	0.365736246
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	72.00397719	-0.984300656	0.386301793
Amino Acids and Derivatives|Arginine_and_Ornithine_Degradation	1.653033637	-1.065292904	0.391822774
Carbohydrates|Entner-Doudoroff_Pathway	28.11399083	-1.007193438	0.402398777
Clustering-based subsystems|Putative_hemin_transporter	104.2230117	0.915875238	0.411756768
Metabolism of Aromatic Compounds|Quinate_degradation	12.98524115	1.044388754	0.415443583
Nitrogen Metabolism|Cyanate_hydrolysis	3.911015641	-1.11280713	0.416661891
Stress Response|Oxidative_stress	129.8399925	0.889833653	0.421650765
Metabolism of Aromatic Compounds|Salicylate_and_gentisate_catabolism	1.209896735	0.896427763	0.444796069
Carbohydrates|D-Galacturonate_and_D-Glucuronate_Utilization	2.833051879	1.024611689	0.449927467
Potassium metabolism|Potassium_homeostasis	62.52891227	-0.850892235	0.454089186
Nucleosides and Nucleotides|Purine_conversions	94.90742017	-0.832554622	0.455673426
Protein Metabolism|Translation_elongation_factors_eukaryotic_and_archaeal	13.13463362	-0.937173568	0.464364658
Carbohydrates|CO2_uptake,_carboxysome	1.239775228	-0.864047201	0.464636744
Iron acquisition and metabolism|Hemin_transport_system	1.239775228	-0.864047201	0.464636744
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	95.09911013	-0.803856689	0.470762449
Cofactors, Vitamins, Prosthetic Groups, Pigments|Chlorophyll_Biosynthesis	65.65066811	0.81294435	0.472215449
Clustering-based subsystems|CBSS-262719.3.peg.410	36.28952353	-0.840594176	0.473639435
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	23.78971678	-0.86858318	0.473886793
Respiration|Biogenesis_of_cytochrome_c_oxidases	37.3923253	0.825284034	0.479841107
Cell Wall and Capsule|Capsular_heptose_biosynthesis	14.5660988	-0.885444791	0.484699038
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	145.4817171	-0.759558252	0.489195761
Carbohydrates|Pentose_phosphate_pathway	24.80792355	-0.815248478	0.499208876
Miscellaneous|YbbK	17.47132667	0.832573465	0.502541126
Clustering-based subsystems|Bacterial_Cell_Division	67.61994616	-0.750338298	0.505820093
Amino Acids and Derivatives|Alanine_biosynthesis	127.5098345	0.713059577	0.516564746
Carbohydrates|Di-Inositol-Phosphate_biosynthesis	10.4633932	-0.844816549	0.517048568

Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_A_Biosynthesis	48.59259985	0.740096469	0.518241735
Protein Metabolism|Programmed_frameshift	8.618669608	-0.846365478	0.522971892
Cell Wall and Capsule|Sialic_Acid_Metabolism	20.525947	0.778919314	0.523948141
Clustering-based subsystems|Cluster-based_Subsystem_Grouping_Hypotheticals_-_perhaps_Proteosome_Related	16.8663783	0.778207615	0.531586584
Potassium metabolism|Glutathione-regulated_potassium-efflux_system_and_associated_functions	4.929222418	-0.853872407	0.532268191
Nitrogen Metabolism|Ammonia_assimilation	25.41287192	-0.740319969	0.537683267
Cell Wall and Capsule|Lipid_A_modifications	4.869465433	0.828663541	0.544237869
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	14.22501638	0.756065084	0.549424774
Amino Acids and Derivatives|Histidine_Biosynthesis	61.0376901	-0.665575828	0.555779534
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	8.205411199	-0.778780833	0.557727303
Respiration|Anaerobic_respiratory_reductases	66.72863187	0.626781548	0.576507626
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	19.31605026	0.681426474	0.578293964
Carbohydrates|Glycolysis_and_Gluconeogenesis	41.98550579	-0.637955805	0.580213646
Protein Metabolism|Ribosome_biogenesis_bacterial	11.06834157	-0.687025606	0.595263304
RNA Metabolism|Rrf2_family_transcriptional_regulators	9.67917388	0.514901143	0.600369788
Respiration|Formate_hydrogenase	16.67468834	0.649073421	0.601137339
Amino Acids and Derivatives|Phenylalanine_and_Tyrosine_Branches_from_Chorismate	0.604948368	0.511188524	0.601829864
Carbohydrates|Maltose_and_Maltodextrin_Utilization	0.604948368	0.511188524	0.601829864
Cofactors, Vitamins, Prosthetic Groups, Pigments|Menaquinone_and_Phylloquinone_Biosynthesis	0.604948368	0.511188524	0.601829864
DNA Metabolism|RuvABC_plus_a_hypothetical	0.604948368	0.511188524	0.601829864
Metabolism of Aromatic Compounds|p-Hydroxybenzoate_degradation	0.604948368	0.511188524	0.601829864
Virulence, Disease and Defense|Multidrug_Resistance_Efflux_Pumps	0.604948368	0.511188524	0.601829864
Membrane Transport|Choline_Transport	3.246310288	0.711860513	0.602812811
Iron acquisition and metabolism|Transport_of_Iron	30.51632481	0.600962891	0.609686729
DNA Metabolism|DNA_Repair_Base_Excision	20.74751545	0.613055116	0.614181894
Cell Division and Cell Cycle|Macromolecular_synthesis_operon	10.97870609	0.643922424	0.617915611
DNA Metabolism|DNA_repair,_UvrABC_system	10.65508316	-0.629480816	0.627268916
Membrane Transport|ABC_transporter_alkylphosphonate_(TC_3.A.1.9.1)	15.36273713	-0.607662877	0.627514988
Motility and Chemotaxis|Flagellum_in_Campylobacter	670.6253313	-0.51391663	0.629480283

Nucleosides and Nucleotides|Ribonucleotide_reduction	195.2419162	-0.519059947	0.631121862
Fatty Acids, Lipids, and Isoprenoids|Glycerolipid_and_Glycerophospholipid_Metabolism_in_Bacteria	18.2256675	-0.573754318	0.641290449
Amino Acids and Derivatives|Arginine_Biosynthesis_extended	153.8912372	-0.504344637	0.642431091
Cell Wall and Capsule|mycolic_acid_synthesis	6.300930619	0.609533985	0.651217178
Cofactors, Vitamins, Prosthetic Groups, Pigments|Riboflavin,_FMN_and_FAD_metabolism	6.300930619	0.609533985	0.651217178
RNA Metabolism|tRNA_processing	13.10475512	-0.531001537	0.675743008
DNA Metabolism|DNA_repair,_bacterial_DinG_and_relatives	1.623155144	0.502414581	0.697520129

MAG: bin_040
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Regulation and Cell signaling|Autoinducer_2_(AI-2)_transport_and_processing_(lsrACDBFGE_operon)	620.6206672	-1.028978702	0.081040494
Carbohydrates|D-ribose_utilization	218.4638414	-0.696731148	0.247214191
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	7.826237921	-0.89492446	0.289119146
Carbohydrates|Inositol_catabolism	5.031152949	-0.745735547	0.34535524
Regulation and Cell signaling|Sex_pheromones_in_Enterococcus_faecalis_and_other_Firmicutes	2.795084972	0.595202057	0.365947967
DNA Metabolism|Plasmid_replication	6.484597135	-0.729854152	0.378798692
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	5.366563146	0.6709351	0.405753687
Protein Metabolism|Universal_GTPases	5.366563146	0.6709351	0.405753687
Amino Acids and Derivatives|Cysteine_Biosynthesis	20.2364152	0.695810635	0.414187082
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	10.50951949	0.699032499	0.420739012
Protein Metabolism|GroEL_GroES	51.76497368	-0.54677734	0.461511122
Clustering-based subsystems|Bacterial_Cell_Division	7.043614129	-0.603881541	0.47358
Protein Metabolism|Proteolysis_in_bacteria,_ATP-dependent	29.18068711	0.529729264	0.514890052

MAG: bin_040
Mesocosm water-source: Offshore

SEED_Subsystem	baseMean	log2FoldChange	pvalue
Stress Response|Heat_shock_dnaK_gene_cluster_extended	17.03756494	2.601919865	0.056535801
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	358.1505285	1.795830648	0.056978462

Respiration|F0F1-type_ATP_synthase	174.7968696	-1.154875905	0.204598591
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	6.371282548	1.751664245	0.224425754
DNA Metabolism|DNA_repair,_bacterial	14.06832115	1.390046368	0.295218282
Carbohydrates|Pentose_phosphate_pathway	2.684374905	1.296755537	0.335807324
Respiration|Ubiquinone_Menaquinone-cytochrome_c_reductase_complexes	2.684374905	1.296755537	0.335807324
Nucleosides and Nucleotides|Ribonucleotide_reduction	71.19080771	-0.833495582	0.394445663
Amino Acids and Derivatives|Methionine_Biosynthesis	1.862631033	-0.965625663	0.445905568
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	1.862631033	-0.965625663	0.445905568
Carbohydrates|TCA_Cycle	1.610624943	0.870938674	0.478022948
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	17.93601766	0.855342984	0.494478121
Protein Metabolism|Proteolysis_in_bacteria,_ATP-dependent	30.57450273	0.723926489	0.520428919
Amino Acids and Derivatives|Chorismate_Synthesis	1.396973274	-0.763256924	0.520571091
Carbohydrates|Lacto-N-Biose_I_and_Galacto-N-Biose_Metabolic_Pathway	1.073749962	0.618283924	0.576354606
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_regulation	1.073749962	0.618283924	0.576354606
Cofactors, Vitamins, Prosthetic Groups, Pigments|Pyridoxin_(Vitamin_B6)_Biosynthesis	1.073749962	0.618283924	0.576354606
Protein Metabolism|Protein_chaperones	1.073749962	0.618283924	0.576354606
Regulation and Cell signaling|Murein_hydrolase_regulation_and_cell_death	1.073749962	0.618283924	0.576354606
RNA Metabolism|RNA_processing_and_degradation,_bacterial	2.613157682	0.763925852	0.582945752
Amino Acids and Derivatives|Cysteine_Biosynthesis	20.87239865	0.655772293	0.587762751
Amino Acids and Derivatives|Polyamine_Metabolism	15.61322049	-0.681663864	0.593798893
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	0.931315516	-0.538359091	0.612166681
Cell Wall and Capsule|Sialic_Acid_Metabolism	0.931315516	-0.538359091	0.612166681
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	2.076282701	0.556304816	0.680655503

Supplementary Table S3.3. Differential expression of specific functional processes based on SEED annotations for MAGs bin004, bin008, bin040, bin093 and bin105 between the t-DOM and control mesocosms at 19h using DESeq2. The multiple sheets correspond to data for the different MAGs and different mesocosms being tested (nearshore/offshore lake-water). A positive value in the column 'log2FoldChange' for a specific functional process indicates a higher expression for that function in the t-DOM sample as compared to the control, and a negative value indicates the opposite. Only those functional processes that are different between the control and t-DOM mesocosms by a fold-change of 0.5 or more are shown. Functional processes that are significantly differentially expressed (Wald test, P value < 0.05) between the control and t-DOM mesocosm are highlighted in yellow.

MAG: bin_004
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Phosphorus Metabolism|Phosphate_metabolism	84.09861541	-2.862720463	0.026625008
Stress Response|Glutaredoxins	91.75572859	2.674054657	0.029546949
Protein Metabolism|Protein_chaperones	228.7523362	2.115831008	0.05271171
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	25.72544296	-2.677991411	0.060643221
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	66.10316651	-2.249841382	0.077214129
Stress Response|Heat_shock_dnaK_gene_cluster_extended	1616.63293	1.667286817	0.086929976
Respiration|Respiratory_dehydrogenases_1	42.10104678	2.201778506	0.09198557
Miscellaneous|YbbK	34.30139511	2.216057024	0.097980303
Stress Response|Hfl_operon	26.95109616	2.107169464	0.123980008
Cofactors, Vitamins, Prosthetic Groups, Pigments|Ubiquinone_Biosynthesis	21.93855868	2.042390994	0.143012226
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	54.05721107	1.672162381	0.175225048
Miscellaneous|ZZ_gjo_need_homes	23.90669182	-1.845743196	0.188046131
Protein Metabolism|Translation_elongation_factor_G_family	177.3372874	1.353951657	0.194572002
Protein Metabolism|Proteasome_bacterial	218.3093156	1.248114795	0.218938572
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	10.55978881	-1.714108757	0.224768642
Amino Acids and Derivatives|Histidine_Biosynthesis	10.22277428	-1.681066993	0.232986446
Carbohydrates|Tricarballylate_Utilization	12.25049825	1.648172612	0.249471595
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	19.52550284	-1.602464335	0.257267685
Sulfur Metabolism|Thioredoxin-disulfide_reductase	54.64024225	1.358591214	0.262699177
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	8.762377953	-1.527778643	0.273385969

RNA Metabolism|Rrf2_family_transcriptional_regulators	33.44537025	1.400073258	0.283429649
Fatty Acids, Lipids, and Isoprenoids|Glycerolipid_and_Glycerophospholipid_Metabolism_in_Bacteria	8.425363416	-1.489851666	0.2839532
Virulence, Disease and Defense|Methicillin_resistance_in_Staphylococci	8.313025237	-1.476978529	0.287590863
DNA Metabolism|DNA_topoisomerases,_Type_I,_ATP-independent	8.200687058	-1.463986864	0.291288102
DNA Metabolism|DNA_repair,_bacterial	52.97650985	1.250206922	0.302620172
Protein Metabolism|tRNA_aminoacylation,_Asp_and_Asn	7.638996164	-1.397191052	0.310708231
Carbohydrates|Acetyl-CoA_fermentation_to_Butyrate	25.1210715	-1.379428965	0.315927028
Protein Metabolism|Programmed_frameshift	6.964967091	-1.312787407	0.336221922
Virulence, Disease and Defense|MLST	6.964967091	-1.312787407	0.336221922
Carbohydrates|D-galactonate_catabolism	6.740290733	-1.283512565	0.345321714
Regulation and Cell signaling|Orphan_regulatory_proteins	6.290938017	-1.223412595	0.364406689
Amino Acids and Derivatives|Methionine_Biosynthesis	14.69496115	-1.264004591	0.37590595
Respiration|Respiratory_Complex_I	113.681807	-0.937049007	0.385710099
Carbohydrates|Glycerol_fermenation_to_1,3-propanediol	5.729247123	-1.144359206	0.390316916
Amino Acids and Derivatives|Chorismate_Synthesis	73.42769516	0.965682506	0.396551031
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	27.79584751	-1.098404607	0.414812722
Clustering-based subsystems|Conserved_gene_cluster_associated_with_Met-tRNA_formyltransferase	63.94297081	0.946663478	0.415662824
Clustering-based subsystems|Bacterial_Cell_Division	34.40245978	-1.060891935	0.416831622
Cell Wall and Capsule|Sialic_Acid_Metabolism	20.73988252	-1.127348524	0.41824924
Respiration|Anaerobic_respiratory_reductases	40.1643207	1.011152228	0.420485475
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	33.72843071	-1.031445489	0.430691518
RNA Metabolism|Transcription_factors_bacterial	38.64997029	-0.942506947	0.46091789
Amino Acids and Derivatives|Methionine_Salvage	4.04417444	-0.880164723	0.483524953
Protein Metabolism|Periplasmic_disulfide_interchange	3.931836261	-0.860936279	0.49070981
Nucleosides and Nucleotides|Ribonucleotide_reduction	137.0324447	-0.714613593	0.491577258
Iron acquisition and metabolism|Transport_of_Iron	28.2294967	0.894700629	0.497803858
Secondary Metabolism|Auxin_biosynthesis	38.94994102	0.833493477	0.506740911
Membrane Transport|ABC_transporter_oligopeptide_(TC_3.A.1.5.1)	10.98780125	-0.928501772	0.516614663
Membrane Transport|Ton_and_Tol_transport_systems	41.77409902	-0.803632964	0.522303086

Carbohydrates|Pyruvate_Alanine_Serine_Interconversions	53.80555767	0.754548872	0.525477687
RNA Metabolism|RNA_processing_and_degradation,_bacterial	17.36973716	-0.89268366	0.525906309
Amino Acids and Derivatives|Glutamine,_Glutamate,_Aspartate_and_Asparagine_Biosynthesis	64.15637366	-0.720118582	0.536967659
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	249.6033502	0.589989977	0.538025825
Respiration|Formate_hydrogenase	47.44496204	0.717100164	0.554181537
Amino Acids and Derivatives|Glycine_cleavage_system	9.864419457	-0.808019948	0.57201533
Cofactors, Vitamins, Prosthetic Groups, Pigments|Pyridoxin_(Vitamin_B6)_Biosynthesis	15.50830544	0.793435133	0.573284093
Carbohydrates|Butanol_Biosynthesis	50.15678186	-0.65883629	0.586832458
DNA Metabolism|DNA_repair,_bacterial_MutL-MutS_system	2.808454472	-0.630772633	0.587011206
Protein Metabolism|Selenocysteine_metabolism	2.808454472	-0.630772633	0.587011206
RNA Metabolism|tRNA_nucleotidyltransferase	2.808454472	-0.630772633	0.587011206
Nucleosides and Nucleotides|Purine_conversions	27.21281633	-0.703914541	0.598109
Photosynthesis|Photosystem_II-type_photosynthetic_reaction_center	21.16789496	-0.705615035	0.608422554
Cell Wall and Capsule|KDO2-Lipid_A_biosynthesis	26.87580179	-0.684415454	0.608717907
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	2.583778114	-0.573228383	0.614729781
Carbohydrates|Propionate-CoA_to_Succinate_Module	16.29467269	0.666240658	0.634703082
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	20.26918952	-0.641697119	0.642421864
Nitrogen Metabolism|Ammonia_assimilation	14.44894451	-0.648454856	0.647949132
Carbohydrates|Pyruvate_metabolism_I:_anaplerotic_reactions,_PEP	47.34832739	-0.555778392	0.648667781
Cell Wall and Capsule|Lipid_A_modifications	9.372386173	0.636256212	0.656625886
DNA Metabolism|DNA_repair,_UvrABC_system	7.842332237	-0.560485378	0.692887747
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_pathogenicity_islands_SaPI	6.13591927	0.552136443	0.693104494
Respiration|Biogenesis_of_cytochrome_c_oxidases	6.13591927	0.552136443	0.693104494
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_A_Biosynthesis	32.38600932	0.502091478	0.696164712
Clustering-based subsystems|CBSS-262719.3.peg.410	18.92113138	-0.539451126	0.697825126
Membrane Transport|Twin-arginine_translocation_system	24.29202368	-0.524066371	0.697910858
Potassium metabolism|Potassium_homeostasis	18.47177866	-0.503483088	0.717565513

MAG: bin_004
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Respiration|Respiratory_dehydrogenases_1	109.7206521	3.040346028	0.022030144
Protein Metabolism|Protein_chaperones	714.7230104	2.91566558	0.024704169
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	194.7839131	-2.965032437	0.024836804
RNA Metabolism|Rrf2_family_transcriptional_regulators	129.2971848	2.567900735	0.047178832
Protein Metabolism|Proteasome_bacterial	656.9006605	2.505888411	0.048651092
Stress Response|Heat_shock_dnaK_gene_cluster_extended	4151.671684	2.299989061	0.066277344
Phosphorus Metabolism|Phosphate_metabolism	129.5265452	-2.178716661	0.087892904
Virulence, Disease and Defense|MLST	9.146621482	-2.307678913	0.110497312
Fatty Acids, Lipids, and Isoprenoids|Glycerolipid_and_Glycerophospholipid_Metabolism_in_Bacteria	8.384403025	-2.231635441	0.122747668
Respiration|F0F1-type_ATP_synthase	203.9608843	-1.896086098	0.127627428
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	68.39729109	1.80281376	0.153789496
Membrane Transport|ABC_transporter_alkylphosphonate_(TC_3.A.1.9.1)	6.288302269	-1.977991659	0.170320895
Regulation and Cell signaling|Orphan_regulatory_proteins	6.288302269	-1.977991659	0.170320895
Respiration|Respiratory_Complex_I	112.0983854	-1.674130006	0.178970796
Stress Response|Glutaredoxins	32.69613109	1.728962968	0.184397203
Miscellaneous|YbbK	27.99803314	1.714903993	0.191802282
Respiration|Biogenesis_of_c-type_cytochromes	21.60686612	1.720731032	0.197859307
Regulation and Cell signaling|Stringent_Response,_(p)ppGpp_metabolism	4.126434116	1.791123771	0.213546408
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	18.63050726	-1.670535238	0.223141574
Cell Wall and Capsule|Lipid_A_modifications	15.40625371	1.663135433	0.223386679
Nucleosides and Nucleotides|De_Novo_Purine_Biosynthesis	90.83214306	1.482258389	0.22918403
RNA Metabolism|Ribonuclease_H	7.322017626	1.719929959	0.230135203
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_pathogenicity_islands_SaPI	13.90373926	1.623381106	0.237627875
Carbohydrates|Pyruvate_Alanine_Serine_Interconversions	89.16099646	1.392377619	0.256822744
Carbohydrates|Pyruvate_metabolism_I:_anaplerotic_reactions,_PEP	39.69437958	-1.452530646	0.259885557
Amino Acids and Derivatives|Glycine_cleavage_system	53.60654615	1.388443407	0.267218242
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	14.28484849	1.487662055	0.276451056

Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	32.66578132	-1.415429027	0.277369845
Protein Metabolism|Ribosome_SSU_bacterial	626.7931173	-1.250304443	0.291607167
Cell Wall and Capsule|Peptidoglycan_Biosynthesis	44.21035023	1.308162092	0.298812328
Amino Acids and Derivatives|Cysteine_Biosynthesis	16.35902678	1.369707865	0.309791145
Clustering-based subsystems|Conserved_gene_cluster_associated_with_Met-tRNA_formyltransferase	100.5065834	1.188729723	0.326646128
Amino Acids and Derivatives|Chorismate:_Intermediate_for_synthesis_of_PAPA_antibiotics,_PABA,_anthranilate,_3-hydroxyanthranilate_and_more.	4.507543344	1.35404986	0.349056227
Amino Acids and Derivatives|Alanine_biosynthesis	132.9092952	1.120187027	0.350282402
Cofactors, Vitamins, Prosthetic Groups, Pigments|Riboflavin,_FMN_and_FAD_metabolism	11.66092882	1.225264783	0.37460999
RNA Metabolism|RNA_polymerase_bacterial	403.7396742	-1.025280322	0.383308163
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	35.09914637	-1.092585771	0.392892439
DNA Metabolism|DNA_repair,_bacterial	52.91009509	1.030776945	0.403653716
RNA Metabolism|tRNA_nucleotidyltransferase	4.698097958	1.197268753	0.407576058
Cofactors, Vitamins, Prosthetic Groups, Pigments|Heme_and_Siroheme_Biosynthesis	63.71943586	-0.992674186	0.419660086
Membrane Transport|Type_IV_pilus	2.096100756	-1.050414758	0.434974483
Photosynthesis|Photosystem_II-type_photosynthetic_reaction_center	49.6841618	-0.96943739	0.436192159
Protein Metabolism|Universal_GTPases	165.499202	-0.910867029	0.442551247
Cofactors, Vitamins, Prosthetic Groups, Pigments|Chlorophyll_Biosynthesis	111.2855744	-0.916912502	0.444543421
Nucleosides and Nucleotides|Ribonucleotide_reduction	105.0191946	-0.886055813	0.460286386
Amino Acids and Derivatives|Chorismate_Synthesis	89.67364048	0.88584539	0.460817957
Nucleosides and Nucleotides|Pyrimidine_utilization	4.888652573	1.064392809	0.461230243
Cofactors, Vitamins, Prosthetic Groups, Pigments|Ubiquinone_Biosynthesis	6.772276249	1.031954262	0.469812155
Miscellaneous|ZZ_gjo_need_homes	43.03667277	-0.890025311	0.477108066
Respiration|Terminal_cytochrome_C_oxidases	47.85281037	0.854835483	0.488288568
Protein Metabolism|Ribosome_biogenesis_bacterial	1.714991528	-0.878884321	0.503530004
Regulation and Cell signaling|Murein_hydrolase_regulation_and_cell_death	1.714991528	-0.878884321	0.503530004
Virulence, Disease and Defense|Resistance_to_fluoroquinolones	29.57306256	-0.83842627	0.513220857
Stress Response|Hfl_operon	14.8784348	0.863121428	0.520643051
Membrane Transport|Twin-arginine_translocation_system	21.42305908	-0.831121696	0.527467404
Carbohydrates|Butanol_Biosynthesis	25.2122289	-0.818082439	0.528342586

Sulfur Metabolism|Thioredoxin-disulfide_reductase	51.26087096	0.759279027	0.535543192
Cofactors, Vitamins, Prosthetic Groups, Pigments|Pyridoxin_(Vitamin_B6)_Biosynthesis	5.694715961	-0.892244893	0.536368081
Amino Acids and Derivatives|Proline_Synthesis	9.674440393	-0.868255157	0.53657569
Carbohydrates|Propionate-CoA_to_Succinate_Module	7.153385478	0.86570476	0.542514499
Cell Wall and Capsule|mycolic_acid_synthesis	1.524436914	-0.78380749	0.544213357
Carbohydrates|L-rhamnose_utilization	21.29152428	0.770472027	0.553891586
Respiration|Formate_hydrogenase	41.65219796	0.73222362	0.554585814
Carbohydrates|TCA_Cycle	157.7893276	-0.668334861	0.569925945
Amino Acids and Derivatives|Aromatic_amino_acid_degradation	5.313606732	-0.813534846	0.573285206
Protein Metabolism|Periplasmic_disulfide_interchange	5.313606732	-0.813534846	0.573285206
RNA Metabolism|tRNA_processing	31.85297035	0.70833476	0.574147294
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	275.4846034	0.629550467	0.587592055
Carbohydrates|Maltose_and_Maltodextrin_Utilization	1.333882299	-0.682051927	0.590020325
DNA Metabolism|DNA_structural_proteins,_bacterial	1.333882299	-0.682051927	0.590020325
RNA Metabolism|Polyadenylation_bacterial	1.333882299	-0.682051927	0.590020325
Carbohydrates|Entner-Doudoroff_Pathway	12.32028253	-0.709866332	0.605465245
Secondary Metabolism|Steroid_sulfates	4.932497504	-0.728797883	0.614128671
Stress Response|Glutathione:_Non-redox_reactions	12.12972791	-0.68722813	0.617298685
Nucleosides and Nucleotides|Nudix_proteins_(nucleoside_triphosphate_hydrolases)	18.09594077	0.642729699	0.625480035
Carbohydrates|Tricarballylate_Utilization	1.143327685	-0.571523052	0.642425943
Carbohydrates|Glycolysis_and_Gluconeogenesis	39.01310341	-0.577914269	0.643050232
Protein Metabolism|tRNA_aminoacylation,_Asp_and_Asn	15.15667927	-0.622051571	0.64432092
DNA Metabolism|DNA_repair,_bacterial_MutL-MutS_system	4.551388275	-0.637013421	0.659549407
Clustering-based subsystems|CBSS-262719.3.peg.410	24.8091972	-0.555753329	0.666461257
Amino Acids and Derivatives|Arginine_and_Ornithine_Degradation	7.959448866	-0.606874294	0.668505021
Amino Acids and Derivatives|Valine_degradation	1.693069062	0.578623329	0.669106805
Protein Metabolism|Selenocysteine_metabolism	4.360833661	-0.588170894	0.684180728
Carbohydrates|Pentose_phosphate_pathway	7.768894251	-0.573582654	0.686009349
DNA Metabolism|DNA_Repair_Base_Excision	18.85815923	0.529653142	0.686053408
DNA Metabolism|RuvABC_plus_a_hypothetical	3.767247353	0.582234126	0.687047146

Metabolism of Aromatic Compounds|Quinate_degradation	3.767247353	0.582234126	0.687047146
Virulence, Disease and Defense|Methicillin_resistance_in_Staphylococci	14.2039062	-0.520229795	0.700434563
Amino Acids and Derivatives|Methionine_Salvage	19.04871384	0.50303806	0.700723243
Protein Metabolism|Peptidyl-prolyl_cis-trans_isomerase	10.37089145	0.509295299	0.712148188
Metabolism of Aromatic Compounds|p-Hydroxybenzoate_degradation	7.387785023	-0.503826338	0.723092319

MAG: bin_008
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Cell Wall and Capsule|mycolic_acid_synthesis	177.2308718	3.428159056	0.006972203
Cofactors, Vitamins, Prosthetic Groups, Pigments|Riboflavin,_FMN_and_FAD_metabolism	20.42218409	2.717921399	0.045703706
Sulfur Metabolism|Thioredoxin-disulfide_reductase	297.1533053	2.308630436	0.050186488
Stress Response|Glutaredoxins	95.79478102	2.400963903	0.050413503
RNA Metabolism|Transcription_initiation,_bacterial_sigma_factors	280.7313429	2.200151879	0.060426862
Miscellaneous|YbbK	33.47553886	2.007651117	0.118570642
Stress Response|Hfl_operon	24.29608292	1.992193625	0.130698144
Amino Acids and Derivatives|Chorismate_Synthesis	73.47775511	1.742569894	0.146573102
Protein Metabolism|Protein_chaperones	298.9218244	1.631796062	0.147261527
Iron acquisition and metabolism|Transport_of_Iron	51.87655834	1.730578742	0.159022146
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	51.49758997	-1.723018779	0.168584504
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	82.57299586	1.616384216	0.172049429
Stress Response|Heat_shock_dnaK_gene_cluster_extended	926.5776512	1.478105828	0.177478366
Cofactors, Vitamins, Prosthetic Groups, Pigments|Ubiquinone_Biosynthesis	60.71915351	1.585754241	0.188613071
Amino Acids and Derivatives|Urea_decomposition	9.095240748	-1.793104209	0.190356168
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	38.23369722	1.607395853	0.199373404
Respiration|Biogenesis_of_cytochrome_c_oxidases	38.36002	1.590246741	0.203774531
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	46.31835566	-1.592626974	0.20411122
Carbohydrates|D-galactonate_catabolism	8.337304019	-1.713772742	0.209218306
Amino Acids and Derivatives|Methionine_Degradation	7.958335654	-1.671722478	0.219664234
Virulence, Disease and Defense|Methicillin_resistance_in_Staphylococci	7.705690078	-1.642730706	0.227046342

DNA Metabolism|RuvABC_plus_a_hypothetical	13.51653833	1.639266693	0.228713201
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	7.326721713	-1.597714078	0.238799008
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	6.947753349	-1.550742421	0.25143649
Motility and Chemotaxis|Flagellum	63.87722321	-1.367940621	0.256660047
Secondary Metabolism|Auxin_biosynthesis	37.26522251	1.407227657	0.258623847
DNA Metabolism|DNA_repair,_bacterial	48.08687469	1.356202932	0.26465188
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	47.28683037	-1.370506593	0.268212184
Protein Metabolism|Selenocysteine_metabolism	13.8955067	1.500322982	0.268857058
Membrane Transport|Na(+)_H(+)_antiporter	14.14815227	1.417864758	0.294913641
Virulence, Disease and Defense|MLST	5.810848255	-1.396587428	0.295562258
Respiration|Terminal_cytochrome_oxidases	6.568784984	1.415315798	0.30298616
Protein Metabolism|Periplasmic_disulfide_interchange	16.88514602	1.369121878	0.305269308
Carbohydrates|Propionyl-CoA_to_Succinyl-CoA_Module	11.66380411	1.396054967	0.307529869
Clustering-based subsystems|Conserved_gene_cluster_associated_with_Met-tRNA_formyltransferase	71.62502089	1.189751944	0.308958177
Clustering-based subsystems|Putative_hemin_transporter	43.24450115	-1.249722842	0.31387704
DNA Metabolism|DNA_structural_proteins,_bacterial	4.210759605	1.302246673	0.3258779
Phosphorus Metabolism|Phosphate_metabolism	55.66624198	-1.175822877	0.330458791
Amino Acids and Derivatives|Glycine_cleavage_system	64.80359033	1.110798737	0.344168644
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	40.5917226	-1.162083193	0.350012444
RNA Metabolism|tRNA_processing	4.673943162	-1.218456779	0.351481179
Cofactors, Vitamins, Prosthetic Groups, Pigments|Folate_Biosynthesis	42.52867201	1.133851926	0.352158043
Nucleosides and Nucleotides|Ribonucleotide_reduction	86.53110989	-1.078490375	0.352431979
Amino Acids and Derivatives|Methionine_Biosynthesis	12.59017122	-1.265097943	0.356671084
Cell Wall and Capsule|Sialic_Acid_Metabolism	19.36949418	-1.201625641	0.367872567
Carbohydrates|Isobutyryl-CoA_to_Propionyl-CoA_Module	12.08488007	-1.219353087	0.374783773
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	12.08488007	-1.219353087	0.374783773
Amino Acids and Derivatives|Polyamine_Metabolism	6.947753349	1.195778168	0.385599766
Protein Metabolism|Proteasome_bacterial	234.6656328	0.941757087	0.385726214
DNA Metabolism|DNA_repair,_UvrABC_system	3.916006433	-1.082483099	0.397589757

Regulation and Cell signaling|Orphan_regulatory_proteins	3.916006433	-1.082483099	0.397589757
DNA Metabolism|DNA-replication	18.02205111	1.115086674	0.399001069
DNA Metabolism|DNA_repair,_bacterial_MutL-MutS_system	3.663360857	-1.03381092	0.414803588
Amino Acids and Derivatives|Lysine_Biosynthesis_DAP_Pathway	24.12765254	-1.059806954	0.416063553
Respiration|Respiratory_dehydrogenases_1	24.00132975	1.040829939	0.418037
RNA Metabolism|Rrf2_family_transcriptional_regulators	15.66402573	1.0311126	0.440272928
Respiration|Respiratory_Complex_I	94.36312276	-0.876403479	0.442399692
Amino Acids and Derivatives|Methionine_Salvage	3.158069704	-0.930329273	0.452624384
Cell Division and Cell Cycle|Macromolecular_synthesis_operon	3.158069704	-0.930329273	0.452624384
Nitrogen Metabolism|Dissimilatory_nitrite_reductase	1.979057015	0.830776331	0.453875495
Carbohydrates|TCA_Cycle	80.80447683	-0.849383266	0.461266678
Amino Acids and Derivatives|Cysteine_Biosynthesis	10.06371546	-1.015042111	0.461495301
Protein Metabolism|Universal_GTPases	215.7172146	-0.776707705	0.473945863
DNA Metabolism|DNA_Repair_Base_Excision	9.81106988	-0.986714978	0.474230648
Miscellaneous|ZZ_gjo_need_homes	33.64396925	-0.890785385	0.477251419
Membrane Transport|Ton_and_Tol_transport_systems	44.17086826	-0.828991755	0.495520606
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_and_NADP_cofactor_biosynthesis_global	2.77910134	-0.827796653	0.495642151
Protein Metabolism|Ribosomal_protein_S12p_Asp_methylthiotransferase	2.77910134	-0.827796653	0.495642151
Respiration|Formate_hydrogenase	28.50684253	0.835866545	0.506291684
Carbohydrates|Di-Inositol-Phosphate_biosynthesis	2.652778551	-0.791036547	0.511825918
RNA Metabolism|ATP-dependent_RNA_helicases,_bacterial	16.54828525	0.858785074	0.517532776
Protein Metabolism|Ribosome_SSU_bacterial	435.4767584	-0.663790693	0.529586934
Nitrogen Metabolism|Ammonia_assimilation	8.674164787	-0.849880487	0.538008058
Protein Metabolism|Translation_elongation_factors_eukaryotic_and_archaeal	8.674164787	-0.849880487	0.538008058
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_pathogenicity_islands_SaPI	7.832012866	0.838772326	0.543300392
Carbohydrates|Propionate-CoA_to_Succinate_Module	23.03285504	0.777447163	0.544952866
Membrane Transport|Choline_Transport	2.400132975	-0.714421999	0.546673746
Amino Acids and Derivatives|Proline_Synthesis	26.40146273	0.736762884	0.560438229
RNA Metabolism|Polyadenylation_bacterial	62.06659658	-0.678314384	0.562299136
Protein Metabolism|GroEL_GroES	941.5258478	0.596650067	0.56575932

Respiration|Biogenesis_of_c-type_cytochromes	20.2958613	0.738808948	0.569839718
RNA Metabolism|RNA_processing_and_degradation,_bacterial	8.161887634	-0.783496427	0.57022424
Respiration|F0F1-type_ATP_synthase	316.3122616	-0.601693528	0.57093608
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_A_Biosynthesis	18.94841822	-0.723571592	0.583796652
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	34.48612117	-0.665783921	0.591567621
RNA Metabolism|Transcription_factors_bacterial	34.35979838	-0.659837742	0.594959337
Motility and Chemotaxis|Bacterial_Chemotaxis	49.77117854	-0.632264385	0.596287608
Amino Acids and Derivatives|Histidine_Biosynthesis	17.81151313	-0.518031044	0.603217941
Amino Acids and Derivatives|Tryptophan_synthesis	2.021164611	-0.592181074	0.605689905
Membrane Transport|ABC_transporter_oligopeptide_(TC_3.A.1.5.1)	18.1904815	-0.66465429	0.615655581
Clustering-based subsystems|Bacterial_Cell_Division	32.84392492	-0.585909561	0.63775457
Clustering-based subsystems|CBSS-262719.3.peg.410	17.55886755	-0.613107848	0.644010614
Carbohydrates|Carboxysome	1.768519034	-0.504426677	0.650729194
DNA Metabolism|DNA_repair,_bacterial_UvrD_and_related_helicases	1.768519034	-0.504426677	0.650729194
Nitrogen Metabolism|Cyanate_hydrolysis	1.768519034	-0.504426677	0.650729194
Secondary Metabolism|Steroid_sulfates	6.905645753	-0.599011093	0.663710454
Carbohydrates|Pyruvate_metabolism_II:_acetyl-CoA,_acetogenesis_from_pyruvate	5.600310275	0.567838561	0.678883664
Protein Metabolism|Protein_degradation	16.80093083	-0.548009103	0.680393324
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_regulation	11.79012689	-0.558662907	0.682280603
Membrane Transport|ABC_transporter_alkylphosphonate_(TC_3.A.1.9.1)	15.45348775	0.527078084	0.692148448
Nucleosides and Nucleotides|Nudix_proteins_(nucleoside_triphosphate_hydrolases)	18.82209544	0.511701576	0.695422043
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	6.526677388	-0.53745325	0.696092331

MAG: bin_008
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Motility and Chemotaxis|Bacterial_Chemotaxis	21.31250047	-3.129931135	0.054960405
Respiration|Respiratory_dehydrogenases_1	27.0489534	2.990804718	0.057594059
Clustering-based subsystems|Putative_hemin_transporter	26.40522735	-2.626771028	0.097306385
Protein Metabolism|Protein_chaperones	309.5678797	2.351385276	0.106359786

Amino Acids and Derivatives|Histidine_Biosynthesis	10.76610848	-2.566457325	0.118318802
RNA Metabolism|Polyadenylation_bacterial	29.30078923	-2.278994399	0.142810263
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	9.008376487	-2.407088046	0.143204642
Protein Metabolism|Proteasome_bacterial	422.5659206	2.100124154	0.144182164
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	19.15458286	-2.317520372	0.145949379
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	105.8258912	2.098518438	0.150975414
Stress Response|Heat_shock_dnaK_gene_cluster_extended	1566.7974	2.040425477	0.153142085
Respiration|Respiratory_Complex_I	45.41858849	-2.15624383	0.154657345
Miscellaneous|YbbK	22.45838646	2.203988032	0.155598283
Amino Acids and Derivatives|Methionine_Degradation	7.909793989	-2.288781023	0.163819683
Protein Metabolism|tRNA_aminoacylation,_Asp_and_Asn	17.17713436	-2.208202353	0.1669839
Carbohydrates|Pyruvate_metabolism_I:_anaplerotic_reactions,_PEP	6.371778491	-2.089835871	0.202647563
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	6.371778491	-2.089835871	0.202647563
Protein Metabolism|Translation_elongation_factor_G_family	40.62407288	-1.817515391	0.226476234
Amino Acids and Derivatives|Glycine_cleavage_system	104.0523914	1.698847711	0.237891343
RNA Metabolism|Transcription_initiation,_bacterial_sigma_factors	264.745025	1.621436551	0.251866065
Respiration|F0F1-type_ATP_synthase	97.92312019	-1.649477429	0.253522112
Stress Response|Glutaredoxins	33.93865967	1.671266315	0.263676491
Sulfur Metabolism|Thioredoxin-disulfide_reductase	280.5411334	1.455258227	0.300274411
Carbohydrates|Butanol_Biosynthesis	10.36592287	-1.676967648	0.300459743
Respiration|Biogenesis_of_c-type_cytochromes	23.29800507	1.570102565	0.302210457
Nucleosides and Nucleotides|Ribonucleotide_reduction	67.70421757	-1.453018711	0.316510868
Secondary Metabolism|Steroid_sulfates	3.735180495	-1.602608488	0.319222791
Protein Metabolism|Universal_GTPases	133.2505175	-1.410017241	0.321003243
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	41.14200066	-1.463923809	0.323474155
Clustering-based subsystems|Conserved_gene_cluster_associated_with_Met-tRNA_formyltransferase	74.9867421	1.411176267	0.325984082
DNA Metabolism|DNA-replication	4.990752547	1.610061203	0.327471946
Carbohydrates|Photorespiration_(oxidative_C2_cycle)	20.11194364	-1.506273113	0.330688776
Cofactors, Vitamins, Prosthetic Groups, Pigments|Chlorophyll_Biosynthesis	24.98495402	-1.470413444	0.335099506

Carbohydrates|Pyruvate_Alanine_Serine_Interconversions	28.46922673	1.403988761	0.348516738
Protein Metabolism|Periplasmic_disulfide_interchange	3.295747495	-1.493034647	0.34938778
Membrane Transport|ABC_transporter_oligopeptide_(TC_3.A.1.5.1)	3.076030996	-1.433637959	0.366299221
Protein Metabolism|Ribosomal_protein_S12p_Asp_methylthiotransferase	3.076030996	-1.433637959	0.366299221
Cofactors, Vitamins, Prosthetic Groups, Pigments|Molybdenum_cofactor_biosynthesis	6.568015434	1.461736368	0.371221957
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_pathogenicity_islands_SaPI	12.437633 98	1.41046333	0.371659558
RNA Metabolism|tRNA_processing	2.856314496	-1.370925852	0.384570665
Clustering-based subsystems|Cluster-based_Subsystem_Grouping_Hypotheticals_-_perhaps_Proteosome_Related	5.210469047	1.403383845	0.393084885
Membrane Transport|ABC_transporter_alkylphosphonate_(TC_3.A.1.9.1)	2.636597996	-1.304346551	0.404422583
Virulence, Disease and Defense|MLST	12.16290226	-1.297324478	0.415108118
Carbohydrates|Di-Inositol-Phosphate_biosynthesis	2.416881497	-1.234304825	0.425809973
Amino Acids and Derivatives|Aromatic_amino_acid_degradation	2.197164997	-1.1593307	0.449247374
Protein Metabolism|Translation_elongation_factors_eukaryotic_and_archaeal	6.850458879	-1.220189719	0.45598933
Protein Metabolism|Ribosome_SSU_bacterial	312.9008082	-1.009064391	0.465567313
Membrane Transport|Ton_and_Tol_transport_systems	30.81532518	-1.07860757	0.468597087
Phosphorus Metabolism|Phosphate_metabolism	34.15032006	-1.046176275	0.479014733
Protein Metabolism|Programmed_frameshift	1.757731997	-0.956633463	0.520785254
Amino Acids and Derivatives|Methionine_Salvage	10.60140721	1.004770384	0.526475926
RNA Metabolism|RNA_polymerase_bacterial	50.43316497	0.904768386	0.528851885
Potassium metabolism|Potassium_homeostasis	13.30073215	-0.982222644	0.532086235
RNA Metabolism|Rrf2_family_transcriptional_regulators	4.29235566	0.960948893	0.558916766
Membrane Transport|Na(+)_H(+)_antiporter	2.715092773	0.944047901	0.55935199
Cell Division and Cell Cycle|Macromolecular_synthesis_operon	1.538015498	-0.842137785	0.564201879
Fatty Acids, Lipids, and Isoprenoids|Glycerolipid_and_Glycerophospholipid_Metabolism_in_Bacteria	1.538015498	-0.842137785	0.564201879
Nitrogen Metabolism|Cyanate_hydrolysis	1.538015498	-0.842137785	0.564201879
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	53.97982025	-0.826696289	0.564573415
Carbohydrates|TCA_Cycle	36.20626333	-0.817158169	0.576760952
Protein Metabolism|GroEL_GroES	778.680785	0.719280991	0.597522098
Virulence, Disease and Defense|Methicillin_resistance_in_Staphylococci	8.427721765	-0.840410108	0.602796891

Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	15.09771153	-0.80144405	0.605659159
Nucleosides and Nucleotides|Purine_conversions	44.45351599	-0.740035519	0.608651836
Cofactors, Vitamins, Prosthetic Groups, Pigments|Riboflavin,_FMN_and_FAD_metabolism	4.512072159	0.826350308	0.615221867
Carbohydrates|Pentose_phosphate_pathway	8.208005265	-0.806453944	0.617890888
Motility and Chemotaxis|Flagellum_in_Campylobacter	113.4290734	-0.677769902	0.626727944
Respiration|Biogenesis_of_cytochrome_c_oxidases	20.24545363	0.733124203	0.627276923
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	11.4802732	0.760577276	0.628870585
Stress Response|Hfl_operon	13.27725259	0.742824365	0.633292282
Cell Wall and Capsule|Lipid_A_modifications	2.934809273	0.759228728	0.640885238
RNA Metabolism|ATP-dependent_RNA_helicases,_bacterial	7.768572266	-0.735249953	0.649947738
Amino Acids and Derivatives|Chorismate_Synthesis	17.09092786	0.680974367	0.655707521
Cofactors, Vitamins, Prosthetic Groups, Pigments|Ubiquinone_Biosynthesis	17.09092786	0.680974367	0.655707521
Cofactors, Vitamins, Prosthetic Groups, Pigments|Heme_and_Siroheme_Biosynthesis	16.89469092	-0.661363095	0.666712935
Amino Acids and Derivatives|Arginine_and_Ornithine_Degradation	1.098582498	-0.584710097	0.67031364
Carbohydrates|Carboxysome	1.098582498	-0.584710097	0.67031364
Clustering-based subsystems|NusA-TFII_Cluster	1.098582498	-0.584710097	0.67031364
Cofactors, Vitamins, Prosthetic Groups, Pigments|Lipoic_acid_metabolism	1.098582498	-0.584710097	0.67031364
DNA Metabolism|DNA_repair,_bacterial_DinG_and_relatives	1.098582498	-0.584710097	0.67031364
DNA Metabolism|YcfH	1.098582498	-0.584710097	0.67031364
Sulfur Metabolism|Utilization_of_glutathione_as_a_sulphur_source	1.098582498	-0.584710097	0.67031364
Carbohydrates|Mannitol_Utilization	4.213860883	-0.678243711	0.679850267
Nucleosides and Nucleotides|Nudix_proteins_(nucleoside_triphosphate_hydrolases)	13.93640209	0.609336954	0.694144317
Miscellaneous|ZZ_gjo_need_homes	30.71335084	-0.552866099	0.706933258
Carbohydrates|Propionate-CoA_to_Succinate_Module	6.748484545	0.606231166	0.709043553
Amino Acids and Derivatives|Alanine_biosynthesis	44.22608776	0.513784707	0.720400199
Amino Acids and Derivatives|Methionine_Biosynthesis	19.98648975	0.520825148	0.729776118
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	6.968201044	0.53215023	0.742811184
Iron acquisition and metabolism|Transport_of_Iron	6.669989768	-0.534337543	0.742910829
Stress Response|Glutathione:_Non-redox_reactions	6.669989768	-0.534337543	0.742910829

MAG: bin_040
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Respiration|F0F1-type_ATP_synthase	171.6571627	-2.226040188	0.097764808
Regulation and Cell signaling|Autoinducer_2_(AI-2)_transport_and_processing_(lsrACDBFGE_operon)	394.3332103	-2.070390106	0.114108939
Carbohydrates|D-ribose_utilization	151.2744949	-1.922341069	0.146960395
Amino Acids and Derivatives|Polyamine_Metabolism	21.17847387	-1.917604519	0.203104981
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	5.079695083	-2.070320968	0.205158744
Protein Metabolism|Ribosome_SSU_bacterial	112.4045155	-1.550766287	0.23717607
Carbohydrates|L-rhamnose_utilization	4.156114159	-1.847581614	0.253831317
Protein Metabolism|Ribosome_LSU_bacterial	119.3470657	-1.45435513	0.263955473
Stress Response|Heat_shock_dnaK_gene_cluster_extended	5.334130159	1.595696366	0.332182286
Phosphorus Metabolism|Phosphate_metabolism	5.62106616	-1.555487173	0.344295337
Amino Acids and Derivatives|Proline,_4-hydroxyproline_uptake_and_utilization	2.770742773	-1.427284065	0.361873015
Secondary Metabolism|Auxin_biosynthesis	2.770742773	-1.427284065	0.361873015
RNA Metabolism|RNA_polymerase_bacterial	43.61454436	-1.238771434	0.366437329
Respiration|Respiratory_Complex_I	5.159275698	-1.456239502	0.376193937
Carbohydrates|TCA_Cycle	2.30895231	-1.256579365	0.41148414
DNA Metabolism|DNA-replication	2.30895231	-1.256579365	0.41148414
Membrane Transport|ABC_transporter_oligopeptide_(TC_3.A.1.5.1)	2.30895231	-1.256579365	0.41148414
Cell Wall and Capsule|Peptidoglycan_Biosynthesis	2.16548431	1.218444649	0.421415315
Amino Acids and Derivatives|Methionine_Degradation	4.235694774	-1.232471778	0.45326309
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	4.235694774	-1.232471778	0.45326309
Protein Metabolism|Proteolysis_in_bacteria,_ATP-dependent	50.5077859	1.004733666	0.454040994
Carbohydrates|Pentose_phosphate_pathway	1.847161848	-1.065031382	0.470906051
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	1.847161848	-1.065031382	0.470906051
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	1.847161848	-1.065031382	0.470906051
DNA Metabolism|Plasmid_replication	5.254549543	1.134902372	0.48994305
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	3.168645849	1.016178767	0.532033809
Carbohydrates|L-fucose_utilization_temp	1.385371386	-0.849853324	0.542428902

Cofactors, Vitamins, Prosthetic Groups, Pigments|Lipoic_acid_metabolism	1.385371386	-0.849853324	0.542428902
Cofactors, Vitamins, Prosthetic Groups, Pigments|Menaquinone_and_Phylloquinone_Biosynthesis	1.385371386	-0.849853324	0.542428902
Stress Response|Glutathione_analogs:_mycothiol	1.385371386	-0.849853324	0.542428902
DNA Metabolism|DNA_repair,_bacterial	10.34993786	0.933893047	0.552825253
Carbohydrates|Inositol_catabolism	6.703808315	-0.932666976	0.566608334
Amino Acids and Derivatives|Glycine_cleavage_system	1.082742155	0.703854065	0.592184719
RNA Metabolism|tRNA_processing	1.082742155	0.703854065	0.592184719
Carbohydrates|Carboxysome	0.923580924	-0.638319301	0.61841343
Carbohydrates|Fermentations:_Mixed_acid	0.923580924	-0.638319301	0.61841343
Fatty Acids, Lipids, and Isoprenoids|Isoprenoid_Biosynthesis	0.923580924	-0.638319301	0.61841343
Nucleosides and Nucleotides|De_Novo_Purine_Biosynthesis	0.923580924	-0.638319301	0.61841343
Potassium metabolism|Glutathione-regulated_potassium-efflux_system_and_associated_functions	0.923580924	-0.638319301	0.61841343
Virulence, Disease and Defense|Resistance_to_fluoroquinolones	0.923580924	-0.638319301	0.61841343
Amino Acids and Derivatives|Cysteine_Biosynthesis	5.780227391	-0.74043319	0.65102414
Amino Acids and Derivatives|Methionine_Biosynthesis	0.541371077	0.539483967	0.669041307
Carbohydrates|Isobutyryl-CoA_to_Propionyl-CoA_Module	0.541371077	0.539483967	0.669041307
Clustering-based subsystems|Cluster-based_Subsystem_Grouping_Hypotheticals_-_perhaps_Proteosome_Related	0.541371077	0.539483967	0.669041307
DNA Metabolism|RuvABC_plus_a_hypothetical	0.541371077	0.539483967	0.669041307
Nucleosides and Nucleotides|Nudix_proteins_(nucleoside_triphosphate_hydrolases)	0.541371077	0.539483967	0.669041307
Nucleosides and Nucleotides|Purine_conversions	0.541371077	0.539483967	0.669041307
Protein Metabolism|Protein_chaperones	0.541371077	0.539483967	0.669041307
Regulation and Cell signaling|Sex_pheromones_in_Enterococcus_faecalis_and_other_Firmicutes	0.541371077	0.539483967	0.669041307
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	11.17824494	-0.636857814	0.682355217
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	87.68865024	0.502895721	0.692448756
Amino Acids and Derivatives|Arginine_Biosynthesis_extended	2.085903694	0.608712251	0.698078558
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	3.853484927	-0.633436963	0.699917942

MAG: bin_040
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
DNA Metabolism|DNA_repair,_bacterial	18.14176572	2.623601146	0.078855365
Respiration|F0F1-type_ATP_synthase	84.03545278	-2.166578865	0.109564051
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	215.0677066	2.045445948	0.119937972
Protein Metabolism|Proteolysis_in_bacteria,_ATP-dependent	38.31291731	2.118237848	0.129189617
Carbohydrates|D-ribose_utilization	192.687828	-1.389716464	0.27503596
Regulation and Cell signaling|Autoinducer_2_(AI-2)_transport_and_processing_(lsrACDBFGE_operon)	191.8100007	-1.328971897	0.295074804
Amino Acids and Derivatives|Cysteine_Biosynthesis	16.85806118	1.489339641	0.302091106
Protein Metabolism|Proteasome_bacterial	2.303116981	1.386905803	0.345342146
Stress Response|Heat_shock_dnaK_gene_cluster_extended	10.8076596	1.273816284	0.392272651
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	1.727337735	1.137068279	0.420872641
Amino Acids and Derivatives|Glycine_cleavage_system	1.736776739	-1.103281317	0.435435615
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	1.302582555	-0.890897036	0.50713852
Virulence, Disease and Defense|MLST	1.302582555	-0.890897036	0.50713852
Amino Acids and Derivatives|Valine_degradation	1.15155849	0.836520157	0.522139315
Cell Wall and Capsule|Sialic_Acid_Metabolism	1.15155849	0.836520157	0.522139315
Potassium metabolism|Glutathione-regulated_potassium-efflux_system_and_associated_functions	1.15155849	0.836520157	0.522139315
RNA Metabolism|RNA_processing_and_degradation,_bacterial	1.15155849	0.836520157	0.522139315
Amino Acids and Derivatives|Polyamine_Metabolism	6.937667954	-0.929152728	0.543892825
Protein Metabolism|Ribosome_LSU_bacterial	118.6010855	-0.738215152	0.555653917
Nucleosides and Nucleotides|Ribonucleotide_reduction	56.53019507	0.755103042	0.558960926
Carbohydrates|Glycolysis_and_Gluconeogenesis	0.86838837	-0.643665879	0.598110875
Carbohydrates|TCA_Cycle	0.86838837	-0.643665879	0.598110875
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	0.86838837	-0.643665879	0.598110875
Phosphorus Metabolism|Phosphate_metabolism	0.86838837	-0.643665879	0.598110875
Carbohydrates|D-galactonate_catabolism	0.57577924	0.529172338	0.645174307
Carbohydrates|Isobutyryl-CoA_to_Propionyl-CoA_Module	0.57577924	0.529172338	0.645174307
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	0.57577924	0.529172338	0.645174307

Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	0.575779245	0.529172338	0.645174307
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_regulation	0.575779245	0.529172338	0.645174307
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	0.575779245	0.529172338	0.645174307
DNA Metabolism|Plasmid_replication	6.645058829	-0.540073848	0.723988248

MAG: bin_093
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Motility and Chemotaxis|Bacterial_Chemotaxis	14.48786751	-3.306297241	0.094898137
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	14.38257778	3.146722375	0.104568272
RNA Metabolism|Transcription_factors_bacterial	8.149425476	-2.883384842	0.147947051
Carbohydrates|Photorespiration_(oxidative_C2_cycle)	60.02216817	2.478360105	0.177739627
Amino Acids and Derivatives|Lysine_Biosynthesis_DAP_Pathway	4.141396237	2.649161005	0.18528642
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	16.94462799	2.39270082	0.205889494
Respiration|F0F1-type_ATP_synthase	14.10882447	-2.411058005	0.210239504
Nitrogen Metabolism|Ammonia_assimilation	5.131119744	-2.500917364	0.211081994
Iron acquisition and metabolism|Iron_acquisition_in_Vibrio	4.225628025	-2.332253233	0.243497307
Protein Metabolism|Ribosome_biogenesis_bacterial	3.621966878	-2.196143783	0.271609681
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	3.018305732	-2.033859429	0.307294217
Protein Metabolism|Universal_GTPases	3.018305732	-2.033859429	0.307294217
Amino Acids and Derivatives|Alanine_biosynthesis	4.44322681	1.999305772	0.313841804
Phosphorus Metabolism|Phosphate_metabolism	8.374043577	-1.884975604	0.332307857
Carbohydrates|Pyruvate_metabolism_I:_anaplerotic_reactions,_PEP	9.490114767	1.76835109	0.355193776
Fatty Acids, Lipids, and Isoprenoids|Fatty_Acid_Biosynthesis_FASII	2.414644586	-1.819370846	0.358368728
Nucleosides and Nucleotides|De_Novo_Purine_Biosynthesis	2.414644586	-1.819370846	0.358368728
Regulation and Cell signaling|Sex_pheromones_in_Enterococcus_faecalis_and_other_Firmicutes	2.414644586	-1.819370846	0.358368728
Protein Metabolism|Ribosome_LSU_bacterial	109.8944059	-1.42456639	0.424429103
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	21.64756948	-1.429590413	0.438512749
Carbohydrates|Butanol_Biosynthesis	9.805983971	-1.46121642	0.4447282
Nucleosides and Nucleotides|Hydantoin_metabolism	2.786668315	1.490632436	0.455943871

RNA Metabolism|tRNA_processing	2.786668315	1.490632436	0.455943871
Respiration|Respiratory_Complex_I	12.74707723	-1.352465325	0.472727653
Carbohydrates|D-ribose_utilization	1.509152866	-1.326886863	0.493389664
Cofactors, Vitamins, Prosthetic Groups, Pigments|Pyridoxin_(Vitamin_B6)_Biosynthesis	1.509152866	-1.326886863	0.493389664
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	1.509152866	-1.326886863	0.493389664
Stress Response|Rubrerythrin	15.08450934	-1.235845339	0.507501407
Amino Acids and Derivatives|Glycine_cleavage_system	0.828279247	1.124062994	0.529953646
DNA Metabolism|Plasmid_replication	0.828279247	1.124062994	0.529953646
Nitrogen Metabolism|Dissimilatory_nitrite_reductase	0.828279247	1.124062994	0.529953646
Amino Acids and Derivatives|Ketoisovalerate_oxidoreductase	1.207322293	-1.106888324	0.560758881
Cell Division and Cell Cycle|Macromolecular_synthesis_operon	1.207322293	-1.106888324	0.560758881
Miscellaneous|Conserved_gene_cluster_possibly_involved_in_RNA_metabolism	1.207322293	-1.106888324	0.560758881
RNA Metabolism|Queuosine-Archaeosine_Biosynthesis	1.207322293	-1.106888324	0.560758881
Stress Response|Glutathione:_Redox_cycle	1.207322293	-1.106888324	0.560758881
Sulfur Metabolism|Thioredoxin-disulfide_reductase	1.207322293	-1.106888324	0.560758881
Amino Acids and Derivatives|Methionine_Degradation	7.391339385	-1.119538735	0.561501991
DNA Metabolism|RuvABC_plus_a_hypothetical	1.958389068	1.107269802	0.57896459
Cofactors, Vitamins, Prosthetic Groups, Pigments|Heme_and_Siroheme_Biosynthesis	3.088498889	1.088212538	0.585133455
Protein Metabolism|Proteasome_bacterial	3.088498889	1.088212538	0.585133455
Clustering-based subsystems|CBSS-262719.3.peg.410	3.846584979	-1.021587969	0.606806897
Amino Acids and Derivatives|Proline_Synthesis	0.90549172	-0.845417428	0.646471596
Carbohydrates|Acetyl-CoA_fermentation_to_Butyrate	0.90549172	-0.845417428	0.646471596
Cell Wall and Capsule|Lipid_A-Ara4N_pathway_(_Polymyxin_resistance_)	0.90549172	-0.845417428	0.646471596
DNA Metabolism|DNA_repair,_bacterial	0.90549172	-0.845417428	0.646471596
Potassium metabolism|Potassium_homeostasis	0.90549172	-0.845417428	0.646471596
Stress Response|Glutathione_analogs:_mycothiol	0.90549172	-0.845417428	0.646471596
Protein Metabolism|Proteolysis_in_bacteria,_ATP-dependent	15.59691939	0.802086102	0.663424306
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_regulation	11.90475935	0.810246876	0.664145457
RNA Metabolism|RNA_polymerase_bacterial	13.19631343	-0.768057322	0.67993878
RNA Metabolism|Polyadenylation_bacterial	10.55705075	-0.752238238	0.689416317

Clustering-based subsystems|Conserved_gene_cluster_associated_with_Met-tRNA_formyltransferase	2.260219641	0.722126027	0.718059703
RNA Metabolism|Rrf2_family_transcriptional_regulators	2.260219641	0.722126027	0.718059703
Respiration|Coenzyme_F420_hydrogenase	2.260219641	0.722126027	0.718059703
Respiration|Succinate_dehydrogenase	2.94109326	-0.706525303	0.723360865
Nitrogen Metabolism|Nitric_oxide_synthase	4.822269855	0.68536778	0.725951126
Protein Metabolism|Ribosome_SSU_bacterial	68.91564111	-0.605977894	0.731827491
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	16.36202479	-0.601507098	0.743714564
Cofactors, Vitamins, Prosthetic Groups, Pigments|Biotin_biosynthesis	7.314126913	-0.613823431	0.748784921
Cell Wall and Capsule|Peptidoglycan_Biosynthesis	25.1080921	-0.578402372	0.749234194
Amino Acids and Derivatives|Arginine_Biosynthesis_extended	0.603661146	-0.538733235	0.756552137
Amino Acids and Derivatives|Valine_degradation	0.603661146	-0.538733235	0.756552137
Clustering-based subsystems|LMPTP_YwlE_cluster	0.603661146	-0.538733235	0.756552137
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	0.603661146	-0.538733235	0.756552137
DNA Metabolism|DNA_Repair_Base_Excision	0.603661146	-0.538733235	0.756552137
DNA Metabolism|DNA_repair,_UvrABC_system	0.603661146	-0.538733235	0.756552137
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	0.603661146	-0.538733235	0.756552137
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	0.603661146	-0.538733235	0.756552137
Protein Metabolism|Translation_elongation_factors_eukaryotic_and_archaeal	0.603661146	-0.538733235	0.756552137
Respiration|Anaerobic_respiratory_reductases	0.603661146	-0.538733235	0.756552137
Stress Response|Oxidative_stress	0.603661146	-0.538733235	0.756552137
Virulence, Disease and Defense|MLST	0.603661146	-0.538733235	0.756552137
Membrane Transport|pVir_Plasmid_of_Campylobacter	4.9766948	-0.602243782	0.758192896
Protein Metabolism|Protein_chaperones	3.692160035	0.596263387	0.763119581
Secondary Metabolism|Auxin_biosynthesis	3.692160035	0.596263387	0.763119581
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	1.130109821	0.545025769	0.778746283
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_F420_synthesis	1.130109821	0.545025769	0.778746283
Secondary Metabolism|Phenylpropionate_Degradation	1.130109821	0.545025769	0.778746283

MAG: bin_093
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Respiration|F0F1-type_ATP_synthase	19.68993521	-3.746319155	0.041462126
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_B12_biosynthesis	13.737164 1	-3.482060835	0.059466394
RNA Metabolism|Polyadenylation_bacterial	10.07392034	-3.227673973	0.082173947
Carbohydrates|Photorespiration_(oxidative_C2_cycle)	178.4422393	2.539224164	0.123737406
DNA Metabolism|DNA-replication	13.47298786	2.406557931	0.174987813
Motility and Chemotaxis|Bacterial_Chemotaxis	20.50007565	-2.104533605	0.221363333
Amino Acids and Derivatives|Cysteine_Biosynthesis	3.663243759	-2.267655384	0.225041558
Protein Metabolism|Universal_GTPases	11.62375424	-2.089338932	0.239715709
Fatty Acids, Lipids, and Isoprenoids|Fatty_Acid_Biosynthesis_FASII	3.205338289	-2.133206624	0.252615484
Membrane Transport|pVir_Plasmid_of_Campylobacter	3.205338289	-2.133206624	0.252615484
Phosphorus Metabolism|Phosphate_metabolism	10.25003783	-1.947494046	0.274992647
Nitrogen Metabolism|Nitric_oxide_synthase	2.747432819	-1.978459562	0.286537067
RNA Metabolism|RNA_polymerase_bacterial	12.62762392	-1.821548747	0.299188364
Nucleosides and Nucleotides|Purine_conversions	15.93863277	1.76977986	0.304832141
Nucleosides and Nucleotides|Ribonucleotide_reduction	40.52463409	-1.641618631	0.318844389
Carbohydrates|TCA_Cycle	11.25390751	-1.68414843	0.338944878
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	14.30074006	1.637206218	0.343918651
RNA Metabolism|Transcription_factors_bacterial	10.33809657	-1.580956035	0.370841556
Stress Response|Heat_shock_dnaK_gene_cluster_extended	104.3320001	1.417577928	0.374966117
Carbohydrates|Acetyl-CoA_fermentation_to_Butyrate	7.502605007	-1.582160542	0.380677292
Regulation and Cell signaling|Sex_pheromones_in_Enterococcus_faecalis_and_other_Firmicutes	4.279654969	1.62793851	0.381293841
Carbohydrates|D-ribose_utilization	1.831621 88	-1.583077904	0.383047943
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	18.95024175	1.461743454	0.388098004
Clustering-based subsystems|Cluster-based_Subsystem_Grouping_Hypotheticals_-_perhaps_Proteosome_Related	6.375453081	1.554889952	0.3929213
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_and_NADP_cofactor_biosynthesis_global	1.637892642	1.516752531	0.398845928
Iron acquisition and metabolism|Siderophore_Pyoverdine	1.637892642	1.516752531	0.398845928
Protein Metabolism|Selenocysteine_metabolism	1.637892642	1.516752531	0.398845928

Respiration|Respiratory_Complex_I	18.47472454	-1.374680604	0.417111263
Amino Acids and Derivatives|Lysine_Biosynthesis_DAP_Pathway	3.733690754	1.472771978	0.429769203
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	1.37371641	-1.321863484	0.453773942
DNA Metabolism|DNA_repair,_bacterial	1.37371641	-1.321863484	0.453773942
Metabolism of Aromatic Compounds|Aromatic_Amin_Catabolism	1.37371641	-1.321863484	0.453773942
Potassium metabolism|Potassium_homeostasis	1.37371641	-1.321863484	0.453773942
Protein Metabolism|Translation_elongation_factors_eukaryotic_and_archaeal	1.37371641	-1.321863484	0.453773942
RNA Metabolism|Rrf2_family_transcriptional_regulators	1.37371641	-1.321863484	0.453773942
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_regulation	8.929156663	1.283354965	0.468618202
Amino Acids and Derivatives|Methionine_Biosynthesis	3.751302503	-1.339391071	0.472475662
Stress Response|Oxidative_stress	3.751302503	-1.339391071	0.472475662
Protein Metabolism|Protein_chaperones	4.737560438	1.194115742	0.516828954
Secondary Metabolism|Phenylpropionate_Degradation	4.737560438	1.194115742	0.516828954
Amino Acids and Derivatives|Valine_degradation	5.213077657	-1.136203345	0.535287291
Protein Metabolism|Ribosome_LSU_bacterial	114.8990494	-0.968770827	0.538802781
Stress Response|Rubrerythrin	14.35357531	-1.027774709	0.54716024
Amino Acids and Derivatives|Histidine_Biosynthesis	8.594533435	-1.024307101	0.562667197
Stress Response|Glutathione_analogs:_mycothiol	2.641762326	1.082018492	0.563256905
Amino Acids and Derivatives|Chorismate:_Intermediate_for_synthesis_of_PAPA_antibiotics,_PABA,_anthranilate,_3-hydroxyanthranilate_and_more.	0.91581094	-0.957929731	0.565483721
Amino Acids and Derivatives|Glycine_cleavage_system	0.91581094	-0.957929731	0.565483721
Carbohydrates|Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utilization	0.91581094	-0.957929731	0.565483721
Clustering-based subsystems|Putative_hemin_transporter	0.91581094	-0.957929731	0.565483721
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_pathogenicity_islands_SaPI	0.91581094	-0.957929731	0.565483721
Virulence, Disease and Defense|Resistance_to_fluoroquinolones	7.291264021	1.014628707	0.570479437
Cofactors, Vitamins, Prosthetic Groups, Pigments|Folate_Biosynthesis	4.755172187	-1.02073113	0.57902599
Cofactors, Vitamins, Prosthetic Groups, Pigments|Thiamin_biosynthesis	2.835491564	-1.020944496	0.585430762
Cofactors, Vitamins, Prosthetic Groups, Pigments|Pyridoxin_(Vitamin_B6)_Biosynthesis	6.216947341	-0.927539116	0.607882574
Nitrogen Metabolism|Ammonia_assimilation	6.216947341	-0.927539116	0.607882574
Carbohydrates|Butanol_Biosynthesis	22.66632076	-0.830447997	0.615761062

DNA Metabolism|RuvABC_plus_a_hypothetical	5.195465908	0.893178628	0.624871801
Carbohydrates|Glycolysis_and_Gluconeogenesis	4.297266717	-0.892382972	0.629040351
Cofactors, Vitamins, Prosthetic Groups, Pigments|Menaquinone_and_Phylloquinone_Biosynthesis	4.297266717	-0.892382972	0.629040351
Protein Metabolism|tRNA_aminoacylation,_Glu_and_Gln	4.297266717	-0.892382972	0.629040351
Carbohydrates|Isobutyryl-CoA_to_Propionyl-CoA_Module	2.095798112	0.827170091	0.657661922
Nucleosides and Nucleotides|Purine_Utilization	2.095798112	0.827170091	0.657661922
Amino Acids and Derivatives|Polyamine_Metabolism	2.377586094	-0.823398661	0.659875866
Amino Acids and Derivatives|Glycine_and_Serine_Utilization	0.545964214	0.662912308	0.669333501
Cell Wall and Capsule|KDO2-Lipid_A_biosynthesis	0.545964214	0.662912308	0.669333501
Fatty Acids, Lipids, and Isoprenoids|Glycerolipid_and_Glycerophospholipid_Metabolism_in_Bacteria	0.545964214	0.662912308	0.669333501
Protein Metabolism|Ribosomal_protein_S12p_Asp_methylthiotransferase	0.545964214	0.662912308	0.669333501
Protein Metabolism|Signal_peptidase	0.545964214	0.662912308	0.669333501
Nucleosides and Nucleotides|De_Novo_Pyrimidine_Synthesis	5.301136401	-0.709614083	0.697014571
Protein Metabolism|Ribosome_biogenesis_bacterial	5.301136401	-0.709614083	0.697014571
Respiration|Succinate_dehydrogenase	5.301136401	-0.709614083	0.697014571
Protein Metabolism|Ribosome_SSU_bacterial	73.54666316	-0.584923031	0.710757333
Carbohydrates|Lacto-N-Biose_I_and_Galacto-N-Biose_Metabolic_Pathway	0.45790547	-0.531437633	0.732588324
Carbohydrates|Mannose_Metabolism	0.45790547	-0.531437633	0.732588324
Cell Division and Cell Cycle|Macromolecular_synthesis_operon	0.45790547	-0.531437633	0.732588324
Cofactors, Vitamins, Prosthetic Groups, Pigments|Lipoic_acid_metabolism	0.45790547	-0.531437633	0.732588324
DNA Metabolism|DNA_repair,_UvrABC_system	0.45790547	-0.531437633	0.732588324
Iron acquisition and metabolism|Iron_acquisition_in_Vibrio	0.45790547	-0.531437633	0.732588324
Membrane Transport|ABC_transporter_branched-chain_amino_acid_(TC_3.A.1.4.1)	0.45790547	-0.531437633	0.732588324
Membrane Transport|ABC_transporter_dipeptide_(TC_3.A.1.5.2)	0.45790547	-0.531437633	0.732588324
Metabolism of Aromatic Compounds|Central_meta-cleavage_pathway_of_aromatic_compound_degradation	0.45790547	-0.531437633	0.732588324
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	0.45790547	-0.531437633	0.732588324
Protein Metabolism|N-linked_Glycosylation_in_Bacteria	0.45790547	-0.531437633	0.732588324
RNA Metabolism|RNA_processing_and_degradation,_bacterial	0.45790547	-0.531437633	0.732588324
Respiration|Biogenesis_of_c-type_cytochromes	0.45790547	-0.531437633	0.732588324
Cell Wall and Capsule|Peptidoglycan_Biosynthesis	34.62469822	-0.538989923	0.738443547

Nucleosides and Nucleotides|De_Novo_Purine_Biosynthesis	3.381455778	-0.583856241	0.753918517
Carbohydrates|Pyruvate_metabolism_I:_anaplerotic_reactions,_PEP	4.103537448	0.574209031	0.755867899
DNA Metabolism|Plasmid_replication	5.107407164	0.513316797	0.778308064
Cofactors, Vitamins, Prosthetic Groups, Pigments|Riboflavin,_FMN_and_FAD_metabolism	1.549833898	0.508560673	0.782543217
DNA Metabolism|DNA_Repair_Base_Excision	1.549833898	0.508560673	0.782543217
Protein Metabolism|Proteasome_bacterial	1.549833898	0.508560673	0.782543217
Respiration|Anaerobic_respiratory_reductases	1.549833898	0.508560673	0.782543217
Sulfur Metabolism|Thioredoxin-disulfide_reductase	1.549833898	0.508560673	0.782543217
Virulence, Disease and Defense|Copper_homeostasis	1.549833898	0.508560673	0.782543217

MAG: bin_105
Mesocosm water-source: Nearshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Protein Metabolism|Glycine_reductase,_sarcosine_reductase_and_betaine_reductase	47	1.143984155	0.065694902
Cofactors, Vitamins, Prosthetic Groups, Pigments|Heme_and_Siroheme_Biosynthesis	15	1.12108083	0.083843703
Amino Acids and Derivatives|Polyamine_Metabolism	10	-0.831074677	0.196087313
Carbohydrates|Mannitol_Utilization	19.5	-0.805768351	0.204134953
DNA Metabolism|DNA_repair,_bacterial	4.5	0.692801758	0.230513871
Cell Division and Cell Cycle|Two_cell_division_clusters_relating_to_chromosome_partitioning	4.5	-0.692801758	0.230513871
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	14.5	0.737009628	0.254474804
RNA Metabolism|RNA_polymerase_bacterial	106.5	0.520924679	0.310756113
Membrane Transport|pVir_Plasmid_of_Campylobacter	6	-0.566933316	0.36554718

MAG: bin_105
Mesocosm water-source: Offshore

SEED Subsystem	baseMean	log2FoldChange	pvalue
Respiration|Ubiquinone_Menaquinone-cytochrome_c_reductase_complexes	104.2836528	1.094895454	0.061817644
Carbohydrates|Mannitol_Utilization	23.17778372	-1.116687419	0.081743618
Amino Acids and Derivatives|Branched-Chain_Amino_Acid_Biosynthesis	65.94330263	0.816487268	0.185113506
Phosphorus Metabolism|Phosphate_metabolism	27.69221016	-0.855275587	0.1917749
Carbohydrates|TCA_Cycle	50.69626658	-0.801886739	0.206831451

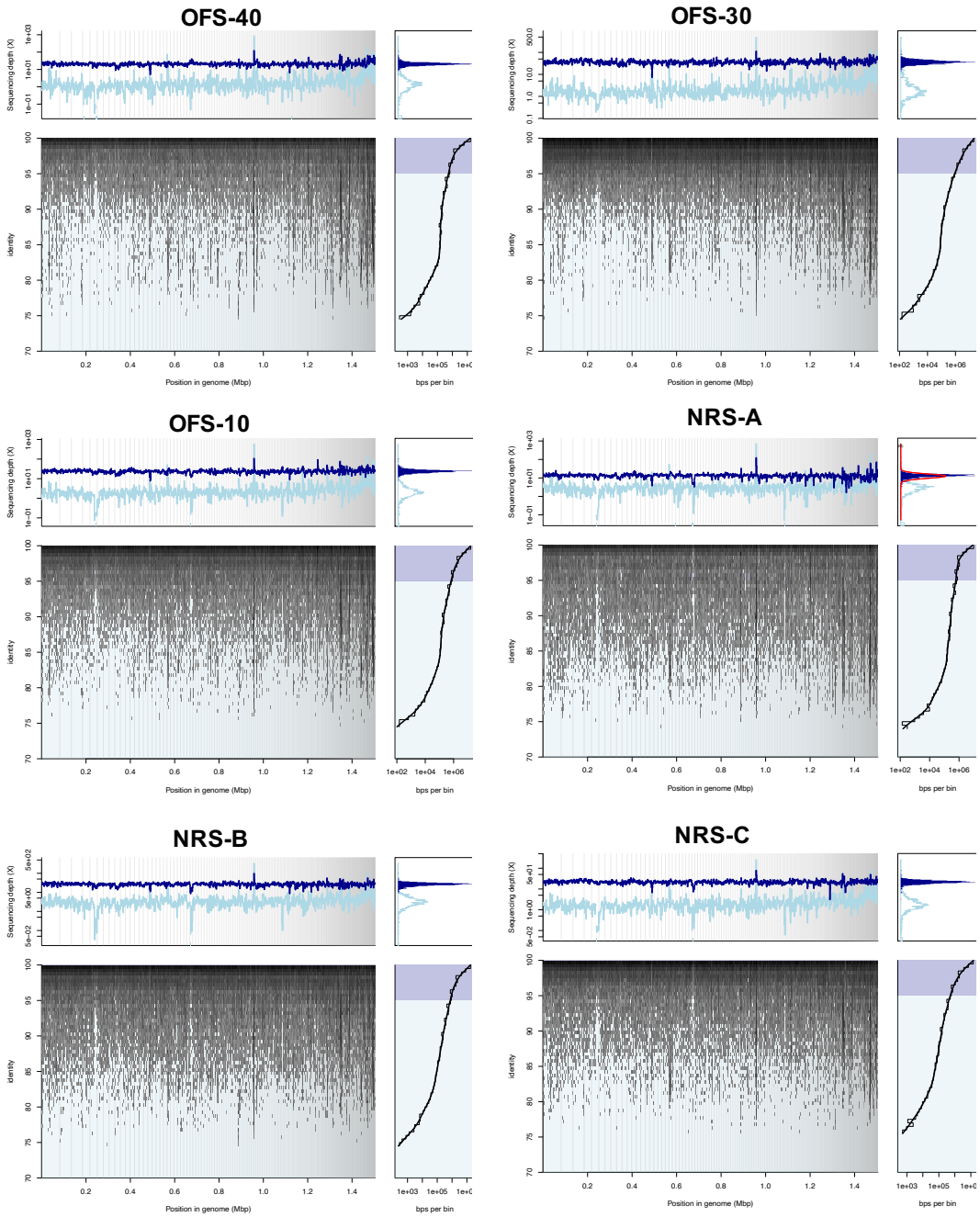
Amino Acids and Derivatives|Polyamine_Metabolism	35.52412411	-0.792093476	0.225352522
RNA Metabolism|RNA_polymerase_bacterial	71.76384607	0.727976336	0.2276538
Nucleosides and Nucleotides|De_Novo_Purine_Biosynthesis	26.72875972	-0.720539277	0.27163488
Cofactors, Vitamins, Prosthetic Groups, Pigments|Coenzyme_A_Biosynthesis	11.83473457	-0.64845006	0.278066465
DNA Metabolism|DNA_repair,_bacterial	16.7563768	-0.59389067	0.350699746
RNA Metabolism|Polyadenylation_bacterial	24.99071851	0.600078802	0.35928453
Membrane Transport|pVir_Plasmid_of_Campylobacter	21.67801903	-0.554353499	0.394992757
Cofactors, Vitamins, Prosthetic Groups, Pigments|Pyridoxin_(Vitamin_B6)_Biosynthesis	21.03252672	0.519956044	0.423695145

Supplementary Table S3.4. Differential expression of specific functional processes based on SEED annotations for MAG bin020 between the t-DOM and control mesocosms for nearshore lake-water at 19h using DESeq2. A positive value in the column 'log2FoldChange' for a specific functional process indicates a higher expression for that function in the t-DOM sample as compared to the control, and a negative value indicates the opposite. Only those functional processes that are different between the control and t-DOM mesocosms by a fold-change of 0.5 or more are shown. Functional processes that are significantly differentially expressed (Wald test, P value < 0.05) between the control and t-DOM mesocosm are highlighted in yellow.

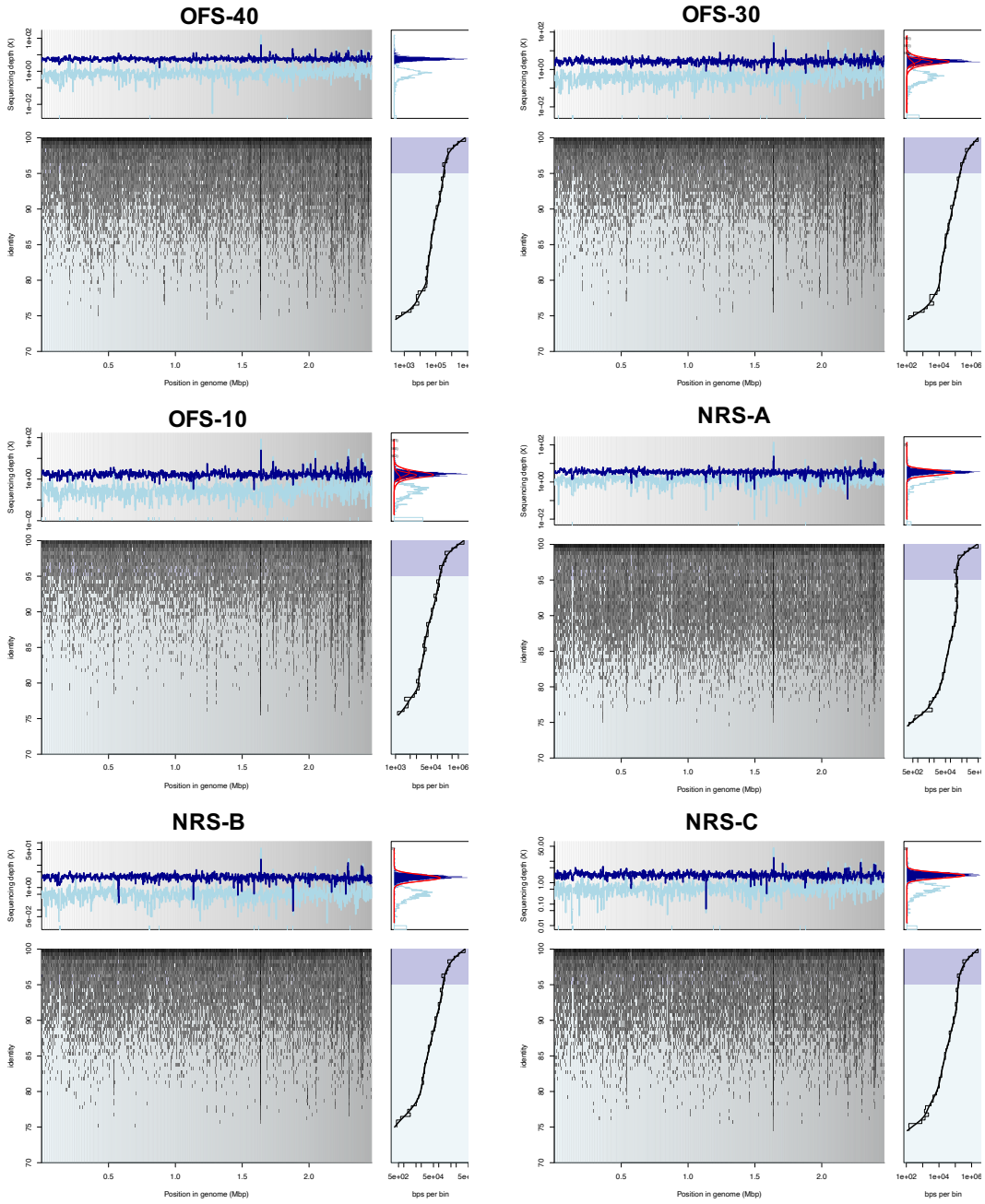
SEED Subsystem	baseMean	log2FoldChange	pvalue
Protein Metabolism|Proteasome_bacterial	270.8993732	1.218316859	0.009706029
Carbohydrates|TCA_Cycle	283.2394522	1.02997452	0.023326327
Stress Response|Heat_shock_dnaK_gene_cluster_extended	1010.340198	1.14445971	0.027389924
Protein Metabolism|GroEL_GroES	1334.601125	1.11639979	0.032399958
Amino Acids and Derivatives|Methionine_Degradation	65.97273283	0.964736786	0.048518885
DNA Metabolism|DNA_repair,_bacterial_RecFOR_pathway	82.58514992	0.862359249	0.064730569
Protein Metabolism|Translation_elongation_factors_eukaryotic_and_archaeal	9.317813875	-0.921872248	0.066590208
Amino Acids and Derivatives|Proline_Synthesis	128.9483025	0.786498512	0.074481385
RNA Metabolism|tRNA_processing	64.60719114	0.853080055	0.077863161
Clustering-based subsystems|Bacterial_Cell_Division	298.8829371	0.706403709	0.095999324
Iron acquisition and metabolism|Transport_of_Iron	42.97942057	0.828498515	0.099572579
Cofactors, Vitamins, Prosthetic Groups, Pigments|NAD_regulation	14.247821	-0.820914747	0.115920225
Amino Acids and Derivatives|Lysine_Biosynthesis_DAP_Pathway	37.44194821	-0.787389937	0.120048131
Metabolism of Aromatic Compounds|Protocatechuate_branch_of_beta-ketoadipate_pathway	13.5951724	0.800732885	0.125037034
Carbohydrates|Acetyl-CoA_fermentation_to_Butyrate	94.18221352	0.682999662	0.125070294
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_pathogenicity_islands_SaPI	24.36387432	0.773510696	0.133952477
Amino Acids and Derivatives|Methionine_Salvage	40.45417252	-0.727944214	0.146948648
Virulence, Disease and Defense|MLST	9.955401355	-0.715268531	0.164128114
Carbohydrates|Propionyl-CoA_to_Succinyl-CoA_Module	17.26004531	-0.702392379	0.177375632
Cell Wall and Capsule|Peptidoglycan_Biosynthesis	87.10850676	-0.581656285	0.190914233
Membrane Transport|pVir_Plasmid_of_Campylobacter	33.78711604	-0.659544734	0.192206222
Cell Wall and Capsule|Alginate_metabolism	5.115760958	-0.562640609	0.197978557
Cofactors, Vitamins, Prosthetic Groups, Pigments|Lipoic_acid_metabolism	18.26412008	-0.666898761	0.199254739

Amino Acids and Derivatives|HMG_CoA_Synthesis	33.85740128	0.613178583	0.223698683
Cell Wall and Capsule|mycolic_acid_synthesis	105.6889104	0.512478564	0.230648638
Amino Acids and Derivatives|Chorismate_Synthesis	59.68722476	0.555594325	0.241434238
Phages, Prophages, Transposable elements, Plasmids|Staphylococcal_phi-Mu50B-like_prophages	30.5891379	-0.526038245	0.298063348

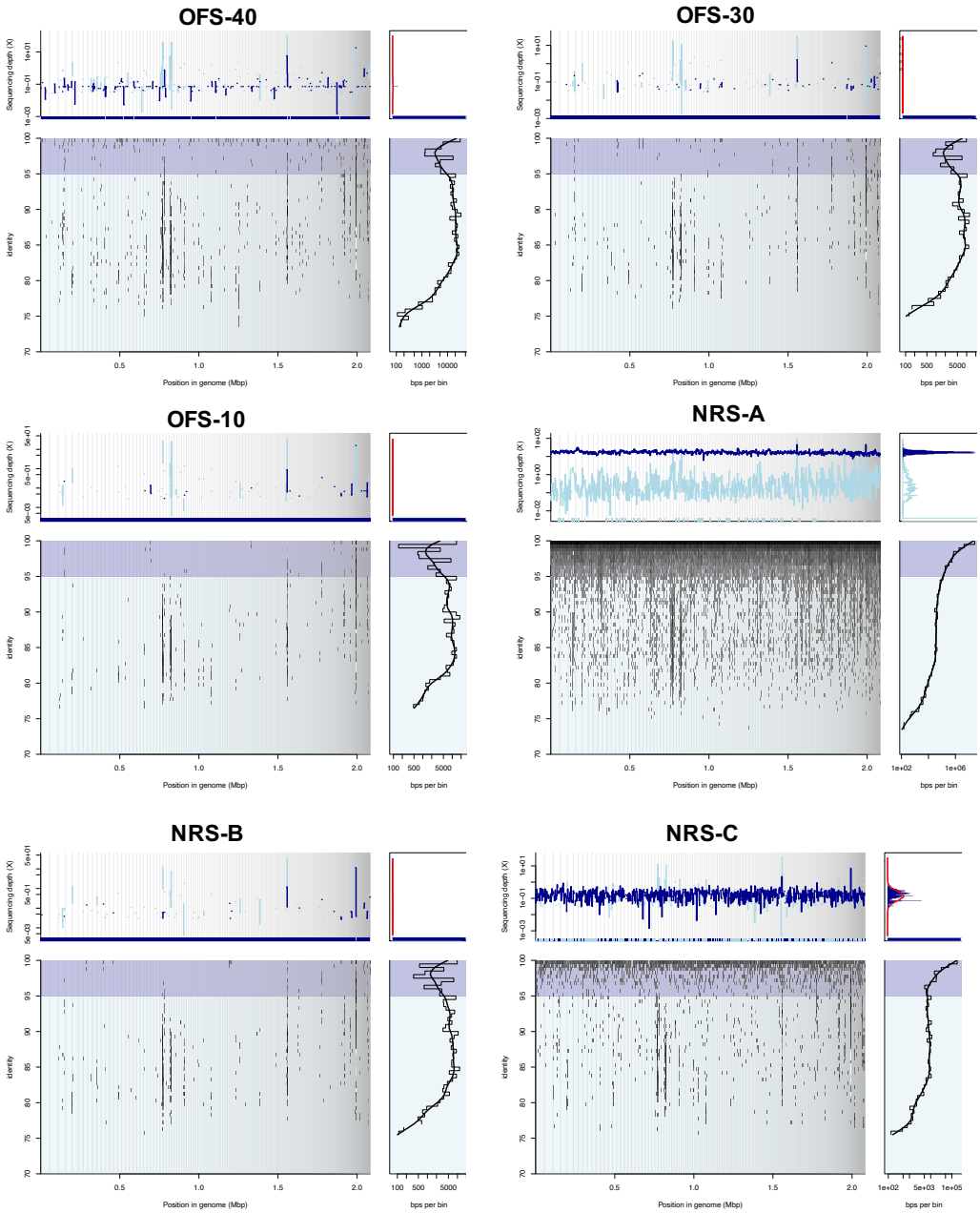
C. Supplementary Materials – Chapter 4



Supplementary Figure S4.1. Read recruitment plots for MAG-based population LMS_bin009 (*Polynucleobacter*) in nearshore (NRS) and offshore (OFS) southern Lake Michigan. The coverage histogram (top left) in each plot shows coverage for the MAG in the corresponding Lake Michigan metagenome from reads that match at ≥ 95% nucleotide identity and ≥ 70 bp in length (dark blue) as well as reads that match at ≥ 70 bp in length and < 95% nucleotide identity (light blue). The recruitment plots (bottom left) show the individual reads mapping to the MAG at each position in the genome. The consistently high coverage of the MAG in both nearshore and offshore metagenomes at high identity (dark blue) can be seen.



Supplementary Figure S4.2. Read recruitment plots for MAG-based population LMS_bin056 (*Cytophagales*) in nearshore (NRS) and offshore (OFS) southern Lake Michigan. The coverage histogram (top left) in each plot shows coverage for the MAG in the corresponding Lake Michigan metagenome from reads that match at ≥ 95% nucleotide identity and ≥ 70 bp in length (dark blue) as well as reads that match at ≥ 70 bp in length and < 95% nucleotide identity (light blue). The recruitment plots (bottom left) show the individual reads mapping to the MAG at each position in the genome. The consistently high coverage of the MAG in both nearshore and offshore metagenomes at high identity (dark blue) can be seen.



Supplementary Figure S4.3. Read recruitment plots for MAG-based population LMS_bin181 (*Fluviicola*) in nearshore (NRS) and offshore (OFS) southern Lake Michigan. The coverage histogram (top left) in each plot shows coverage for the MAG in the corresponding Lake Michigan metagenome from reads that match at ≥ 95% nucleotide identity and ≥ 70 bp in length (dark blue) as well as reads that match at ≥ 70 bp in length and < 95% nucleotide identity (light blue). The recruitment plots (bottom left) show the individual reads mapping to the MAG at each position in the genome. The low/insufficient coverage of the MAG in the offshore metagenomes in comparison to nearshore metagenomes can be seen.

VI. VITA

Adit Chaudhary

EDUCATION

University of Illinois at Chicago, Illinois 2014 – present
Ph.D. Biological Sciences

BITS Pilani, KK Birla Goa Campus, India 2007 – 2012
B.E. Electronics and Instrumentation + M.Sc. Biological Sciences

RESEARCH EXPERIENCE

University of Illinois at Chicago, USA
Doctoral Research 2014 – present
Advisor: Dr. Rachel Poretsky
Multi-omics-based investigation of the impact of environmental perturbations on aquatic microbial community structure and function.

The Energy and Resources Institute, India
Project Trainee 2013 – 2014
Advisors: Dr. Manab Das and Dr. Shanuja Beri
Sampling, isolation and functional characterization of plant growth promoting rhizobacteria associated with sugarcane crops grown in alkaline soil conditions.

BITS Pilani, KK Birla Goa Campus, India
Master's Thesis Research 2012
Advisor: Dr. Judith M. Braganca
Evaluating cadmium tolerance in haloarchaeal strains isolated from solar salterns of Goa, India.

PUBLICATIONS

Chaudhary A, Turner S, Macam R, Poretsky R. Bacterioplankton response to allochthonous dissolved organic matter across a coastal to offshore transect in Lake Michigan. Limnol Oceanogr. (*In revision*)

Chaudhary A, Kauser I, Ray A, Poretsky R. 2018. Taxon-driven functional shifts associated with storm flow in an urban stream microbial community. mSphere. 3: e00194-18.

Binh CTT, Petrovich ML, **Chaudhary A**, Wright D, Murphy BT, Wells G, Poretsky R. 2018. Metagenomics analysis reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. Appl Environmental Microbiol. 84: e02168-17.

SELECTED CONFERENCE PRESENTATIONS

Chaudhary A, Turner S, Macam R, Poretsky R. "Bacterioplankton response to allochthonous dissolved organic carbon across a coastal to offshore transect in Lake Michigan" at ASM Microbe, San Francisco, California, 2019 [*Poster*]

Chaudhary A, Kauser I, Ray A, Poretsky R. 2018. "Stormflow associated taxa-driven functional shifts in an urban stream microbial community" at ASM Microbe, Atlanta, Georgia, 2018 [*Oral Presentation*]

Chaudhary A, Turner S, Macam R, Poretsky R. "Lake Michigan bacterioplankton metagenomics and response to allochthonous dissolved organic matter" at the 60th IAGLR conference, Detroit, Michigan, 2017 [*Poster*]

Chaudhary A, Turner S, Macam R, Poretsky R. "Freshwater microbial community response to autochthonous and allochthonous dissolved organic carbon" at the 16th International Symposium on Microbial Ecology, Montreal, Canada, 2016 [*Poster*]

Chaudhary A, Kauser I, Ray A, Poretsky R. "Impact of rainfall induced combined sever overflow event on the urban stream microbiome" at Microbes in the City: Mapping the Urban Genome, New York Academy of Sciences, NY, 2015. [*Poster*]

TEACHING AND MENTORING EXPERIENCE

Teaching Assistant, University of Illinois at Chicago 2014 – present

Microbiology Laboratory (BIOS 351)
Microbiology Lecture (BIOS 350)
Cell Biology Laboratory (BIOS 312)

Undergraduate Research Mentor:

Jennifer Arista (2015, Honors College Capstone Project)
Hillary Pham (2015, Honors College Capstone Project)
Rachel Macam (2015-16)

GRANTS AND AWARDS

- **Biological Sciences Graduate Research Award**, University of Illinois at Chicago: 2019
- **Outstanding Abstract Award**: ASM Microbe 2018
- **Elmer Hadley Research Grants**, University of Illinois at Chicago: 2017-2019 ($8865)
- **Illinois-Indiana SeaGrant** Research Assistantship: 2015-2016
- **Graduate Student Travel Awards**, University of Illinois at Chicago: 2015-2019

TECHNICAL SKILLS

- **Omics techniques:** Extensive experience with sample processing and library preparation for environmental metagenomics and metatranscriptomics
- **Programming languages and statistical software:** Perl, Unix shell scripting, R
- **Knowledge of multiple bioinformatics tools and databases including:** QIIME, Velvet, Newbler, Megahit, SPAdes, DIAMOND, BLAST, BLAT, MetaGeneMark, Swiss-Prot, Gene Ontology, SEED, GenBank, MaxBin, CheckM

FIELDWORK EXPERIENCE

2015 – 2018	R/V Lake Guardian Doctoral research project studying bacterial community dynamics across a nearshore-to-offshore transect in southern Lake Michigan.
2013	Ugar Khurd, Karnataka, India I was part of the Water4Crops project team at TERI that studied rhizospheric bacteria associated with sugarcane grown in alkaline soil conditions.

SERVICE AND OUTREACH

Lake Michigan CSMI 2020 Workshop, University of Wisconsin-Milwaukee, October 2018: Recommendations for research and monitoring for the next Lake Michigan CSMI intensive field year 2020

Professional societies

- American Society for Microbiology
- Water Environment Federation