THE UNIVERSITY OF CHICAGO

UNRAVELING MICROBIAL CONTRIBUTIONS TO NITROGEN AND CARBON CYCLING IN THE LAURENTIAN GREAT LAKES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
DEPARTMENT OF THE GEOPHYSICAL SCIENCES

BY
JUSTIN CONNOR PODOWSKI

CHICAGO, ILLINOIS
JUNE 2022

TABLE OF CONTENTS

2.7 Supplemental Text

2.8 Supplemental Figures

2.9 Supplemental Data

4.7 Supplemental Figures

4.8 Supplemental Data

Supplemental Files Are Available Online

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Abstract

Biogeochemical cycles in aquatic systems are largely controlled by single cell microbes that carry out oxidation-reduction reactions either to generate energy or to convert compounds to biologically accessible oxidation states. These transformations are key to regulating availability of nutrients that limit growth and primary productivity, but the rates of these transformations and the microbes responsible for them are not well characterized, especially in large freshwater systems. In order to understand biogeochemical transformations and the microbes responsible for them in the Laurentian Great Lakes, we surveyed microbial communities using amplicon and whole community sequencing, and stable isotope probing to determine rates of nitrogen oxidation.

First, we used amplicon and whole community sequencing to survey the taxonomic identity and metabolism of nitrogen oxidizing microbes (nitrifiers) in the Laurentian Great Lakes. This survey showed that taxonomic composition of nitrifier communities was distinct across lakes, with nitrite oxidizers showing particularly dynamic distribution potentially related to substrate availability. We demonstrated that nitrifiers in the Great Lakes have streamlined genomes as compared to soil relatives, and contain metabolic adaptations that allow for access to organic nitrogen sources. We found ammonia oxidizing bacteria strains with adaptations to photic environments including proteorhodopsin, a novel protein in nitrifiers which may suggest the ability to generate energy from light. These results suggest specific adaptations to this large

freshwater ecosystem, and highlight the potential importance of organic nitrogen to nitrification in the Great Lakes.

Next, we employed stable isotope tracer methods to measure rates of nitrification, along with further whole community sequencing. We found that Lake Erie was distinct from the other Great Lakes in many respects. Rates of nitrification were up to an order of magnitude higher, importance of urea relative to ammonium was considerably lower, and nitrification is in some cases associated with a larger size fraction. Consistent with this, we find that the genomic identity of nitrifiers in Lake Erie are distinct from the other Great Lakes. Together this work confirms predictions made using previous whole community DNA sequencing, and established Lake Erie as a distinct biogeochemical regime with respect to nitrogen cycling.

Finally, we employed whole community DNA sequencing to survey the taxonomy and metabolism of autotrophs in the Great Lakes. We discovered a surprising number of carbon fixation genes associated with order Burkholderiales, and this taxonomic group contributed the majority of carbon fixation genes in the surface and deep chlorophyll layer of most of the Great Lakes. We reconstructed the metabolism of these microbes which demonstrated that these carbon fixation genes are associated with microbes that also contain sulfur oxidation and anoxygenic aerobic phototrophy genes. We further demonstrated that carbon fixation genes and organization of the carbon fixation operon is not congruent with the overall genome phylogeny, suggesting multiple acquisitions of carbon fixation genes. Overall these results emphasize the potential importance of chemoautotrophs to organic carbon inventories, and highlight the need for measurement of dark carbon fixation in the Great Lakes.

1.2 Introduction

The Laurentian Great Lakes make up 20% of the Earth's unfrozen freshwater and provide drinking water and ecosystem services to more than 30 million people. The health of the Great Lakes as an ecosystem is currently facing a number of threats, many directly anthropogenic in nature. Concentrations of nitrate in Lake Superior, Michigan, Huron and Ontario have been on the rise since at least the 1970's (Finlay, Small, and Sterner 2013), while well documented the sources of nitrate are enigmatic. A recent surge in anthropogenic eutrophication in the form of nitrogenous inputs into Lake Erie (Scavia et al. 2014) has driven a resurgence in algal blooms (Michalak et al. 2013) whose intensity is likely to worsen in the face of increasing global temperatures. More broadly, temperature increases felt throughout the lakes have driven a decrease in average ice cover (Wang et al. 2012) and water temperature increases felt even in the deepest waters in Lake Michigan (Anderson et al. 2021). Longer ice free periods and higher temperatures have already led to increases in primary productivity in Lake Superior (O'Beirne et al. 2017). And longer periods of stratification brought on by warming (Anderson et al. 2021) could lead to depletion of hypolimnetic oxygen (Butcher et al. 2015) and mobilization of phosphorus from the sediment, a dynamic already seen in Lake Erie's central basin (Scavia et al. 2014). Though the causes of these large scale changes are varied, these changes are mediated through changing nutrient concentrations and felt first at the base of the food web. To understand both the short and long term consequences of these large scale changes, we must understand how organisms at the base of the Great Lakes food web respond to these changes, and how these organisms interact with and process the nutrients through which these changes are mediated.

Single celled bacterial and archaeal microorganisms- or microbes- are the base of the aquatic food web and carry out virtually all the oxidation-reduction reactions present in so called 'biogeochemical cycles'- the transformations of elements such as nitrogen, carbon and sulfur from one oxidation state to another (Falkowski, Fenchel, and Delong 2008). Transformation of these elements from one oxidation state to another controls their biological availability, such as the reduction of atmospheric dinitrogen by diazotrophic cyanobacteria to ammonium which is biologically available and can then be accessed by other members of the aquatic community (Ohlendieck, Stuhr, and Siegmund 2000). Understanding the capacity of microbes in an ecosystem to perform these transformations informs predictions of how an ecosystem may respond to changing availability of nutrients. An ecosystem which lacks diazotrophic microbes would not be able to access atmospheric dinitrogen as a source of bioavailable nitrogen, even in the face of scarcity of bioavailable nitrogen. Understanding the limitations and capacities of microbes to respond to changing availability of nutrients is especially important in the face of anthropogenic inputs. Two ecosystems with different microbial communities may respond to the same set of anthropogenic inputs in different ways. Inputs of urea, a common anthropogenic nitrogen input (Glibert et al. 2006) may in some instances promote algal blooms and growth of bloom forming phytoplankton (Ivey et al. 2020; Belisle et al. 2016). However, metabolic processes present in some ecosystems (Kitzinger et al. 2019) allow for the oxidation of urea into nitrate which is less available to bloom forming algae. A complete understanding of how an ecosystem may respond to anthropogenic perturbations requires an understanding of the metabolic capacities of the microbes that make up the base of the food web. And, answers as to why two ecosystems respond differently to the same anthropogenic perturbation may lie in the metabolic capacities of the microbes that live in one ecosystem but not the other.

1.3 Changing Nutrient Cycling in the Great Lakes

The impacts of large scale anthropogenic influences on the Great Lakes are felt most directly through inputs and changes to the nitrogen cycle, and to a lesser extent the carbon cycle.

As in many large lakes around the world, all of the Great Lakes with the exception of Lake Erie are experiencing increases in nitrate concentrations (Finlay, Small, and Sterner 2013). Broadly these increases in nitrate in large lakes across the globe have been explained by decreases in anthropogenic phosphorus inputs without a concomitant decrease in nitrogen inputs (Bernhardt 2013). However, specific investigation into increases in nitrate in Lake Superior provide several reasons why this dynamic may not be so simple. While evidence of increasing nitrate in large lakes globally begins in the 1970’s (Finlay, Small, and Sterner 2013), nitrate in Lake Superior has been increasing for more than a century (Sterner et al. 2007), predating both the rise of anthropogenic phosphorus inputs, and the tapering of those inputs following the establishment that phosphorus inputs largely drove eutrophication in lakes (Schindler 1974). Investigation into the accumulating nitrate in Lake Superior has also demonstrated that the nitrate was not deposited in the form of nitrate from the atmosphere, but instead was deposited in some reduced form and oxidized in lake to nitrate (Finlay, Sterner, and Kumar 2007). This finding suggests that not only is deposition of nitrogen potentially increasing, but also that an in lake, microbially driven process compensates to facilitate this increase. And while decreasing ability to bury nitrate in the sediment or incorporate it into biomass may also be a compelling explanation for nitrate increases, recent evidence that primary productivity (and consequently the sinks in which to store nitrate) has increased in Lake Superior over the last century (O’Beirne et al. 2017) cast some doubt on this explanation. Potential sources of ammonium inputs remain unclear, as does the degree to which lessons learned from Lake Superior translate to Lake Michigan, Huron and Ontario.

Nitrogen cycling dynamics in Lake Erie are distinct from the rest of the Great Lakes, due to many anthropogenic and physical factors. Inputs of reduced nitrogen to Lake Erie are greater, both allochthonous (Robertson and Saad 2011) and autochthonous (Natwora and Sheik 2021). Nitrogen outputs are also greater, largely as a result of higher productivity in Lake Erie. Unlike Lake Superior, anoxic sediments in Lake Erie act as a sink of nitrate through denitrification (Small et al. 2014) which allows nitrogen to leave the system as a gas in the form of N_2O or N_2. Water column anoxia is now frequent in Erie's central basin (Scavia et al. 2014) but stable anoxic regions are rare in the other Great Lakes (Tellier et al. 2021). The greater ability to remove nitrogen from the system means that despite much greater inputs as compared to the other Great Lakes, Lake Erie has lower nitrate concentrations (Dove and Chapra 2015) and has not seen increases in nitrate like the other Great Lakes. That Lake Erie responds differently to anthropogenic inputs than the other Great Lakes is a testament to how metabolic capacity of microbial communities determines how an ecosystem responds to perturbation. A better understanding of the metabolic capacity of microbes in Lake Erie and how they contrast with microbes in the other Great Lakes may further illuminate why Lake Erie has a greater capacity to remove anthropogenic nitrogen.

With respect to organic carbon availability and the rate of carbon fixation (primary productivity), the long established dynamic is of much higher availability of and production of organic carbon in Lake Erie and Ontario than Lake Superior, Michigan and Huron (Vollenweider, Munawar, and Stadelmann 1974). In Lake Superior, primary productivity appears to have been increasing over the last century, but especially so since 1950 (O'Beirne et al. 2017). In contrast, primary productivity in Lake Ontario and Erie peaked in the 1970's upon abatement of phosphorus inputs and has since declined (Schelske and Hodell 1995). An understanding of the differences in carbon fixing microbes between Superior and Erie and Ontario may illuminate why these lakes have taken different trajectories with respect to rates of primary productivity. A great deal of research has been dedicated to how phytoplankton community composition changes across the Great Lakes (Vollenweider, Munawar, and Stadelmann 1974; Bramburger and Reavie 2016)**.** Less effort has been devoted to description of smaller microbial carbon fixers, which may be especially important in Lake Superior where the smallest organisms are implicated in performing the majority of carbon fixation (Munawar et al. 1978; Fahnenstiel et al. 1986).

1.4 Microbial contributions to nitrogen and carbon cycling in the Great Lakes

While the importance of microbes in the Great Lakes has been long recognized (Munawar et al. 1994; Reed and Hicks 2011), measurements of the processes performed by microbes are more common than taxonomic or functional characterization of microbial communities. Primary productivity measurements investigating what size of organisms perform what fraction of total carbon fixation demonstrated the overwhelming importance of autotrophic bacteria to primary productivity (Fahnenstiel et al. 1986; Munawar et al. 1978; 1994), and that the importance of autotrophic bacteria and the 'microbial' food web is greater in oligotrophic regions than in eutrophic ones (Heath, Hwang, and Munawar 2010; Biddanda, Ogdahl, and Cotner 2001). A focus on microbially mediated transformations of nitrogen both recently (Small et al. 2013; Clevinger, Heath, and Bade 2014) and post recovery from heavy pollution in the 1970's (Lean and Knowles 1987) demonstrates that oxidation of nitrogen in Lake Superior is considerably slower than in more productive Erie and Ontario (Lean and Knowles 1987; Clevinger, Heath, and Bade 2014; Small et al. 2013). Though these measurements are instrumental to understanding general trends in rates of this microbially mediated process, inconsistent methodologies and the use of proxies in some cases complicates comparisons. Further, a focus on the process itself without subsequent investigation into the differences between microbial communities and their functions limits potential explanatory power for differences in the rates of this process.

More recent study of microbes in the Great Lakes has focused on broadly describing the taxonomic makeup of microbial communities, or on specifically characterizing certain functional groups. Study of microbial community taxonomy has demonstrated that microbes in the surface water of the Great Lakes are distinct from those that live deeper in the water column (Rozmarynowycz et al. 2019) and that communities in Lake Superior, Michigan and Huron are distinct from those in Erie and Ontario (Paver, Newton, and Coleman 2020). Functional surveys have demonstrated that within these distinct deep water communities, degradation of recalcitrant organic carbon may be important to sustain growth (Denef et al. 2016). Targeted surveys of microbes involved with the oxidation of nitrogen have emerged alongside measurements of the rates of this oxidation of nitrogen, suggesting that these taxonomic makeup of this functional groups is distinct between Lake Erie and Lake Superior, both in the water column (Mukherjee et al. 2016; Small et al. 2013) and sediment (Bollmann, Bullerjahn, and McKay 2014). However these surveys are limited by their methodologies, which use taxon specific probes to detect these microbes. This method limits the discovery of any unexpected potential members of this functional group. As a result, untargeted taxonomic surveys of the Great Lakes at large appear to indicate that important members of this functional group were missed, and others which were detected appear to considerably less abundant than implied by targeted methods (Paver, Newton, and Coleman 2020; Fujimoto et al. 2016; Rozmarynowycz et al. 2019).

1.5 Nitrification, the microbially mediated oxidation of ammonium to nitrate

Nitrification is the microbially mediated oxidation of ammonium to nitrate, an energetically favorable reaction which provides energy to the organisms that perform this process. Nitrification is a two-step process, the oxidation of ammonium to nitrite, and then the oxidation of nitrite to nitrate. This process is largely carried out by two distinct groups of microbes referred to as nitrifiers (Costa, Pérez, and Kreft 2006), though a taxonomically limited group of microbes exist that carry out both parts of this process (Daims et al. 2015; Kits et al. 2017). These two distinct groups are called ammonia oxidizers (oxidizing ammonia to nitrite) and nitrite oxidizers (oxidizing nitrite to nitrate). Nitrifiers are important members of aquatic ecosystems as they are responsible for the removal of ammonia (Kuypers, Marchant, and Kartal 2018) and generation of nitrate, which is less available as a source of nitrogen to photosynthetic microbes (Berube et al. 2015). Nitrifiers are chemoautotrophs, using energy from the oxidation of nitrogen to reduce inorganic carbon, contributing in aphotic environments a large fraction of total carbon fixation (Pachiadaki et al. 2017). While unified by their ability to oxidize nitrogen and fix carbon, nitrifiers contain diverse metabolic content allowing them to access organic nitrogen (Palatinszky et al. 2015; Hanna Koch et al. 2015) and even reduced inorganic compounds beyond nitrogen (Füssel et al. 2017; H. Koch et al. 2014) as energy sources. As these metabolisms are not distributed throughout all nitrifiers, describing both the taxonomy and functional capacity of nitrifiers in an ecosystem is vital to understanding how an ecosystem may respond to input of both reduced nitrogen and other reduced inorganic compounds.

Well studied especially in marine systems (Francis et al. 2005; Alves et al. 2018; Daims, Lücker, and Wagner 2016), more recent attention has focused on the importance of taxonomy and distribution of nitrifiers in freshwater ecosystems. Distribution of ammonia oxidizing taxa appears to be driven by ecosystem productivity (Hugoni et al. 2013), with ammonia oxidizing bacteria dominant in high productivity environments and ammonia oxidizing archaea dominant in low productivity environments. This relationship to productivity is likely a result of higher reduced nitrogen in higher productivity ecosystems, and is a dynamic seen in marine and soil ecosystems (Hatzenpichler 2012) and related to differences in enzyme kinetics and growth rate in archaeal and bacterial ammonia oxidizers (Schleper 2010; Martens-Habbena et al. 2009). A similar clear cut dynamic has not been established in freshwater nitrite oxidizers, but bacteria in genera *Nitrospira* (Alfreider et al. 2018) and *Ca*. Nitrotoga (Boddicker and Mosier 2018) appear globally important in lakes, and *Nitrobacter* may also play a role (Mukherjee et al. 2016; Ward, Arp, and Klotz 2011). Functional capacity of freshwater nitrifiers is poorly understood, and is largely inferred from the functional capacity of isolated nitrifiers.

1.6 References

Alfreider, Albin, Victoria Grimus, Martin Luger, Anja Ekblad, Michaela M Salcher, and Monika Summerer. 2018. "Autotrophic Carbon Fixation Strategies Used by Nitrifying Prokaryotes in Freshwater Lakes." *FEMS Microbiology Ecology* 94 (10). https://doi.org/10.1093/femsec/fiy163.

Alves, Ricardo J. Eloy, Bui Quang Minh, Tim Urich, Arndt von Haeseler, and Christa Schleper. 2018. "Unifying the Global Phylogeny and Environmental Distribution of Ammonia-Oxidising Archaea Based on AmoA Genes." *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-03861-1.

Anderson, Eric J., Craig A. Stow, Andrew D. Gronewold, Lacey A. Mason, Michael J. McCormick, Song S. Qian, Steven A. Ruberg, Kyle Beadle, Stephen A. Constant, and Nathan Hawley. 2021. "Seasonal Overturn and Stratification Changes Drive Deep-Water Warming in One of Earth's Largest Lakes." *Nature Communications* 12 (1): 1688. https://doi.org/10.1038/s41467-021-21971-1.

Belisle, B. Shafer, Morgan M. Steffen, Helena L. Pound, Sue B. Watson, Jennifer M. DeBruyn, Richard A. Bourbonniere, Gregory L. Boyer, and Steven W. Wilhelm. 2016. "Urea in Lake Erie: Organic Nutrient Sources as Potentially Important Drivers of Phytoplankton Biomass." *Journal of Great Lakes Research* 42 (3): 599–607. https://doi.org/10.1016/j.jglr.2016.03.002.

Bernhardt, Emily S. 2013. "Cleaner Lakes Are Dirtier Lakes." *Science* 342 (6155): 205–6. https://doi.org/10.1126/science.1245279.

Berube, Paul M, Steven J Biller, Alyssa G Kent, Jessie W Berta-Thompson, Sara E Roggensack, Kathryn H Roache-Johnson, Marcia Ackerman, et al. 2015. "Physiology and Evolution of Nitrate Acquisition in Prochlorococcus." *The ISME Journal* 9 (5): 1195–1207. https://doi.org/10.1038/ismej.2014.211.

Biddanda, Bopaiah, Megan Ogdahl, and James Cotner. 2001. "Dominance of Bacterial Metabolism in Oligotrophic Relative to Eutrophic Waters." *Limnology and Oceanography* 46 (3): 730–39. https://doi.org/10.4319/lo.2001.46.3.0730.

Boddicker, Andrew M., and Annika C. Mosier. 2018. "Genomic Profiling of Four Cultivated Candidatus Nitrotoga Spp. Predicts Broad Metabolic Potential and Environmental Distribution." *The ISME Journal* 12 (12): 2864–82. https://doi.org/10.1038/s41396-018-0240-8.

Bollmann, Annette, George S. Bullerjahn, and Robert Michael McKay. 2014. "Abundance and Diversity of Ammonia-Oxidizing Archaea and Bacteria in Sediments of Trophic End Members of the Laurentian Great Lakes, Erie and Superior." Edited by Jose Luis Balcazar. *PLoS ONE* 9 (5): e97068. https://doi.org/10.1371/journal.pone.0097068.

Bramburger, Andrew J., and Euan D. Reavie. 2016. "A Comparison of Phytoplankton

Communities of the Deep Chlorophyll Layers and Epilimnia of the Laurentian Great Lakes." *Journal of Great Lakes Research* 42 (5): 1016–25. https://doi.org/10.1016/j.jglr.2016.07.004.

Butcher, Jonathan B., Daniel Nover, Thomas E. Johnson, and Christopher M. Clark. 2015. "Sensitivity of Lake Thermal and Mixing Dynamics to Climate Change." *Climatic Change* 129 (1): 295–305. https://doi.org/10.1007/s10584-015-1326-1.

Clevinger, Curtis C., Robert T. Heath, and Darren L. Bade. 2014. "Oxygen Use by Nitrification in the Hypolimnion and Sediments of Lake Erie." *Journal of Great Lakes Research* 40 (1): 202–7. https://doi.org/10.1016/j.jglr.2013.09.015.

Costa, Engràcia, Julio Pérez, and Jan-Ulrich Kreft. 2006. "Why Is Metabolic Labour Divided in Nitrification?" *Trends in Microbiology* 14 (5): 213–19. https://doi.org/10.1016/j.tim.2006.03.006.

Daims, Holger, Elena V. Lebedeva, Petra Pjevac, Ping Han, Craig Herbold, Mads Albertsen, Nico Jehmlich, et al. 2015. "Complete Nitrification by Nitrospira Bacteria." *Nature* 528 (7583): 504–9. https://doi.org/10.1038/nature16461.

Daims, Holger, Sebastian Lücker, and Michael Wagner. 2016. "A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria." *Trends in Microbiology* 24 (9): 699–712. https://doi.org/10.1016/j.tim.2016.05.004.

Denef, Vincent J., Ryan S. Mueller, Edna Chiang, James R. Liebig, and Henry A. Vanderploeg. 2016. "Chloroflexi CL500-11 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton Rely on Nitrogen-Rich Dissolved Organic Matter Metabolism and C_1 Compound Oxidation." Edited by A. M. Spormann. *Applied and Environmental Microbiology* 82 (5): 1423–32. https://doi.org/10.1128/AEM.03014-15.

Dove, Alice, and Steven C. Chapra. 2015. "Long-Term Trends of Nutrients and Trophic Response Variables for the Great Lakes." *Limnology and Oceanography* 60 (2): 696–721. https://doi.org/10.1002/lno.10055.

Fahnenstiel, Gary L., Linda Sicko-Goad, Donald Scavia, and Eugene F. Stoermer. 1986. "Importance of Picoplankton in Lake Superior." *Canadian Journal of Fisheries and Aquatic Sciences* 43 (1): 235–40. https://doi.org/10.1139/f86-028.

Falkowski, P. G., T. Fenchel, and E. F. Delong. 2008. "The Microbial Engines That Drive Earth's Biogeochemical Cycles." *Science* 320 (5879): 1034–39. https://doi.org/10.1126/science.1153213.

Finlay, Jacques C., Gaston E. Small, and Robert W. Sterner. 2013. "Human Influences on Nitrogen Removal in Lakes." *Science* 342 (6155): 247–50. https://doi.org/10.1126/science.1242575.

Finlay, Jacques C., Robert W. Sterner, and Sanjeev Kumar. 2007. "ISOTOPIC EVIDENCE FOR IN-LAKE PRODUCTION OF ACCUMULATING NITRATE IN LAKE SUPERIOR." *Ecological Applications* 17 (8): 2323–32. https://doi.org/10.1890/07-0245.1.

Francis, Christopher A., Kathryn J. Roberts, J. Michael Beman, Alyson E. Santoro, and Brian B. Oakley. 2005. "Ubiquity and Diversity of Ammonia-Oxidizing Archaea in Water Columns and Sediments of the Ocean." *Proceedings of the National Academy of Sciences* 102 (41): 14683–88. https://doi.org/10.1073/pnas.0506625102.

Fujimoto, Masanori, Joann Cavaletto, James R. Liebig, Ann McCarthy, Henry A. Vanderploeg, and Vincent J. Denef. 2016. "Spatiotemporal Distribution of Bacterioplankton Functional Groups along a Freshwater Estuary to Pelagic Gradient in Lake Michigan." *Journal of Great Lakes Research* 42 (5): 1036–48. https://doi.org/10.1016/j.jglr.2016.07.029.

Füssel, Jessika, Sebastian Lücker, Pelin Yilmaz, Boris Nowka, Maartje A. H. J. van Kessel, Patric Bourceau, Philipp F. Hach, et al. 2017. "Adaptability as the Key to Success for the Ubiquitous Marine Nitrite Oxidizer *Nitrococcus*." *Science Advances* 3 (11): e1700807. https://doi.org/10.1126/sciadv.1700807.

Glibert, Patricia M., John Harrison, Cynthia Heil, and Sybil Seitzinger. 2006. "Escalating Worldwide Use of Urea – A Global Change Contributing to Coastal Eutrophication." *Biogeochemistry* 77 (3): 441–63. https://doi.org/10.1007/s10533-005-3070-5.

Hatzenpichler, Roland. 2012. "Diversity, Physiology, and Niche Differentiation of Ammonia-Oxidizing Archaea." *Applied and Environmental Microbiology* 78 (21): 7501–10. https://doi.org/10.1128/AEM.01960-12.

Heath, Robert, S-J Hwang, and M. Munawar. 2010. "A Hypothesis for the Assessment of the Importance of Microbial Food Web Linkages in Nearshore and Offshore Habitats of the Laurentian Great Lakes." *Aquatic Ecosystem Health & Management* 6 (November): 231–39. https://doi.org/10.1080/14634980301495.

Hugoni, Mylène, Sandrine Etien, Antoine Bourges, Cécile Lepère, Isabelle Domaizon, Clarisse Mallet, Gisèle Bronner, Didier Debroas, and Isabelle Mary. 2013. "Dynamics of Ammonia-Oxidizing Archaea and Bacteria in Contrasted Freshwater Ecosystems." *Research in Microbiology* 164 (4): 360–70. https://doi.org/10.1016/j.resmic.2013.01.004.

Ivey, James E., Jennifer L. Wolny, Cynthia A. Heil, Susan M. Murasko, Julie A. Brame, and Ashley A. Parks. 2020. "Urea Inputs Drive Picoplankton Blooms in Sarasota Bay, Florida, U.S.A." *Water* 12 (10): 2755. https://doi.org/10.3390/w12102755.

Kits, K. Dimitri, Christopher J. Sedlacek, Elena V. Lebedeva, Ping Han, Alexandr Bulaev, Petra Pjevac, Anne Daebeler, et al. 2017. "Kinetic Analysis of a Complete Nitrifier Reveals an Oligotrophic Lifestyle." *Nature* 549 (7671): 269–72. https://doi.org/10.1038/nature23679.

Kitzinger, Katharina, Cory C. Padilla, Hannah K. Marchant, Philipp F. Hach, Craig W. Herbold, Abiel T. Kidane, Martin Könneke, et al. 2019. “Cyanate and Urea Are Substrates for Nitrification by Thaumarchaeota in the Marine Environment.” *Nature Microbiology* 4 (2): 234. https://doi.org/10.1038/s41564-018-0316-2.

Koch, H., A. Galushko, M. Albertsen, A. Schintlmeister, C. Gruber-Dorninger, S. Lucker, E. Pelletier, et al. 2014. “Growth of Nitrite-Oxidizing Bacteria by Aerobic Hydrogen Oxidation.” *Science* 345 (6200): 1052–54. https://doi.org/10.1126/science.1256985.

Koch, Hanna, Sebastian Lücker, Mads Albertsen, Katharina Kitzinger, Craig Herbold, Eva Spieck, Per Halkjaer Nielsen, Michael Wagner, and Holger Daims. 2015. “Expanded Metabolic Versatility of Ubiquitous Nitrite-Oxidizing Bacteria from the Genus *Nitrospira*.” *Proceedings of the National Academy of Sciences* 112 (36): 11371–76. https://doi.org/10.1073/pnas.1506533112.

Kuypers, Marcel M. M., Hannah K. Marchant, and Boran Kartal. 2018. “The Microbial Nitrogen-Cycling Network.” *Nature Reviews Microbiology* 16 (5): 263–76. https://doi.org/10.1038/nrmicro.2018.9.

Lean, D. R. S., and R. Knowles. 1987. “Nitrogen Transformations in Lake Ontario.” *Canadian Journal of Fisheries and Aquatic Sciences* 44 (12): 2133–43. https://doi.org/10.1139/f87-262.

Martens-Habbena, Willm, Paul M. Berube, Hidetoshi Urakawa, José R. de la Torre, and David A. Stahl. 2009. “Ammonia Oxidation Kinetics Determine Niche Separation of Nitrifying Archaea and Bacteria.” *Nature* 461 (7266): 976–79. https://doi.org/10.1038/nature08465.

Michalak, Anna M., Eric J. Anderson, Dmitry Beletsky, Steven Boland, Nathan S. Bosch, Thomas B. Bridgeman, Justin D. Chaffin, et al. 2013. “Record-Setting Algal Bloom in Lake Erie Caused by Agricultural and Meteorological Trends Consistent with Expected Future Conditions.” *Proceedings of the National Academy of Sciences* 110 (16): 6448–52. https://doi.org/10.1073/pnas.1216006110.

Mukherjee, Maitreyee, Anirban Ray, Anton F. Post, Robert Michael McKay, and George S. Bullerjahn. 2016. “Identification, Enumeration and Diversity of Nitrifying Planktonic Archaea and Bacteria in Trophic End Members of the Laurentian Great Lakes.” *Journal of Great Lakes Research* 42 (1): 39–49. https://doi.org/10.1016/j.jglr.2015.11.007.

Munawar, M., I. F. Munawar, T. Weisse, G. G. Leppard, and M. Legner. 1994. “The Significance and Future Potential of Using Microbes for Assessing Ecosystem Health: The Great Lakes Example.” *Journal of Aquatic Ecosystem Health* 3 (4): 295–310. https://doi.org/10.1007/BF00115288.

Munawar, M., I.f. Munawar, L.r. Culp, and G. Dupuis. 1978. “Relative Importance of Nannoplankton in Lake Superior Phytoplankton Biomass and Community Metabolism.” *Journal of Great Lakes Research* 4 (January): 462–80. https://doi.org/10.1016/S0380-

1330(78)72215-X.

Natwora, Kaela E., and Cody S. Sheik. 2021. “Assessment of Nitrogen Fixation Rates in the Laurentian Great Lakes.” *Journal of Great Lakes Research* 47 (5): 1288–95. https://doi.org/10.1016/j.jglr.2021.07.005.

O’Beirne, M. D., J. P. Werne, R. E. Hecky, T. C. Johnson, S. Katsev, and E. D. Reavie. 2017. “Anthropogenic Climate Change Has Altered Primary Productivity in Lake Superior.” *Nature Communications* 8 (June): 15713. https://doi.org/10.1038/ncomms15713.

Ohlendieck, Ute, Annegret Stuhr, and Heike Siegmund. 2000. “Nitrogen Fixation by Diazotrophic Cyanobacteria in the Baltic Sea and Transfer of the Newly Fixed Nitrogen to Picoplankton Organisms.” *Journal of Marine Systems* 25 (3): 213–19. https://doi.org/10.1016/S0924-7963(00)00016-6.

Pachiadaki, Maria G., Eva Sintes, Kristin Bergauer, Julia M. Brown, Nicholas R. Record, Brandon K. Swan, Mary Elizabeth Mathyer, et al. 2017. “Major Role of Nitrite-Oxidizing Bacteria in Dark Ocean Carbon Fixation.” *Science* 358 (6366): 1046–51. https://doi.org/10.1126/science.aan8260.

Palatinszky, Marton, Craig Herbold, Nico Jehmlich, Mario Pogoda, Ping Han, Martin von Bergen, Ilias Lagkouvardos, et al. 2015. “Cyanate as an Energy Source for Nitrifiers.” *Nature* 524 (7563): 105–8. https://doi.org/10.1038/nature14856.

Paver, Sara F., Ryan J. Newton, and Maureen L. Coleman. 2020. “Microbial Communities of the Laurentian Great Lakes Reflect Connectivity and Local Biogeochemistry.” *Environmental Microbiology* 22 (1): 433–46. https://doi.org/10.1111/1462-2920.14862.

Reed, Andrew J., and Randall E. Hicks. 2011. “Microbial Ecology of Lake Superior Bacteria and Archaea: An Overview.” *Aquatic Ecosystem Health & Management* 14 (4): 386–95. https://doi.org/10.1080/14634988.2011.630282.

Robertson, Dale M., and David A. Saad. 2011. “Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models1.” *JAWRA Journal of the American Water Resources Association* 47 (5): 1011–33. https://doi.org/10.1111/j.1752-1688.2011.00574.x.

Rozmarynowycz, Mark J., Benjamin F. N. Beall, George S. Bullerjahn, Gaston E. Small, Robert W. Sterner, Sandra S. Brovold, Nigel A. D’souza, Susan B. Watson, and Robert Michael L. McKay. 2019. “Transitions in Microbial Communities along a 1600 km Freshwater Trophic Gradient.” *Journal of Great Lakes Research* 45 (2): 263–76. https://doi.org/10.1016/j.jglr.2019.01.004.

Scavia, Donald, J. David Allan, Kristin K. Arend, Steven Bartell, Dmitry Beletsky, Nate S. Bosch, Stephen B. Brandt, et al. 2014. “Assessing and Addressing the Re-Eutrophication of Lake Erie: Central Basin Hypoxia.” *Journal of Great Lakes Research* 40 (2): 226–46.

https://doi.org/10.1016/j.jglr.2014.02.004.

Schelske, Claire L., and David A. Hodell. 1995. “Using Carbon Isotopes of Bulk Sedimentary Organic Matter to Reconstruct the History of Nutrient Loading and Eutrophication in Lake Erie.” *Limnology and Oceanography* 40 (5): 918–29. https://doi.org/10.4319/lo.1995.40.5.0918.

Schindler, D. W. 1974. “Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management.” *Science (New York, N.Y.)* 184 (4139): 897–99. https://doi.org/10.1126/science.184.4139.897.

Schleper, Christa. 2010. “Ammonia Oxidation: Different Niches for Bacteria and Archaea?” *The ISME Journal* 4 (9): 1092–94. https://doi.org/10.1038/ismej.2010.111.

Small, Gaston E., George S. Bullerjahn, Robert W. Sterner, Benjamin F. N. Beall, Sandra Brovold, Jacques C. Finlay, Robert M. L. McKay, and Maitreyee Mukherjee. 2013. “Rates and Controls of Nitrification in a Large Oligotrophic Lake.” *Limnology and Oceanography* 58 (1): 276–86. https://doi.org/10.4319/lo.2013.58.1.0276.

Small, Gaston E., James B. Cotner, Jacques C. Finlay, Rebecca A. Stark, and Robert W. Sterner. 2014. “Nitrogen Transformations at the Sediment–Water Interface across Redox Gradients in the Laurentian Great Lakes.” *Hydrobiologia* 731 (1): 95–108. https://doi.org/10.1007/s10750-013-1569-7.

Sterner, Robert W., Eleni Anagnostou, Sandra Brovold, George S. Bullerjahn, Jacques C. Finlay, Sanjeev Kumar, R. Michael L. McKay, and Robert M. Sherrell. 2007. “Increasing Stoichiometric Imbalance in North America’s Largest Lake: Nitrification in Lake Superior.” *Geophysical Research Letters* 34 (10). https://doi.org/10.1029/2006GL028861.

Tellier, Joshua M., Nicholas I. Kalejs, Benjamin S. Leonhardt, David Cannon, Tomas O. Höök, and Paris D. Collingsworth. 2021. “Widespread Prevalence of Hypoxia and the Classification of Hypoxic Conditions in the Laurentian Great Lakes.” *Journal of Great Lakes Research*, November. https://doi.org/10.1016/j.jglr.2021.11.004.

Vollenweider, R. A., M. Munawar, and P. Stadelmann. 1974. “A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes.” *Journal of the Fisheries Research Board of Canada* 31 (5): 739–62. https://doi.org/10.1139/f74-100.

Wang, Jia, Xuezhi Bai, Haoguo Hu, Anne Clites, Marie Colton, and Brent Lofgren. 2012. “Temporal and Spatial Variability of Great Lakes Ice Cover, 1973–2010.” *Journal of Climate* 25 (4): 1318–29. https://doi.org/10.1175/2011JCLI4066.1.

Ward, B. B., D J Arp, and Martin G Klotz. 2011. *Nitrification*. IWA Publishing.

Chapter 2

Genome streamlining, proteorhodopsin, and organic nitrogen metabolism in freshwater nitrifiers

2.1 Abstract

Microbial nitrification is a critical process governing nitrogen availability in aquatic systems. Freshwater nitrifiers have received little attention, leaving many unanswered questions about their taxonomic distribution, functional potential, and ecological interactions. Here, we reconstructed genomes to infer the metabolism and ecology of free-living nitrifiers across Earth's largest freshwater system, the Laurentian Great Lakes. Surprisingly, ammonia oxidizing Bacteria (AOB) related to *Nitrosospira* dominated over ammonia oxidizing Archaea (AOA) at nearly all stations, with distinct ecotypes prevailing in the transparent, oligotrophic upper lakes compared to Lakes Erie and Ontario. Unexpectedly, one ecotype of *Nitrosospira* encodes proteorhodopsin, which could enhance survival in conditions where ammonia oxidation is inhibited or substrate-limited. Nitrite oxidizing Bacteria (NOB) *Ca.* Nitrotoga and *Nitrospira* fluctuated in dominance, with the latter prevailing in deeper, less productive basins. Genome reconstructions reveal highly reduced genomes and features consistent with genome streamlining, along with diverse adaptations to sunlight and oxidative stress and widespread capacity for organic nitrogen use. Our findings expand the known functional diversity of nitrifiers and establish their ecological genomics in freshwater systems. As lakes worldwide undergo rapid environmental change, understanding the relationship between microbial biodiversity and nitrogen cycling is essential for predicting future ecosystem health and services.

2.2 Introduction

The oxidation of ammonia to nitrate powers the growth of nitrifying microorganisms and represents a critical flux in the global nitrogen cycle. In aquatic systems, microbial nitrification of ammonia released from organic matter degradation produces nitrate, which can then be removed from the system by denitrification (Canfield, Glazer, and Falkowski 2010). As chemolithoautotrophs, nitrifiers are also a major source of dark carbon fixation (Pachiadaki et al. 2017), which may contribute significant organic carbon to the microbial food web of the ocean's interior (Reinthaler, van Aken, and Herndl 2010; Swan et al. 2011; Baltar and Herndl 2019) and of deep freshwater lakes (Callieri et al. 2014).

Microbial nitrifiers are found in Archaea and several phyla of Bacteria, spanning diverse physiology and ecology; understanding the drivers and consequences of this diversity across ecosystems is a fundamental challenge. Ammonia-oxidizing Archaea (AOA) in the phylum *Thaumarchaeota* dominate the mesopelagic oceans (Santoro, Richter, and Dupont 2019), likely due to their high affinity for ammonia (Schleper 2010) and streamlined genomes (Santoro et al. 2015). In freshwater systems, AOA are abundant in some oligotrophic lakes, while ammonia-oxidizing Bacteria (AOB) affiliated with the *Nitrosomonadaceae* (Betaproteobacteria) are highly variable but tend to dominate more eutrophic systems (Auguet et al. 2012; Herber et al. 2020; Urbach et al. 2001; Okazaki et al. 2017; Mukherjee et al. 2016; Hugoni et al. 2013; Hayden and Beman 2014). Complicating this picture, however, there is considerable physiological variation within each group, such as low-nutrient-adapted clades of AOB (A. Bollmann, Bar-Gilissen, and Laanbroek 2002; Sedlacek et al. 2019) and the ability of some strains to use alternative substrates like urea (Sedlacek et al. 2019; Alonso-Sáez et al. 2012). Within the AOA, there are also distinct ecotypes that appear to segregate with depth in the water column, in both marine (Santoro, Richter, and Dupont 2019) and freshwater systems (Auguet et al. 2012). In freshwaters especially — which are poorly characterized compared to the oceans — it remains difficult to predict which AOA and AOB taxa are likely to dominate in a given system (Hayden and Beman 2014).

For aquatic nitrite oxidizers, which span the phyla *Nitrospira, Nitrospinae,* and *Proteobacteria*, niche differentiation is even less clear. The oceans are dominated by exclusively marine lineages (Daims, Lücker, and Wagner 2016; Pachiadaki et al. 2017), consistent with ancient salinity-associated divergence. Cultivated strains of NOB show variation in substrate affinity and physiology (Daims, Lücker, and Wagner 2016; Nowka, Daims, and Spieck 2015; Wegen, Nowka, and Spieck 2019), but the phylogenetic conservation of these traits, and their influence on environmental distributions, are poorly understood. Moreover, recent studies have discovered that NOB are capable of alternative energy metabolisms (H. Koch et al. 2014; Füssel et al. 2017) and can access nitrogen from cyanate and urea (Palatinszky et al. 2015; Hanna Koch et al. 2015), expanding their ecological potential. In freshwater systems, the NOB *Ca.* Nitrotoga (Betaproteobacteria) was only recently discovered to be widespread (Boddicker and Mosier 2018), and the diversity of this genus and factors favoring its success are unknown.

Here, we use the Laurentian Great Lakes as a model system to examine niche partitioning among planktonic freshwater nitrifiers. The Great Lakes hold 20% of Earth's surface freshwater and more than half of this volume receives little to no light (< 1% surface irradiance). This system, while hydrologically connected, spans strong trophic and chemical gradients: ultraoligotrophic Lake Superior supports rates of primary production and nitrification comparable to the ocean gyres (Sterner 2010; Small et al. 2013), while Lake Erie supports greater production (Vollenweider, Munawar, and Stadelmann 1974) and more than 70-fold higher nitrification rate (Clevinger, Heath, and Bade 2014). Between these extremes, Lake

Ontario has low ambient ammonium concentrations like Lake Superior (Neilson and Stevens 1987) but nitrification rates up to four times higher (Lean and Knowles 1987). While previous studies reported that AOA and AOB dominate Lakes Superior and Erie, respectively (Small et al. 2013; Mukherjee et al. 2016), recent community profiling has revealed broader diversity in both ammonia-oxidizing and nitrite-oxidizing lineages (Paver, Newton, and Coleman 2020; Rozmarynowycz et al. 2019; Fujimoto et al. 2016). We sought to link taxonomic, genomic, and metabolic diversity of nitrifiers with the varied biogeochemistry of the Great Lakes, using genome reconstructions and abundance profiling. Our results uncover novel lineages and metabolic capabilities, and provide the first large-scale assessment of freshwater nitrifier genomics.

2.3 Results and Discussion

2.3.1 Niche partitioning of nitrifiers across the Great Lakes

To map free-living (here defined as cells that pass through a 1.6µm filter) nitrifiers across the Great Lakes, we searched our recent 16S rRNA datasets for known nitrifying taxa (Paver, Newton, and Coleman 2020). We detected putative AOB in the genus *Nitrosospira* (Betaproteobacteria, family *Nitrosomonadaceae*) and AOA similar to *Nitrosoarchaeum* (family *Nitrosopumilaceae*), along with putative NOB in the genera *Ca.* Nitrotoga (Betaproteobacteria, family *Gallionellaceae*) and *Nitrospira* (family *Nitrospiraceae*). We did not detect 16S rRNA amplicons from *Nitrosococcus*, *Nitrococcus*, *Nitrospina* or *Nitrobacter*. The relative abundance of nitrifiers was negatively correlated with photosynthetically active radiation (PAR; Spearman's rho = -0.89, $p < 2.2e-16$) and reached a maximum below the depth of 1% PAR in each lake, up to 20% of amplicon sequences (Fig S2.1a). The relative abundances of ammonia- and nitrite-oxidizing taxa were strongly correlated (Spearman's rho = 0.918, p < 2.2e-16; Fig S2.1b). Free-living nitrifiers were rare (<0.1% relative abundance) in bottom water samples from the southern basin of Huron (HU15M) and western basin of Erie (HU91M); these two stations are the shallowest in our dataset and have relatively high light penetration to the bottom (~1% PAR). Chlorophyll-*a* concentration was also negatively correlated with the relative abundance of nitrifiers (Spearman's rho = -0.677, p < 1.7e-7; Fig 2S1c). These findings are consistent with previous work demonstrating photoinhibition of nitrification (Hooper and Terry 1974; Horrigan and Springer 1990; Guerrero and Jones 1996; Merbt et al. 2012), as well as potential competition with phototrophs for ammonium (Smith, Chavez, and Francis 2014).

Figure 1

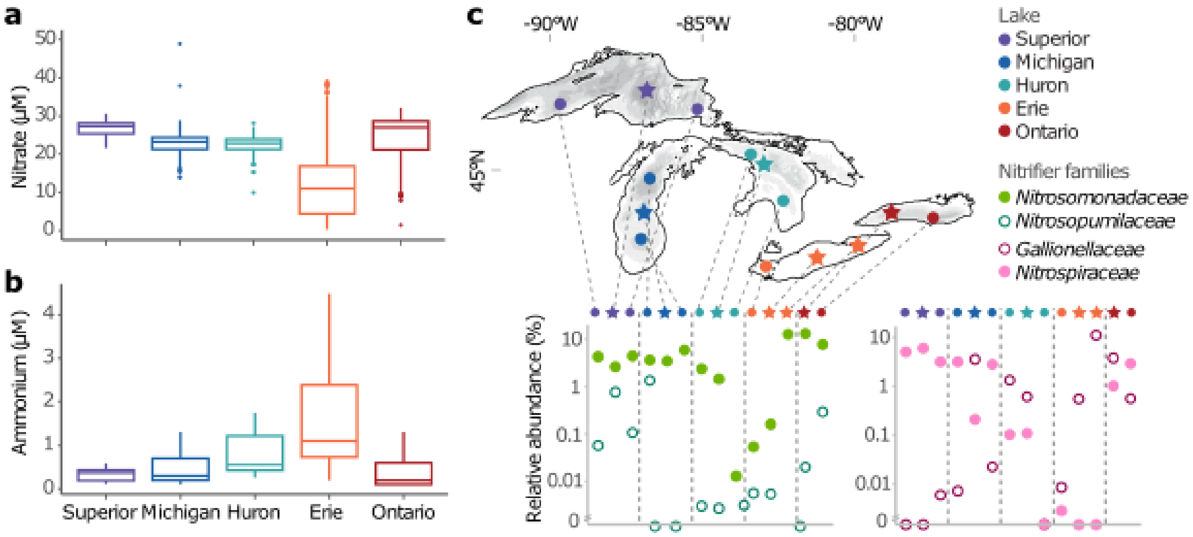


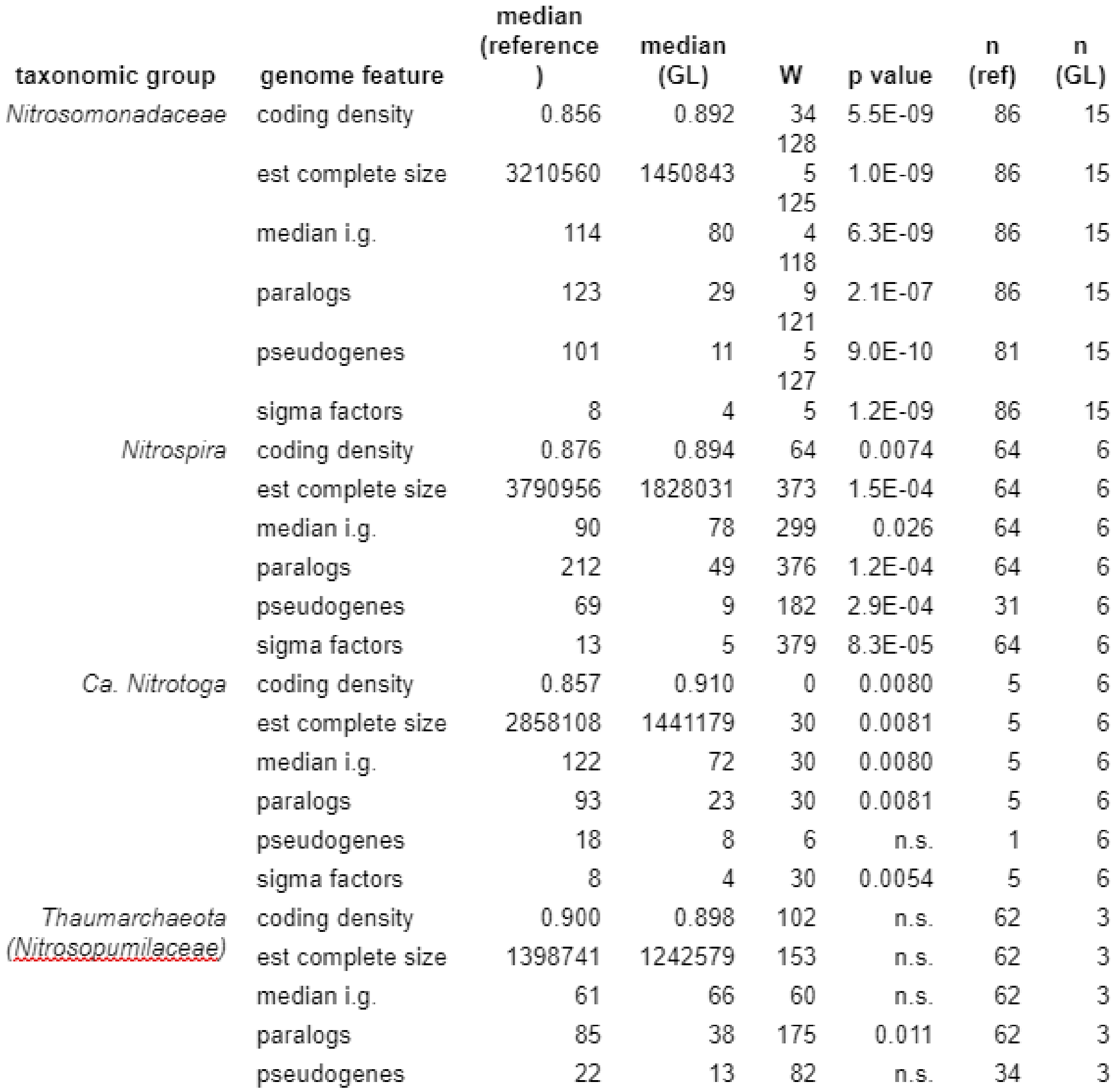
Figure 2.1 Dissolved inorganic nitrogen availability and distribution of nitrifiers across the Great Lakes. (a) Oxidized nitrogen concentrations. Values include NOx concentrations from published studies (n=128; (Mukherjee et al. 2016; Lean and Knowles 1987; Rozmarynowycz et al. 2019; Murphy 1980; Gardner et al. 2004)), US EPA Water Quality Surveys in 2012 and 2013 (n=1626 from GLENDA database), and this study (n=20). (b) Ammonium concentrations. Values are derived from the literature as in (a) (n=118) and from this study (n=20). (c) Distribution of nitrifiers across the Great Lakes. Top panel: map of sampling stations; stars indicate stations chosen for metagenome analysis. Bottom panel: relative abundance of ammonia-oxidizing (green) and nitrite-oxidizing (pink) families based on 16S rRNA V4-V5 amplicon sequencing,

sampled in the mid-hypolimnion (except western Erie, sampled 1m from bottom). Data is plotted roughly west to east as indicated on the map.

The taxonomic assemblage of nitrifiers differed across lakes and even among stations within a lake (Fig 2.1, Dataset S2.1), in association with variable productivity and nitrogen availability. Surface ammonium is typically below 300 nM except in Erie, where it is several-fold higher and spatially variable; nitrate, on the other hand, is very high across the lakes but lowest in Erie due to biological uptake (Kumar et al. 2007; Dove and Chapra 2015). Few measurements of urea exist but it can exceed ammonium (Belisle et al. 2016) (Dataset S2.2). AOB (*Nitrosomonadaceae*) were observed across all lakes. In contrast, AOA (*Nitrosopumilaceae*) sequences only exceeded 0.5% relative abundance at the three deepest stations (SU08M, MI41M, ON55M) where the ratio of AOB: AOA ranged from 10:1 to 1:3. We found pronounced shifts in the dominant NOB across stations (Fig 2.1), and all stations except those in Lake Ontario showed strong dominance (greater than ten-fold) of either *Ca.* Nitrotoga (family *Gallionellaceae*) or *Nitrospira*. *Nitrospira* was the only nitrite oxidizer detected in Superior and the dominant nitrite oxidizer in parts of Michigan (MI41M, MI18M). By contrast, *Ca.* Nitrotoga was the only nitrite oxidizer observed in Erie and Huron and the dominant nitrite oxidizer at the shallowest station in Michigan (MI27M). Within each taxon, a single 16S rRNA oligotype dominated the AOA, *Ca*. Nitrotoga, and *Nitrospira,* while several oligotypes of *Nitrosomonadaceae* shifted abundance across samples (Fig S2.2), consistent with ecotypic diversity as discussed below.

2.3.2 Ecotypic variation in abundant streamlined Nitrosospira

We reconstructed 15 genomes of the AOB *Nitrosospira*, substantially expanding genome descriptions for this genus (Rice et al. 2016; Garcia et al. 2013; Norton et al. 2008). Based on a phylogenomic tree, free-living Great Lakes *Nitrosospira* fall into two major clades, both of which are distinct from published species; each of these clades also includes MAGs recovered from lakes Biwa and Baikal, suggesting novel globally distributed freshwater lineages (Fig S2.3). One clade, which we call NspGL1, has a highly reduced genome (median 1.42 Mb) and low G+C content (40.7%) (Fig 2.2, Dataset S2.3). The second clade was resolved into three subclades (denoted NspGL2a, 2b, and 3; Fig S2.3) based on phylogeny and average nucleotide identity (ANI), all with small genome sizes of 1.45-1.68 Mb and 50% G+C content (Fig 2.2, Dataset S2.3). Compared to 86 reference *Nitrosomonadaceae* genomes, Great Lakes *Nitrosospira* genomes are not only smaller (median estimated complete genome size: reference = 3.21 Mb, GL = 1.45 Mb; Table 2.1), but also have shorter intergenic spacers, fewer paralogs, fewer pseudogenes, and fewer sigma factors (Table 2.1, Fig S2.4, Dataset S2.4), consistent with genome streamlining to reduce resource demands (Giovannoni, Cameron Thrash, and Temperton 2014). Based on short read mapping, these subclades are ecologically distinct: NspGL1 and NspGL2b — with the smallest genomes — are the dominant AOB in the upper oligotrophic lakes, while NspGL2a is only abundant in Ontario and NspGL3 is only abundant in Erie (Fig 2.2). Hereafter we refer to these subclades as ecotypes due to their phylogenetic and ecological divergence.

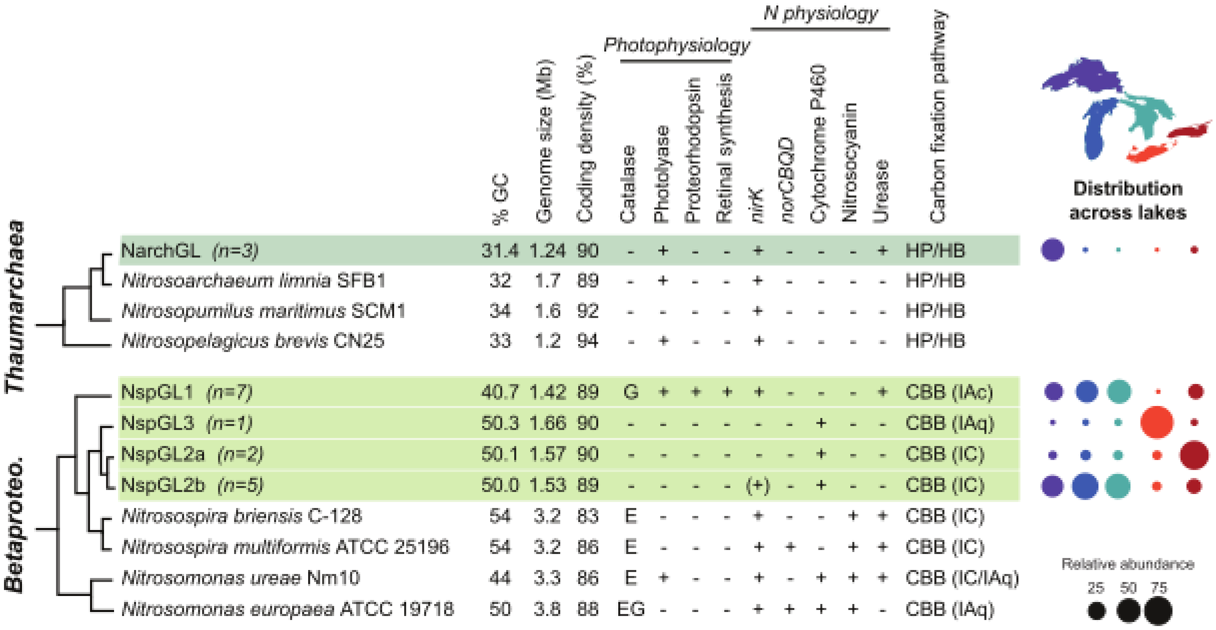


taxonomic group	genome feature	median (reference)	median (GL)	W	p value	n (ref)	n (GL)
Nitrosomonadaceae	coding density	0.856	0.892	34	5.5E-09	86	15
	est complete size	3210560	1450843	1285	1.0E-09	86	15
	median i.g.	114	80	1254	6.3E-09	86	15
	paralogs	123	29	1189	2.1E-07	86	15
	pseudogenes	101	11	1215	9.0E-10	81	15
	sigma factors	8	4	1275	1.2E-09	86	15
Nitrospira	coding density	0.876	0.894	64	0.0074	64	6
	est complete size	3790956	1828031	373	1.5E-04	64	6
	median i.g.	90	78	299	0.026	64	6
	paralogs	212	49	376	1.2E-04	64	6
	pseudogenes	69	9	182	2.9E-04	31	6
	sigma factors	13	5	379	8.3E-05	64	6
Ca. Nitrotoga	coding density	0.857	0.910	0	0.0080	5	6
	est complete size	2858108	1441179	30	0.0081	5	6
	median i.g.	122	72	30	0.0080	5	6
	paralogs	93	23	30	0.0081	5	6
	pseudogenes	18	8	6	n.s.	1	6
	sigma factors	8	4	30	0.0054	5	6
Thaumarchaeota (Nitrosopumilaceae)	coding density	0.900	0.898	102	n.s.	62	3
	est complete size	1398741	1242579	153	n.s.	62	3
	median i.g.	61	66	60	n.s.	62	3
	paralogs	85	38	175	0.011	62	3
	pseudogenes	22	13	82	n.s.	34	3

Table 2.1 Evidence for genome streamlining in nitrifiers from the Laurentian Great Lakes. Genome features were compared between Great Lakes MAGs and reference genomes using a two-sided Wilcoxon/Mann-Whitney test. n.s., not significant at 0.05 level. Only genomes with >70% completion and <10% contamination are included.

We next compared gene content between our Great Lakes *Nitrosospira* and 86 *Nitrosomonadaceae* reference genomes. On average, Great Lakes *Nitrosospira* encode far fewer two-component signal transduction systems (NspGL = 5-8, mean reference = 19), transposases (NspGL = 0-7, mean reference = 39), motility genes (NspGL = 0-4, mean reference = 52), pilus and secretion genes (NspGL = 2-9, mean reference = 27), and defense-related genes (NspGL = 4-11, mean reference = 39) (Dataset S2.5). They also lack functions related to biofilm formation such as polysaccharide matrix production (e.g. *pel* genes) and extracellular protein targeting (exosortase and PEP-CTERM motifs). We note that while our 15 new *Nitrosospira* MAGs have high estimated completion (median 98.6%; Dataset S2.3), some gene absences may reflect incomplete assemblies. Nevertheless, this overall picture of gene content in Great Lakes AOB contrasts with *Nitrosospira* isolates from soil (Norton et al. 2008; Rice et al. 2016) and even to oligotrophic *Nitrosomonas* isolates (Annette Bollmann et al. 2013), and is consistent with a passive planktonic lifestyle in extremely low-nutrient systems.

Figure 2



	% GC	Genome size (Mb)	Coding density (%)	Catalase	Photolyase	Proteorhodopsin	Retinal synthesis	*nirK*	*norCBQD*	Cytochrome P460	Nitrosocyanin	Urease	Carbon fixation pathway
NarchGL *(n=3)*	31.4	1.24	90	-	+	-	-	+	-	-	-	+	HP/HB
Nitrosoarchaeum limnia SFB1	32	1.7	89	-	+	-	-	+	-	-	-	-	HP/HB
Nitrosopumilus maritimus SCM1	34	1.6	92	-	-	-	-	+	-	-	-	-	HP/HB
Nitrosopelagicus brevis CN25	33	1.2	94	-	+	-	-	+	-	-	-	-	HP/HB
NspGL1 *(n=7)*	40.7	1.42	89	G	+	+	+	+	-	-	-	+	CBB (IAc)
NspGL3 *(n=1)*	50.3	1.66	90	-	-	-	-	-	-	+	-	-	CBB (IAq)
NspGL2a *(n=2)*	50.1	1.57	90	-	-	-	-	-	-	+	-	-	CBB (IC)
NspGL2b *(n=5)*	50.0	1.53	89	-	-	-	-	(+)	-	+	-	-	CBB (IC)
Nitrosospira briensis C-128	54	3.2	83	E	-	-	-	+	-	-	+	+	CBB (IC)
Nitrosospira multiformis ATCC 25196	54	3.2	86	E	-	-	-	+	+	-	+	+	CBB (IC)
Nitrosomonas ureae Nm10	44	3.3	86	E	+	-	-	+	-	+	+	+	CBB (IC/IAq)
Nitrosomonas europaea ATCC 19718	50	3.8	88	EG	-	-	-	+	+	+	+	-	CBB (IAq)

Figure 2.2 Genome properties and cross-lake distribution of ammonia oxidizing organisms, showing both Archaea (top) and Betaproteobacteria (bottom). Rows highlighted in green represent clusters of genomes reconstructed from the Great Lakes and median values are shown for genome size, GC content, and coding density. For catalase, “E” indicates monofunctional catalase *katE*; “G” indicates bifunctional catalase-peroxidase *katG*. For carbon fixation, RuBisCO type is shown in parentheses (Badger and Bek 2008); HP/HB, 3-Hydroxypropionate/4-

Hydroxybutyrate cycle; CBB, Calvin-Benson-Bassham cycle. Bubble plot shows composition of ammonia oxidizers in hypolimnion samples, using MAGs as probes to recruit metagenomic reads (values sum to 100% for each lake column). Genes identified in only a subset of genomes are shown as (+).

We next compared metabolic potential among Great Lakes AOB ecotypes to understand their ecological preferences for upper lakes (NspGL1, NspGL2b), Ontario (NspGL2a), and Erie (NspGL3). Surprisingly, all seven NspGL1 MAGs encode proteorhodopsin, a light-driven proton pump that supports bacterial energy production (Beja 2000; Béjà et al. 2001). They also carry the genes necessary to synthesize its chromophore retinal, including 15,15'-Beta-carotene dioxygenase (*blh*), lycopene cyclase (*crtY*), phytoene dehydrogenase (*crtI*), phytoene synthase (*crtB*), and GGPP synthase (*crtE*) ((Sabehi et al. 2005; Martinez et al. 2007); Fig 2.3a). We also identified proteorhodopsin in a single-cell amplified genome representing NspGL1 from Lake Michigan (Fig 2.3a; 99.8% ANI with NspGL1 MAGs), demonstrating that it is not an artifact of metagenome assembly. To our knowledge, this is the first example of a nitrifier with proteorhodopsin. All NspGL1 proteorhodopsins share residues H95, D127 and E138 along with a short beta-turn (G111-P116) between helices B and C, which are characteristic features of proteorhodopsin as distinct from sensory and other rhodopsins (Reckel et al. 2011), and the presence of leucine at position 135 suggests green light tuning (Kralj et al. 2008) (Fig 2.3b). All of the genes in this module have highest similarity to homologs from *Polynucleobacter*, but are flanked by Nitrosomonadaceae-like genes, suggesting recent horizontal gene transfer (Fig 2.3a). The predicted NspGL1 proteorhodopsins cluster with *Polynucleobacter*, *Methylopumilus*, and other freshwater Betaproteobacteria in Supercluster III as defined by MicRhoDE ((Boeuf et al. 2015); Fig 2.3c). We compared the homologous genome region in two highly similar MAGs from Lakes Biwa and Baikal (Fig S2.5); these contigs lack the proteorhodopsin module but appear to flank a variable region where the contig assembly ends. A proteorhodopsin photosystem could support survival of NspGL1 in the presence of sunlight, which has been shown to inhibit ammonia oxidation (Hooper and Terry 1974; Hyman and Arp 1992). In the upper lakes where NspGL1 is abundant, light penetration is high well below the thermocline in stratified periods (Yousef et al. 2017), and deep water taxa are seasonally advected to the surface by water column mixing (Paver, Newton, and Coleman 2020). In addition to proteorhodopsin, NspGL1 — but not the other three ecotypes of Great Lakes *Nitrosospira* — encode a class I cyclopyrimidine dimer photolyase, which uses light energy to repair UV-induced DNA damage, and the catalase-peroxidase *katG*, suggesting that the NspGL1 ecotype is adapted to relatively shallow depths in the water column (Fig 2.2).

Figure 3

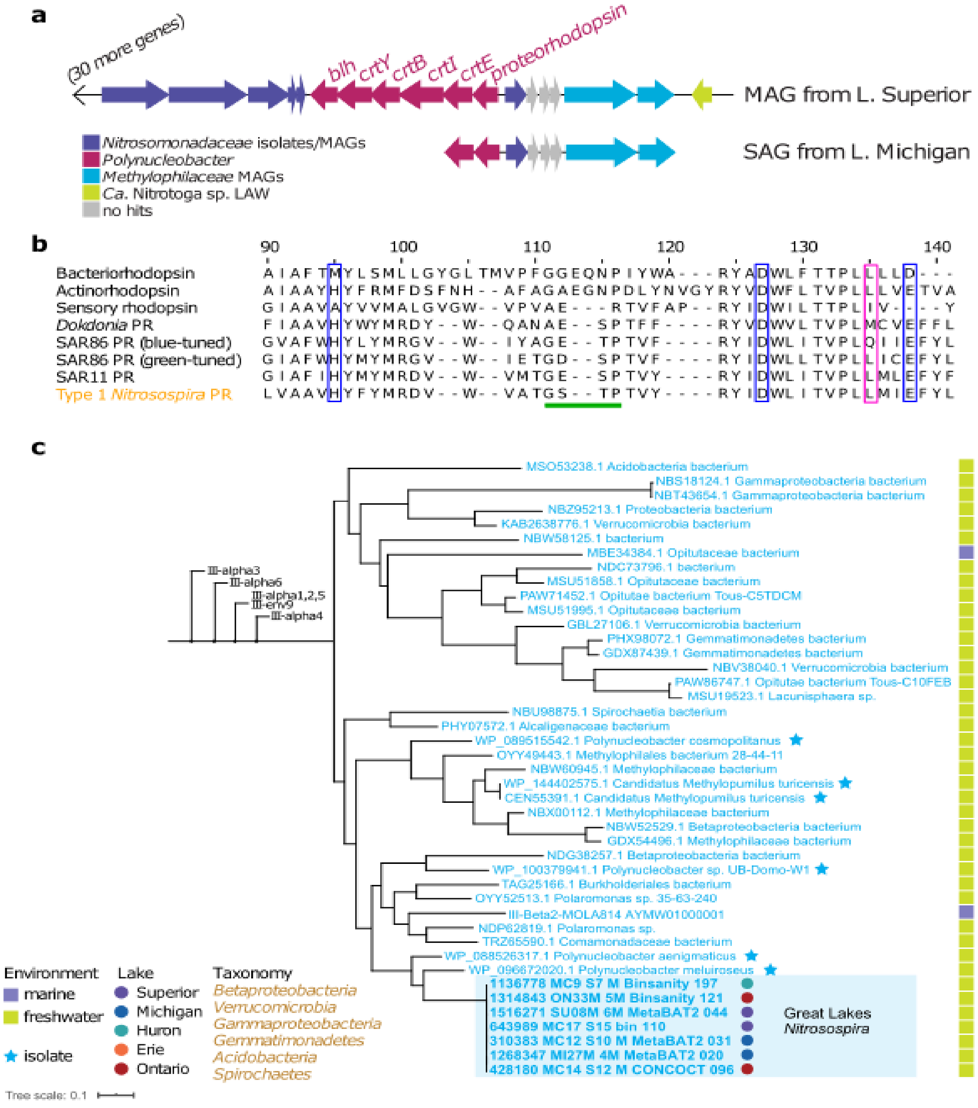


Figure 2.3 Evidence for proteorhodopsin (PR) in *Nitrosospira* from the Great Lakes. (a) Gene neighborhood surrounding PR in *Nitrosospira* MAG MC17_S15_bin_110 and SAG 207399. Genes are colored according to the best BLAST hit taxonomy in the NCBI nr database. (b) Alignment of predicted *Nitrosospira* PR with reference sequences. Diagnostic features are highlighted (Reckel et al. 2011; Kralj et al. 2008): blue boxes, diagnostic residues for PR; pink box, residue indicative of blue or green tuning; green underline, shorter beta-sheet region in PR. Sequence accession numbers: bacteriorhodopsin P02945, actinorhodopsin A0A1D9E0H1, sensory rhodopsin P42196, *Dokdonia* PR EAQ40507.1, SAR86 blue-tuned PR Q4PP54, SAR86 green-tuned PR Q9F7P4, SAR11 PR A6YQL7. (c) Phylogenetic tree showing close relatives of *Nitrosospira* PR within Supercluster III, as defined by MicRhoDE database (Boeuf et al. 2015). Neighboring clusters have been collapsed for clarity.

Great Lakes *Nitrosospira* carry a reduced, ecotype-specific complement of nitrogen metabolism genes compared to reference AOB (Fig 2.2, Dataset S2.5; gene absences were verified as described in Materials & Methods). All are presumed to have the core ammonia oxidation enzymes ammonia monooxygenase and hydroxylamine dehydrogenase; these genes were assembled and binned as expected in some MAGs, and were manually identified on short unbinned contigs in other cases (Dataset S2.6; see Materials & Methods). Surprisingly, all Great Lakes *Nitrosospira* MAGs lack the copper protein nitrosocyanin, whose precise function is unknown but so far has been found in all AOB except one member of the *N. oligotropha* clade (Annette Bollmann et al. 2013). The lack of nitrosocyanin extends beyond the Great Lakes MAGs to closely related freshwater and marine strains, along with five more members of the *N. oligotropha* clade (Fig S2.3); its absence may be related to the divergence of these clades. Only NspGL1 and NspGL2b encode NO-forming nitrite reductase (NirK), which confers nitrite tolerance (Beaumont et al. 2002); this result is surprising given that these two clades dominate the upper lakes where productivity and reduced N are lowest. None of the Great Lakes ecotypes encode NO reductase (NorCBQD), and NspGL1 lacks cytochrome P460 family proteins; these genomic patterns may influence N_2O emissions across the Great Lakes. Nitrogen acquisition is also distinct among Great Lakes AOB: NspGL1 lacks an apparent ammonium transporter, but encodes urease structural and accessory genes (*ureABCEFG*) and a high-affinity urea transporter (*urtABCDE)*. Further, all Great Lakes ecotypes encode a high-affinity amino acid transporter (*livFGHM*); these genes are rare (<5%) in reference genomes and could supply reduced nitrogen and/or organic carbon. Finally, NspGL1 and NspGL3 have genes for producing cyanophycin, an intracellular storage compound for nitrogen (Norton et al. 2008; Watzer and Forchhammer

2018). Together, the distinctive gene complements present in Great Lakes *Nitrosospira* illustrate the variability and adaptability of AOB gene content, even across a connected freshwater habitat.

As with nitrogen metabolism, carbon metabolism is also distinct between Great Lakes and reference AOB, and among Great Lakes ecotypes (Fig 2.2, Dataset S2.5). Unlike most reference AOB, Great Lakes *Nitrosospira* lack two key enzymes of the oxidative pentose phosphate pathway, glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase. All ecotypes except Erie-specific NspGL3 also lack genes for glycogen synthesis and degradation. These findings suggest continuous autotrophy, rather than cycles of carbon storage and utilization. The key enzyme for carbon fixation, RuBisCO, has evolved several kinetically distinct forms whose distribution likely reflects ecological pressures (Badger and Bek 2008). NspGL1 and NspGL3 both contain Form IA RuBisCO, while NspGL2a and NspGL2b contain Form IC RuBisCO (Fig 2.2; (Rae et al. 2013; Badger and Bek 2008)). NspGL1 genomes also possess an alpha carboxysome-like *cso* operon, similar to *Nitrosomonas eutropha* C91 (Rae et al. 2013), though our draft assembly lacks the expected carbonic anhydrase (*csoS3*/*csoSCA*). Carboxysome-associated RuBisCO may allow NspGL1, the ecotype most strongly adapted to energy and nutrient limitation, to more efficiently fix CO_2 by minimizing the wasteful oxygenation reaction and reducing the cellular nitrogen allocation to RuBisCO (Badger and Bek 2008). The ranges of kinetic properties observed in other autotrophs for Form IAq (found in NspGL3) and Form IC (found in NspGL2a and NspGL2b) overlap, and therefore more work is needed to understand the fitness advantages, if any, that this RuBisCO diversity confers on Great Lakes nitrifiers.

2.3.3 Streamlined freshwater Thaumarchaeota

We reconstructed three similar genomes (>99% ANI) of *Nitrosoarchaeum* (NarchGL; Fig S2.6, Dataset S2.3) from three separate samples (two from Superior, one from Ontario), consistent with our low observed 16S rRNA diversity for *Thaumarchaeota*. These NarchGL genomes are very similar (~99% ANI) to two genomes reconstructed from Lake Baikal, located thousands of kilometers away (Cabello-Yeves et al. 2018). Their next closest relatives are also from freshwater environments, and phylogenetic clustering suggests that salinity is an important driver of divergence throughout the *Nitrosopumilaceae* (Fig S2.6). As a group, the *Thaumarchaeota* tend to have smaller genomes, lower G+C content, higher coding density, and fewer paralogs and pseudogenes than nitrifying bacterial taxa; even within this group, NarchGL genomes fall below the 30th percentile in size and have significantly fewer paralogs than average (Table 2.1, Fig S2.4, Dataset S2.4). Using our reconstructed genomes as probes for metagenomic read recruitment, NarchGL were detected in Superior, Michigan, and Ontario; they represented roughly one-third of ammonia oxidizers in the mid-hypolimnion of station SU08M (Fig 2.2).

NarchGL share nearly 90% of their predicted proteins with close relatives including *Ca.* Nitrosoarchaeum limnia; at the same time, they show distinctive patterns in gene content that pinpoint the key selective pressures of deep lakes (Fig 2.2, Dataset S2.5). All three NarchGL genomes encode urease and a urea transporter, implicating urea as a vital source of nitrogen for energy and/or biosynthesis. Consistent with phosphorus scarcity in much of the Great Lakes (Sterner et al. 2004), NarchGL encode high affinity transport systems for phosphate and potentially phosphonates, though we did not identify a phosphonate lyase. In addition to both CRISPR/Cas enzymes cas1 and cas4, NarchGL genomes contain several phage proteins, suggesting viral infection and integration events may be common. DNA photolyases, which have been found in epipelagic clades of marine *Thaumarchaeota* (Santoro, Richter, and Dupont 2019), are present in all low salinity *Nitrosoarchaeum* including NarchGL, suggesting NarchGL are exposed to sunlight due to high water clarity (Yousef et al. 2017) and/or annual mixing in the Great Lakes. NarchGL also lack the common tRNA modification 4-thiouridylation (indicated by K04487 and PF02568-PF18297 (Ryals et al. 1982)); we propose that the absence of this modification, which is susceptible to near-UV radiation (Ryals et al. 1982), is also related to sunlight exposure.

Genomes of NarchGL reveal striking reduction in environmental sensing, response, and regulatory functions, relative to most other *Nitrosoarchaeum* and *Nitrosopumilaceae* (Dataset S2.5). NarchGL encode 9-12 domains representing the general archaeal transcription factors, TATA binding protein (TBPs, PF00352) and transcription factor B (TFBs, PF00382 and PF08271), compared to 21 in *Ca.* N. limnia. NarchGL lack common domains found in two-component systems that transmit environmental signals to control gene expression or protein activity (domains PF02743, PF00672, PF00512, PF00072; NarchGL = 0, *Ca.* N. limnia = 53-54 copies per genome). Further, they are depleted in ArsR family transcription factors (PF01022; 0 copies in NarchGL vs. 2-3 in *Ca.* N. limnia), P-II proteins for regulation of nitrogen metabolism (PF00543; 1 copy per NarchGL genome vs. 5 in *Ca.* N. limnia), and other potential regulatory domains (CBS PF00571: 5 in NarchGL vs. 18-19 in *Ca.* N. limnia; USP PF00582: 1 in NarchGL vs. 15 in *Ca.* N. limnia). This extremely limited regulatory capacity in NarchGL stands in sharp contrast to closely related *Ca.* N. limnia, but instead parallels the oceanic minimalist *Ca.* Nitrosopelagicus brevis (Santoro et al. 2015).

2.3.4 Expanded diversity of Ca. Nitrotoga with reduced genomes

Despite the broad distribution of *Ca.* Nitrotoga in freshwater systems and beyond, only six genomes are available, derived from rivers heavily impacted by urban and agricultural influence, a wastewater treatment plant, and coastal sediment (Boddicker and Mosier 2018; Kitzinger et al. 2018; Ishii et al. 2020). Hence the metabolic and phylogenetic diversity of this group is virtually unexplored. We reconstructed six new MAGs of *Ca*. Nitrotoga, which form two clusters with >99% ANI within each cluster and ~97% ANI between clusters (NtogaGL1a and NtogaGL1b; Fig S2.3). These new *Ca*. Nitrotoga MAGs are far smaller than published genomes (median GL = 1.44 Mb, reference = 2.61-2.98 Mb), have shorter intergenic regions, fewer sigma factors, and fewer paralogs (Table 2.1, Fig S2.4, Dataset S2.4), consistent with genome streamlining (Giovannoni, Cameron Thrash, and Temperton 2014). They lack functions such as motility and chemotaxis, pilus biogenesis, and DNA repair (*mutLS*) (Dataset S2.5). Great Lakes *Ca.* Nitrotoga also encode markedly fewer two-component systems for sensing and responding to environmental cues than four river-derived genomes (NtogaGL = 2-6 per genome vs. 30-35 in four reference genomes; reference strain KNB has 7). Compared to reference genomes, NtogaGL have fewer defense-related genes (restriction-modification, toxin-antitoxin, and CRISPR-Cas systems; mean NtogaGL = 11 vs. 39 for references), and transposases (mean NtogaGL = 3 vs. 19 for references) (Dataset S2.5). While incomplete assembly of hypervariable genome regions may explain some of these absences, the overall genome properties are consistent with a relatively stable low-nutrient environment and planktonic lifestyle.

The reduced genomes of NtogaGL1a/b help clarify core features of the genus *Nitrotoga*, along with accessory functions that may enable local adaptation in specific populations. To date, sequenced *Nitrotoga* including NtogaGL1a/b encode similar electron transport pathways, including NADH dehydrogenase-Complex I, succinate dehydrogenase-Complex II, and

Alternative Complex III, along with high-affinity cbb3-type cytochrome oxidases suggesting adaptation to low oxygen conditions. They also share the Calvin cycle for carbon fixation, a complete TCA cycle, and an evolutionarily distinct nitrite oxidoreductase (NXR) from other NOB (Boddicker and Mosier 2018; Kitzinger et al. 2018; Ishii et al. 2020). All *Nitrotoga* to date also share transporters for amino acids and peptides, potential sources of C and/or N. *Nitrotoga* can also potentially access reduced sulfur compounds for energy via sulfite dehydrogenase, suggesting metabolic flexibility beyond nitrite oxidation.

Beyond these similarities, the small genomes of NtogaGL1a/b are distinct from previously described *Nitrotoga* in many ways. NtogaGL1a/b lack NiFe hydrogenase to use hydrogen as an energy source. They also lack nitrogen metabolism functions including assimilatory nitrite reductase (*nirBD*) and nitrite reductase to NO (*nirK*). Based on gene content, NtogaGL1a/b appear unable to use hexoses like glucose, since they lack the glycolytic enzyme phosphofructokinase and the Entner-Doudoroff pathway, similar to *Nitrobacter winogradskyi* (Starkenburg et al. 2006). Consistent with this, they also lack genes for storage and breakdown of glycogen (Dataset S2.5). All but one of the NtogaGL1a/1b genomes encode cyanate lyase (*cynS*), which is found in other NOB but no *Nitrotoga* to date (Palatinszky et al. 2015; S. Lücker et al. 2010; Sebastian Lücker et al. 2013). The *cynS* gene, adjacent to *glnK-amtB* for ammonium sensing and transport, likely functions in N assimilation, as recently described for *Nitrospinae* (Kitzinger et al. 2020). While cyanase has been shown to mediate reciprocal feeding between some NOB and ammonia oxidizers (Palatinszky et al. 2015), it remains to be seen whether such an interaction occurs in the free-living (<1.6um) size-fraction and dilute environment sampled here. Notably, cyanase from NtogaGL1a/1b, along with predicted *Nitrospirae* proteins from

Lake Baikal and soil, form a distinct phylogenetic cluster from most nitrifier cyanase proteins observed to date (Fig S2.7).

The two ANI-based clusters we detected, NtogaGL1a and NtogaGL1b, appear to be phylogenetically and ecologically distinct ecotypes. Based on short-read mapping, NtogaGL1b dominates Erie, while NtogaGL1a dominates all other *Ca.* Nitrotoga-containing samples (Fig 2.4). We found several metabolic genes that differentiate the two ecotypes. ER-specific NtogaGL1b genomes share a region encoding thiosulfate dehydrogenase (*tsdA*), cytochromes, transport of sulfur-containing compounds, lactate dehydrogenase (ldh*)*, a two-component system, and a Crp-family transcription factor (Fig S2.8). This region may be involved in oxidizing thiosulfate as an energy source, and sensing and responding to redox changes that accompany seasonal hypoxia in Lake Erie. The corresponding region in NtogaGL1a encodes an integrase and photolyase, consistent with greater DNA photodamage in the more transparent waters of Michigan, Huron, and Ontario where NtogaGL1a is abundant.

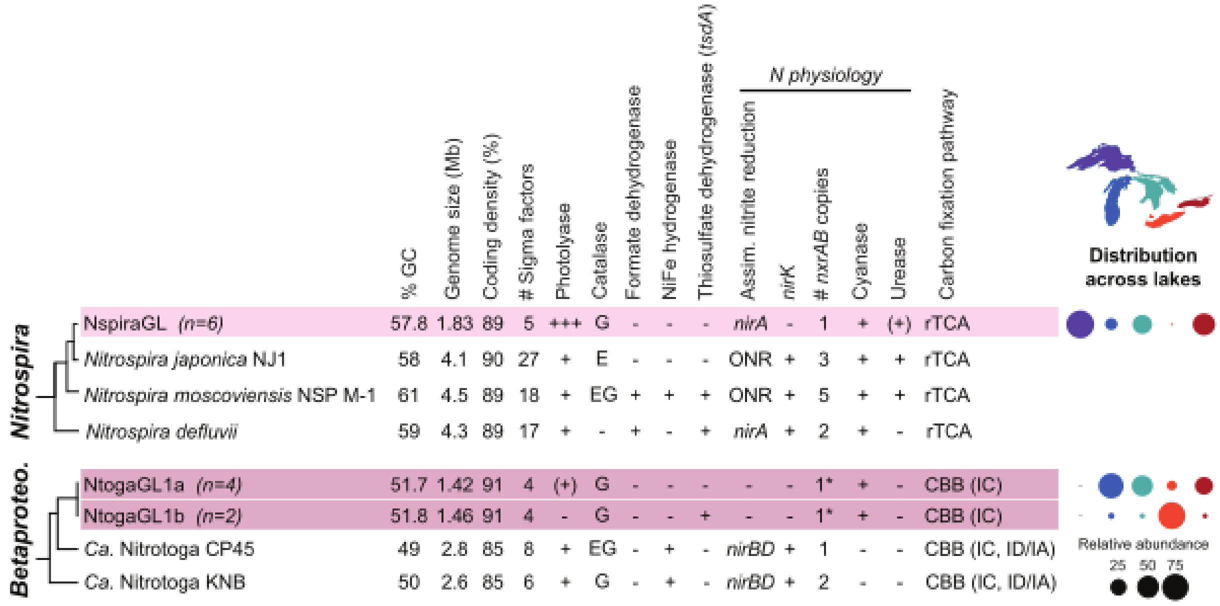
2.3.5 Great Lakes Nitrospira reveal adaptations to sunlit oxic environment

We reconstructed six closely related genomes of *Nitrospira* (~99% ANI; Fig S2.9, Dataset S2.3), representing the predominant NOB throughout Lake Superior and in parts of Michigan, Ontario, and Huron (Fig. 2.4; Dataset S2.1). These genomes, which we refer to as NspiraGL, fall within lineage II (Fig S2.9), which is broadly distributed across soil, freshwater, and engineered habitats (Daims, Lücker, and Wagner 2016); however, genome analyses to date have focused on strains from wastewater and engineered systems, leaving major blind spots. NspiraGL share core features of *Nitrospira* metabolism, including a periplasmic-facing NXR that is advantageous under substrate-limiting conditions, multiple cytochrome *bd*-like oxidases, and the reverse TCA cycle for carbon fixation (S. Lücker et al. 2010). However, as with *Ca.* Nitrotoga, the *Nitrospira* genomes we reconstructed in the Great Lakes are markedly smaller than published reference genomes (median NspiraGL = 1.83 Mb, median reference = 3.72 Mb), with higher coding density and fewer paralogs, sigma factors and pseudogenes (Fig. S2.4; Dataset S2.4), consistent with genome streamlining theory (Giovannoni, Cameron Thrash, and Temperton 2014). Compared to 75 lineage II *Nitrospira* reference genomes, NspiraGL have reduced capacity for environmental sensing (two-component systems: NspiraGL = 7, mean reference = 26), transport (NspiraGL = 76-83, mean reference = 140), defense (NspiraGL = 7-8, mean reference = 26), and transposition (NspiraGL = 0-2, mean reference = 15), and lack pilus or flagellar motility (Dataset S2.5). NspiraGL encode just five sigma factors, compared to 18 in *N. moscoviensis.* Further, NspiraGL genomes encode a single NXR, while *N. moscoviensis* carries five copies that are differentially regulated (Hanna Koch et al. 2015; Mundinger et al. 2019). NspiraGL also lack the *glnE* gene for glutamine synthetase (GS) adenylyltransferase, suggesting that GS activity is not repressed by this mechanism. Together, these features suggest limited regulatory and ecological flexibility, consistent with a relatively constant, oligotrophic environment.

Compared to other *Nitrospira,* NspiraGL exhibit limited energetic flexibility, but can access diverse nitrogen sources (Fig. 2.4, Dataset S2.5). We predict that NspiraGL are unable to grow on hydrogen or formate as alternative energy sources (Hanna Koch et al. 2015; H. Koch et al. 2014), as they lack NiFe-hydogenase and formate dehydrogenase. The glycolysis and oxidative TCA cycles appear to be incomplete, lacking phosphofructokinase and citrate synthase, respectively; this suggests a limited capacity for organic carbon utilization. NspiraGL lack *nirK,* encoding NO-forming nitrite reductase, which is found in a majority of reference genomes. To obtain nitrogen for biosynthesis, NspiraGL encode a high-affinity nitrate/nitrite/cyanate transporter (*nrtABC*), assimilatory nitrite reductase (*nirA*), and cyanase (*cynS*), along with *amt* family ammonium transporter. Although none of the NspiraGL MAGs include urease (*ureCBA*), one does contain urease accessory proteins (*ureEFGD*) and two contain a urea transporter (*urtABCD*), suggesting incomplete assembly of the urea utilization pathway. As with *Ca.* Nitrotoga, we suggest that cyanase, along with urease where present, functions in nitrogen assimilation rather than cross-feeding, given the dilute environment and free-living planktonic cells.

Beyond energy, carbon, and nitrogen metabolism, we discovered striking differences between NspiraGL and reference *Nitrospira* related to DNA repair. NspiraGL encode two additional photolyase-related proteins, along with a class I cyclopyrimidine dimer (CPD) photolyase found in most reference *Nitrospira* (Fig 2.5). Photolyases use blue light energy to repair DNA lesions caused by UV radiation (Sancar 2003). The two additional genes in NspiraGL are adjacent and share best hits with *Betaproteobacteria*, suggesting recent horizontal transfer (Fig S2.10). One likely encodes an FeS-BCP photolyase, which repairs (6-4) dipyrimidine lesions (Zadow et al. 2016; F. Zhang et al. 2013). The other shares an FAD-binding domain with photolyases but the C-terminal region has no recognizable domains (Fig 2.5). This protein is widespread in aquatic bacteria and has not been functionally characterized, though an actinobacterial homolog was suggested to be involved in light sensing and regulation (Maresca et al. 2019). Beyond photolyases, NspiraGL also encode uracil-DNA glycosylase (UNG), which removes misincorporated uracil from DNA. Uracil results from deamination of cytosine, which can occur spontaneously or be induced by NO (Wink et al. 1991). In addition to the photolyases and UNG that repair DNA lesions, NspiraGL encode translesion DNA polymerase V (*umuCD*) which enables replication to proceed past lesions. Together, these genes indicate that *Nitrospira* in the Great Lakes experience significant DNA damage, including UV-induced damage that also requires light for the repair process, in hypolimnion waters with high transparency (Yousef et al. 2017) and/or during seasonal mixing.

Figure 4



										N physiology					
	% GC	Genome size (Mb)	Coding density (%)	# Sigma factors	Photolyase	Catalase	Formate dehydrogenase	NiFe hydrogenase	Thiosulfate dehydrogenase (*tsdA*)	Assim. nitrite reduction	*nirK*	# *nxrAB* copies	Cyanase	Urease	Carbon fixation pathway
NspiraGL (*n=6*)	57.8	1.83	89	5	+++	G	-	-	-	*nirA*	-	1	+	(+)	rTCA
Nitrospira japonica NJ1	58	4.1	90	27	+	E	-	-	-	ONR	+	3	+	+	rTCA
Nitrospira moscoviensis NSP M-1	61	4.5	89	18	+	EG	+	+	+	ONR	+	5	+	+	rTCA
Nitrospira defluvii	59	4.3	89	17	+	-	+	-	+	*nirA*	+	2	+	-	rTCA
NtogaGL1a (*n=4*)	51.7	1.42	91	4	(+)	G	-	-	-	-	-	1*	+	-	CBB (IC)
NtogaGL1b (*n=2*)	51.8	1.46	91	4	-	G	-	-	+	-	-	1*	+	-	CBB (IC)
Ca. Nitrotoga CP45	49	2.8	85	8	+	EG	-	+	-	*nirBD*	+	1	-	-	CBB (IC, ID/IA)
Ca. Nitrotoga KNB	50	2.6	85	6	+	G	-	+	-	*nirBD*	+	2	-	-	CBB (IC, ID/IA)

Figure 2.4 Genome properties and cross-lake distribution of nitrite oxidizing taxa *Nitrospira* (top) and *Ca.* Nitrotoga (Betaproteobacteria*;* bottom). Rows highlighted in pink represent clusters of genomes reconstructed from the Great Lakes and median values are shown for genome size, GC content, and coding density. rTCA, reductive tricarboxylic acid cycle; CBB, Calvin-Benson-Bassham cycle; ONR, octaheme nitrite reductase. Values in parentheses indicate RuBisCO type (Badger and Bek 2008). Bubble plot shows composition of NOB per lake based on metagenomic read mapping. Genes identified in only a subset of genomes are indicated by (+). The asterisk (*) indicates for *Ca.* Nitrotoga, one *nxrAB* copy was recovered in genome assemblies, but short read analysis suggests two copies per genome (Supplemental Text).

Other major differences between NspiraGL and reference *Nitrospira* are related to reactive oxygen species (ROS) (Dataset S2.5). Surprisingly, despite their oxic habitat, NspiraGL lack superoxide dismutase (SOD), monofunctional catalase (*katE*), and bacterioferritin, which limits the Fenton reaction by sequestering free iron. However, all six NspiraGL MAGs, but few reference genomes (7% of 75), have recently acquired bifunctional catalase-peroxidase *katG*; interestingly we also observed *katG* in Great Lakes *Ca.* Nitrotoga and *Nitrosospira* (Fig 2.2, Fig 2.4). The absence of SOD suggests that NspiraGL does not produce damaging levels of endogenous superoxide, perhaps because NspiraGL lack the major respiratory and non-respiratory flavoproteins that produce ROS in other SOD-containing *Nitrospira* (Imlay 2013). Unlike superoxide, H_2O_2 can cross membranes, and is known to be produced by both photooxidation of dissolved organic matter and dark heterotrophic activity (T. Zhang et al. 2016). The lakes where NspiraGL dominate have high water clarity (Yousef et al. 2017) and low productivity, consistent with abiotic photochemistry as the primary source of exogenous ROS; this stress may have selected for *katG* as a defense. NspiraGL also lack cytochrome *c* peroxidase, which is found in 70 of 75 reference genomes; this protein is proposed to function in anaerobic respiration of H_2O_2 (Khademian and Imlay 2017) and therefore its absence in NspiraGL is consistent with a constant oxic environment. Together, these results indicate that *Nitrospira* in the Great Lakes face distinct ROS pressures that have shaped their gene content.

Figure 5

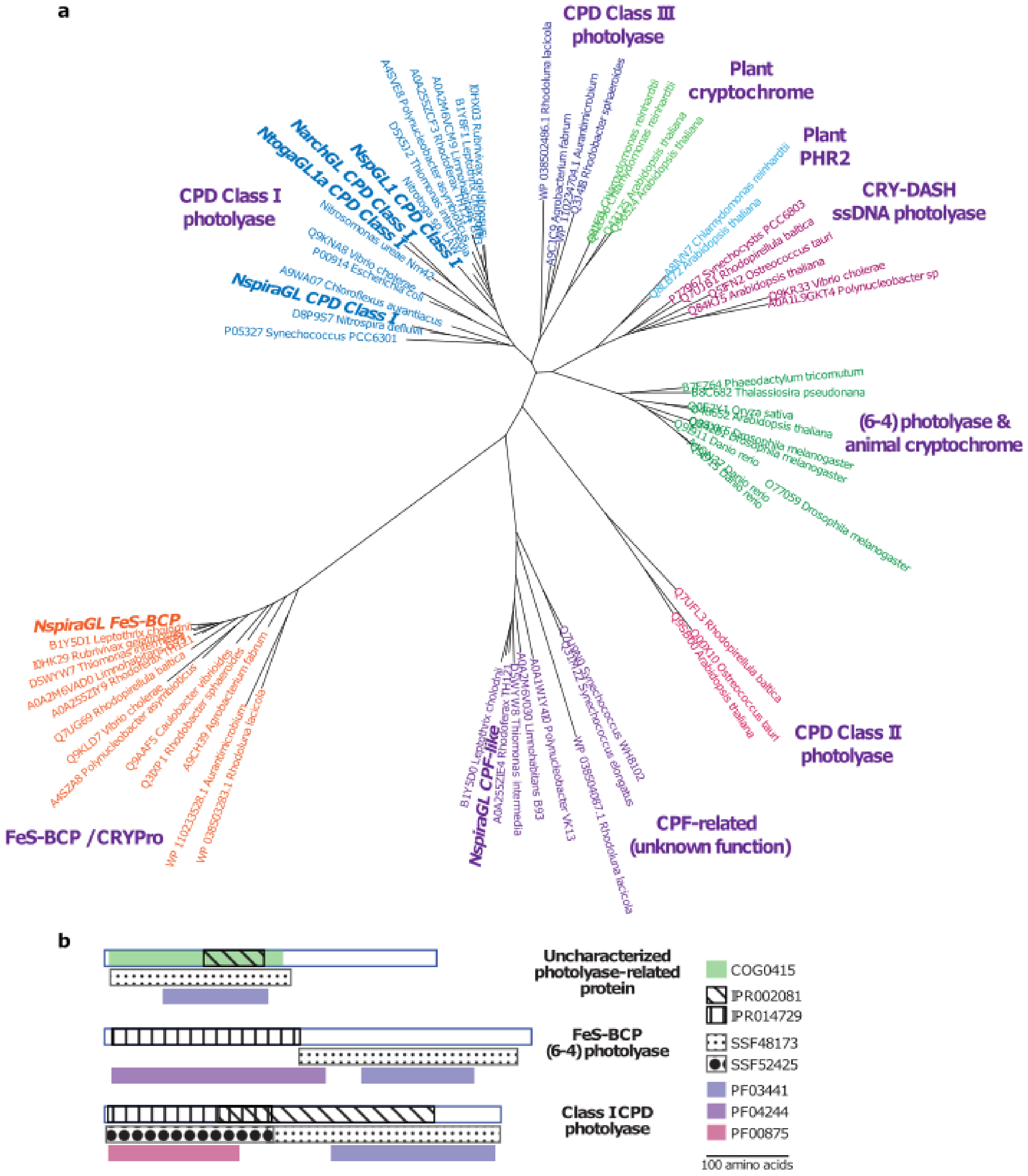


Figure 2.5 Distinct photolyase proteins in NspiraGL. **a)** Phylogenetic tree showing families of photolyases. Three families are found in NspiraGL: CPD class I photolyase, FeS-BCP / CRYPro family, and an uncharacterized CPF-related family found in diverse Bacteria. CPD Class I photolyases are also found in other nitrifiers including *Ca*. Nitrotoga NtogaGL1a, *Nitrosospira*

NspGL1, and *Nitrosoarchaeum* NarchGL. **b)** Domain structure of the three photolyase families present in NspiraGL.

2.4 Conclusions

The Laurentian Great Lakes harbor nitrifiers that are phylogenetically related, but markedly different in genome size and functional capacity, from their well-studied relatives inhabiting wastewater systems, soils, and even other freshwater systems. By examining the entire nitrifier assemblage at once, we detected common features across taxa that illuminate the selective pressures faced by microbes in deep lakes. All the lineages we describe show small genome sizes (1.3-1.7 Mb), reduced capacity for environmental sensing and response, and adaptation to a passive (i.e. non-motile) planktonic lifestyle, expanding the paradigm of streamlining observed for oceanic AOA (Santoro et al. 2015) to AOB, *Nitrospira*, and *Ca.* Nitrotoga in oligotrophic freshwaters. Within the AOB *Nitrosospira*, we found ecotypes with a gradient of genome reduction that maps onto their habitats' trophic gradient: from NspGL1 (1.4 Mb, low GC, upper lakes) to NspGL2b (1.5 Mb, upper lakes) to NspGL2a (1.6 Mb, Ontario) to NspGL3 (1.7 Mb, Erie) (Fig 2.2). The thaumarchaeal NarchGL have markedly reduced regulatory capacity like the open ocean strain *Nitrosopelagicus brevis* (Santoro et al. 2015). The NOB NspiraGL have genomes 50-60% smaller than described *Nitrospira* and dominate, along with the AOB NspGL1 and the AOA NarchGL, the deeper more oligotrophic basins, while *Ca.* Nitrotoga favor shallower, more productive basins. The emergence of Erie-specific ecotypes of both *Nitrosospira* (NspGL3) and *Ca.* Nitrotoga (NtogaGL1b) demonstrates how distinct this habitat is compared to the other lakes. Importantly, our findings here represent planktonic cells in the smallest size fraction (<1.6 µm); it is likely, especially in Erie, that particle-associated nitrifiers may be abundant and genetically distinct.

Nitrifiers inhabiting the transparent waters of the upper Great Lakes show distinctive adaptations to light including diverse photolyases, ROS detoxification, and even proteorhodopsin. This discovery is surprising, given that nitrifiers are rare in the surface mixed layer of the Great Lakes (Fig. S2.1) and that photoinhibition of ammonia oxidation and nitrifier growth is well documented (Hooper and Terry 1974; Merbt et al. 2012; Hyman and Arp 1992). We propose that proteorhodopsin could be used to augment energy metabolism when ammonia oxidation is photoinhibited and/or ammonia is substrate limited. Water clarity has increased over the past several decades in Lakes Michigan and Huron, now surpassing that of Lake Superior (Yousef et al. 2017). High light penetration along with seasonal mixing likely exposes deep water cells to damaging levels of light and oxidative stress. Future cultivation and physiological studies should examine photoinhibition and potential phototrophy in Great Lakes nitriifiers.

The capacity for nitrification is found across multiple phyla, and our work unveils new clues to understanding the ecological and evolutionary potential of these diverse lineages. This collective nitrifier diversity undoubtedly influences the cycling of carbon and nitrogen across this ecosystem, and future work will explore the differential contributions to nitrification by the distinct lineages we described here. Understanding what controls the diversity of nitrifiers and other key functional groups, and the consequences of this diversity for biogeochemistry, are essential for forecasting the effects of rapid environmental change across the large lakes of the world (e.g.(O’Reilly et al. 2015)) and predicting impacts on the critical ecosystem services they provide (Sterner et al. 2020).

2.5 Materials & Methods

2.5.1 Sample collection

Water samples were collected from the Laurentian Great Lakes aboard the R/V *Lake Guardian*, during the biannual Water Quality Surveys conducted by the U.S. EPA Great Lakes National Program Office (Barbiero et al. 2018). Station information is provided in Supplemental Dataset S2.8. Data presented here were collected in April and August 2012. Samples were collected using a CTD rosette sampler (Sea-Bird Scientific) at the surface (2 m), deep chlorophyll maximum (if present), the mid-hypolimnion, and near the bottom of the water column (10 m above the lake bottom at most stations, 1 m above bottom at shallow stations). For each sample, 5-8L of water was pre-filtered through a GF/A glass fiber filter (Whatman 1820-047; nominal pore size 1.6 µm) to exclude eukaryotic phytoplankton and particle associated microbes, and cells were collected on 0.22 µm Sterivex filters (Millipore SVGP01050). Filters were stored at -80°C. For dissolved nutrient analysis, 0.22 µm filtrate was collected in 125 ml acid-clean HDPE bottles (Nalgene) and stored at -20°C. Samples for single cell amplified genomes (SAG) were collected in August 2014. For each sample, 1 ml of raw water was incubated with 100 µl of glycerol-TE buffer (20 ml 100X TE pH 8 + 100 ml glycerol + 60 ml water; final concentration after sample addition is 10mM Tris, 1mM EDTA, 5% glycerol) for 10 minutes in the dark, then flash frozen in liquid nitrogen and stored at -80°C until processing.

2.5.2 Physicochemical data

CTD profiles, water chemistry and chlorophyll-*a* data were collected by the U.S. EPA according to standard protocols (US Environmental Protection Agency 2003) and retrieved from the Great Lakes Environmental Database (https://cdx.epa.gov/) for 2012 and 2013. In addition, we measured dissolved nitrogen species from August 2013 samples. Ammonium concentrations were measured using the OPA method in a 96-well plate (Holmes et al. 1999). Nitrate and nitrite concentrations were measured using the Greiss reaction method in a 96-well plate (Miranda, Espey, and Wink 2001). Urea concentrations were measured in a 24 well plate using a colorimetric reaction (Revilla, Alexander, and Glibert 2005).

2.5.3 16S rRNA analysis

The full 16S rRNA amplicon dataset was described by Paver and colleagues (Paver, Newton, and Coleman 2020). Here we focus on data from the V4-V5 region (primers: 515F-Y, 926R (Parada, Needham, and Fuhrman 2016)), collected in 2012 in tandem with metagenome samples from select stations. We classified sequences using the Silva v. 132 database (Quast et al. 2013) and the wang method (Wang et al. 2007) as implemented by mothur (Kozich et al. 2013). Sequences classified to each detected family of nitrifiers (ammonia oxidizer families *Nitrosomonadaceae* and *Nitrosopumilaceae*; nitrite oxidizer families *Gallionellaceae* and *Nitrospiraceae*) with a mothur-assigned confidence score above 90 were delineated into taxonomic units using minimum entropy decomposition with a minimum substantive abundance of 10 (Eren et al. 2013).

2.5.4 Metagenome and single-cell genome sequencing

One station per lake in Superior, Michigan, Huron, and Ontario, and two stations in Erie, were selected for metagenome sequencing. DNA was extracted using a modified phenol:chloroform extraction protocol (Paver, Newton, and Coleman 2020) and libraries prepared according to the Illumina TruSeq protocol. Samples from spring 2012 were sequenced at the Joint Genome Institute using Illumina HiSeq (2x150bp). Samples from summer 2012 were sequenced at the University of Chicago Functional Genomics Core Facility using Illumina HiSeq 2500 (2x250bp).

To confirm the presence of proteorhodopsin, we analyzed a single cell amplified genome from *Nitrosospira* collected from Lake Michigan and sequenced by the Joint Genome Institute. Quality filtered reads were downloaded from JGI IMG/ER and normalized using bbnorm.sh with target=100 and mindepth=2. Normalized reads were assembled using SPAdes 3.1.11 in single cell mode (Bankevich et al. 2012) with flags --sc and --careful. Resulting scaffolds were annotated identically to MAGs as described below.

2.5.5 Obtaining metagenome-assembled genomes

Raw reads for spring surface samples were quality controlled at the Joint Genome Institute, using bbduk.sh for adapter trimming (ktrim=r, minlen=40, minlenfraction=0.6,mink=11, tbo, tpe, k=23, hdist=1, hdist2=1, ftm=5) and quality filtering (maq=8, maxns=1, minlen=40, minlenfraction=0.6, k=27, hdist=1, trimq=12, qtrim=rl). Raw reads for summer hypolimnion samples were adapter trimmed, quality filtered, and interleaved using bbduk (parameters: ktrim=r, mink=8, hdist=2, k=21, forcetrimleft=10, forcetrimright=199, minlen=150) using BBTools suite version 35.74 (https://sourceforge.net/projects/bbmap/). Separate assemblies of quality filtered reads were carried out for each metagenome using metaSPAdes 3.1.11 --meta mode using default k sizes of 21, 33, 55 (Nurk et al. 2017). To enable binning based on sequence coverage, forward and reverse reads were merged using bbmerge in BBtools, using qtrim2=r trimq=10,13,16 and adapter=default. Merged short reads were then mapped onto each assembly using bowtie2 2.2.9 in --sensitive mode (Langmead and Salzberg 2012, 2), and this coverage information was used to bin assembled contigs. Binning was performed using MetaBAT2 2.12.1 (Kang et al. 2015), Binsanity 0.2.6.3 (Graham, Heidelberg, and Tully 2017) and CONCOCT 1.0.0 (Alneberg et al. 2014) using default parameters. The resulting bins were scored, aggregated, and de-replicated using DAS_Tool 1.1.1 (Sieber et al. 2018) followed by manual curation using Anvi’o 4.0 (Eren et al. 2015). We assessed genome completion and contamination of manually curated bins using CheckM 1.1.0 lineage_wf (Parks et al. 2015), and all new MAGs presented here are greater than 70% complete with less than 10% contamination (Dataset S2.3). Potential nitrifiers were screened by searching for ammonia monooxygenase, hydroxylamine oxidoreductase and nitrite oxidoreductase within reconstructed genomes using blastp 2.5.0 (Camacho et al. 2009). For bins where any of these genes were detected, we identified bacterial single copy core genes (Campbell et al. 2011) or archaeal single copy core genes (Rinke et al. 2013) using HMMER (Mistry et al. 2013), as implemented in Anvi’o. Single copy core genes were queried against proteins predicted from bacterial and archaeal genomes in RefSeq (NCBI) (O’Leary et al. 2016), and taxonomic identity of these core genes was ascertained based on a least common ancestor approach using a 0.1% window around the bit score of the best hit using KronaTools 2.7.1 (Ondov, Bergman, and Phillippy 2011). Taxonomic assignment was further validated using GTDB-tk 1.0.0 (Chaumeil et al., n.d.). Grouping of MAGs into clades and subclades based on ANI was carried out using fastANI 1.1.0 (Jain et al. 2018). Genome characteristics for each genome group were calculated as the median of those values for the group. Estimated complete genome size was calculated for MAGs and for references in the pangenome analysis using CheckM (Parks et al. 2015) completion and contamination, as follows: Estimated = Actual*(1-Contamination) / Completion. To quantify the abundance of each clade/ecotype across samples, we used competitive mapping of merged short reads using bowtie2 in sensitive mode against all nitrifier MAGs, summing up mapped read count across all MAGs in a given clade/ecotype, and dividing by total mapped nitrifier reads in a sample; these values are shown in bubble plots (Fig 2.2, 2.4).

2.5.6 Annotation and gene cluster analysis

Reference genomes were obtained from GenBank (accession numbers listed in Supplemental Dataset S2.4). The full pangenome analyses included all the genomes listed in Supplemental Dataset S2.4, but we only report results from the subset of genomes most closely related to our MAGs. This subset consists of 86 *Nitrosomonadaceae*, 5 *Ca.* Nitrotoga, 78 *Nitrosopumilaceae* within Thaumarchaeota, and 75 *Nitrospira* that fall within Lineage II. Reference genomes were treated consistently with GL MAGs, with *de novo* gene calling by prodigal 2.6.3 (Hyatt et al. 2010) via Anvi'o. Unless otherwise noted, default settings were used for all software. Genes were annotated using InterProScan 5.30-69.0 (Jones et al. 2014), GhostKOALA (Kanehisa, Sato, and Morishima 2016) and eggnog-mapper 1.0.3 against the bactNOG database (Huerta-Cepas et al. 2017). Gene cluster analysis was carried out using the Anvi'o pangenome pipeline (Delmont and Eren 2018), using blastp to determine sequence similarity, ITEP to eliminate weak similarity (Benedict et al. 2014) and MCL to cluster, using a minbit of 0.5, MCL inflation of 2 and minimum gene occurrence of 1 (van Dongen and Abreu-Goodger 2012). Sigma factors were tallied by identifying gene clusters annotated with the following PFAMs: PF00309, PF03979, PF00140, PF04542, PF04539, PF04545, PF08281. Pseudogene counts were retrieved where available from NCBI PGAP annotated genomes (Tatusova et al. 2016). Paralog counts are reported as the number of gene clusters with more than one gene per genome. Intergenic spacers were calculated using bedtools complementBed function (Quinlan and Hall 2010). Prokka 1.14.5 (Seemann 2014) was used to generate GenBank-format files from MAGs and SAGs, and genoPlotR 0.8.9 (Guy, Roat Kultima, and Andersson 2010) was used to generate initial gene neighborhood maps.

2.5.7 Gene tree construction

The NspGL1 proteorhodopsin sequence was inserted into the MicRhoDE rhodopsin tree using pplacer (Matsen, Kodner, and Armbrust 2010) through the MicRhoDE Galaxy pipeline (Boeuf et al. 2015). We then constructed a more targeted phylogenetic tree using aligned reference sequences of Supercluster III from MicRhoDE, filtered to exclude fragments shorter than 220 amino acids. To this alignment, we added NspGL1 sequences using MAFFT 7.310 (Katoh et al. 2002) along with high similarity sequences from NCBI nr that were not present in MicRhoDE. The tree was inferred using RaxML 8.2.12 with model PROTGAMMALG (Stamatakis 2014). The tree was visualized in iTOL (Letunic and Bork 2016) and more distant clusters were collapsed for clarity.

A cyanase phylogenetic tree was created using sequences drawn from querying NtogaGL cyanase against NCBI nr using blastp, as well as sequences from references (Pachiadaki et al. 2017; Spang et al. 2012; Palatinszky et al. 2015). Sequences were aligned using MAFFT (Katoh et al. 2002) and the tree was inferred using RaxML 8.2.12 with model PROTGAMMALG (Stamatakis 2014). Tree was visualized in iTOL (Letunic and Bork 2016) and branches were colored based on the taxonomy of the parent genome.

Photolyase-related proteins in GL MAGs were identified by searching for the following features: K01669, COG0415, PF03441, PF00875, PF04244, SSF48173, SSF52425. Reference proteins (n=56) spanning the previously defined families of photolyases and cryptochromes (Vechtomova, Telegina, and Kritsky 2020) were obtained from UniProt, along with aquatic bacterial sequences described by Maresca and colleagues (Maresca et al. 2019). The reference sequences were aligned using MAFFT (Katoh et al. 2002), and sequences from GL MAGs were added using the MAFFT --addfragments option. The tree was estimated using IQ-TREE 2 1.6.11 (Minh et al. 2020) and visualized using iTOL (Letunic and Bork 2016).

2.5.8 Phylogenomic tree construction

Nitrospirae, *Thaumarchaeota*, *Gallionellaceae* and *Nitrosomonadaceae* genomes were downloaded from Genbank (NCBI) (Benson et al. 2005) and included in phylogenomic trees for their respective family. Phylogenomic analyses were carried out within Anvi'o. Briefly, single copy core genes were extracted as described above, individually aligned at the protein level using muscle (Edgar 2004), and concatenated for each genome. Concatenated alignments were trimmed using Gblocks 0.91b (Castresana 2000) and analyzed by RAxML 8.2.12 (Stamatakis 2014) to create a phylogenetic tree using the PROTGAMMALG model and 50 bootstraps. Trees were visualized in iTOL (Letunic and Bork 2016).

2.5.9 Proteorhodopsin assembly verification

We used several approaches to validate the presence of proteorhodopsin in assembled *Nitrosospira* genomes, to rule out the possibility of chimeric assemblies from different species. We note that proteorhodopsin-containing contigs were independently assembled and binned together with core *Nitrosospira* contigs from seven different samples (i.e. each sample was assembled and binned separately, rather than co-assembled). In five of seven cases, proteorhodopsin and retinal biosynthesis genes were assembled together with core *Nitrosospira* genes on the same contig. To rule out a systematic reproducible error in assembly and/or binning, we compared these seven MAGs to a single cell amplified *Nitrosospira* genome (SAG) from Lake Michigan, obtained as part of another project with the JGI. This SAG was processed through JGI's standard decontamination pipeline and manually investigated to ensure lack of contamination. We found no evidence of contaminating core genes, as all core genes had best hits to either *Nitrosospira* or more generally *Nitrosomonadaceae* in nr. SAG contigs were matched to homologous contigs from NspGL1 MAGs to determine if any SAG contigs were unique using FastANI 1.1.0 (Jain et al. 2018) with --visualize flag. All contigs from this *Nitrosospira* SAG were found within an NspGL1 MAG. Bandage 0.8.1 (Wick et al. 2015) was used to manually inspect the assembly graph around the contig that contained the NspGL1 *Nitrosospira* proteorhodopsin to ensure that the assembled contig did not represent a chimeric contig or inappropriate scaffolding. We verified that a single, unique path exists from the beginning to the end of the NspGL1 contig containing proteorhodopsin (Fig 2.3). Further, we verified that consistent coverage across this contig existed by mapping short reads from the original sample using bowtie2 (Langmead and Salzberg 2012) and viewing results using Integrated Genomics Viewer 2.7.0 (Thorvaldsdóttir, Robinson, and Mesirov 2013) A closely related assembly of the same genomic region from Lake Biwa did not show evidence of proteorhodopsin; to confirm this difference between the Biwa and Great Lakes MAGs, we mapped reads from Lake Biwa (Okazaki et al. 2019) (BioProject PRJDB6644) onto the assembled contig described above using bowtie2 (Langmead and Salzberg 2012, 2)**.** This analysis demonstrated that while a large fraction of the NspGL1 contig in question recruited reads from Lake Biwa at high identity (98-99%), starting upstream of proteorhodopsin and retinal biosynthesis, this contig no longer recruited reads from Lake Biwa.

2.5.10 Manual identification of key nitrification genes

Despite recovery of 15 high completion MAGs in NspGL1/2a/2b/3, many of these MAGs lacked key nitrification genes in *amo* and *hao* operons. This was largely due to the fact that *amo* and *hao* operons were often assembled on small contigs below the minimum size cutoff we imposed for binning contigs. Difficulty in assembling these contigs was likely in part due to the several *amo* and *hao* operons with extremely high identity to one another in each genome, a phenomenon which has been observed in other *Nitrosospira* (Sedlacek et al. 2019). Manual assembly graph inspection with Bandage (Wick et al. 2015) supported this hypothesis, as did assessment of abundance of short reads associated with *amo* operons from NspGL and comparison of abundance of short reads associated with core gene *rpoB* from NspGL, using ROCker (Orellana, Rodriguez-R, and Konstantinidis 2017). Still, an exemplar MAG from at least one representative of each ecotype (NspGL1/2a/2b/3) was found with both *amo* and *hao* operons. Further, manual inspection of unbinned contigs confirmed that *amo* and *hao* operons existed on contigs in every sample from which a MAG for a particular ecotype was recovered. That is, for every time that an NspGL1 MAG was recovered from a sample, we were able to determine that an *amo* and *hao* operon which could be affiliated with NspGL1 existed, even if it was not correctly binned. Affiliation for these unbinned key nitrification genes was carried out by alignment of *amoAB* and *haoAB* sequences to *amoAB* and *haoAB* sequences correctly binned in NspGL ecotypes. This process was also carried out for two NtogaGL1a MAGs for *nxrAB*, which were poorly assembled in those two samples. Dataset S2.6 summarizes the presence of genes related to nitrification and nitrogen metabolism across all our MAGs.

2.5.11 Verification of gene absences

Metagenome-assembled genomes typically comprise tens or even hundreds of contigs, and this fragmented nature makes it impossible to say with certainty whether a particular gene is truly absent. To substantiate our claims of gene absence based on MAGs, we used several lines of evidence. First, we note that our MAGs have high estimated completion (median 96.4%, mean 94.3%), based on the presence of universal core gene markers. Second, for all new lineages described here except NspGL3, we assembled multiple similar MAGs independently from different samples, and we inferred gene absences only if the absence was replicated in multiple assemblies. Together these two factors provide strong support for cases where a missing gene would be expected to occur in a region of predominantly core genes; however these factors are less informative for cases where a missing gene might occur in a genomic island, because we have no way of assessing the completion of regions lacking core genes, and islands tend to have systematic poor assemblies across samples. A third line of evidence that we considered is chromosome organization: if a single gene is deleted from an otherwise conserved region of synteny, then this deletion should be apparent in a gene neighborhood diagram (e.g. Figures S2.5, S2.7, S2.8, S2.10). Unfortunately in many cases, our MAGs are too dissimilar from reference genomes and share little synteny with them, so this approach is not always informative.

We used a fourth approach based on quantitative analysis of short reads to verify gene absences. If a suspected missing gene were actually present in the population, but failed to assemble and/or bin with the rest of the genome, then it should be detectable in the unassembled short reads. The frequency of a gene in the population can be estimated from its abundance in the short reads, compared to the abundance of core marker genes in the short reads. We implemented this approach as follows. We searched unassembled short reads for each gene of interest that we identified as absent from MAGs (e.g. nitrosocyanin) using tblastn. Short reads with significant similarity were then filtered by best-hit taxonomy to the appropriate nitrifier group (i.e. *Nitrosomonadaceae*, *Ca.* Nitrotoga, *Nitrospira, Thaumarchaea*). These filtered short reads were enumerated and length-normalized (1000 * number of short reads / length of the target gene of interest). The same procedure was repeated for genes expected to be present in every cell (e.g. *amoAB*, *hao*, *nxrAB*, ribosomal protein genes) for comparison. If a putative missing gene (based on MAGs) has near-zero detection in the short reads, we can be confident that the gene is truly missing (or has undetectable sequence similarity, or was so recently acquired from another lineage that its best hit points to a different taxon). By contrast, if a putative missing gene (based on MAGs) does recover short reads, then the gene may be present in genomes related to our MAGs but was unassembled/unbinned, or the gene may be present in another lineage of nitrifiers that is not represented by our MAGs. Short read-based quantification of select genes is presented in Supplemental Dataset S2.7 and described in Supplemental Text.

2.5.12 Statistical analysis and plots

All statistical comparisons were carried out in R version 3.5.3 (Team 2018) and plots were generated using ggplot2 3.2.0 (Wickham 2009).

2.6 References

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz, Leo Lahti, Nicholas J Loman, Anders F Andersson, and Christopher Quince. 2014. "Binning Metagenomic Contigs by Coverage and Composition." *Nature Methods* 11 (11): 1144–46. https://doi.org/10.1038/nmeth.3103.

Alonso-Sáez, Laura, Alison S. Waller, Daniel R. Mende, Kevin Bakker, Hanna Farnelid, Patricia L. Yager, Connie Lovejoy, et al. 2012. "Role for Urea in Nitrification by Polar Marine Archaea." *Proceedings of the National Academy of Sciences* 109 (44): 17989–94. https://doi.org/10.1073/pnas.1201914109.

Auguet, Jean-Christophe, Xavier Triadó-Margarit, Natalya Nomokonova, Lluís Camarero, and Emilio O Casamayor. 2012. "Vertical Segregation and Phylogenetic Characterization of Ammonia-Oxidizing Archaea in a Deep Oligotrophic Lake." *The ISME Journal* 6 (9): 1786–97. https://doi.org/10.1038/ismej.2012.33.

Badger, Murray Ronald, and Emily Jane Bek. 2008. "Multiple Rubisco Forms in Proteobacteria: Their Functional Significance in Relation to CO2 Acquisition by the CBB Cycle." *Journal of Experimental Botany* 59 (7): 1525–41. https://doi.org/10.1093/jxb/erm297.

Baltar, Federico, and Gerhard J. Herndl. 2019. "Ideas and Perspectives: Is Dark Carbon Fixation Relevant for Oceanic Primary Production Estimates?" *Biogeosciences* 16 (19): 3793–99. https://doi.org/10.5194/bg-16-3793-2019.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Barbiero, Richard P., Barry M. Lesht, Elizabeth K. Hinchey, and Todd G. Nettesheim. 2018. "A Brief History of the U.S. EPA Great Lakes National Program Office's Water Quality Survey." *Journal of Great Lakes Research* 44 (4): 539–46. https://doi.org/10.1016/j.jglr.2018.05.011.

Beaumont, H. J. E., N. G. Hommes, L. A. Sayavedra-Soto, D. J. Arp, D. M. Arciero, A. B. Hooper, H. V. Westerhoff, and R. J. M. van Spanning. 2002. "Nitrite Reductase of Nitrosomonas Europaea Is Not Essential for Production of Gaseous Nitrogen Oxides and Confers Tolerance to Nitrite." *Journal of Bacteriology* 184 (9): 2557–60. https://doi.org/10.1128/JB.184.9.2557-2560.2002.

Beja, O. 2000. "Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea." *Science* 289 (5486): 1902–6. https://doi.org/10.1126/science.289.5486.1902.

Béjà, Oded, Elena N. Spudich, John L. Spudich, Marion Leclerc, and Edward F. DeLong. 2001. "Proteorhodopsin Phototrophy in the Ocean." *Nature* 411 (6839): 786–89. https://doi.org/10.1038/35081051.

Belisle, B. Shafer, Morgan M. Steffen, Helena L. Pound, Sue B. Watson, Jennifer M. DeBruyn, Richard A. Bourbonniere, Gregory L. Boyer, and Steven W. Wilhelm. 2016. “Urea in Lake Erie: Organic Nutrient Sources as Potentially Important Drivers of Phytoplankton Biomass.” *Journal of Great Lakes Research* 42 (3): 599–607. https://doi.org/10.1016/j.jglr.2016.03.002.

Benedict, Matthew N, James R Henriksen, William W Metcalf, Rachel J Whitaker, and Nathan D Price. 2014. “ITEP: An Integrated Toolkit for Exploration of Microbial Pan-Genomes.” *BMC Genomics* 15 (1): 8. https://doi.org/10.1186/1471-2164-15-8.

Benson, Dennis A., Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and David L. Wheeler. 2005. “GenBank.” *Nucleic Acids Research* 33 (suppl_1): D34–38. https://doi.org/10.1093/nar/gki063.

Boddicker, Andrew M., and Annika C. Mosier. 2018. “Genomic Profiling of Four Cultivated Candidatus Nitrotoga Spp. Predicts Broad Metabolic Potential and Environmental Distribution.” *The ISME Journal* 12 (12): 2864–82. https://doi.org/10.1038/s41396-018-0240-8.

Boeuf, Dominique, Stéphane Audic, Loraine Brillet-Guéguen, Christophe Caron, and Christian Jeanthon. 2015. “MicRhoDE: A Curated Database for the Analysis of Microbial Rhodopsin Diversity and Evolution.” *Database: The Journal of Biological Databases and Curation* 2015. https://doi.org/10.1093/database/bav080.

Bollmann, A., M.-J. Bar-Gilissen, and H. J. Laanbroek. 2002. “Growth at Low Ammonium Concentrations and Starvation Response as Potential Factors Involved in Niche Differentiation among Ammonia-Oxidizing Bacteria.” *Applied and Environmental Microbiology* 68 (10): 4751–57. https://doi.org/10.1128/AEM.68.10.4751-4757.2002.

Bollmann, Annette, Christopher J. Sedlacek, Jeanette Norton, Hendrikus J. Laanbroek, Yuichi Suwa, Lisa Y. Stein, Martin G. Klotz, et al. 2013. “Complete Genome Sequence of Nitrosomonas Sp. Is79, an Ammonia Oxidizing Bacterium Adapted to Low Ammonium Concentrations.” *Standards in Genomic Sciences* 7 (3): 469–82. https://doi.org/10.4056/sigs.3517166.

Cabello-Yeves, Pedro J., Tamara I. Zemskaya, Riccardo Rosselli, Felipe H. Coutinho, Alexandra S. Zakharenko, Vadim V. Blinov, and Francisco Rodriguez-Valera. 2018. “Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal.” Edited by Harold L. Drake. *Applied and Environmental Microbiology* 84 (1). https://doi.org/10.1128/AEM.02132-17.

Callieri, Cristiana, Manuela Coci, Ester M. Eckert, Michaela M. Salcher, and Roberto Bertoni. 2014. “Archaea and Bacteria in Deep Lake Hypolimnion: In Situ Dark Inorganic Carbon Uptake.” *Journal of Limnology* 73 (1). https://doi.org/10.4081/jlimnol.2014.937.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10 (1): 421. https://doi.org/10.1186/1471-2105-10-421.

Campbell, B. J., L. Yu, J. F. Heidelberg, and D. L. Kirchman. 2011. “Activity of Abundant and Rare Bacteria in a Coastal Ocean.” *PNAS* 108 (31): 12776–81. https://doi.org/10.1073/pnas.1101405108.

Canfield, Donald E., Alexander N. Glazer, and Paul G. Falkowski. 2010. “The Evolution and Future of Earth’s Nitrogen Cycle.” *Science* 330 (6001): 192–96. https://doi.org/10.1126/science.1186120.

Castresana, J. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution* 17 (4): 540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chaumeil, Pierre-Alain, Aaron J. Mussig, Philip Hugenholtz, and Donovan H. Parks. n.d. “GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btz848.

Clevinger, Curtis C., Robert T. Heath, and Darren L. Bade. 2014. “Oxygen Use by Nitrification in the Hypolimnion and Sediments of Lake Erie.” *Journal of Great Lakes Research* 40 (1): 202–7. https://doi.org/10.1016/j.jglr.2013.09.015.

Daims, Holger, Sebastian Lücker, and Michael Wagner. 2016. “A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria.” *Trends in Microbiology* 24 (9): 699–712. https://doi.org/10.1016/j.tim.2016.05.004.

Delmont, Tom O., and A. Murat Eren. 2018. “Linking Pangenomes and Metagenomes: The Prochlorococcus Metapangenome.” *PeerJ* 6 (January): e4320. https://doi.org/10.7717/peerj.4320.

Dongen, Stijn van, and Cei Abreu-Goodger. 2012. “Using MCL to Extract Clusters from Networks.” *Methods in Molecular Biology (Clifton, N.J.)* 804: 281–95. https://doi.org/10.1007/978-1-61779-361-5_15.

Dove, Alice, and Steven C. Chapra. 2015. “Long-Term Trends of Nutrients and Trophic Response Variables for the Great Lakes.” *Limnology and Oceanography* 60 (2): 696–721. https://doi.org/10.1002/lno.10055.

Edgar, R. C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. 2015. “Anvi’o: An Advanced Analysis and Visualization Platform for ‘omics Data.” *PeerJ* 3 (October): e1319.

https://doi.org/10.7717/peerj.1319.

Eren, A. Murat, Loïs Maignien, Woo Jun Sul, Leslie G. Murphy, Sharon L. Grim, Hilary G. Morrison, and Mitchell L. Sogin. 2013. "Oligotyping: Differentiating between Closely Related Microbial Taxa Using 16S RRNA Gene Data." Edited by Robert Freckleton. *Methods in Ecology and Evolution* 4 (12): 1111–19. https://doi.org/10.1111/2041-210X.12114.

Fujimoto, Masanori, Joann Cavaletto, James R. Liebig, Ann McCarthy, Henry A. Vanderploeg, and Vincent J. Denef. 2016. "Spatiotemporal Distribution of Bacterioplankton Functional Groups along a Freshwater Estuary to Pelagic Gradient in Lake Michigan." *Journal of Great Lakes Research* 42 (5): 1036–48. https://doi.org/10.1016/j.jglr.2016.07.029.

Füssel, Jessika, Sebastian Lücker, Pelin Yilmaz, Boris Nowka, Maartje A. H. J. van Kessel, Patric Bourceau, Philipp F. Hach, et al. 2017. "Adaptability as the Key to Success for the Ubiquitous Marine Nitrite Oxidizer *Nitrococcus*." *Science Advances* 3 (11): e1700807. https://doi.org/10.1126/sciadv.1700807.

Garcia, J. C., H. Urakawa, V. Q. Le, L. Y. Stein, M. G. Klotz, and J. L. Nielsen. 2013. "Draft Genome Sequence of Nitrosospira Sp. Strain APG3, a Psychrotolerant Ammonia-Oxidizing Bacterium Isolated from Sandy Lake Sediment." *Genome Announcements* 1 (6): e00930-13, 1/6/e00930-13. https://doi.org/10.1128/genomeA.00930-13.

Gardner, Wayne S., Peter J. Lavrentyev, Joann F. Cavaletto, Mark J. McCarthy, Brian J. Eadie, Thomas H. Johengen, and James B. Cotner. 2004. "Distribution and Dynamics of Nitrogen and Microbial Plankton in Southern Lake Michigan during Spring Transition 1999-2000." *Journal of Geophysical Research: Oceans* 109 (C3). https://doi.org/10.1029/2002JC001588.

Giovannoni, Stephen J, J Cameron Thrash, and Ben Temperton. 2014. "Implications of Streamlining Theory for Microbial Ecology." *The ISME Journal* 8 (8): 1553–65. https://doi.org/10.1038/ismej.2014.60.

Graham, Elaina D., John F. Heidelberg, and Benjamin J. Tully. 2017. "BinSanity: Unsupervised Clustering of Environmental Microbial Assemblies Using Coverage and Affinity Propagation." *PeerJ* 5 (March): e3035. https://doi.org/10.7717/peerj.3035.

Guerrero, Ma, and Rd Jones. 1996. "Photoinhibition of Marine Nitrifying Bacteria. I. Wavelength-Dependent Response." *Marine Ecology Progress Series* 141: 183–92. https://doi.org/10.3354/meps141183.

Guy, Lionel, Jens Roat Kultima, and Siv G. E. Andersson. 2010. "GenoPlotR: Comparative Gene and Genome Visualization in R." *Bioinformatics* 26 (18): 2334–35. https://doi.org/10.1093/bioinformatics/btq413.

Hayden, Curtis J., and J. Michael Beman. 2014. "High Abundances of Potentially Active

Ammonia-Oxidizing Bacteria and Archaea in Oligotrophic, High-Altitude Lakes of the Sierra Nevada, California, USA.” Edited by Melanie R. Mormile. *PLoS ONE* 9 (11): e111560. https://doi.org/10.1371/journal.pone.0111560.

Herber, Janina, Franziska Klotz, Benjamin Frommeyer, Severin Weis, Dietmar Straile, Allison Kolar, Johannes Sikorski, Markus Egert, Michael Dannenmann, and Michael Pester. 2020. “A Single Thaumarchaeon Drives Nitrification in Deep Oligotrophic Lake Constance.” *Environmental Microbiology* 22 (1): 212–28. https://doi.org/10.1111/1462-2920.14840.

Holmes, Robert M, Alain Aminot, Roger Kérouel, Bethanie A Hooker, and Bruce J Peterson. 1999. “A Simple and Precise Method for Measuring Ammonium in Marine and Freshwater Ecosystems.” *Canadian Journal of Fisheries and Aquatic Sciences* 56 (10): 1801–8. https://doi.org/10.1139/f99-128.

Hooper, Alan B, and Kathleen R Terry. 1974. “Photoinactivation of Ammonia Oxidation in Nitrosomonas.” *Journal of Bacteriology* 119: 8.

Horrigan, S. G., and A. L. Springer. 1990. “Oceanic and Estuarine Ammonium Oxidation: Effects of Light.” *Limnology and Oceanography* 35 (2): 479–82. https://doi.org/10.4319/lo.1990.35.2.0479.

Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. “Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper.” *Molecular Biology and Evolution* 34 (8): 2115–22. https://doi.org/10.1093/molbev/msx148.

Hugoni, Mylène, Sandrine Etien, Antoine Bourges, Cécile Lepère, Isabelle Domaizon, Clarisse Mallet, Gisèle Bronner, Didier Debroas, and Isabelle Mary. 2013. “Dynamics of Ammonia-Oxidizing Archaea and Bacteria in Contrasted Freshwater Ecosystems.” *Research in Microbiology* 164 (4): 360–70. https://doi.org/10.1016/j.resmic.2013.01.004.

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” *BMC Bioinformatics* 11 (1). https://doi.org/10.1186/1471-2105-11-119.

Hyman, M R, and D J Arp. 1992. “14C2H2- and 14CO2-Labeling Studies of the de Novo Synthesis of Polypeptides by Nitrosomonas Europaea during Recovery from Acetylene and Light Inactivation of Ammonia Monooxygenase.” *Journal of Biological Chemistry* 267 (3): 1534–45. https://doi.org/10.1016/S0021-9258(18)45979-0.

Imlay, James A. 2013. “The Molecular Mechanisms and Physiological Consequences of Oxidative Stress: Lessons from a Model Bacterium.” *Nature Reviews Microbiology* 11 (7): 443–54. https://doi.org/10.1038/nrmicro3032.

Ishii, Kento, Hirotsugu Fujitani, Yuji Sekiguchi, and Satoshi Tsuneda. 2020. “Physiological and

Genomic Characterization of a New ‘Candidatus Nitrotoga’ Isolate.” *Environmental Microbiology* 22 (6): 2365–82. https://doi.org/10.1111/1462-2920.15015.

Jain, Chirag, Luis M. Rodriguez-R, Adam M. Phillippy, Konstantinos T. Konstantinidis, and Srinivas Aluru. 2018. “High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries.” *Nature Communications* 9 (1): 1–8. https://doi.org/10.1038/s41467-018-07641-9.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, et al. 2014. “InterProScan 5: Genome-Scale Protein Function Classification.” *Bioinformatics* 30 (9): 1236–40. https://doi.org/10.1093/bioinformatics/btu031.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. 2016. “BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences.” *Journal of Molecular Biology*, Computation Resources for Molecular Biology, 428 (4): 726–31. https://doi.org/10.1016/j.jmb.2015.11.006.

Kang, Dongwan D., Jeff Froula, Rob Egan, and Zhong Wang. 2015. “MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities.” *PeerJ* 3 (August): e1165. https://doi.org/10.7717/peerj.1165.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66. https://doi.org/10.1093/nar/gkf436.

Khademian, Maryam, and James A. Imlay. 2017. “*Escherichia Coli* Cytochrome *c* Peroxidase Is a Respiratory Oxidase That Enables the Use of Hydrogen Peroxide as a Terminal Electron Acceptor.” *Proceedings of the National Academy of Sciences* 114 (33): E6922–31. https://doi.org/10.1073/pnas.1701587114.

Kitzinger, Katharina, Hanna Koch, Sebastian Lücker, Christopher J. Sedlacek, Craig Herbold, Jasmin Schwarz, Anne Daebeler, et al. 2018. “Characterization of the First ‘Candidatus Nitrotoga’ Isolate Reveals Metabolic Versatility and Separate Evolution of Widespread Nitrite-Oxidizing Bacteria.” *MBio* 9 (4). https://doi.org/10.1128/mBio.01186-18.

Kitzinger, Katharina, Hannah K. Marchant, Laura A. Bristow, Craig W. Herbold, Cory C. Padilla, Abiel T. Kidane, Sten Littmann, et al. 2020. “Single Cell Analyses Reveal Contrasting Life Strategies of the Two Main Nitrifiers in the Ocean.” *Nature Communications* 11 (1): 767. https://doi.org/10.1038/s41467-020-14542-3.

Koch, H., A. Galushko, M. Albertsen, A. Schintlmeister, C. Gruber-Dorninger, S. Lucker, E. Pelletier, et al. 2014. “Growth of Nitrite-Oxidizing Bacteria by Aerobic Hydrogen Oxidation.” *Science* 345 (6200): 1052–54. https://doi.org/10.1126/science.1256985.

Koch, Hanna, Sebastian Lücker, Mads Albertsen, Katharina Kitzinger, Craig Herbold, Eva

Spieck, Per Halkjaer Nielsen, Michael Wagner, and Holger Daims. 2015. “Expanded Metabolic Versatility of Ubiquitous Nitrite-Oxidizing Bacteria from the Genus *Nitrospira*.” *Proceedings of the National Academy of Sciences* 112 (36): 11371–76. https://doi.org/10.1073/pnas.1506533112.

Kozich, James J., Sarah L. Westcott, Nielson T. Baxter, Sarah K. Highlander, and Patrick D. Schloss. 2013. “Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform.” *Appl. Environ. Microbiol.* 79 (17): 5112–20. https://doi.org/10.1128/AEM.01043-13.

Kralj, Joel M., Vladislav B. Bergo, Jason J. Amsden, Elena N. Spudich, John L. Spudich, and Kenneth J. Rothschild. 2008. “Protonation State of Glu142 Differs in the Green- and Blue-Absorbing Variants of Proteorhodopsin.” *Biochemistry* 47 (11): 3447–53. https://doi.org/10.1021/bi7018964.

Kumar, Sanjeev, Robert W. Sterner, Jacques C. Finlay, and Sandra Brovold. 2007. “Spatial and Temporal Variation of Ammonium in Lake Superior.” *Journal of Great Lakes Research* 33 (3): 581–91. https://doi.org/10.3394/0380-1330(2007)33[581:SATVOA]2.0.CO;2.

Langmead, Ben, and Steven L Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Lean, D. R. S., and R. Knowles. 1987. “Nitrogen Transformations in Lake Ontario.” *Canadian Journal of Fisheries and Aquatic Sciences* 44 (12): 2133–43. https://doi.org/10.1139/f87-262.

Letunic, Ivica, and Peer Bork. 2016. “Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees.” *Nucleic Acids Research* 44 (W1): W242–45. https://doi.org/10.1093/nar/gkw290.

Lücker, S., M. Wagner, F. Maixner, E. Pelletier, H. Koch, B. Vacherie, T. Rattei, et al. 2010. “A Nitrospira Metagenome Illuminates the Physiology and Evolution of Globally Important Nitrite-Oxidizing Bacteria.” *Proceedings of the National Academy of Sciences* 107 (30): 13479–84. https://doi.org/10.1073/pnas.1003860107.

Lücker, Sebastian, Boris Nowka, Thomas Rattei, Eva Spieck, and Holger Daims. 2013. “The Genome of Nitrospina Gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer.” *Frontiers in Microbiology* 4. https://doi.org/10.3389/fmicb.2013.00027.

Maresca, Julia A., Jessica L. Keffer, Priscilla P. Hempel, Shawn W. Polson, Olga Shevchenko, Jaysheel Bhavsar, Deborah Powell, Kelsey J. Miller, Archana Singh, and Martin W. Hahn. 2019. “Light Modulates the Physiology of Nonphototrophic Actinobacteria.” *Journal of Bacteriology* 201 (10). https://doi.org/10.1128/JB.00740-18.

Martinez, A., A. S. Bradley, J. R. Waldbauer, R. E. Summons, and E. F. DeLong. 2007.

“Proteorhodopsin Photosystem Gene Expression Enables Photophosphorylation in a Heterologous Host.” *Proceedings of the National Academy of Sciences* 104 (13): 5590–95. https://doi.org/10.1073/pnas.0611470104.

Matsen, Frederick A., Robin B. Kodner, and E Virginia Armbrust. 2010. “Pplacer: Linear Time Maximum-Likelihood and Bayesian Phylogenetic Placement of Sequences onto a Fixed Reference Tree.” *BMC Bioinformatics* 11 (1): 538. https://doi.org/10.1186/1471-2105-11-538.

Merbt, Stephanie N., David A. Stahl, Emilio O. Casamayor, Eugènia Martí, Graeme W. Nicol, and James I. Prosser. 2012. “Differential Photoinhibition of Bacterial and Archaeal Ammonia Oxidation.” *FEMS Microbiology Letters* 327 (1): 41–46. https://doi.org/10.1111/j.1574-6968.2011.02457.x.

Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. 2020. “IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.” *Molecular Biology and Evolution* 37 (5): 1530–34. https://doi.org/10.1093/molbev/msaa015.

Miranda, K. M., M. G. Espey, and D. A. Wink. 2001. “A Rapid, Simple Spectrophotometric Method for Simultaneous Detection of Nitrate and Nitrite.” *Nitric Oxide: Biology and Chemistry* 5 (1): 62–71. https://doi.org/10.1006/niox.2000.0319.

Mistry, Jaina, Robert D. Finn, Sean R. Eddy, Alex Bateman, and Marco Punta. 2013. “Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions.” *Nucleic Acids Research* 41 (12): e121–e121. https://doi.org/10.1093/nar/gkt263.

Mukherjee, Maitreyee, Anirban Ray, Anton F. Post, Robert Michael McKay, and George S. Bullerjahn. 2016. “Identification, Enumeration and Diversity of Nitrifying Planktonic Archaea and Bacteria in Trophic End Members of the Laurentian Great Lakes.” *Journal of Great Lakes Research* 42 (1): 39–49. https://doi.org/10.1016/j.jglr.2015.11.007.

Mundinger, Aniela B., Christopher E. Lawson, Mike S. M. Jetten, Hanna Koch, and Sebastian Lücker. 2019. “Cultivation and Transcriptional Analysis of a Canonical Nitrospira Under Stable Growth Conditions.” *Frontiers in Microbiology* 10. https://doi.org/10.3389/fmicb.2019.01325.

Murphy, T. P. 1980. “Ammonia and Nitrate Uptake in the Lower Great Lakes.” *Canadian Journal of Fisheries and Aquatic Sciences* 37 (9): 1365–72. https://doi.org/10.1139/f80-175.
Neilson, M. A., and R. J. J. Stevens. 1987. “Spatial Heterogeneity of Nutrients and Organic Matter in Lake Ontario.” *Canadian Journal of Fisheries and Aquatic Sciences* 44 (12): 2192–2203. https://doi.org/10.1139/f87-269.

Norton, Jeanette M., Martin G. Klotz, Lisa Y. Stein, Daniel J. Arp, Peter J. Bottomley, Patrick S.

G. Chain, Loren J. Hauser, et al. 2008. “Complete Genome Sequence of Nitrosospira Multiformis, an Ammonia-Oxidizing Bacterium from the Soil Environment.” *Appl. Environ. Microbiol.* 74 (11): 3559–72. https://doi.org/10.1128/AEM.02722-07.

Nowka, Boris, Holger Daims, and Eva Spieck. 2015. “Comparison of Oxidation Kinetics of Nitrite-Oxidizing Bacteria: Nitrite Availability as a Key Factor in Niche Differentiation.” Edited by G. Voordouw. *Applied and Environmental Microbiology* 81 (2): 745–53. https://doi.org/10.1128/AEM.02734-14.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. “MetaSPAdes: A New Versatile Metagenomic Assembler.” *Genome Research* 27 (5): 824–34. https://doi.org/10.1101/gr.213959.116.

Okazaki, Yusuke, Shohei Fujinaga, Atsushi Tanaka, Ayato Kohzu, Hideo Oyagi, and Shin-ichi Nakano. 2017. “Ubiquity and Quantitative Significance of Bacterioplankton Lineages Inhabiting the Oxygenated Hypolimnion of Deep Freshwater Lakes.” *The ISME Journal* 11 (10): 2279–93. https://doi.org/10.1038/ismej.2017.89.

Okazaki, Yusuke, Yosuke Nishimura, Takashi Yoshida, Hiroyuki Ogata, and Shin-Ichi Nakano. 2019. “Genome-Resolved Viral and Cellular Metagenomes Revealed Potential Key Virus-Host Interactions in a Deep Freshwater Lake.” *Environmental Microbiology*, October. https://doi.org/10.1111/1462-2920.14816.

O’Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation.” *Nucleic Acids Research* 44 (D1): D733–45. https://doi.org/10.1093/nar/gkv1189.

Ondov, Brian D, Nicholas H Bergman, and Adam M Phillippy. 2011. “Interactive Metagenomic Visualization in a Web Browser.” *BMC Bioinformatics* 12 (1). https://doi.org/10.1186/1471-2105-12-385.

O’Reilly, Catherine M., Sapna Sharma, Derek K. Gray, Stephanie E. Hampton, Jordan S. Read, Rex J. Rowley, Philipp Schneider, et al. 2015. “Rapid and Highly Variable Warming of Lake Surface Waters around the Globe.” *Geophysical Research Letters* 42 (24): 10,773-10,781. https://doi.org/10.1002/2015GL066235.

Orellana, Luis H., Luis M. Rodriguez-R, and Konstantinos T. Konstantinidis. 2017. “ROCker: Accurate Detection and Quantification of Target Genes in Short-Read Metagenomic Data Sets by Modeling Sliding-Window Bitscores.” *Nucleic Acids Research* 45 (3): e14. https://doi.org/10.1093/nar/gkw900.

Pachiadaki, Maria G., Eva Sintes, Kristin Bergauer, Julia M. Brown, Nicholas R. Record, Brandon K. Swan, Mary Elizabeth Mathyer, et al. 2017. “Major Role of Nitrite-Oxidizing Bacteria in Dark Ocean Carbon Fixation.” *Science* 358 (6366): 1046–51. https://doi.org/10.1126/science.aan8260.

Palatinszky, Marton, Craig Herbold, Nico Jehmlich, Mario Pogoda, Ping Han, Martin von Bergen, Ilias Lagkouvardos, et al. 2015. “Cyanate as an Energy Source for Nitrifiers.” *Nature* 524 (7563): 105–8. https://doi.org/10.1038/nature14856.

Parada, Alma E., David M. Needham, and Jed A. Fuhrman. 2016. “Every Base Matters: Assessing Small Subunit RRNA Primers for Marine Microbiomes with Mock Communities, Time Series and Global Field Samples.” *Environmental Microbiology* 18 (5): 1403–14. https://doi.org/10.1111/1462-2920.13023.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55. https://doi.org/10.1101/gr.186072.114.

Paver, Sara F., Ryan J. Newton, and Maureen L. Coleman. 2020. “Microbial Communities of the Laurentian Great Lakes Reflect Connectivity and Local Biogeochemistry.” *Environmental Microbiology* 22 (1): 433–46. https://doi.org/10.1111/1462-2920.14862.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.” *Nucleic Acids Research* 41 (D1): D590–96. https://doi.org/10.1093/nar/gks1219.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Rae, Benjamin D., Benedict M. Long, Murray R. Badger, and G. Dean Price. 2013. “Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria.” *Microbiology and Molecular Biology Reviews* 77 (3): 357–79. https://doi.org/10.1128/MMBR.00061-12.

Reckel, Sina, Daniel Gottstein, Jochen Stehle, Frank Löhr, Mirka-Kristin Verhoefen, Mitsuhiro Takeda, Robert Silvers, et al. 2011. “Solution NMR Structure of Proteorhodopsin.” *Angewandte Chemie International Edition* 50 (50): 11942–46. https://doi.org/10.1002/anie.201105648.

Reinthaler, Thomas, Hendrik M. van Aken, and Gerhard J. Herndl. 2010. “Major Contribution of Autotrophy to Microbial Carbon Cycling in the Deep North Atlantic’s Interior.” *Deep Sea Research Part II: Topical Studies in Oceanography* 57 (16): 1572–80. https://doi.org/10.1016/j.dsr2.2010.02.023.

Revilla, Marta, Jeffrey Alexander, and Patricia M. Glibert. 2005. “Urea Analysis in Coastal Waters: Comparison of Enzymatic and Direct Methods.” *Limnology and Oceanography: Methods* 3 (7): 290–99. https://doi.org/10.4319/lom.2005.3.290.

Rice, Marlen C., Jeanette M. Norton, Frederica Valois, Annette Bollmann, Peter J. Bottomley, Martin G. Klotz, Hendrikus J. Laanbroek, et al. 2016. “Complete Genome of Nitrosospira Briensis C-128, an Ammonia-Oxidizing Bacterium from Agricultural Soil.” *Standards in Genomic Sciences* 11 (July). https://doi.org/10.1186/s40793-016-0168-4.

Rinke, Christian, Patrick Schwientek, Alexander Sczyrba, Natalia N. Ivanova, Iain J. Anderson, Jan-Fang Cheng, Aaron Darling, et al. 2013. “Insights into the Phylogeny and Coding Potential of Microbial Dark Matter.” *Nature* 499 (7459): 431–37. https://doi.org/10.1038/nature12352.

Rozmarynowycz, Mark J., Benjamin F. N. Beall, George S. Bullerjahn, Gaston E. Small, Robert W. Sterner, Sandra S. Brovold, Nigel A. D’souza, Susan B. Watson, and Robert Michael L. McKay. 2019. “Transitions in Microbial Communities along a 1600 km Freshwater Trophic Gradient.” *Journal of Great Lakes Research* 45 (2): 263–76. https://doi.org/10.1016/j.jglr.2019.01.004.

Ryals, J., R. Y. Hsu, M. N. Lipsett, and H. Bremer. 1982. “Isolation of Single-Site Escherichia Coli Mutants Deficient in Thiamine and 4-Thiouridine Syntheses: Identification of a NuvC Mutant.” *Journal of Bacteriology* 151 (2): 899–904.

Sabehi, Gazalah, Alexander Loy, Kwang-Hwan Jung, Ranga Partha, John L. Spudich, Tal Isaacson, Joseph Hirschberg, Michael Wagner, and Oded Béjà. 2005. “New Insights into Metabolic Properties of Marine Bacteria Encoding Proteorhodopsins.” *PLoS Biology* 3 (8): e273. https://doi.org/10.1371/journal.pbio.0030273.

Sancar, Aziz. 2003. “Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors.” *Chemical Reviews* 103 (6): 2203–38. https://doi.org/10.1021/cr0204348.

Santoro, Alyson E., Christopher L. Dupont, R. Alex Richter, Matthew T. Craig, Paul Carini, Matthew R. McIlvin, Youngik Yang, William D. Orsi, Dawn M. Moran, and Mak A. Saito. 2015. “Genomic and Proteomic Characterization of ‘ *Candidatus* Nitrosopelagicus Brevis’: An Ammonia-Oxidizing Archaeon from the Open Ocean.” *Proceedings of the National Academy of Sciences* 112 (4): 1173–78. https://doi.org/10.1073/pnas.1416223112.

Santoro, Alyson E., R. Alexander Richter, and Christopher L. Dupont. 2019. “Planktonic Marine Archaea.” *Annual Review of Marine Science* 11 (1): 131–58. https://doi.org/10.1146/annurev-marine-121916-063141.

Schleper, Christa. 2010. “Ammonia Oxidation: Different Niches for Bacteria and Archaea?” *The ISME Journal* 4 (9): 1092–94. https://doi.org/10.1038/ismej.2010.111.

Sedlacek, Christopher J., Brian McGowan, Yuichi Suwa, Luis Sayavedra-Soto, Hendrikus J. Laanbroek, Lisa Y. Stein, Jeanette M. Norton, Martin G. Klotz, and Annette Bollmann. 2019. “A Physiological and Genomic Comparison of Nitrosomonas Cluster 6a and 7

Ammonia-Oxidizing Bacteria." *Microbial Ecology* 78 (4): 985–94. https://doi.org/10.1007/s00248-019-01378-8.

Seemann, Torsten. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics (Oxford, England)* 30 (14): 2068–69. https://doi.org/10.1093/bioinformatics/btu153.

Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess, Susannah G. Tringe, and Jillian F. Banfield. 2018. "Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy." *Nature Microbiology* 3 (7): 836–43. https://doi.org/10.1038/s41564-018-0171-1.

Small, Gaston E., George S. Bullerjahn, Robert W. Sterner, Benjamin F. N. Beall, Sandra Brovold, Jacques C. Finlay, Robert M. L. McKay, and Maitreyee Mukherjee. 2013. "Rates and Controls of Nitrification in a Large Oligotrophic Lake." *Limnology and Oceanography* 58 (1): 276–86. https://doi.org/10.4319/lo.2013.58.1.0276.

Smith, Jason M., Francisco P. Chavez, and Christopher A. Francis. 2014. "Ammonium Uptake by Phytoplankton Regulates Nitrification in the Sunlit Ocean." Edited by Connie Lovejoy. *PLoS ONE* 9 (9): e108173. https://doi.org/10.1371/journal.pone.0108173.

Spang, Anja, Anja Poehlein, Pierre Offre, Sabine Zumbrägel, Susanne Haider, Nicolas Rychlik, Boris Nowka, et al. 2012. "The Genome of the Ammonia-Oxidizing Candidatus Nitrososphaera Gargensis: Insights into Metabolic Versatility and Environmental Adaptations." *Environmental Microbiology* 14 (12): 3122–45. https://doi.org/10.1111/j.1462-2920.2012.02893.x.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Starkenburg, Shawn R., Patrick S. G. Chain, Luis A. Sayavedra-Soto, Loren Hauser, Miriam L. Land, Frank W. Larimer, Stephanie A. Malfatti, et al. 2006. "Genome Sequence of the Chemolithoautotrophic Nitrite-Oxidizing Bacterium Nitrobacter Winogradskyi Nb-255." *Applied and Environmental Microbiology* 72 (3): 2050–63. https://doi.org/10.1128/AEM.72.3.2050-2063.2006.

Sterner, Robert W. 2010. "In Situ-Measured Primary Production in Lake Superior." *Journal of Great Lakes Research* 36 (1): 139–49. https://doi.org/10.1016/j.jglr.2009.12.007.

Sterner, Robert W., Bonnie Keeler, Stephen Polasky, Rajendra Poudel, Kirsten Rhude, and Maggie Rogers. 2020. "Ecosystem Services of Earth's Largest Freshwater Lakes." *Ecosystem Services* 41 (February): 101046. https://doi.org/10.1016/j.ecoser.2019.101046.

Sterner, Robert W., Tanya M. Smutka, R. Michael L. McKay, Qin Xiaoming, Erik T. Brown, and Robert M. Sherrell. 2004. "Phosphorus and Trace Metal Limitation of Algae and Bacteria in Lake Superior." *Limnology and Oceanography* 49 (2): 495–507.

https://doi.org/10.4319/lo.2004.49.2.0495.

Swan, B. K., M. Martinez-Garcia, C. M. Preston, A. Sczyrba, T. Woyke, D. Lamy, T. Reinthaler, et al. 2011. “Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean.” *Science* 333 (6047): 1296–1300. https://doi.org/10.1126/science.1203690.

Tatusova, Tatiana, Michael DiCuccio, Azat Badretdin, Vyacheslav Chetvernin, Eric P. Nawrocki, Leonid Zaslavsky, Alexandre Lomsadze, Kim D. Pruitt, Mark Borodovsky, and James Ostell. 2016. “NCBI Prokaryotic Genome Annotation Pipeline.” *Nucleic Acids Research* 44 (14): 6614–24. https://doi.org/10.1093/nar/gkw569.

Team, R. Core. 2018. “R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing V, Austria. 2016.”

Thorvaldsdóttir, Helga, James T. Robinson, and Jill P. Mesirov. 2013. “Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration.” *Briefings in Bioinformatics* 14 (2): 178–92. https://doi.org/10.1093/bib/bbs017.

Urbach, Ena, Kevin L. Vergin, Lei Young, Ariel Morse, Gary L. Larson, and Stephen J. Giovannoni. 2001. “Unusual Bacterioplankton Community Structure in Ultra-Oligotrophic Crater Lake.” *Limnology and Oceanography* 46 (3): 557–72. https://doi.org/10.4319/lo.2001.46.3.0557.

US Environmental Protection Agency. 2003. *Sampling and Analytical Procedures for GLNPO’s Open Lake Water Quality Survey of the Great Lakes*. Great Lakes National Program Office Illinois,, USA.

Vechtomova, Yu. L., T. A. Telegina, and M. S. Kritsky. 2020. “Evolution of Proteins of the DNA Photolyase/Cryptochrome Family.” *Biochemistry (Moscow)* 85 (1): 131–53. https://doi.org/10.1134/S0006297920140072.

Vollenweider, R. A., M. Munawar, and P. Stadelmann. 1974. “A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes.” *Journal of the Fisheries Research Board of Canada* 31 (5): 739–62. https://doi.org/10.1139/f74-100.

Wang, Qiong, George M. Garrity, James M. Tiedje, and James R. Cole. 2007. “Naïve Bayesian Classifier for Rapid Assignment of RRNA Sequences into the New Bacterial Taxonomy.” *Applied and Environmental Microbiology* 73 (16): 5261–67. https://doi.org/10.1128/AEM.00062-07.

Watzer, Björn, and Karl Forchhammer. 2018. “Cyanophycin Synthesis Optimizes Nitrogen Utilization in the Unicellular Cyanobacterium Synechocystis Sp. Strain PCC 6803.” *Appl. Environ. Microbiol.* 84 (20): e01298-18.

Wegen, Simone, Boris Nowka, and Eva Spieck. 2019. “Low Temperature and Neutral PH Define

‘Candidatus Nitrotoga Sp.’ as a Competitive Nitrite Oxidizer in Coculture with Nitrospira Defluvii.” *Applied and Environmental Microbiology* 85 (9). https://doi.org/10.1128/AEM.02569-18.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. “Bandage: Interactive Visualization of de Novo Genome Assemblies.” *Bioinformatics* 31 (20): 3350–52. https://doi.org/10.1093/bioinformatics/btv383.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Use R! New York: Springer-Verlag.

Wink, D. A., K. S. Kasprzak, C. M. Maragos, R. K. Elespuru, M. Misra, T. M. Dunams, T. A. Cebula, et al. 1991. “DNA Deaminating Ability and Genotoxicity of Nitric Oxide and Its Progenitors.” *Science* 254 (5034): 1001–3. https://doi.org/10.1126/science.1948068.

Yousef, Foad, Robert Shuchman, Michael Sayers, Gary Fahnenstiel, and Azad Henareh. 2017. “Water Clarity of the Upper Great Lakes: Tracking Changes between 1998–2012.” *Journal of Great Lakes Research* 43 (2): 239–47. https://doi.org/10.1016/j.jglr.2016.12.002.

Zadow, Andrea von, Elisabeth Ignatz, Richard Pokorny, Lars-Oliver Essen, and Gabriele Klug. 2016. “Rhodobacter Sphaeroides CryB Is a Bacterial Cryptochrome with (6–4) Photolyase Activity.” *The FEBS Journal* 283 (23): 4291–4309. https://doi.org/10.1111/febs.13924.

Zhang, Fan, Patrick Scheerer, Inga Oberpichler, Tilman Lamparter, and Norbert Krauß. 2013. “Crystal Structure of a Prokaryotic (6-4) Photolyase with an Fe-S Cluster and a 6,7-Dimethyl-8-Ribityllumazine Antenna Chromophore.” *Proceedings of the National Academy of Sciences* 110 (18): 7217–22. https://doi.org/10.1073/pnas.1302377110.

Zhang, Tong, Colleen M. Hansel, Bettina M. Voelker, and Carl H. Lamborg. 2016. “Extensive Dark Biological Production of Reactive Oxygen Species in Brackish and Freshwater Ponds.” *Environmental Science & Technology* 50 (6): 2983–93. https://doi.org/10.1021/acs.est.5b03906.

Chapter 3

Distinct nitrification regime in Lake Erie compared to other Laurentian Great Lakes

3.1 Abstract

Microbial nitrification is a key process controlling nitrogen availability in aquatic ecosystems. Despite this, few measurements of nitrification rate exist for large freshwater lakes. Here we present direct measurements of nitrification rate from Lakes Superior, Michigan, Erie and Ontario. Nitrification rates in Lake Superior resemble open ocean rates, while rates in Lakes Michigan and Ontario resemble coastal marine systems. High rates of nitrification in Lake Erie resemble smaller eutrophic lakes. Nitrification was associated with the picoplanktonic size fraction across the Great Lakes, except in Lake Erie's central basin, where it was entirely associated with the > 1.6-micron size fraction. Lake Erie harbored genetically and functionally distinct nitrifying organisms from the other lakes, but nitrifying taxa were not distinct between size fractions. We show that urea is readily nitrified in all lakes except Erie, consistent with lack of urease genes in nitrifiers in Lake Erie. Altogether, we show that nitrification and nitrifiers are distinct in Lake Erie as compared to the other Laurentian Great Lakes, reflecting Erie's higher productivity and availability of ammonium.

3.2 Introduction

The Laurentian Great Lakes are five interconnected freshwater ecosystems that span strong biogeochemical gradients in productivity (Vollenweider et al., 1974), phosphorus availability (Dove & Chapra, 2015; Sterner, 2021) and nitrogen availability (Dove & Chapra, 2015; Mukherjee et al., 2016; Rozmarynowycz et al., 2019; Sterner, 2021). In many aquatic systems, nitrogen is actively oxidized and reduced by microbial processes, which control both its biological availability and flux into and out of ecosystems (Kuypers et al., 2018). While differences in nitrogen species concentrations across the Great Lakes are fairly well documented, it is less clear whether these differences are the result of distinct in-lake nitrogen cycling dynamics, or differences in inputs (Robertson & Saad, 2011).

Variation in nitrifying taxa across the Great Lakes suggests a potential for differences in nitrogen cycling. Clear variation in microbial community composition is present across the Great Lakes (Paver et al., 2020; Rozmarynowycz et al., 2019), especially in taxonomic groups associated with the oxidation of reduced nitrogen species (Mukherjee et al., 2016; Paver et al., 2020; Podowski et al., 2021). The composition of communities of nitrogen oxidizing microbes (nitrifiers) may have consequences for the rate of nitrification, in part due to differing affinities for reduced nitrogen (Kits et al., 2017; Martens-Habbena et al., 2009; Nowka et al., 2015) and differences in auxiliary metabolic capabilities (Kitzinger et al., 2019; Palatinszky et al., 2015). Taken together, the diversity of nitrifying communities found across the Great Lakes, along with differences in productivity (Vollenweider et al., 1974) and ammonium concentrations (Mukherjee et al., 2016; Rozmarynowycz et al., 2019) suggests that nitrification rates across the Great Lakes are likely geographically variable.

Sparse measurements of nitrification in the Great Lakes limit the current understanding of how nitrogen cycling differs across lakes. Nitrification rates have only been measured directly in Lake Superior (Small et al., 2013) and Lake Erie (Hoffman, 2020), though proxies such as oxygen consumption (Clevinger et al., 2014) and carbon fixation (Lean & Knowles, 1987) have also been used in Erie and Ontario. These estimates suggest large differences in rates between Superior, Erie, and Ontario, with Erie and Ontario two and four times higher than Superior, respectively. However, lack of methodological and sampling consistency makes comparisons difficult across existing measurements. Published rate measurements are lacking for Lakes Michigan and Huron, which have unique bathymetry (Cuhel & Aguilar, 2013) and natural history (Barbiero et al., 2012).

Organic nitrogen in the form of urea may represent an important but poorly studied substrate for nitrification. While recently acknowledged as an important substrate for marine nitrifiers (Kitzinger et al., 2019; Tolar et al., 2017), the contribution of organic nitrogen sources to freshwater nitrification has not been widely considered. Measurements of urea concentrations in the Great Lakes are sparse, but limited data show that Superior (Podowski et al., 2021) and Erie (Belisle et al., 2016; Podowski et al., 2021) have concentrations more than an order of magnitude higher than in the Gulf of Mexico (Kitzinger et al., 2019) and other marine coastal environments (Tolar et al., 2017). Thus, the rate of urea-derived nitrification should be much higher in Lake Erie and Superior than in the Gulf of Mexico or coastal environments (Kitzinger et al., 2019; Tolar et al., 2017), considering positive relationships between urea concentration and rates of urea-derived nitrification (Tolar et al., 2017). Surprisingly, however, urease genes were not recovered in genome reconstructions of ammonia oxidizing bacteria from Lake Erie (Podowski et al., 2021). By contrast, in the other Great Lakes where urea concentrations are poorly constrained, urease was detected in reconstructed genomes of ammonia oxidizing archaea and some ammonia oxidizing bacteria, suggesting an important and complex role for urea in nitrogen cycling across the Great Lakes (Podowski et al., 2021).

In characterizing the taxonomic and genomic diversity of picoplanktonic nitrifiers (0.2-1.6 µm) across the Great Lakes, we uncovered associations between nitrifier diversity and biogeochemical factors. Ammonia oxidizing archaea were found in deep unproductive basins, along with *Nitrospira* and some lineages of *Nitrosospira* specific to these basins. Other distinct lineages of *Nitrosospira* were found in productive and shallow basins along with *Ca*. Nitrotoga (Podowski et al., 2021). But this genomic survey of Great Lakes nitrifiers left several unanswered questions. First, relative and absolute abundances of nitrifiers are highest in Lakes Erie and Ontario, reaching up to 20% of the total picoplankton microbial community, compared to <10% in other lakes (Podowski et al., 2021); microbial cell concentrations are also two to fivefold higher in Erie and Ontario (Paver et al., 2020). The causes of this difference in nitrifier abundance, as well as the consequences, are unclear. Second, the southern basin of Lake Huron and the western and central basins of Lake Erie appear to lack picoplanktonic nitrifiers, a surprising result considering high measured rates of nitrification in Lake Erie (Clevinger et al., 2014; Hoffman, 2020). We investigated the possibility that nitrifiers are particle-associated rather than picoplanktonic in these basins. Finally, metabolic reconstructions of ammonia oxidizers imply that urea plays a variable role as a nitrification substrate across lakes, and is negligibly important in Lake Erie, a result which is perhaps surprising considering the availability of urea in Lake Erie (Belisle et al., 2016). Here we address these questions by coupling genomic data with rate measurements across size fractions and with both ammonium and urea as substrates.

3.3 Methods

3.3.1 Nitrification rate measurements from $^{15}NH_4^+$ and ^{15}N Urea

Nitrification rates were measured during three cruises in the summer and fall of 2019. Measurements from Lake Michigan (stations OR1 and MI14M) and Lake Superior (station SU08M) were taken in June aboard the R/V *Blue Heron* (Fig 3.1). Additional nitrification measurements from Lake Michigan, Erie, Ontario and Superior were taken aboard the R/V *Lake Guardian* in August and September 2019. $^{15}NH_4^+$ or ^{15}N-urea nitrification rates were determined following previously published protocols (Carini & Joye, 2008; Small et al., 2013; Tolar et al., 2017), using incubations to track accumulation of $^{15}NO_3^-$. Water was collected approximately 10 meters from the bottom of each site in Niskin bottles. Assays were carried out in 250 ml square polycarbonate bottles (Nalgene, ThermoFisher Scientific MA, USA), filled until overflowing (approximately 330 ml). For all incubations in Superior, Michigan and Ontario, substrates were added at concentrations of 0.03 μM ^{15}N-urea or $^{15}NH_4^+$. In Erie, final concentrations of 0.3 μM ^{15}N-urea or $^{15}NH_4^+$ were used, except in September 2019 assays, where ^{15}N-urea was added at a concentration of 0.03 μM. Concentrations of added $^{15}NH_4^+$ or ^{15}N-urea were chosen to be <=10% of ambient concentrations in order to not stimulate nitrification activity. This was achieved for $^{15}NH_4^+$ incubations, but ^{15}N-urea additions frequently exceeded 10% of ambient (Table 3.1). For 1.6 μm filtered water treatments, water was gravity filtered through a Whatman GF/A filter prior to being added to 250 ml assay bottles. Bottles were then stored in the dark at 4 °C to mimic *in situ* conditions. All assays were performed in triplicate. After 24 hours, water was filtered through a 0.2 μm syringe filter into HDPE bottles (Wheaton, Millville, New Jersey, USA #209546) and then stored at -20 °C until downstream analysis.

Atom % ^{15}N in NO_3^- was determined using the nitrifier denitrification method (Sigman et al., 2001) at the University of California, Davis Stable Isotope Facility. Nitrification rate was calculated using the following equation (Small et al., 2013):

$$((^{15}NO_3^{-}\text{excess}) \times [NO_3^-])/((^{15}NH_4^+)/[NH_4^+] \times \text{incubation time})$$

where $^{15}NO_3^-$ excess is the difference between ambient $^{15}NO_3^-$ atom % and the $^{15}NO_3^-$ atom % after incubation. $[NO_3^-]$ is the final concentration of NO_3^- as determined by Griess reagent assay. $(^{15}NH_4^+)$ was the concentration of $^{15}NH_4^+$ in the incubation bottle after addition (either 0.3 μM or 0.03 μM) and $[NH_4^+]$ was the sum of ambient NH_4^+ in the bottle plus the added $^{15}NH_4^+$. For ^{15}N Urea measurements, the same formula was used, except the $(^{15}NH_4^+)$ and $[NH_4^+]$ were replaced with (^{15}N-urea) and [urea] respectively, and the denominator was multiplied by 2 to account for the two nitrogen atoms in urea (Tolar et al., 2017). If ambient concentration of urea was below the limit of detection, the ambient urea term in [urea] was set to zero (Tolar et al., 2017). While a spike of urea which was much greater than ambient concentrations may potentially stimulate use of urea beyond basal rates, the relationship between ^{15}N-urea–derived nitrification rate and the ratio of [^{15}N-urea] to [urea] was both negative and non-significant (Pearson's R : -0.45, $p > .15$).

3.3.2 Ambient nutrient concentrations

Collection and processing of nutrient samples was identical across all cruises. Water was collected from the same Niskin bottle as was used for nitrification rate incubations. Water was filtered through a 0.2 μm syringe filter into acid washed Wheaton HDPE bottles (209546) and then stored at -20 °C. For concentrations of NOx, NO_3^- and NO_2^-, a modified version of the Griess reagent assay was used, optimized for a 96 well plate reader (Miranda et al., 2001). NH_4^+ concentrations were determined using the OPA method, measured in small volumes in a 96 well plate (Holmes et al., 1999). Urea concentrations were determined using the room temperature DAM method, measured in 24 well plates (Chen et al., 2015; Revilla et al., 2005).

3.3.3 Cell Abundance Quantification

Samples for cell abundance quantification were taken from raw unfiltered water samples across all cruises, where 1ml of raw unfiltered water was collected and 5 µL of 25% glutaraldehyde was added (final glutaraldehyde concentration 0.125%). Samples were incubated in the dark for 10 minutes then flash frozen in liquid nitrogen and stored at -80 °C until analysis. Cell abundance was quantified using flow cytometry on a CytoFLEX S (Beckman Coulter CA, USA). Samples were thawed and 100 microliters of sample was stained with 0.5 uL of 100x SYBR Green to achieve a final concentration of 0.5x SYBR GREEN I (Thermofisher). Samples were incubated in the dark for 30 minutes and enumerated using the CytoFLEX S flow cytometer.

3.3.4 Metagenome and Single Cell Genome Sequencing

Metagenomic samples from Lake Erie were collected in September 2019 aboard the R/V *Lake Guardian* concurrent with rate measurements. For 0.2-1.6 micron samples, approximately 4L of water was passed through a 25 micron mesh and then pre-filtered through a GF/A glass fiber filter (Whatman 1820-047; nominal pore size 1.6 µm) before being collected on a 0.22 Millipore Express Plus 47mm membrane filter (GPWP04700,ThermoFisher Scientific MA, USA). For 25-0.2 micron samples, approximately 4L of water was passed through a 25 micron mesh and then collected on a 0.22 µm Millipore Express Plus 47mm membrane filter (GPWP04700, ThermoFisher Scientific MA, USA). All samples were stored at -80 °C until DNA extraction using FastDNA Spin Kit for Soil (116560-200, MP Biomedical, Irvine, CA, USA), then further purified using Genomic DNA Clean and Concentrator (D4010, Zymo Research, Irvine, CA, USA).

Sequencing and library preparation were performed at MiGS Microbial Genome Sequencing Center (Pittsburgh, PA, USA), using the Illumina Library Prep and NextSeq 2000 platform (Illumina). DNA sequences were quality controlled using bbduk qtrim=r trimq=30 and then normalized using bbnorm target=100 mindepth=2 in BBTools suite version 35.74 (https://sourceforge.net/projects/bbmap/). Assembly was carried out in metaSPAdes 3.15.2 (Nurk et al., 2017). metaWRAP using the binning protocol (Uritskiy et al., 2018) facilitated mapping of quality controlled, non-normalized reads using BWA (Li & Durbin, 2009), and binning using MetaBAT2 (Kang et al., 2019, p. 2), CONCOCT (Alneberg et al., 2014) and MaxBin2 (Wu et al., 2016, p. 2). metaWRAP Bin_refinement protocol compared and consolidated binning output using CheckM (Parks et al., 2015)**.** metaWRAP Reassemble_bins was then used to map reads back to high quality bins using BWA (Li & Durbin, 2009) and then reassembled using metaSPAdes 3.15.2 (Nurk et al., 2017)**.** Final bins with greater than 75% completion and less than 5% contamination as assigned by CheckM (Parks et al., 2015) were deemed metagenome assembled genomes (MAGs).

Metagenomic samples from the bottom of the water column in Lakes Superior, Michigan Huron and Ontario were collected in summer of 2012 aboard the R/V *Lake Guardian*, at the same time and in the same manner as described in (Podowski et al., 2021). Sample collection was conducted identically, except that library preparation was conducted using Illumina Nextera XT and sequenced at Harvard Bauer Core Facility. Quality control of DNA sequences was carried out using bbduk qtrim=rl trimq=13 minlength=40 in BBTools suite version 35.74

(https://sourceforge.net/projects/bbmap/). The assembly, mapping, binning and generation of MAGs was carried out as described in (Podowski et al., 2021).

Samples for single cell amplified genomes (SAG) were collected in August 2014 and processed as described in (Podowski et al., 2021). SAGs presented here were collected from Lakes Michigan, Superior, Erie and Ontario. SAG assemblies were downloaded from JGI IMG after JGI carried out normalization using bbnorm target=100 mindepth=2, reformat.sh samplereadstarget=20000000 using BBTools suite version 38.39 (https://sourceforge.net/projects/bbmap/) and assembly using SPAdes (Bankevich et al., 2012) -phred-offset 33 -t 16 -m 64 --sc -k 25,55,95.

MAGs and SAGs were all annotated and analyzed identically once generated. CheckM (Parks et al., 2015) was used to assess completion and contamination of all SAGs and MAGs. All MAGs generated were greater than 75% complete and less than 5% contaminated, but SAGs were of lower completion and ranged between 20 and 40% complete. GTDB-tk (Chaumeil et al., 2020) was used to assign taxonomy to MAGs and fastANI (Jain et al., 2018) was used to assign MAGs to nitrifier genome groups. Anvi'o v7 (Eren et al., 2015) was used to call genes using prodigal (Hyatt et al., 2010) and to annotate gene calls against COG (Tatusov et al., 2000), PFAM (Finn et al., 2014) and KEGG (Kanehisa et al., 2016) databases. Anvi'o v7 pangenome pipeline (Delmont & Eren, 2018) was used to generate gene clusters using blastp (Camacho et al., 2009) to determine sequence similarity, ITEP (Benedict et al., 2014) to eliminate weak similarity and MCL (van Dongen & Abreu-Goodger, 2012) to generate clusters. In order to assess abundance of genomes and genome groups in metagenomic samples, bowtie2 (Langmead & Salzberg, 2012, p. 2) was used to map quality controlled, non normalized paired short reads to MAG fasta files.

3.3.5 Metaproteomic Sample Preparation and Analysis

Filter samples for proteomics analysis were extracted by vortexing, heating (95˚C, 20 min) and ultrasonication ((QSonica Q500; 10 min, 1s pulse/1s pause, 85% amplitude) in a reducing and denaturing SDS(1%)/Tris(200mM, pH 8.0)/DTT(10 mM) buffer, and cysteine thiols alkylated with 40mM iodoacetamide. Proteins were precipitated overnight in acetone (-20˚C), redissolved in 8M urea, and purified by a modified eFASP (enhanced filter-aided sample preparation) protocol (Erde, Loo, and Loo 2014), using Sartorius Vivacon 500 concentrators (30kDa nominal cutoff). Proteins were digested with MS-grade trypsin (37˚C, overnight), and peptides were eluted from the concentrator dried by vacuum centrifugation. For quantitative analysis, peptides were isotopically labeled at both N- and C-termini using the diDO-IPTL methodology (Waldbauer et al. 2017). Briefly, C-termini were labeled with either oxygen-16 or -18 by enzymatic exchange in isotopic water of >98 atom% enrichment. N-termini were labeled with either un- or dideuterated formaldehyde via reductive alkylation using sodium cyanoborohydride. Peptide extracts from each sample were split and aliquots labeled separately with CD2O/16O and CH2O/18O; the latter were pooled to serve as a common internal standard for quantification. Aliquots of the 16O-labeled peptides and 18O-labeled internal standard were mixed 1:1 v/v and analyzed by LC-MS for protein expression quantification.

For LC-MS analysis, peptide samples were separated on a monolithic capillary C18 column (GL Sciences Monocap Ultra, 100µm I.D. × 200cm length) using a water-acetonitrile + 0.1% formic acid gradient (2-50% AcN over 180 min) at 360nl/min using a Dionex Ultimate 3000 LC system with nanoelectrospray ionization (Proxeon Nanospray Flex source). Mass spectra were collected on an Orbitrap Elite mass spectrometer (Thermo) operating in a data-dependent acquisition (DDA) mode, with one high-resolution (120,000 m/Δm) MS1 parent ion full scan triggering Rapid-mode 15 MS2 CID fragment ion scans of selected precursors.

Proteomic mass spectral data were analyzed using MorpheusFromAnotherPlace (MFAP; (Waldbauer et al. 2017). Precursor and product ion mass tolerances for MFAP searches were set to 20ppm and 0.6Da, respectively. Static cysteine carbamidomethylation and variable methionine oxidation, N-terminal (d4)-dimethylation, and C-terminal 18O2 were included as modifications. False discovery rate for peptide-spectrum matches was controlled by target-decoy searching. Protein-level relative abundances and standard errors were calculated in R using the Arm postprocessing scripts for diDO-IPTL data (Waldbauer et al. 2017); github.com/waldbauerlab).

3.4 Results

3.4.1 Bulk nitrification from ammonium and urea

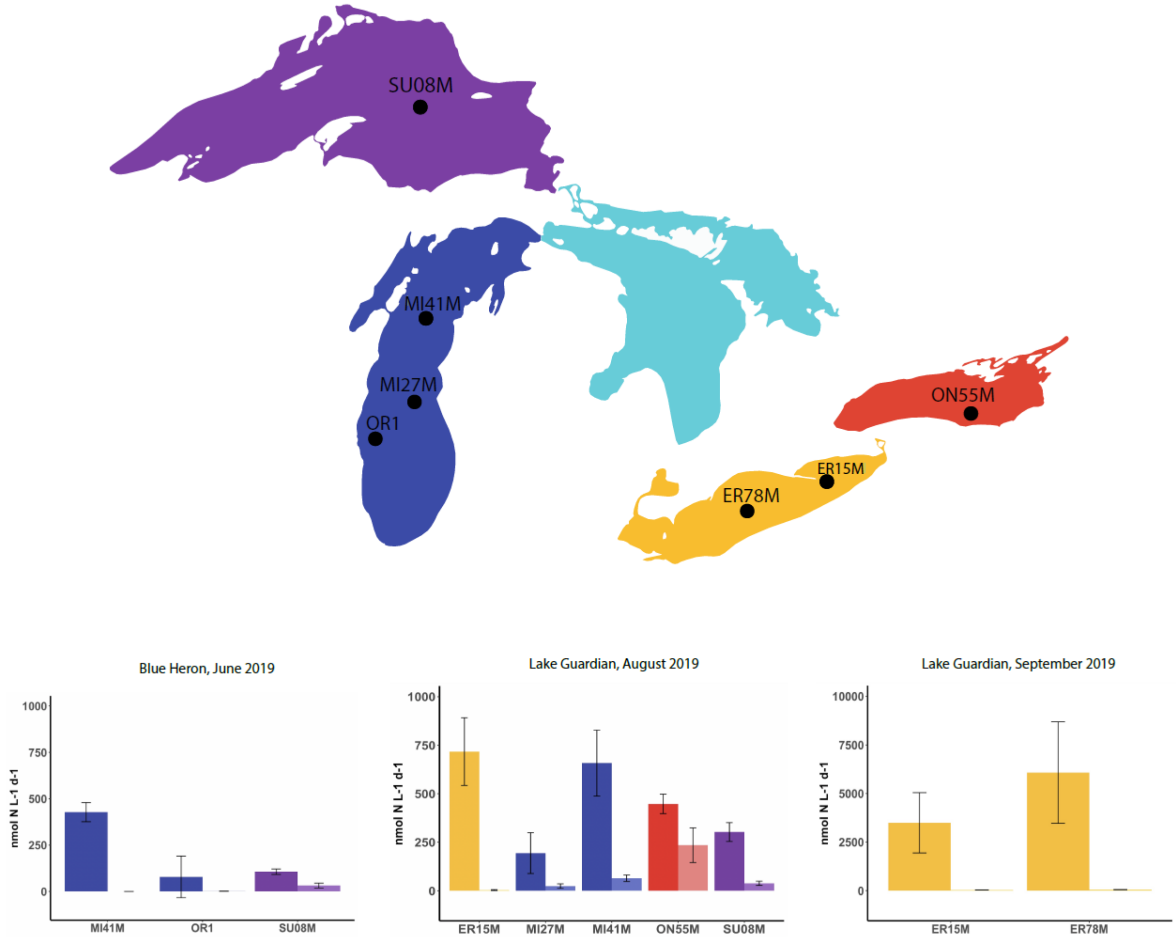
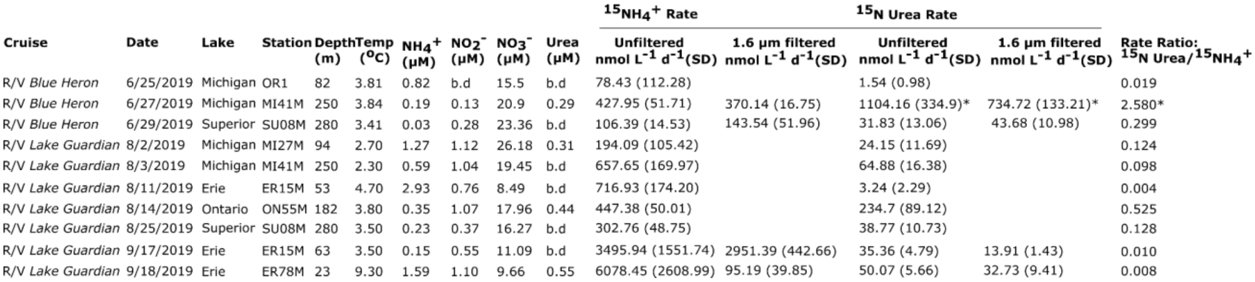


Figure 3.1: Map of sampling stations across the Laurentian Great Lakes. Rates of nitrification for each station are presented in the accompanying bar plots. For each station, the rate of nitrification from ammonia is presented on the left, and the rate of nitrification from urea is presented on the right. Error bars show one standard deviation. Scales on the bar plots are to 1000 nM nitrogen per liter per day, except Lake Guardian September 2019 where scales go to 10000 nM nitrogen per liter per day.

Mean measured rates of nitrification spanned from 78 nmol L^{-1} d^{-1} to 6078 nmol L^{-1} d^{-1} (Table 3.1) with the highest rates from Lake Erie (Fig 3.1). Rates of nitrification in Lake Erie were higher than Lake Superior (Mann Whitney U; W = 54, p = 0.0004), Michigan (W = 106, p

= 2.722e-05) and Ontario (W = 27, p = 0.009). Rates in Lake Ontario were higher than Lake Superior (W = 18, p = 0.02381). Significant differences in nitrification rates between basins were found between stations OR1 and MI27M (both central basin) and MI41M (northern basin) in Lake Michigan (Fig 3.1) (W = 36, p = 0.002). Rates across Lake Erie differed seasonally, with September being 2 to 3 times higher than August (W = 0, p = 0.024). We found a positive relationship between nitrification rate and cell concentration (Fig 3.2) (Spearman's rho = 0.74, S = 22, p = 0.045). Otherwise, no significant correlations between measured nitrification rates and measured environmental parameters (NH_4^+, NO_2^- , NO_3^-, temperature, depth) were found.



										$^{15}NH_4^+$ Rate		^{15}N Urea Rate		
Cruise	**Date**	**Lake**	**Station**	**Depth (m)**	**Temp (°C)**	**NH_4^+ (μM)**	**NO_2^- (μM)**	**NO_3^- (μM)**	**Urea (μM)**	**Unfiltered nmol L^{-1} d^{-1}(SD)**	**1.6 μm filtered nmol L^{-1} d^{-1}(SD)**	**Unfiltered nmol L^{-1} d^{-1}(SD)**	**1.6 μm filtered nmol L^{-1} d^{-1}(SD)**	**Rate Ratio: ^{15}N Urea/$^{15}NH_4^+$**
R/V *Blue Heron*	6/25/2019	Michigan	OR1	82	3.81	0.82	b.d	15.5	b.d	78.43 (112.28)		1.54 (0.98)		0.019
R/V *Blue Heron*	6/27/2019	Michigan	MI41M	250	3.84	0.19	0.13	20.9	0.29	427.95 (51.71)	370.14 (16.75)	1104.16 (334.9)*	734.72 (133.21)*	2.580*
R/V *Blue Heron*	6/29/2019	Superior	SU08M	280	3.41	0.03	0.28	23.36	b.d	106.39 (14.53)	143.54 (51.96)	31.83 (13.06)	43.68 (10.98)	0.299
R/V *Lake Guardian*	8/2/2019	Michigan	MI27M	94	2.70	1.27	1.12	26.18	0.31	194.09 (105.42)		24.15 (11.69)		0.124
R/V *Lake Guardian*	8/3/2019	Michigan	MI41M	250	2.30	0.59	1.04	19.45	b.d	657.65 (169.97)		64.88 (16.38)		0.098
R/V *Lake Guardian*	8/11/2019	Erie	ER15M	53	4.70	2.93	0.76	8.49	b.d	716.93 (174.20)		3.24 (2.29)		0.004
R/V *Lake Guardian*	8/14/2019	Ontario	ON55M	182	3.80	0.35	1.07	17.96	0.44	447.38 (50.01)		234.7 (89.12)		0.525
R/V *Lake Guardian*	8/25/2019	Superior	SU08M	280	3.50	0.23	0.37	16.27	b.d	302.76 (48.75)		38.77 (10.73)		0.128
R/V *Lake Guardian*	9/17/2019	Erie	ER15M	63	3.50	0.15	0.55	11.09	b.d	3495.94 (1551.74)	2951.39 (442.66)	35.36 (4.79)	13.91 (1.43)	0.010
R/V *Lake Guardian*	9/18/2019	Erie	ER78M	23	9.30	1.59	1.10	9.66	0.55	6078.45 (2608.99)	95.19 (39.85)	50.07 (5.66)	32.73 (9.41)	0.008

Table 3.1 Table of physical and chemical data measured alongside nitrification rates. Chemical measurements (NH4+, NO2-, NO3-, urea) were taken from the same water that was used in the incubations to determine rate. Temperature was measured from the CTD at time of water collection. b.d= below detection. Standard deviation (SD) was calculated using three replicates for each rate measured.

Mean measured rates of nitrification from urea ranged from 1.54 nmol L^{-1} d^{-1} to 234.7 nmol L^{-1} d^{-1} (Table 3.1), with the exception of a single high value which was excluded due to likely overestimation. Rates of nitrification derived from urea were higher in Lake Ontario compared to Lake Erie (Mann Whitney U; W = 0, p = 0.009), Lake Michigan (W = 0, p = 0.009) and Lake Superior (W = 0, p = 0.024) (Fig 3.1). Unlike NH_4^+-derived nitrification, urea-derived nitrification showed no seasonal changes in rate. We found no significant correlations between urea-derived nitrification rates and measured environmental parameters (NH_4^+ , NO_2^- , NO_3^- , urea, temperature, depth, microbial cell concentration).

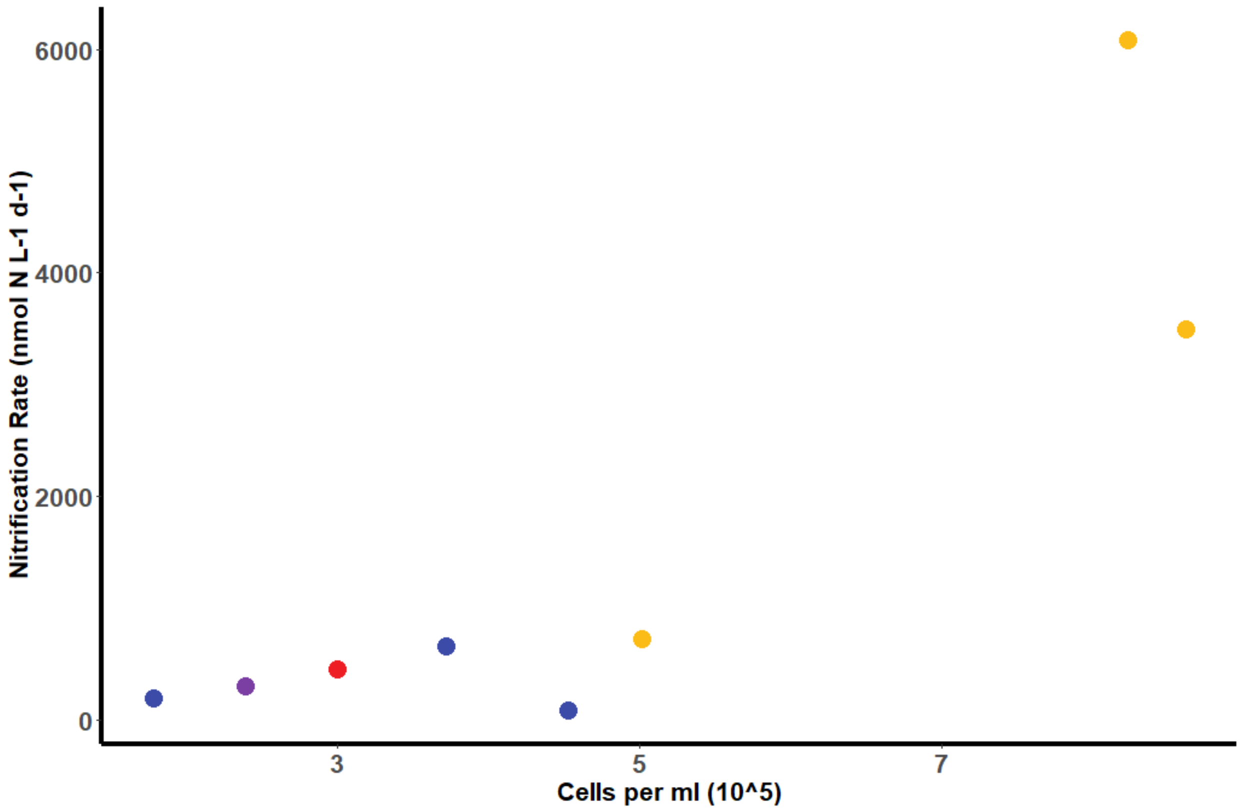


Figure 3.2 Nitrification rates from ammonia compared to cell abundance measured in cells per ml. Nitrification rates are an average of three bottles.

Lake	Basin	Depth (m)	Rate (nmol N L^{-1} d^{-1})	Method	Reference
Superior	Western	150	12-50	^{15}N Tracer	Small et al. 2013
Superior	Central	280	106-303	^{15}N Tracer	This study
Michigan	Central	80	78-194	^{15}N Tracer	This study
Michigan	Northern	250	427-657	^{15}N Tracer	This study
Erie	Eastern	63	350-1,509	Nitrapyrin-sensitive O_2 consumption	Clevinger et al. 2014
Erie	Eastern	63	717- 3495	^{15}N Tracer	This study
Erie	Central	23	218-4,526	Nitrapyrin-sensitive O_2 consumption	Clevinger et al. 2014
Erie	Central	23	6078	^{15}N Tracer	This study
Ontario	Eastern	182	447	^{15}N Tracer	This study
Ontario	Eastern	176	71	Nitrapyrin-sensitive dark 14C bicarbonate incorporation	Lean and Knowles 1987

Table 3.2 Table of previous measurements of nitrification in the Laurentian Great Lakes, along with depth the sample was taken from, the basin and lake the sample was taken from, the method used for calculating the rate, and the reference for the rate measurement.

3.4.2 Nitrification by picoplankton fraction

Mean measured rates of nitrification from NH_4^+ by the picoplanktonic fraction (<1.6 μm) ranged from 95 nmol L^{-1} d^{-1} to 2951 nmol L^{-1} d^{-1} (Table 3.1). These rates represented 86% of bulk (total unfiltered) nitrification rates in the northern basin of Lake Michigan and 84% in the eastern basin of Lake Erie, but only 2% of the bulk nitrification rates in the central basin of Lake Erie (Fig 3.3). Urea-derived nitrification by free-living microbes displayed different trends, making up 66% of bulk rates in the northern basin of Michigan, 39% of bulk rates in the eastern basin of Erie, and 65% of bulk rates in the central basin of Erie (Fig 3.3).

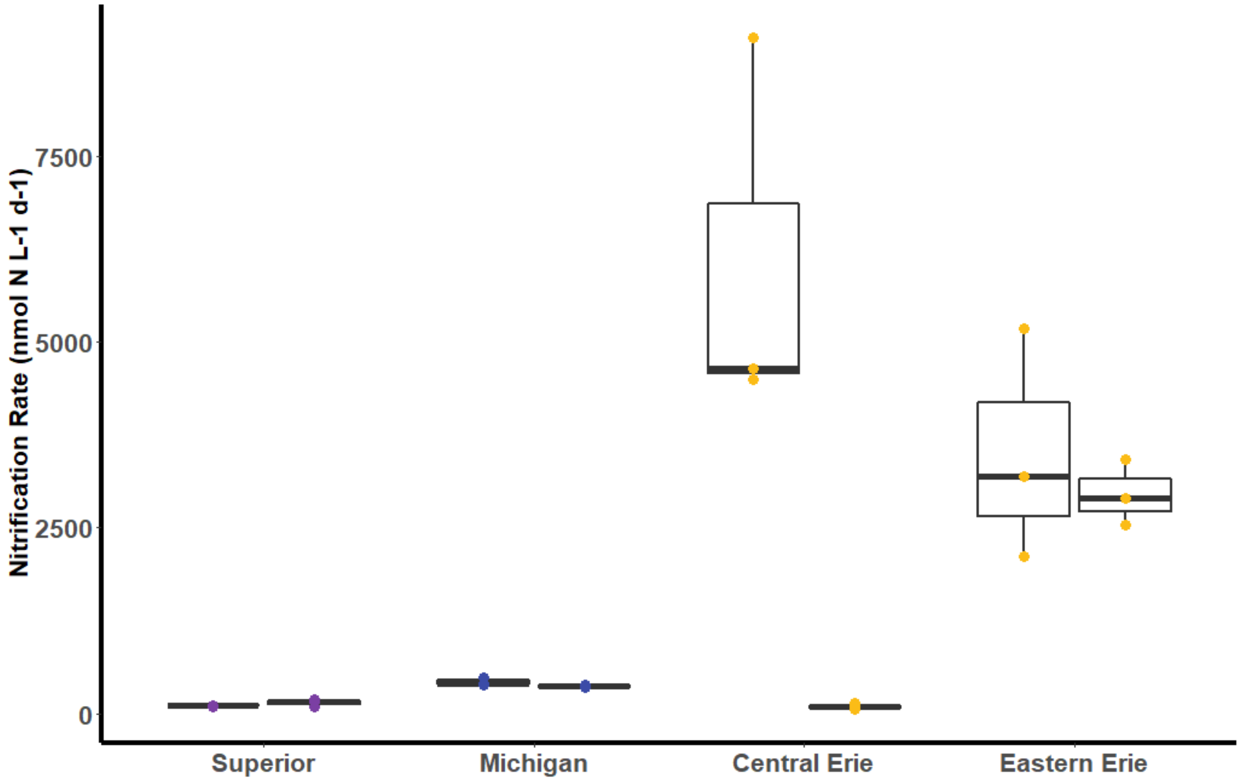


Figure 3.3: Nitrification rates derived from raw water incubations on left, and 1.6 μm-filtered water incubations on right. Each point represents a single bottle.

3.4.3 Novel nitrifiers in genera Nitrosospira and *Ca*. Nitrotoga

A novel *Nitrosospira* MAG from Lake Erie's western basin had metabolic adaptations unique among NspGL groups. NspGL4, which was generated from a 1.6-25 micron size fraction sample, shares 97% average nucleotide identity with NspGL3, the previously identified dominant ammonia oxidizing bacteria in samples from Lake Erie's western basin (Podowski et al., 2021). Like NspGL3, NspGL4 is predicted to have a genome size of approximately 1.6 Mb with 50.3% GC content. NspGL4 contains cytochrome P460, and lacks NirK and urease genes, and appears to contain a type IAq CBB carbon fixation pathway (Badger & Bek, 2008; Rae et al., 2013). Most notable in NspGL4 are genes encoding flagella, chemotaxis, and two component systems which are not found in other Great Lakes ammonia oxidizing bacteria. Similarly, NspGL4 also contains glycogen phosphorylase (*glgP*), glucose-1-phosphate adenylyltransferase (*glgC*), glycogen synthase (*glgA*) and 1,4-alpha-glucan branching enzyme (*glgB*), all of which are missing from other Great Lakes *Nitrosospira* found outside Lake Erie. The presence of glycogen storage and degradation genes suggests that fluctuation in energy availability may select for this energy storage mechanism.

We discovered an additional novel lineage of *Nitrosospira* in Lake Erie using single cell amplified genomes (SAGs). We identified two low completion SAGs (approximately 40% completion) that appear, based upon nucleotide identity of shared regions, to be from the same new group we termed NspGL5, both isolated from the same sample from the western basin Lake Erie. Each of these NspGL5 SAGs share 97% ANI with NspGL4 and 95% ANI with NspGL3, and have 50% G+C content. Like NspGL4, these two NspGL5 SAGs contain many flagella, chemotaxis, and two component system genes otherwise absent from Great Lakes *Nitrosospira* (Podowski et al., 2021), as well as *glgB* and *glgC* for energy storage.

We also reconstructed a novel lineage of *Ca*. Nitrotoga from Lake Michigan, phylogenetically distinct from recently described NtogaGL1a and NtogaGL1b and reference *Ca*. Nitrotoga (Podowski et al., 2021). Generated from a sample from the bottom of Lake Michigan's central basin, this new MAG shares less than a 77% ANI to any NtogaGL genomes. We assign this distinct *Ca.* Nitrotoga MAG as NtogaGL2. *Ca*. Nitrotoga AM1P (Ishii et al., 2020) is the most closely related existing genome to NtogaGL2, and these two genomes share 89% ANI. The predicted genome size for NtogaGL2 is approximately 2 Mb, larger than NtogaGL1a or NtogaGL1b, but still smaller than reference *Ca*. Nitrotoga genomes.

NtogaGL2 is metabolically more similar to reference *Ca*. Nitrotoga than to NtogaGL1a and NtogaGL1b. As in previously denoted publications (Boddicker & Mosier, 2018; Ishii et al., 2020; Kitzinger et al., 2018), NtogaGL2 lacks cyanase (*cynS*) and contains nitrite reductase (*nirK*). NtogaGL2 also contains glycogen storage and breakdown genes *glgABCP* as well as phosphofructokinase (*pfk*) and portions of the Entner-Doudoroff pathway (*zwf, pgl*) present in reference *Ca.* Nitrotoga but absent in NtogaGL1a and NtogaGL1b. In addition, NtogaGL2 contains a number of chemotaxis and flagella genes absent in NtogaGL1a and NtogaGL1b. Together this gene complement suggests that NtogaGL2 is adapted to a more dynamic heterogeneous environment than its relatives NtogaGL1a and NtogaGL1b.

The RuBisCo operon in NtogaGL2 is distinct from reference *Ca.* Nitrotoga and NtogaGL1a and NtogaGL1b. We detected carboxysome genes near *rbcL* and *rbcS* which suggests that NtogaGL2 contains an alpha-carboxysome *cso* operon, more similar to *Nitrosospira* NspGL1 or *Nitrosomonas eutropha* C91 (Badger & Bek, 2008; Rae et al., 2013) than to any known *Ca*. Nitrotoga. Though the region is fragmented, the ordering of *rbcL*, *rbcS, csoS2 and csoS3/csoSCA* suggests a type IAc operon. A second contig containing several other typical alpha-carboxysome type *cso* genes provides further support for this assignment. Interestingly, *csoS3/csoSCA* was absent in the NspGL1 assembly, the one deviation from canonical type IAc organization.

3.4.4 Phylogenetically similar nitrifiers across size fractions in Lake Erie

Distinct size fractions did not contain distinct taxonomic groups of nitrifiers in Lake Erie's central or eastern basin. Quantification of nitrifier genome groups described previously (Podowski et al., 2021) as well as those newly identified (NtogaGL2, NspGL4, NspGL5) demonstrated that while active nitrification may be nearly exclusive to the > 1.6 micron fraction in central Lake Erie (Fig 3.1), the genomic identity of the nitrifiers in the > 1.6 micron size fraction is identical to those elsewhere in Lake Erie in the picoplanktonic size fraction (Fig 3.4). Nitrite oxidizers assembled from Lake Erie's central basin samples that contain both particle associated and free living microbes (25 - 0.2 micron samples) were clearly affiliated with NtogaGL1b, and this is reflected in the fact that NtogaGL1b is highly abundant in the 25-0.2 micron size fraction samples, as well as all other samples from Lake Erie (Fig 3.4). Newly discovered NtogaGL2 was rare, and NtogaGL1a and NtogaGL1b ranged from 10 to 400 times more abundant. NspGL3, NspGL4 and NspGL5 have approximately equal abundances in all Lake Erie samples, both those that represent the 25-0.2 micron size fraction and those representing only the 'free living' < 1.6 micron size fraction described previously (Podowski et al., 2021). These results suggest that rather than two phylogenetically distinct groups of nitrifiers, one 'free living' and one 'particle associated', that the two communities are one and the same.

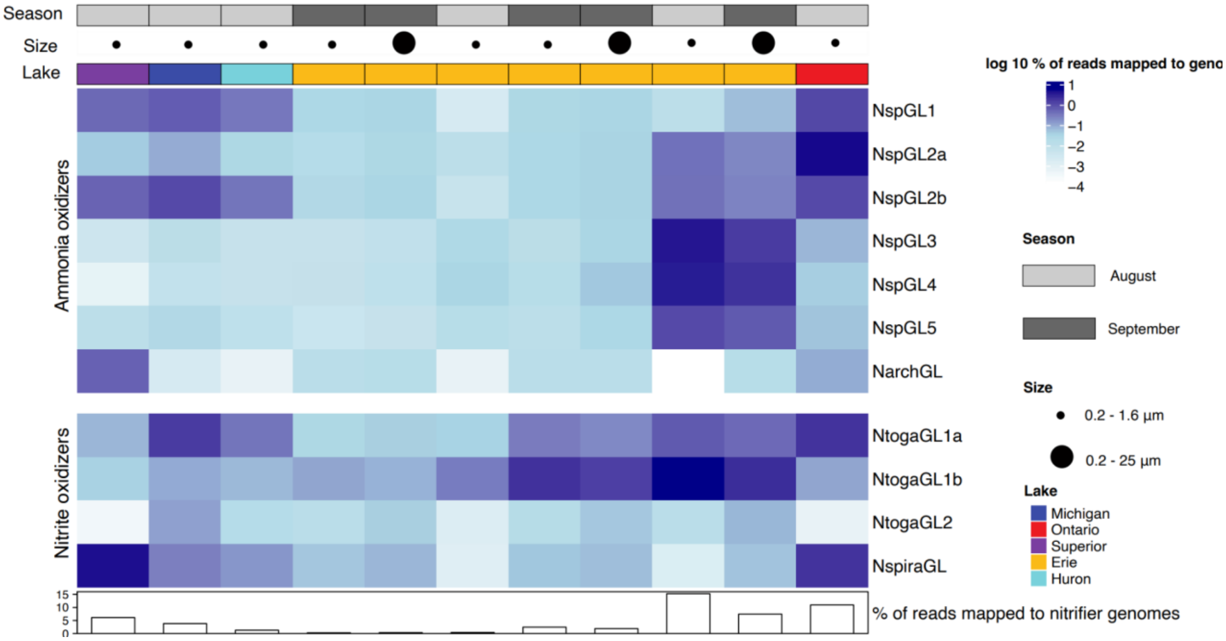


Figure 3.4 Heatmap of the percent of the total microbial community represented by nitrifier genomes, broken down by genome group of the nitrifier. Each column represents a single metagenomic sample, and each row represents a single genome group of nitrifiers. Plotted values are log10 transformed percentages of total reads that are mapped to genomes in that genome group. Barplot below each column plots the percentage of total reads in that sample that were recruited to all nitrifier genomes. Blocks above each sample denote features of that sample.

3.4.5 Detection of nitrification enzymes confirms activity of *Ca*. Nitrotoga and *Nitrosospira*

We used proteomics to search for proteins involved in nitrification and attempted to attribute proteins to specific lineages of nitrifiers. We detected peptides that are diagnostic of AmoA and AmoB, confirming the activity of *Nitrosospira* and suggesting greater activity in Erie's western basin. We found a single peptide belonging to the protein AmoA which matches the sequence of NspGL4, but is distinct from known AmoA sequences in all other Great Lakes *Nitrosospira* and *Nitrosarchaeum*. We also identified four unique peptides belonging to protein AmoB, all of which exclude AmoB from *Nitrosoarchaeum* and two of which match sequences from *Nitrosospira* NspGL2a and NspGL2b (Fig 5). However, *amoB* from NspGL3 and NspGL5 are fragmentary and this region of the gene is not present, and so production by these two groups cannot be ruled out. Altogether, these peptides are likely derived from AmoA and AmoB proteins produced by *Nitrosospira*, but the fragmentary *amoA* and *amoB* gene sequences, as well as peptides arising from conserved regions, limits conclusions about which specific Great Lakes lineage of *Nitrosospira* is producing these proteins.

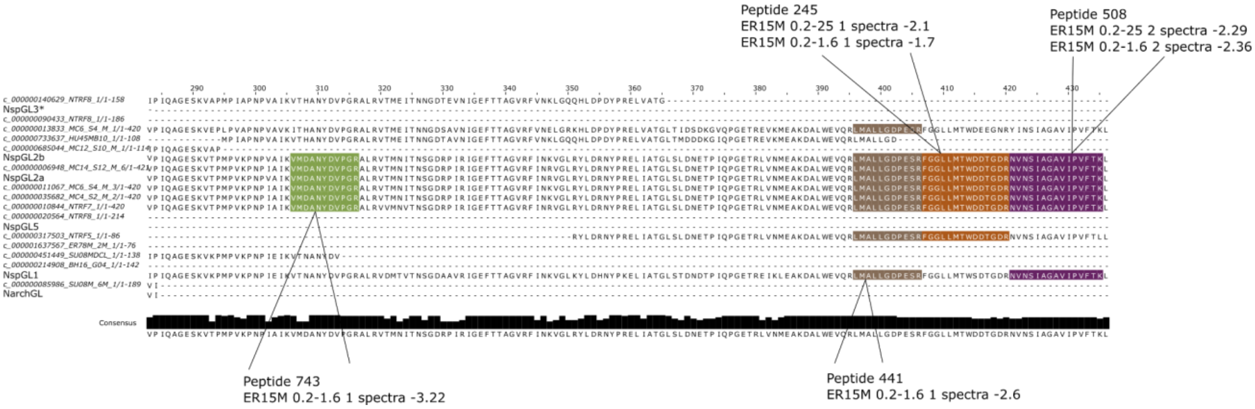


Figure 3.5 Alignment of AmoB sequences from samples in the Great Lakes. AmoB sequences originating from Nitrosospira (NspGL) or Nitrosarchaeum (Narch) groups are denoted, while those not specifically denoted come from unbinned fragments whose phylogenetic affiliation could not be determined. Unique peptides detected by mass spectrometry are highlighted. Unique peptide name is projected from sequence, along with samples in which it was detected (station name, size fraction, number of spectra, normalized log2 abundance value).

Identification of peptides diagnostic to proteins NxrA and NxrB from *Ca*. Nitrotoga suggests specific activity of NtogaGL1b. We detected two unique peptides diagnostic to protein NxrB, which rule out NxrB from NspiraGL but do specifically match a single NtogaGL group. The five unique peptides diagnostic for the NxrA protein provide a greater degree of specificity (Fig 3.6). All five match *Ca.* Nitrotoga (NtogaGL), but not *Nitrospira.* Two of the five cannot discriminate between the three NtogaGL groups. One peptide matches both NtogaGL1a and NtogaGL1b, while one other peptide matches both NtogaGL1b and NtogaGL2. The single remaining peptide is diagnostic to NtogaGL1b. Together, these results are consistent with NxrA and NxrB expression from NtogaGL1b, which is supported by the numerical dominance of NtogaGL1b in these samples at the DNA level.

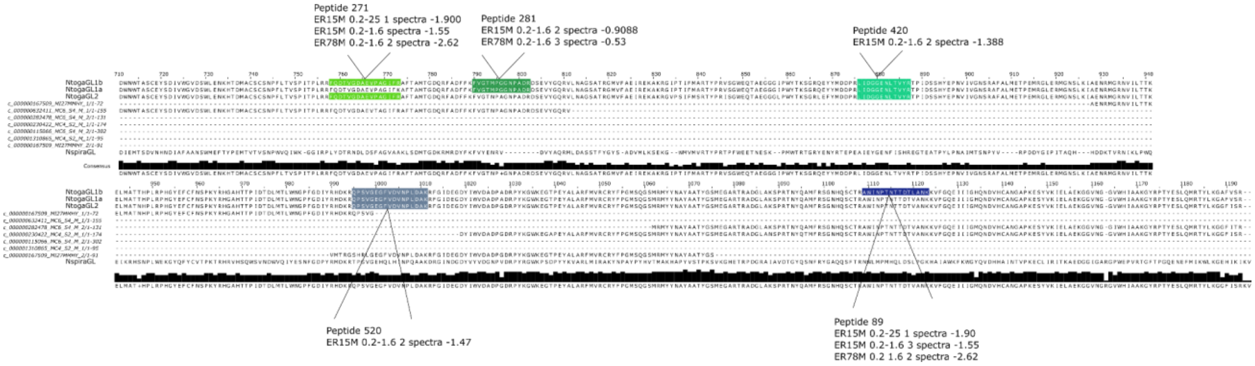


Figure 3.6 Alignment of NxrA sequences from Great Lakes metagenome samples. NxrA sequences originating from *Ca*. Nitrotoga (NtogaGL) or Nitrospira (NspiraGL) groups are denoted, while those not specifically denoted come from unbinned fragments whose phylogenetic affiliation could not be determined. Unique peptides detected are highlighted. Unique peptide name is projected from sequence, along with samples in which it was detected (station name,, size fraction, number of spectra, normalized log2 abundance value).

3.5 Discussion

Rates of nitrification distinguish the Great Lakes from one another, and allow comparisons to other aquatic ecosystems. Our reported nitrification rates from Lake Superior are higher than past measurements (Small et al., 2013) perhaps owing to methodological differences during incubation. Previous measurements used *in-situ* incubations, while our incubations were carried out on deck in light and temperature conditions mimicking *in situ* conditions. Differences between rates of carbon fixation measured on deck vs *in situ* have been previously noted (Callieri et al., 2014). Our reported nitrification rates from Lake Erie are on par with previous direct and indirect measurements (Clevinger et al., 2014; Hoffman, 2020). Nitrification in Lake Erie was distinct from Lake Superior, Michigan and Ontario, and rates of nitrification in Lake Erie more closely resemble eutrophic Lake Taihu (Hampel et al., 2018) and Lake Mendota (Small et al., 2013). Rates of nitrification in Lake Michigan and Ontario more closely resemble coastal marine environments (Tolar et al., 2017), while rates in Superior resemble values from the higher end of the range of open ocean environments (Santoro et al., 2010). Rates of nitrification from urea also resembled those in coastal environments (Tolar et al., 2017), except in Lake Ontario where we reported very high urea-derived nitrification rates.

Nitrification was concentrated in different size fractions across the Great Lakes. In Lakes Superior, Michigan and the eastern basin of Erie, the picoplanktonic size fraction primarily drove nitrification. By contrast, in the central basin of Erie, the picoplanktonic fraction contributed very little to total nitrification (Fig 3.3). There are several potential explanations for this finding. One explanation would be that the large free-living cells (> 1.6 µm) in Erie are the primary nitrifiers there. This seems unlikely, based on published micrographs that show *Nitrosospira* and *Ca*. Nitrotoga to be small cells less than 1 µm in size (Jiang & Bakken, 1999) (Kitzinger et al., 2018). Another explanation could be that nitrifiers in Lake Erie's central basin form aggregates using extracellular polysaccharides (EPS), which is common in wastewater treatment plants (Schramm et al., 1996) and drinking water systems (Lipponen et al., 2004) and ensures close proximity between ammonia- and nitrite-oxidizing partners. Finally, nitrifiers may be particle-associated in Lake Erie's central basin but free living elsewhere. This hypothesis is supported by higher productivity in Lake Erie's central basin (Glooschenko et al., 1974; Vollenweider et al., 1974), as higher productivity has been shown to influence the particle association of other processes such as respiration (Biddanda et al., 2001). Single-cell or single-particle imaging will be required to determine the spatial context of nitrifiers across Lake Erie and elsewhere.

Nitrifier assemblages distinguish the Great Lakes from both marine and other freshwater systems. Nitrifiers from the Great Lakes – particularly ammonia oxidizers in *Nitrosarchaeum* and *Nitrosospira,* and nitrite oxidizers in *Nitrospira* – are strikingly similar at the genomic level to related taxa recovered from other large, deep lakes such as Biwa and Baikal (Cabello-Yeves et al., 2019; Okazaki et al., 2019). However, unlike in smaller French lakes, lower ammonia concentrations and oligotrophic conditions did not predict dominance of ammonia oxidizing archaea (Hugoni et al., 2013). Instead, all of the Great Lakes are dominated by ammonia oxidizing bacteria, while ammonia oxidizing archaea occur only as a minority in oligotrophic Lake Superior and Michigan (Podowski et al., 2021). This also distinguishes the Great Lakes from marine systems, where ammonia oxidizing archaea are largely dominant offshore (Schleper, 2010). Nitrite oxidizers in the Great Lakes fall more in line with traditional freshwater lineages, with *Nitrospira* dominant in Superior and some basins of Michigan and Ontario (Mukherjee et al., 2016; Paver et al., 2020; Podowski et al., 2021; Small et al., 2013). Lake Erie and Huron, as well as some basins of Michigan and Ontario are dominated by *Ca*. Nitrotoga (Fig 3.4, (Podowski et al., 2021)) which are found broadly distributed in freshwater environments (Boddicker & Mosier, 2018) but whose identity as important freshwater nitrite oxidizers was only recognized recently. Like all other freshwater systems, the Great Lakes have distinct nitrite oxidizers from marine systems, which in open ocean environments are dominated by *Nitrospinae* (Daims et al., 2016; Pachiadaki et al., 2017). Together, these nitrifier communities position the Great Lakes as a unique system.

Our work suggests that urea plays a variable role across the Great Lakes, based on genomic capacity to hydrolyze urea and rates of urea-derived nitrification. Urea could potentially be an important source of nitrogen for heterotrophic and autotrophic picoplankton in the Great Lakes, especially as use of urea as an agricultural fertilizer has increased 100-fold (Glibert et al., 2006)**.** The capacity to use urea is enabled by the enzyme urease, which is distributed throughout all three domains of life (Solomon et al., 2010) but not uniformly, and microcosm experiments have demonstrated that urea addition does not affect community members equally and may be used preferentially by eukaryotic phytoplankton (Wheeler & Kirchman, 1986). Mesocosm experiments in nitrogen-limited Lake Erie have demonstrated that urea is broadly bioavailable to phytoplankton and can relieve nutrient limitation of growth (Belisle et al., 2016). While phytoplankton and heterotrophs use urea for nitrogen assimilation, urea could serve as a source of ammonia for oxidation by nitrifiers (Allison & Prosser, 1991). Ammonia oxidizers in Lake Erie appear to lack urease, suggesting nitrifiers cannot use urea as a substrate for nitrification (Podowski et al., 2021). Our measurements of urea-derived nitrification confirm these genomic predictions, suggesting a reduced role for urea in nitrification in Lake Erie. This reduced role may be driven by competition with phytoplankton for urea, which would be more important to phytoplankton growth in this nitrogen-limited system. A similar dynamic of phytoplankton–-nitrifier competition for substrate is thought to exclude nitrification from the photic zone of the ocean (Smith et al., 2014).

While our work advances our understanding of nitrification, many open questions about nitrogen cycling in the Great Lakes remain. The concentration of oxidized nitrogen in Lakes Superior, Michigan and Ontario has been increasing since the 1970s (Finlay et al., 2013) and at least in Lake Superior this nitrate is thought to be largely produced in-lake by nitrifiers (Finlay et al., 2007). Our results do not speak to the cause of this increase, but are consistent with past discussion that nitrate increases are driven by a decrease of nitrate loss through either decreasing incorporation into biomass or decreasing loss due to denitrification (Small et al., 2014), rather than an increase in rates of nitrification and therefore nitrate accumulation. This continual nitrate increase is not occurring in Erie (Dove & Chapra, 2015), likely due to increased nitrogen loss via denitrification in the sediment (Small et al., 2014) and potentially in the water column due to hypolimnetic hypoxia (Scavia et al., 2014). Our results reinforce the uniqueness of Lake Erie's nitrogen cycle, in terms of both rates of nitrification (Fig 3.1) and identity of nitrifiers (Fig 3.4, (Podowski et al., 2021)). This uniqueness of Erie's nitrogen cycle may play a role in the intensity of *Microcystis* blooms and microcystin production, which are influenced by nitrogen availability (Wilhelm et al., 2020) and in particular urea (Belisle et al., 2016). Additional future work is urgently needed to better understand the distinctive patterns of nitrogen cycling across the Great Lakes, their role in drinking water quality, and potential alterations due to warming and environmental change.

3.6 References

Allison, S. M., & Prosser, J. I. (1991). Urease activity in neutrophilic autotrophic ammonia-oxidizing bacteria isolated from acid soils. *Soil Biology and Biochemistry*, *23*(1), 45–51. https://doi.org/10.1016/0038-0717(91)90161-C

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle. *Journal of Experimental Botany*, *59*(7), 1525–1541. https://doi.org/10.1093/jxb/erm297

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barbiero, R. P., Lesht, B. M., & Warren, G. J. (2012). Convergence of trophic state and the lower food web in Lakes Huron, Michigan and Superior. *Journal of Great Lakes Research*, *38*(2), 368–380. https://doi.org/10.1016/j.jglr.2012.03.009

Belisle, B. S., Steffen, M. M., Pound, H. L., Watson, S. B., DeBruyn, J. M., Bourbonniere, R. A., Boyer, G. L., & Wilhelm, S. W. (2016). Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*, *42*(3), 599–607. https://doi.org/10.1016/j.jglr.2016.03.002

Benedict, M. N., Henriksen, J. R., Metcalf, W. W., Whitaker, R. J., & Price, N. D. (2014). ITEP: An integrated toolkit for exploration of microbial pan-genomes. *BMC Genomics*, *15*(1), 8. https://doi.org/10.1186/1471-2164-15-8

Biddanda, B., Ogdahl, M., & Cotner, J. (2001). Dominance of bacterial metabolism in oligotrophic relative to eutrophic waters. *Limnology and Oceanography*, *46*(3), 730–739. https://doi.org/10.4319/lo.2001.46.3.0730

Boddicker, A. M., & Mosier, A. C. (2018). Genomic profiling of four cultivated Candidatus Nitrotoga spp. Predicts broad metabolic potential and environmental distribution. *The ISME Journal*, *12*(12), 2864–2882. https://doi.org/10.1038/s41396-018-0240-8

Cabello-Yeves, P. J., Zemskaya, T. I., Zakharenko, A. S., Sakirko, M. V., Ivanov, V. G., Ghai, R., & Rodriguez-Valera, F. (2019). Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnology and Oceanography*.

Callieri, C., Coci, M., Eckert, E. M., Salcher, M. M., & Bertoni, R. (2014). Archaea and Bacteria in deep lake hypolimnion: In situ dark inorganic carbon uptake. *Journal of Limnology*, *73*(1). https://doi.org/10.4081/jlimnol.2014.937

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Carini, S. A., & Joye, S. B. (2008). Nitrification in Mono Lake, California: Activity and community composition during contrasting hydrological regimes. *Limnology and Oceanography*, *53*(6), 2546–2557. https://doi.org/10.4319/lo.2008.53.6.2546

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chen, L., Ma, J., Huang, Y., Dai, M., & Li, X. (2015). Optimization of a colorimetric method to determine trace urea in seawater. *Limnology and Oceanography: Methods*, *13*(6), 303–311. https://doi.org/10.1002/lom3.10026

Clevinger, C. C., Heath, R. T., & Bade, D. L. (2014). Oxygen use by nitrification in the hypolimnion and sediments of Lake Erie. *Journal of Great Lakes Research*, *40*(1), 202–207. https://doi.org/10.1016/j.jglr.2013.09.015

Cuhel, R. L., & Aguilar, C. (2013). Ecosystem Transformations of the Laurentian Great Lake Michigan by Nonindigenous Biological Invaders. *Annual Review of Marine Science*, *5*(1), 289–320. https://doi.org/10.1146/annurev-marine-120710-100952

Daims, H., Lücker, S., & Wagner, M. (2016). A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends in Microbiology*, *24*(9), 699–712. https://doi.org/10.1016/j.tim.2016.05.004

Delmont, T. O., & Eren, A. M. (2018). Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ*, *6*, e4320. https://doi.org/10.7717/peerj.4320

Dove, A., & Chapra, S. C. (2015). Long-term trends of nutrients and trophic response variables for the Great Lakes. *Limnology and Oceanography*, *60*(2), 696–721. https://doi.org/10.1002/lno.10055

Erde, Jonathan, Rachel R. Ogorzalek Loo, and Joseph A. Loo. 2014. "Enhanced FASP (EFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments." *Journal of Proteome Research* 13 (4): 1885–95. https://doi.org/10.1021/pr4010019.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319

Finlay, J. C., Small, G. E., & Sterner, R. W. (2013). Human Influences on Nitrogen Removal in

Lakes. *Science*, *342*(6155), 247–250. https://doi.org/10.1126/science.1242575

Finlay, J. C., Sterner, R. W., & Kumar, S. (2007). ISOTOPIC EVIDENCE FOR IN-LAKE PRODUCTION OF ACCUMULATING NITRATE IN LAKE SUPERIOR. *Ecological Applications*, *17*(8), 2323–2332. https://doi.org/10.1890/07-0245.1

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., & Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Research*, *42*(Database issue), D222–D230. https://doi.org/10.1093/nar/gkt1223

Glibert, P. M., Harrison, J., Heil, C., & Seitzinger, S. (2006). Escalating Worldwide use of Urea – A Global Change Contributing to Coastal Eutrophication. *Biogeochemistry*, *77*(3), 441–463. https://doi.org/10.1007/s10533-005-3070-5

Glooschenko, W. A., Moore, J. E., Munawar, M., & Vollenweider, R. A. (1974). Primary Production in Lakes Ontario and Erie: A Comparative Study. *Journal of the Fisheries Research Board of Canada*, *31*(3), 253–263. https://doi.org/10.1139/f74-045

Hampel, J. J., McCarthy, M. J., Gardner, W. S., Zhang, L., Xu, H., Zhu, G., & Newell, S. E. (2018). Nitrification and ammonium dynamics in Taihu Lake, China: Seasonal competition for ammonium between nitrifiers and cyanobacteria. *Biogeosciences*, *15*(3), 733–748. https://doi.org/10.5194/bg-15-733-2018

Hoffman, D. K. (2020). *The Fate of Anthropogenic Nitrogen Along Hydrologic Continuums: Patterns of Transformation and Recycling in a Eutrophic Lake and Coastal Marine Sediments* [Wright State University]. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=wright1598886275031918

Holmes, R. M., Aminot, A., Kérouel, R., Hooker, B. A., & Peterson, B. J. (1999). A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(10), 1801–1808. https://doi.org/10.1139/f99-128

Hugoni, M., Etien, S., Bourges, A., Lepère, C., Domaizon, I., Mallet, C., Bronner, G., Debroas, D., & Mary, I. (2013). Dynamics of ammonia-oxidizing Archaea and Bacteria in contrasted freshwater ecosystems. *Research in Microbiology*, *164*(4), 360–370. https://doi.org/10.1016/j.resmic.2013.01.004

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1). https://doi.org/10.1186/1471-2105-11-119

Ishii, K., Fujitani, H., Sekiguchi, Y., & Tsuneda, S. (2020). Physiological and genomic characterization of a new "Candidatus Nitrotoga" isolate. *Environmental Microbiology*, *22*(6), 2365–2382. https://doi.org/10.1111/1462-2920.15015

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-07641-9

Jiang, Q. Q., & Bakken, L. R. (1999). Comparison of Nitrosospira strains isolated from terrestrial environments. *FEMS Microbiology Ecology*, *30*(2), 171–186. https://doi.org/10.1111/j.1574-6941.1999.tb00646.x

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*, e7359. https://doi.org/10.7717/peerj.7359

Kits, K. D., Sedlacek, C. J., Lebedeva, E. V., Han, P., Bulaev, A., Pjevac, P., Daebeler, A., Romano, S., Albertsen, M., Stein, L. Y., Daims, H., & Wagner, M. (2017). Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature*, *549*(7671), 269–272. https://doi.org/10.1038/nature23679

Kitzinger, K., Koch, H., Lücker, S., Sedlacek, C. J., Herbold, C., Schwarz, J., Daebeler, A., Mueller, A. J., Lukumbuzya, M., Romano, S., Leisch, N., Karst, S. M., Kirkegaard, R., Albertsen, M., Nielsen, P. H., Wagner, M., & Daims, H. (2018). Characterization of the First "Candidatus Nitrotoga" Isolate Reveals Metabolic Versatility and Separate Evolution of Widespread Nitrite-Oxidizing Bacteria. *MBio*, *9*(4). https://doi.org/10.1128/mBio.01186-18

Kitzinger, K., Padilla, C. C., Marchant, H. K., Hach, P. F., Herbold, C. W., Kidane, A. T., Könneke, M., Littmann, S., Mooshammer, M., Niggemann, J., Petrov, S., Richter, A., Stewart, F. J., Wagner, M., Kuypers, M. M. M., & Bristow, L. A. (2019). Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nature Microbiology*, *4*(2), 234. https://doi.org/10.1038/s41564-018-0316-2

Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, *16*(5), 263–276. https://doi.org/10.1038/nrmicro.2018.9

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lean, D. R. S., & Knowles, R. (1987). Nitrogen Transformations in Lake Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, *44*(12), 2133–2143. https://doi.org/10.1139/f87-262

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lipponen, M. T. T., Martikainen, P. J., Vasara, R. E., Servomaa, K., Zacheus, O., & Kontro, M. H. (2004). Occurrence of nitrifiers and diversity of ammonia-oxidizing bacteria in developing drinking water biofilms. *Water Research*, *38*(20), 4424–4434. https://doi.org/10.1016/j.watres.2004.08.021

Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R., & Stahl, D. A. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature*, *461*(7266), 976–979. https://doi.org/10.1038/nature08465

Miranda, K. M., Espey, M. G., & Wink, D. A. (2001). A rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. *Nitric Oxide: Biology and Chemistry*, *5*(1), 62–71. https://doi.org/10.1006/niox.2000.0319

Mukherjee, M., Ray, A., Post, A. F., McKay, R. M., & Bullerjahn, G. S. (2016). Identification, enumeration and diversity of nitrifying planktonic archaea and bacteria in trophic end members of the Laurentian Great Lakes. *Journal of Great Lakes Research*, *42*(1), 39–49. https://doi.org/10.1016/j.jglr.2015.11.007

Nowka, B., Daims, H., & Spieck, E. (2015). Comparison of Oxidation Kinetics of Nitrite-Oxidizing Bacteria: Nitrite Availability as a Key Factor in Niche Differentiation. *Applied and Environmental Microbiology*, *81*(2), 745–753. https://doi.org/10.1128/AEM.02734-14

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

Okazaki, Y., Nishimura, Y., Yoshida, T., Ogata, H., & Nakano, S.-I. (2019). Genome-resolved viral and cellular metagenomes revealed potential key virus-host interactions in a deep freshwater lake. *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.14816

Pachiadaki, M. G., Sintes, E., Bergauer, K., Brown, J. M., Record, N. R., Swan, B. K., Mathyer, M. E., Hallam, S. J., Lopez-Garcia, P., Takaki, Y., Nunoura, T., Woyke, T., Herndl, G. J., & Stepanauskas, R. (2017). Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. *Science*, *358*(6366), 1046–1051. https://doi.org/10.1126/science.aan8260

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., Lagkouvardos, I., Karst, S. M., Galushko, A., Koch, H., Berry, D., Daims, H., & Wagner, M. (2015). Cyanate as an energy source for nitrifiers. *Nature*, *524*(7563), 105–108. https://doi.org/10.1038/nature14856

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Paver, S. F., Newton, R. J., & Coleman, M. L. (2020). Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry. *Environmental Microbiology*, *22*(1), 433–446. https://doi.org/10.1111/1462-2920.14862

Podowski, J. C., Paver, S. F., Newton, R. J., & Coleman, M. L. (2021). *Genome streamlining, proteorhodopsin, and organic nitrogen metabolism in freshwater nitrifiers* (p. 2021.01.19.427344). bioRxiv. https://doi.org/10.1101/2021.01.19.427344

Rae, B. D., Long, B. M., Badger, M. R., & Price, G. D. (2013). Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiology and Molecular Biology Reviews*, *77*(3), 357–379. https://doi.org/10.1128/MMBR.00061-12

Revilla, M., Alexander, J., & Glibert, P. M. (2005). Urea analysis in coastal waters: Comparison of enzymatic and direct methods. *Limnology and Oceanography: Methods*, *3*(7), 290–299. https://doi.org/10.4319/lom.2005.3.290

Robertson, D. M., & Saad, D. A. (2011). Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models1. *JAWRA Journal of the American Water Resources Association*, *47*(5), 1011–1033. https://doi.org/10.1111/j.1752-1688.2011.00574.x

Rozmarynowycz, M. J., Beall, B. F. N., Bullerjahn, G. S., Small, G. E., Sterner, R. W., Brovold, S. S., D'souza, N. A., Watson, S. B., & McKay, R. M. L. (2019). Transitions in microbial communities along a 1600 km freshwater trophic gradient. *Journal of Great Lakes Research*, *45*(2), 263–276. https://doi.org/10.1016/j.jglr.2019.01.004

Santoro, A. E., Casciotti, K. L., & Francis, C. A. (2010). Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environmental Microbiology*, *12*(7), 1989–2006. https://doi.org/10.1111/j.1462-2920.2010.02205.x

Scavia, D., David Allan, J., Arend, K. K., Bartell, S., Beletsky, D., Bosch, N. S., Brandt, S. B., Briland, R. D., Daloğlu, I., DePinto, J. V., Dolan, D. M., Evans, M. A., Farmer, T. M., Goto, D., Han, H., Höök, T. O., Knight, R., Ludsin, S. A., Mason, D., … Zhou, Y. (2014). Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. *Journal of Great Lakes Research*, *40*(2), 226–246. https://doi.org/10.1016/j.jglr.2014.02.004

Schleper, C. (2010). Ammonia oxidation: Different niches for bacteria and archaea? *The ISME Journal*, *4*(9), 1092–1094. https://doi.org/10.1038/ismej.2010.111

Schramm, A., Larsen, L. H., Revsbech, N. P., Ramsing, N. B., Amann, R., & Schleifer, K. H. (1996). Structure and function of a nitrifying biofilm as determined by in situ hybridization and the use of microelectrodes. *Applied and Environmental Microbiology*, *62*(12), 4641–4647. https://doi.org/10.1128/aem.62.12.4641-4647.1996

Sigman, D. M., Casciotti, K. L., Andreani, M., Barford, C., Galanter, M., & Böhlke, J. K. (2001). A Bacterial Method for the Nitrogen Isotopic Analysis of Nitrate in Seawater and Freshwater. *Analytical Chemistry*, *73*(17), 4145–4153. https://doi.org/10.1021/ac010088e

Small, G. E., Bullerjahn, G. S., Sterner, R. W., Beall, B. F. N., Brovold, S., Finlay, J. C., McKay, R. M. L., & Mukherjee, M. (2013). Rates and controls of nitrification in a large oligotrophic lake. *Limnology and Oceanography*, *58*(1), 276–286. https://doi.org/10.4319/lo.2013.58.1.0276

Small, G. E., Cotner, J. B., Finlay, J. C., Stark, R. A., & Sterner, R. W. (2014). Nitrogen transformations at the sediment–water interface across redox gradients in the Laurentian Great Lakes. *Hydrobiologia*, *731*(1), 95–108. https://doi.org/10.1007/s10750-013-1569-7

Smith, J. M., Chavez, F. P., & Francis, C. A. (2014). Ammonium Uptake by Phytoplankton Regulates Nitrification in the Sunlit Ocean. *PLoS ONE*, *9*(9), e108173. https://doi.org/10.1371/journal.pone.0108173

Solomon, C., Collier, J., Berg, G., & Glibert, P. (2010). Role of urea in microbial metabolism in aquatic systems: A biochemical and molecular review. *Aquatic Microbial Ecology*, *59*, 67–88. https://doi.org/10.3354/ame01390

Sterner, R. W. (2021). The Laurentian Great Lakes: A Biogeochemical Test Bed. *Annual Review of Earth and Planetary Sciences*, *49*(1), null. https://doi.org/10.1146/annurev-earth-071420-051746

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33–36.

Tolar, B. B., Wallsgrove, N. J., Popp, B. N., & Hollibaugh, J. T. (2017). Oxidation of urea-derived nitrogen by thaumarchaeota-dominated marine nitrifying communities. *Environmental Microbiology*, *19*(12), 4838–4850. https://doi.org/10.1111/1462-2920.13457

Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, *6*(1), 158. https://doi.org/10.1186/s40168-018-0541-1

van Dongen, S., & Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Methods in Molecular Biology (Clifton, N.J.)*, *804*, 281–295. https://doi.org/10.1007/978-

1-61779-361-5_15

Vollenweider, R. A., Munawar, M., & Stadelmann, P. (1974). A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes. *Journal of the Fisheries Research Board of Canada*, *31*(5), 739–762. https://doi.org/10.1139/f74-100

Waldbauer, Jacob, Lichun Zhang, Adriana Rizzo, and Daniel Muratore. 2017. “DiDO-IPTL: A Peptide-Labeling Strategy for Precision Quantitative Proteomics.” *Analytical Chemistry* 89 (21): 11498–504. https://doi.org/10.1021/acs.analchem.7b02752.

Wheeler, P. A., & Kirchman, D. L. (1986). Utilization of inorganic and organic nitrogen by bacteria in marine systems1. *Limnology and Oceanography*, *31*(5), 998–1009. https://doi.org/10.4319/lo.1986.31.5.0998

Wilhelm, S. W., Bullerjahn, G. S., & McKay, R. M. L. (2020). The Complicated and Confusing Ecology of Microcystis Blooms. *MBio*. https://doi.org/10.1128/mBio.00529-20

Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, *32*(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638

Chapter 4

Phylogenetically and functionally diverse autotrophs are abundant in the Laurentian Great Lakes and widespread in other freshwater ecosystems

4.1 Abstract

Carbon fixation is the fundamental process underlying biomass production on earth. Most carbon fixation is done by plants on land and phytoplankton in aquatic systems, using energy conserved by oxygenic photosynthesis. Additionally, a variety of Bacteria and Archaea are known to couple carbon fixation to chemolithotrophic (non-photosynthetic) energy metabolism or anoxygenic photosynthesis. However, the contributions of these taxa and their associated energy conservation pathways to primary productivity are poorly constrained. Here we investigate the taxa and energetic pathways associated with carbon fixation in the picoplankton size fraction of the Laurentian Great Lakes. The Calvin-Benson-Bassham pathway was by far the most prevalent carbon fixation pathway. RuBisCo, the key enzyme in this pathway, was predominantly associated with eukaryotic phytoplankton in the spring. In stratified summer conditions, RuBisCo genes from nitrifiers *Nitrosospira* and *Ca*. Nitrotoga were abundant in deep water. Unexpectedly, we found that RuBisCo genes associated with Burkholderiales constituted the majority of taxonomically resolvable RuBisCo genes in many summer samples, even from sunlit surface and deep chlorophyll-maximum layers where phytoplankton are presumed to dominate. To characterize these abundant bacterial autotrophs, we reconstructed genomes from metagenomes and discovered a diverse array of lineages in Burkholderiales that encode RuBisCo, photosystem II, and sulfur oxidation. The distribution of RuBisCo genes within these MAGs, the phylogeny of *rbcL* itself, and the organization of the RuBisCo operon suggest a

complicated evolutionary history for this metabolic complement. We find this metabolic complement in genomes from freshwater systems around the world, suggesting a potentially important role for uncharacterized autotrophs in a wide range of systems. A more complete understanding of energy metabolism in these organisms will help inform their potential impact on carbon fixation, in the Laurentian Great Lakes and in other freshwater systems globally.

4.2 Introduction

Carbon fixation in oxygenated aquatic ecosystems is classically considered to be carried out almost exclusively by photosynthetic phytoplankton. However, recent inventories in marine systems suggest that dark (non-photosynthetic) carbon fixation in the euphotic zone (Baltar and Herndl 2019) and below the euphotic zone (Pachiadaki et al. 2017; Reinthaler, van Aken, and Herndl 2010) both contribute substantially to global primary productivity, with a total flux on the order of one-quarter the photosynthetic flux. This dark carbon fixation is frequently coupled to the oxidation of reduced nitrogen and sulfur substrates (Kitzinger et al. 2020; Tuttle and Jannasch 1977; Shah et al. 2019; Callbeck et al. 2018) and is predicted to be carried out by diverse taxa, many of which have not yet been isolated or physiologically characterized (E. D. Graham, Heidelberg, and Tully 2018; Pachiadaki et al. 2017; Swan et al. 2011). These discoveries emphasize the need to better understand the breadth of organisms capable of carbon fixation and to specifically assess their contributions to global primary productivity, as they are overlooked by estimates based on chlorophyll *a* (Carr et al. 2006).

The Laurentian Great Lakes are a unique system for quantifying microbial carbon fixation and respiration budgets. They are unusual among lakes in that autochthonous carbon fixation is a much larger contribution to total organic carbon than allochthonous inputs, owing to their large volume and relatively small watershed area (Zhou, Guo, and Minor 2016). Primary productivity from photosynthesizers has been well studied in the Great Lakes (Vollenweider, Munawar, and Stadelmann 1974; Glooschenko et al. 1974; Sterner 2010; Reavie et al. 2021) and the upper lakes (Superior, Michigan, and Huron) have generally much lower productivity than the lower lakes (Erie and Ontario). Microbes – bacteria, archaea and small single cell eukaryotes – are vital to both heterotrophic respiration and carbon fixation in offshore regions of the Great Lakes (Heath, Hwang, and Munawar 2010), especially in Lake Superior where bacterial respiration (Biddanda, Ogdahl, and Cotner 2001) is more than 80% of total planktonic respiration and microbes smaller than 3 microns perform more than 50% of all carbon fixation (Fahnenstiel et al. 1986). While the diversity and potential contribution of picoplanktonic cyanobacteria to carbon fixation has been explored (Ivanikova et al. 2007; 2008), a more holistic understanding of photosynthetic and chemolithotrophic microbial contributions to carbon fixation in the Great Lakes is lacking.

Recent genomic surveys of freshwater systems have highlighted the potential for non-photosynthetic bacteria to contribute to carbon fixation via chemolithotrophy. In the Great Lakes and elsewhere, autotrophic nitrifiers are abundant and likely important contributors to carbon fixation (Cabello-Yeves et al. 2019; Alfreider et al. 2018; Podowski et al. 2021). Beyond nitrifiers, recent surveys have found a number of novel potential chemolithotrophs in the water column of lakes (Arora-Williams et al. 2018; Tran et al. 2021; Peura et al. 2018). These organisms belong to the taxa Chlorobiaceae, Gallionellaceae and Desulfobulbaceae among others, and are predicted to use reduced compounds such as hydrogen and iron as electron donors, and are therefore associated with oxygen depleted zones of stratified water columns. As water column oxygen depletion is rare in offshore regions in the Great Lakes (Tellier et al. 2021) apart from some parts of Lake Erie (Scavia et al. 2014), these types of chemolithotrophy are expected to be rare in the Great Lakes, except perhaps in anoxic microzones on particles (Bianchi et al. 2018).

Among the most abundant lineages in freshwater systems are Betaproteobacteria, which are generally heterotrophs (Newton et al. 2011). Two genera in particular, *Limnohabitans* and *Rhodoferax*, are ubiquitous and abundant in the Great Lakes (Paver, Newton, and Coleman 2020; Rozmarynowycz et al. 2019). Recent genomics work has demonstrated that both of these genera may contain aerobic anoxygenic phototrophs (AAPs) which use sunlight energy to drive photophosphorylation and thereby supplement energy derived from organic carbon catabolism (Salka et al. 2011; Kasalický et al. 2018). These genera may also include chemolithotrophs that conserve energy entirely from inorganic sources (Baker et al. 2017; Zeng et al. 2012). Whether these metabolic capabilities are found in the abundant lineages of *Limnohabitans* and *Rhodoferax* that inhabit the Great Lakes, and could therefore influence the lakes' carbon and energy budgets, is unclear.

Here we assess carbon fixation potential in the Laurentian Great Lakes, using metagenomic and metatranscriptomic sequencing of the picoplanktonic size fraction (< 1.6 µm). We identified the taxonomic lineages that encode carbon fixation pathways and characterized the associated metabolic potential of these organisms, particularly with respect to energy metabolism. Surprisingly, we found widespread genomic potential for carbon fixation among the Betaproteobacterial family *Burkholderiales* associated with photosystem II genes and a suite of sulfur oxidation genes that were evidently assembled via a complex evolutionary history including repeated gene transfer events. We demonstrate that this metabolic complement is widespread geographically among global lakes. Together these results suggest that uncharacterized bacterial lineages may make significant and previously undescribed contributions to carbon fixation in oxygenated lakes across the globe.

4.3 Methods

4.3.1 Sample collection and sequencing

Samples from 2012 were collected as described in (Podowski et al. 2021). Water samples were collected in spring and summer 2012 during the EPAs Water Quality Survey aboard the R/V *Lake Guardian*. Samples from Lake Michigan were collected aboard the R/V *Blue Heron* in June of 2016 during the NSF-UNOLS Chief Scientist Training Cruise. For metagenome samples collected on the R/V *Blue Heron*, approximately 5-8 L of water was collected using a CTD rosette sampler (Sea-Bird Scientific) from the surface and the deep chlorophyll layer. Water was filtered through a GF/A glass fiber filter (Whatman 1820-047; nominal pore size 1.6 µm) to exclude larger cells and particles, and cells were collected onto a 0.22 µm Millipore Express Plus 47mm membrane filter (GPWP04700, ThermoFisher Scientific MA, USA). Filters were stored at -80°C. For metatranscriptome samples collected on the R/V *Blue Heron*, approximately 2L of water was collected using a CTD rosette sampler (Sea-Bird Scientific) from the surface. Water was filtered through a GF/A glass fiber filter (Whatman 1820-047; nominal pore size 1.6 µm) to exclude larger cells and particles, and cells were collected onto a 0.22 µm Millipore Express Plus 25mm membrane filter (GPWP02500, ThermoFisher Scientific MA, USA). Filters were flash frozen in liquid nitrogen no more than 15 minutes after water was collected from the rosette. Filters were then stored at -80°C.

DNA for all metagenomic samples was extracted using the phenol:chloroform extraction protocol (Paver, Newton, and Coleman 2020) as in (Podowski et al. 2021). RNA was extracted from metatranscriptomic samples using a modified version of the Zymo Quick-RNA Fungal/Bacterial Microprep Kit. Six surface spring metagenomes from 2012 were sequenced at the Joint Genome Institute (JGI) using Illumina HiSeq (2x150bp) as in (Podowski et al. 2021). 15 summer metagenomes from 2012 were sequenced at the University of Chicago Functional Genomics Core Facility using Illumina HiSeq 2500 (2x250bp) as in (Podowski et al. 2021). 14 summer metagenomes were sampled identically to (Podowski et al. 2021) except library preparation was conducted using Illumina Nextera XT and sequenced at Harvard Bauer Core Facility. Four Lake Michigan metagenomes from 2016 were sequenced at the JGI using Illumina NovaSeq S4 (2x151bp). Two Lake Michigan metatranscriptomes from 2016 were sequenced at the JGI using Illumina NovaSeq S4 (2x151bp).

Samples for single-cell sequencing were collected in August 2014 from the bottom of the water column as described in (Podowski et al. 2021). Samples were collected in summer 2014 during the EPAs Water Quality Survey aboard the R/V *Lake Guardian*. For each sample, 1 ml of raw water was incubated with 100 µl of glycerol-TE buffer (20 ml 100X TE pH 8 + 100 ml glycerol + 60 ml water; final concentration after sample addition is 10mM Tris, 1mM EDTA, 5% glycerol) for 10 minutes in the dark, then flash frozen in liquid nitrogen and stored at -80°C until processing.

4.3.2 Metagenome and single-cell genome sequencing

Raw interleaved reads for the 6 spring surface samples were quality-controlled at the Joint Genomes Institute (JGI) using bbduk.sh for adapter trimming (ktrim=r, minlen=40, minlenfraction=0.6,mink=11, 519 tbo, tpe, k=23, hdist=1, hdist2=1, ftm=5) and quality filtering (maq=8, maxns=1, minlen=40, 520 minlenfraction=0.6, k=27, hdist=1, trimq=12, qtrim=rl). Raw interleaved reads for the 15 summer samples were adapter trimmed, quality filtered, and interleaved using bbduk (parameters: ktrim=r, mink=8, hdist=2, k=21, forcetrimleft=10, forcetrimright=199, minlen=150) using BBTools suite version 35.74 (https://sourceforge.net/projects/bbmap/). 14 summer samples that were sequenced at Harvard Bauer Core Facility were quality controlled using bbduk qtrim=rl trimq=13 minlength=40 in BBTools suite version 35.74 (https://sourceforge.net/projects/bbmap/). 4 2016 Lake Michigan metagenome samples were quality controlled by JGI using bbduk.sh for adapter trimming (ktrim=r ordered minlen=51 minlenfraction=0.33 mink=11 tbo tpe rcomp=f overwrite=true k=23 hdist=1 hdist2=1 ftm=5) and quality filtering (maq=3.0,0 trimq=0.0 qtrim=r ordered overwrite=true maxns=3 minlen=51 minlenfraction=0.33 k=31 hdist=1). 2 2016 Lake Michigan metatranscriptome samples were quality controlled by JGI using bbduk.sh for adapter trimming (ktrim=r ordered minlen=51 minlenfraction=0.33 mink=11 tbo tpe rcomp=f overwrite=true k=23 hdist=1 hdist2=1 ftm=5) and quality filtering (maq=10.0,0 trimq=0.0 qtrim=r ordered overwrite=true maxns=1 minlen=51 minlenfraction=0.33 k=31 hdist=1). Interleaved short read metagenomes for 6 spring summer surface samples and 15 summer samples were assembled individually using metaSPAdes v 3.11.1 (Nurk et al. 2017). Interleaved short read metagenomes for 14 Nextera metagenomes were assembled individually using metaSPAdes v 3.12.0 (Nurk et al. 2017).

SAG genomes were downloaded from JGI IMG/M after quality control and normalization by JGI, normalization using bbnorm target=100 mindepth=2, reformat.sh samplereadstarget=20000000 using BBTools suite version 38.39 (https://sourceforge.net/projects/bbmap/) and assembly using SPAdes (Bankevich et al. 2012) -phred-offset 33 -t 16 -m 64 --sc -k 25,55,95.

4.3.3 Taxonomic and functional analysis of short reads

Normalized reads per kilobase for *rbcL* genes was generated using a short read centric, blast alignment based approach. First, open reading frames (ORFs) in amino acid format were generated from all 35 assemblies using prodigal v2.6.3 (Hyatt et al. 2010), which amounted to approximately 49 million ORFs. These ORFs were annotated functionally using Kofamscan (Aramaki et al. 2020) to assign KEGG Orthology IDs (KOs). KOs were assigned to an ORF that was the best hit given by Kofamscan as long as the bitscore was greater than 50. This same pool of 49 million ORFs were aligned against RefSeq proteins from complete genomes (downloaded on 10/27/2020) (O'Leary et al. 2016) using DIAMOND v2.0.11.149 (Buchfink, Reuter, and Drost 2021) in blastp mode. KronaTools (Ondov, Bergman, and Phillippy 2011) ktClassifyBlast with a -t 0.1 was used to generate lowest common ancestor assignments around the best hit in RefSeq. TaxIDs as assigned by KronaTools were converted to lineages by an in house script. This pool of 49 million ORFs, with both KO assignments and taxonomy assignments, was then used downstream to characterize short reads, and will be referred to as the LGL Gene Catalog.

We next took the 21 merged short read metagenomes from 2012, the forward reads from the 4 short read metagenomes from BH2016 set, and the forward reads from the 2 short read metatranscriptome from the BH2016 set, and aligned each of them independently against the LGL Gene Catalog using DIAMOND v2.0.11.149 (Buchfink, Reuter, and Drost 2021) in blastx mode. From this output, the best hit for each of the merged short reads against the LGL Gene Catalog was extracted using scripts from enveomics (Rodriguez-R and Konstantinidis 2016) and each merged read inherited the KO assignment of the best hit ORF from the LGL Gene Catalog. These 21 merged short read metagenome samples, 4 forward read metagenome samples, and 2 forward read metatranscriptome samples were also independently aligned against the same RefSeq proteins database described above, and taxonomy was assigned to each merged short read like above, using KronaTools (Ondov, Bergman, and Phillippy 2011) with -t 0.1 and in house script to convert TaxIDs to lineages.

For all KOs assigned at least once in the LGL Gene Catalog, the number of reads in each KO was tallied. To normalize each KO count by length, for each KO all genes were taken from the LGL Gene Catalog that were assigned that KO, and the lengths of those genes were determined in nucleotide bases. The maximum length of all these genes was taken to normalize the KO by length. This was to account for the numerous very short, fragmentary ORFs. Length normalized KO counts were then normalized by sample by determining a set of ribosomal protein KOs, and finding the mean of the length normalized ribosomal protein KO counts. The KO counts in each sample were then divided by the mean of the length normalized ribosomal protein KO counts. This normalized KO count for *rbcL* (K01601), normalized by length across samples and ribosomal protein counts within a sample, is plotted in Figure 1.

Taxonomic composition of a pool of reads was determined in a similar fashion. KO assignments for merged short reads from the 21 merged short read metagenomes were taken as above, inherited from best hit in LGL Gene Catalog. Taxonomic assignments for short reads were inherited from the lowest common ancestor of the hits within 0.1 bitscore from the best hit in RefSeq. For a given sample, read names from reads assigned as K01601 were taken and then the names of these reads were queried against the best hit summarized merged read taxonomy for that sample. For each sample, the fraction of this total pool of reads assigned to K01601 that were taxonomically assigned to “Eukaryota” “Nitrosomonadales” “Burkholderiales” and “Synechococcales” were generated by dividing the number of reads assigned to each group in a sample, by the total number of reads assigned to K01601 from that sample. This was performed identically for forward reads for the 4 BH2016 metagenomes, and the 2 BH2016 metatranscriptomes. These values are presented in Figure 1, 2a and 2b.

In order to assess expression and compare that to DNA level abundance, we specifically identified reads and transcripts associated with Burkholderiales. This was done on the basis of short reads annotated against the LGL Gene Catalog as described previously. First reads which had a best hit to an ORF that was taxonomically annotated as Burkholderiales were identified. From those reads, KO counts were summarized based on the KOs assigned to those short reads. To determine the degree to which high transcript count was a product of high abundance at the DNA level, for each of these KOs the counts of transcripts in a sample were divided by the counts of KOs at the DNA level in the metagenome sample taken from the surface of station MI27M.

4.3.4 Generating metagenomic assembled genomes (MAGs)

Mapping and binning was carried out on a sample by sample basis. Short reads were merged using bbmerge (Bushnell, Rood, and Singer 2017) qtrim2=r trimq=10,13,16 and adapter=default. Merged short reads from 15 metagenomes sequenced at University of Chicago Functional Genomics Facility and 6 metagenomes sequenced at the JGI were mapped against each assembly using bowtie2 (Langmead and Salzberg 2012, 2) in –sensitive mode. These coverage values were used to bin each assembly using MetaBAT2 (Kang et al. 2019), Binsanity (Elaina D. Graham, Heidelberg, and Tully 2017) and CONCOCT (Alneberg et al. 2014). DASTools (Sieber et al. 2018) was used to compare and score these binning results, and generate optimal bins. CheckM lineage_wf (Parks et al. 2015) was used to assess completion and contamination of these bins. Anvi'o v5 (Eren et al. 2015) was then used to manually verify bins. Bins greater than 75% complete and less than 5% contaminated as determined by CheckM were considered MAGs. MAGs from all samples were then combined and processed together. MAG taxonomy was determined using GTDB-tk 1.7.0 (Chaumeil et al. 2020). FastANI (Jain et al. 2018) was used to compare average nucleotide identity (ANI) between each MAG and generate 95% ANI clusters.

To determine relative abundance of MAGs and MAG clusters in samples, bowtie2 (Langmead and Salzberg 2012, 2) in –sensitive mode was used to map merged short reads onto MAGs in a competitive fashion. Abundance of MAG clusters was determined by summing up all reads mapped to any MAG in that cluster. To identify MAGs and SAGs with *rbcL,* prodigal v 2.6.3 (Hyatt et al. 2010) was used to find ORFs on MAG and SAG contig files and reference *rbcL* genes were searched against ORFs using NCBI blastp (Camacho et al. 2009). MAGs and SAGs which directly contained these *rbcL* sequences were considered downstream, but also any MAGs and SAGs in 95% ANI clusters with a MAG or SAG with *rbcL*. Further, sequence comparisons were made using fastANI (Jain et al. 2018) between any MAG with *rbcL* or in a cluster with a *rbcL* containing MAG, and all SAGs. This was done in order to identify any SAGs that may represent a fragment of a MAG, considering low completion in many cases.

4.3.5 Annotation and comparative genomic analysis

SAGs and MAGs identified above were considered for downstream analysis. Anvi'o v 7.1 (Eren et al. 2015) was used to generate contigs databases for each MAG and SAG, and run prodigal v 2.6.3 to find ORFs (Hyatt et al. 2010). Anvi'o was then used to search these ORFs against KEGG (Aramaki et al. 2020), PFAM (Finn et al. 2014) and COG (Tatusov et al. 2000) databases.

In order to search for MAGs from diverse environments that contain similar metabolic complements to what was presented here, we took a hybrid approach. First we identified any genomes in the same genus or closely related genus as any of the MAGs or SAGs that were found to contain *rbcL* and *pufLM*, using taxonomy from GTDB (Parks et al. 2022). These were Rhodoferax_A, Rhodoferax, Rhodoferax_B, Pseudorhodoferax, Limnohabitans_A, Limnohabitans, Hylemonella, CAINMN01, Hydrogenophaga, CAISIP01, and CAIKVZ01. We then took genomes from large-scale metagenomic studies of freshwater systems (Cabello-Yeves et al. 2018; Cabello-Yeves et al. 2019; Okazaki et al. 2019; Linz et al. 2018; Tran et al. 2021; Arora-Williams et al. 2018) and specifically ran them through GTDB-tk v 1.7.0 (Chaumeil et al. 2020) to ensure no potential matches were missed. We further used existing GTDB based taxonomy assignments from a large scale freshwater metagenome study (Buck et al. 2021) to search for MAGs and SAGs from these genera. The total of all of these reference MAGs and SAGs, plus selected references NAT185 (E. D. Graham, Heidelberg, and Tully 2018), LSUCC0115 (Henson et al. 2020), and several MAGs from Lake Michigan (Props and Denef 2020) combined with all MAGs and SAGs presented in this paper, was 971 genomes. Anvi'o v7.1 was used to generate contigs databases for all references, and to annotate ORFs from these references against KEGG, PFAM and COG databases. Anvi'o was then used to run this set of 971 genomes through the pangenome pipeline (Delmont and Eren 2018). DIAMOND blastp (Buchfink, Xie, and Huson 2015) was used to determine sequence similarity, ITEP (Benedict et al. 2014) to eliminate weak similarity and MCL (van Dongen and Abreu-Goodger 2012) to generate clusters.

A phylogenomic tree of these 971 genomes was generated using the Anvi'o phylogenomic pipeline. Bacterial core genes were found in Anvi'o contigs databases using HMMer (Mistry et al. 2013) and for each genome a concatenated set of 71 bacterial core genes were aligned using muscle (Edgar 2004)**.** This alignment was trimmed using Gblocks v0.91b

(Castresana 2000) and a phylogenomic tree was constructed using RaxML v8.2.12 (Stamatakis 2014) using the following parameters: -k -N 50 -p 83 -x 54 -f a.

Phylogenetic tree of *rbcL* was generated using sequences from references (F. R. Tabita et al. 2007) and from *rbcL* sequence found in MAGs described above. Sequences were aligned using mafft (Katoh et al. 2002) v7.310 in L-INS-i mode.Then, RaxML v8.2.12 (Stamatakis 2014) was used to construct a phylogenetic tree using -k -N 100 -p 29 -x 72 -f a. *rbcL* phylogenetic groups were assigned to this tree using reference phylogeny (F. R. Tabita et al. 2007).

4.4 Results

4.4.1 Distribution of rbcL throughout the water column

We first quantified the distribution of the gene encoding the large subunit of RuBisCO, *rbcL*, as a marker of carbon fixation potential. We identified abundant *rbcL* genes in both surface and deep waters of the Great Lakes. At five of six stations, *rbcL* had a higher normalized abundance in the hypolimnion compared to the surface in stratified summer conditions (Fig 4.1). The exception was Lake Superior, where normalized abundance of *rbcL* was more than two fold greater in the surface compared to the deep chlorophyll layer or the mid-hypolimnion. In Lakes Erie and Ontario, deep water samples had two to four fold greater *rbcL* abundance than surface water samples in the summer. Spring surface samples had two to four-fold higher *rbcL* abundance than summer surface samples, except in Lake Superior, where *rbcL* was more abundant in the summer.

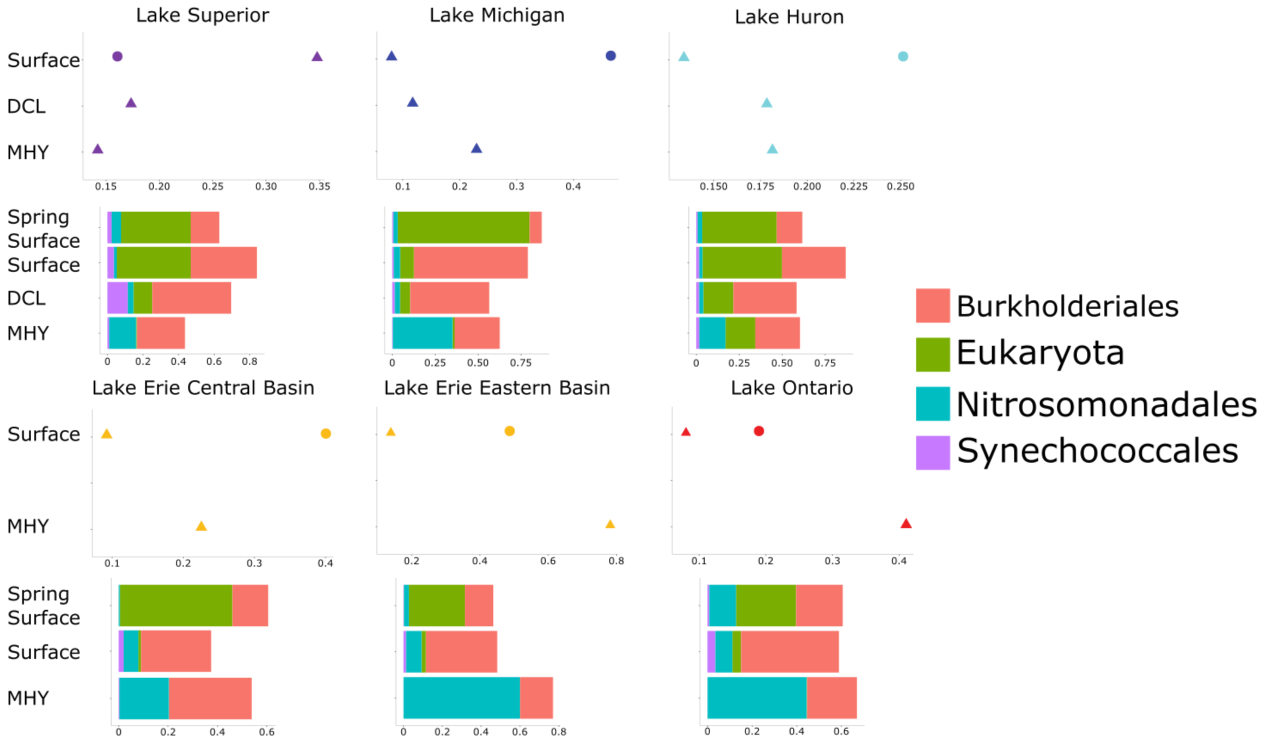


Figure 4.1 Distribution and taxonomy of *rbcL* genes across lakes and depths in the Great Lakes in spring and summer 2012. Top panel dot plot shows the reads per kilobase value for *rbcL*, normalized across samples according to the reads per kilobase for ribosomal reads in that sample. Circles represent summer samples, triangles represent spring samples. Key for sample naming: Surface= Surface Sample, DCL= Deep Chlorophyll Layer Sample, MHY= Mid Hypolimnion Sample. The X axis is not fixed between lakes. Bottom plot shows a stacked bar plot of the taxonomic makeup of merged short reads assigned as *rbcL*. Sample key is the same as above, with Spring Surface= Samples collected at the surface during the spring. Each bar is a fraction of total *rbcL* reads- and the remainder of the total is taxonomically ambiguous or does not belong to one of these 4 major groups. DCL samples were not collected in Lake Erie or Lake Ontario

4.4.2 Taxonomic affiliation of rbcL genes and transcripts reveals unexpected contributions from Burkholderiales

We assigned taxonomy to all *rbcL* genes in assemblies and assessed the contribution of major taxonomic groups to the total pool of *rbcL* genes. Burkholderiales associated *rbcL* genes comprised the majority of taxonomically assignable *rbcL* genes in many parts of the stratified water column, even outnumbering those of photosynthetic taxa in the photic zone. In Lake Michigan, Lake Ontario, and both basins of Lake Erie, *rbcL* genes associated with

Burkholderiales made up the majority of taxonomically assignable *rbcL* genes in summer surface samples, two-fold greater than other taxonomic groups (Fig 4.1). In Lakes Superior and Huron, Burkholderiales *rbcL* and eukaryotic *rbcL* genes were approximately equally abundant in summer surface waters. In all deep chlorophyll layer samples from Lakes Superior, Michigan, and Huron, Burkholderiales *rbcL* genes accounted for the majority of taxonomically assignable *rbcL* genes. Burkholderiales *rbcL* genes also made up the majority of taxonomically assignable *rbcL* genes in the mid-hypolimnion of Lake Superior and Lake Erie's central basin.

In deep water samples, Nitrosomonadales dominate *rbcL* pools, especially in Erie and Ontario. In Lake Michigan, Ontario, and Erie's eastern basin, *rbcL* sequences associated with *Nitrosomonadales* are the majority of taxonomically described *rbcL* sequences (Fig 4.1). In Erie's eastern basin and Ontario, Nitrosomonadales associated *rbcL* sequences are two times greater than *rbcL* from any other taxa. These *rbcL* sequences in Nitrosomonadales are contributed by nitrifiers from *Nitrosospira* and Ca. *Nitrotoga* which both fix carbon via the CBB pathway (Boddicker and Mosier 2018; Kitzinger et al. 2018; Norton et al. 2008; Podowski et al. 2021), and reach very high relative abundance in Erie and Ontario (Paver, Newton, and Coleman 2020; Podowski et al. 2021).

Taxonomic composition of *rbcL* transcripts suggest transcriptional dynamics decoupled from DNA abundance. Nearly all *rbcL* transcripts in Lake Michigan's surface waters in June 2016 appeared to come from eukaryotic cells. While *rbcL* genes associated with eukaryotes made up approximately 40% of the total *rbcL* genes in our Lake Michigan samples in June 2016 (Fig 4.2a), *rbcL* transcripts associated with eukaryotes made up more than 90% of the total *rbcL* transcripts we measured (Fig 4.2b). Burkholderiales associated *rbcL* genes made up approximately 20% of total *rbcL* genes (Fig 4.2a), but only about 1% of *rbcL* transcripts (Fig

4.2b), owing to the transcript dominance of eukaryotes. Nevertheless, *rbcL* appeared to be highly expressed compared to other Burkholderiales genes, based on the ratio of RNA to DNA sequencing reads. The genome-wide median ratio of Burkholderiales RNA to DNA reads was approximately 0.3, while the ratio of RNA to DNA reads for Burkholderiales was approximately 3.5 (Table S4.1). This suggests that *rbcL* was highly expressed in Burkholderiales at the time of sampling, above the general level of transcriptional activity of Burkholderiales.

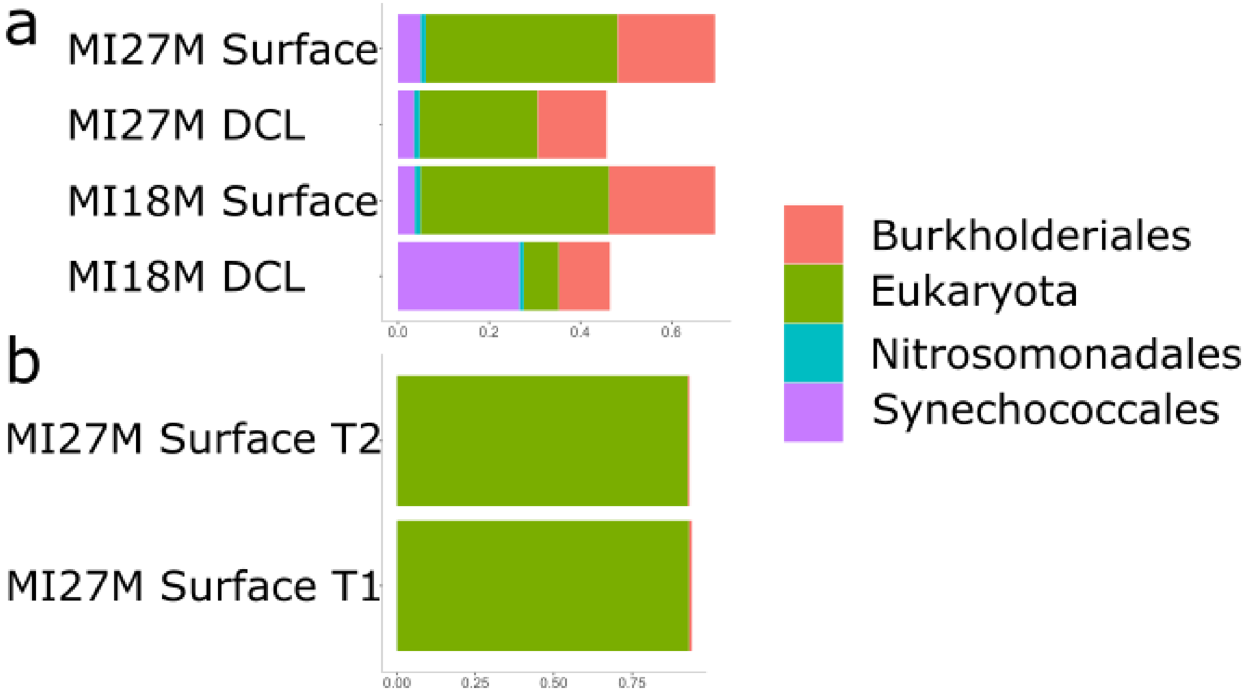


Figure 4.2 Taxonomy of *rbcL* genes in Lake Michigan at stations MI27M and MI18M in 2016. Surface=Sample collected at the surface, DCL= Deep Chlorophyll layer. Each bar is a fraction of total *rbcL* reads- and the remainder of the total is taxonomically ambiguous or does not belong to one of these 4 major groups. **4.2b**:Taxonomy of *rbcL* transcripts taken from two timepoints at the surface of station MI27M

4.4.3 Diverse MAGs in Burkholderiaceae encode RuBisCO with varied evolutionary histories

To explore the metabolism of potential autotrophs within Burkholderiales, we searched for genomes (MAGs and SAGs) containing *rbcL* sequences. We identified 8 phylogenetically distinct clusters that contain *rbcL* (Fig 4.3), within the genera CAINMN01, CAISIP01, and *Rhodoferax* according to GTDB release 202 (Parks et al. 2022). All of these MAGs also contain genes involved in anoxygenic phototrophy, including *pufLM* and bacteriochlorophyll biosynthesis genes.

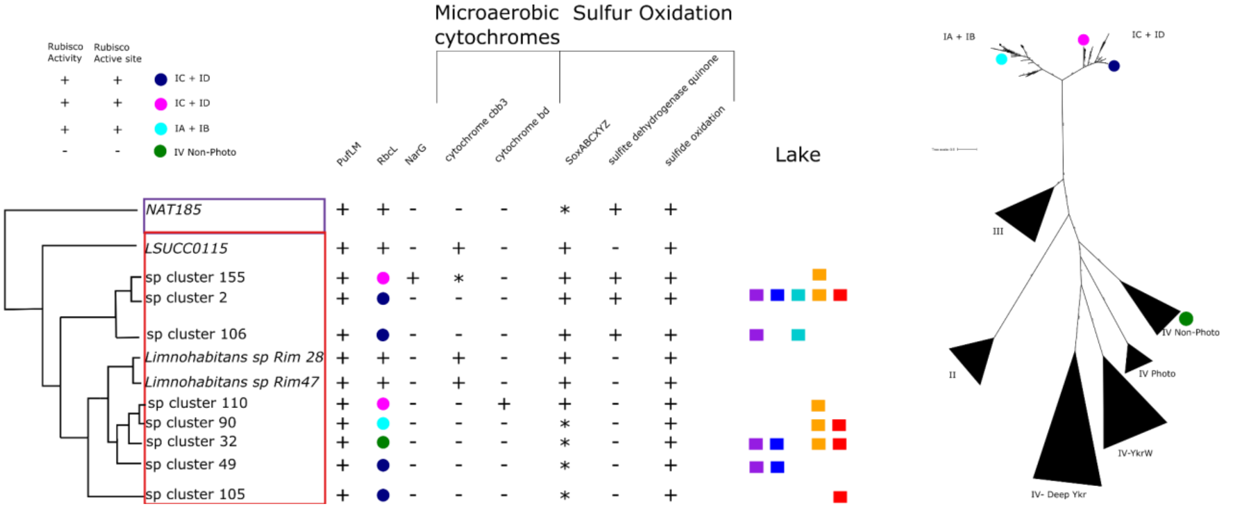


Figure 4.3 Schematic representation of various metabolic features of MAGs containing *rbcL* and *pufLM*. Bottom left: schematic tree displaying approximate phylogenetic relationships between distinct 95% clusters of MAGs found containing *pufLM* and *rbcL* and references. Groups within the red box are in Betaproteobacteria, NAT185 within the purple box is in Alphaproteobacteria. Gene content is described to the right, for *pufLM*, *rbcL*, *narG* and a number of genes involved in microaerobic growth and sulfur oxidation. Asterisk indicates that associated genes are not found in full in any single genome in that cluster. For *rbcL*, colored circles indicate presence of *rbcL* and mapped onto table to top left and tree to far right to indicate form of *rbcL*, only for MAGs presented here. Lake designation at the far right of the schematic indicates in which lake a MAG from that cluster was assembled. Far right: Tree of *rbcL* sequences adapted from (F. R. Tabita et al. 2007) with *rbcL* from these MAGs added to demonstrate the form of *rbcL* sequence present in each cluster of MAGs. Top left: Table adapted from (F. R. Tabita et al. 2007) summarizing confirmed activity and present active site of different forms of *rbcL*, for *rbcL* sequence found in MAGs.

We reconstructed the phylogeny of *rbcL* to understand its evolutionary history and to classify it as one of several known functional classes (F. R. Tabita et al. 2007). Of the 8 MAG clusters that contained *rbcL* and *rbcS*, 6 of these clusters contained *rbcL* sequences that fell within the group IC + ID, 1 contained *rbcL* sequences within group IA + IB, and 1 contained *rbcL* sequences within group IV Non-Photosynthetic (Fig 4.3) (F. R. Tabita et al. 2007).

Enzymatic activity of RuBisCO from *rbcL* type IC + ID and IA + IB has been verified, but not from *rbcL* from IV Non-Photosynthetic, whose function remains unknown (F. Robert Tabita et al. 2007). Phylogeny of *rbcL* sequences was not coherent with phylogenomic relationships among the MAG clusters. MAG cluster 110, for instance, was more closely related to MAG cluster 90 than to MAG cluster 49, based on the concatenated core gene phylogeny. However, the *rbcL* gene in cluster 110 was more closely related to the *rbcL* in cluster 49 than the one in cluster 90 (Fig 4.3). This was also the case for cluster 155, cluster 2 and 106. This suggests multiple acquisitions of the RuBisCO operon, rather than a single acquisition event in the common ancestor of all 8 of these clusters.

Diverse RuBisCO operon organization supports a complex evolutionary history. Reconstruction of the RuBisCO operon of all genome clusters with genuine (i.e., IA-ID) *rbcL* sequences revealed a diversity of operon organizations for genome clusters (Fig 4.4). Clusters 2, 110, and 155 show more similar RuBisCO operon organization despite not having closely related core gene phylogeny. *rbcL* phylogeny is more consistent with RuBisCO operon similarity than core gene phylogeny for these genome clusters (Fig 4.4). Better understanding of the RuBisCO operon is hampered by more fragmentary assemblies for clusters 49, 90, 105, and 106 (Fig 4.4). Interestingly, operon organization for clusters 2, 110, and 155 was very similar to organization in cultivated strain LSUCC0115 from coastal Gulf of Mexico (Fig 4.4) (Henson et al. 2020), and much less similar to more closely related references *Limnohabitans* Rim 28 or Rim 47 (Fig 4.4) (Zeng et al. 2012).

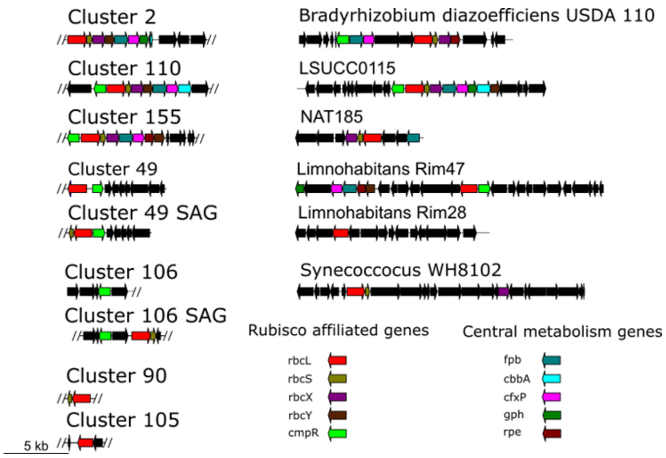


Figure 4.4 Gene operon map around *rbcL* gene for clusters of MAGs that contain genuine *rbcL* sequences. Operon map presented once for each cluster for presented MAGs, and also for several references both as presented in Figure 4 and other well described Rubisco containing bacteria. In cases where a SAG was identified for the cluster and contains *rbcL*, a map was presented both for this SAG and the representative MAG for that cluster.. Key genes are color coded as presented in bottom left. Forward slash in the map of MAGs indicate the end of the contig.

4.4.4 RuBisCO-containing MAGs appear metabolically flexible with potential for both chemolithoautotrophic and organoheterotrophic lifestyles

Gene content suggests potential for anaerobic or microaerobic growth in some but not all RuBisCO-containing MAG clusters. Several genes which indicate potential for anaerobic or microaerobic growth were investigated. *narG*, which is involved in dissimilatory nitrate reduction during anaerobic growth, was found in species cluster 155. A cbb_3-type cytochrome oxidase, which is involved in microaerobic growth (Preisig et al. 1996), was found in some MAGs in cluster 15. Cytochrome bd oxidase, also involved in microaerobic growth (D'mello, Hill, and Poole 1996), was found in species cluster 110. *coxABC*, involved in production of aa_3-cytochrome oxidase, was found in all clusters. This low affinity terminal oxidase can be involved in aerobic growth with replete oxygen (Morris and Schmidt 2013).

Sulfur oxidation genes in many MAG clusters suggest a potential energy source, but this function is unevenly distributed across clusters. Genes involved in thiosulfate oxidation, *soxABCXYZ*, were found in all MAG clusters studied here (Fig 4.3). However, in clusters 32, 49, 90, and 105, some but not all of these genes were present. *soeABC*, which is involved in oxidation of sulfite (Boughanemi et al. 2020), was found in clusters 2, 106, and 155. *fccB*, which is associated with sulfide oxidation, was found in all genome clusters.

Bacteriochlorophyll synthesis and photosystem genes imply capacity for phototrophy. A complete photosynthetic gene cluster (PGC) was found in at least one MAG for all 8 clusters, but the content of these operons differed (Table S4.2). Most notably, the PGC in cluster 2 lacked *pucAB* and *pufA* similar to *Limnohabitans* B9-3 (Kasalický et al. 2018), while these three genes were present in other clusters and many other *Limnohabitans* isolates (Kasalický et al. 2018). Genes *bluF* and *bchD*, found in *Limnohabitans* (PCG) were absent in all PCGs in our 8 clusters. *acsF*, the oxygen dependent ring cyclase (G. E. Chen, Canniffe, and Hunter 2017) was found in all 8 clusters, suggesting bacteriochlorophyll synthesis in oxygenated conditions.

Central carbon metabolism suggests potential for heterotrophic or mixotrophic growth. While a complete Embden–Meyerhof–Parnas (EMP) glycolysis pathway was not found in any MAGs in the 8 clusters, a complete Entner–Doudoroff (ED) pathway was found in clusters 2 and 110, while portions of this pathway were found in clusters 32, 106 and 155 (Table S4.2). This pathway allows for growth on glucose, and while less efficient than the EMP pathway, may be better optimized for mixotrophic growth during carbon fixation using the Calvin-Benson-Basham pathway (X. Chen et al. 2016). A complete TCA cycle was found in clusters 2, 32 and 155, and partial pathways were found in 105, 106, 49, and 90 (Table S4.2). The glyoxylate shunt was found complete in all 8 clusters, and this pathway is distributed throughout AAPs (Tang, Tang, and Blankenship 2011) and allows for use of carbon sources such as acetate or butyrate

(Dolan and Welch 2018; Vargas Asensio 2020). Together these results indicate the potential for use of several organic carbon sources.

Glycogen biosynthesis and degradation capacity suggests a need for long term energy storage. Complete glycogen biosynthesis pathway (*glgABC*) as well as glycogen degradation to glucose-6-phosphate (*glgP* and *malQ*) were found in clusters 2, 49 and 106 (Table S4.2). Complete glycogen biosynthesis and glycogen degradation to maltodextrin (*treX*) was found in cluster 155 (Table S4.2). The capacity to generate and breakdown glycogen suggests a need for long term energy storage, or transient conditions in which organic carbon availability does not limit growth (Wilson et al. 2010).

4.4.5 Distribution of rbcL-containing AAPs within the Great Lakes and across freshwater ecosystems

Distinct distribution of clusters suggests different environmental niches. Cluster 2 reached the highest relative abundance of all 8 genome clusters and was most abundant in deep chlorophyll or mid hypolimnion samples (Fig 4.5). Meanwhile, clusters 90, 110 and 155 were least abundant in these samples. Two of these clusters, 110 and 155, were those that displayed a potential for microaerobic or anaerobic growth, and the distinct distribution may be related to this potential physiological difference. Further, the high relative abundance of cluster 2 in deeper water may suggest chemolithotrophic or chemoorganotrophic growth as opposed to phototrophic growth.

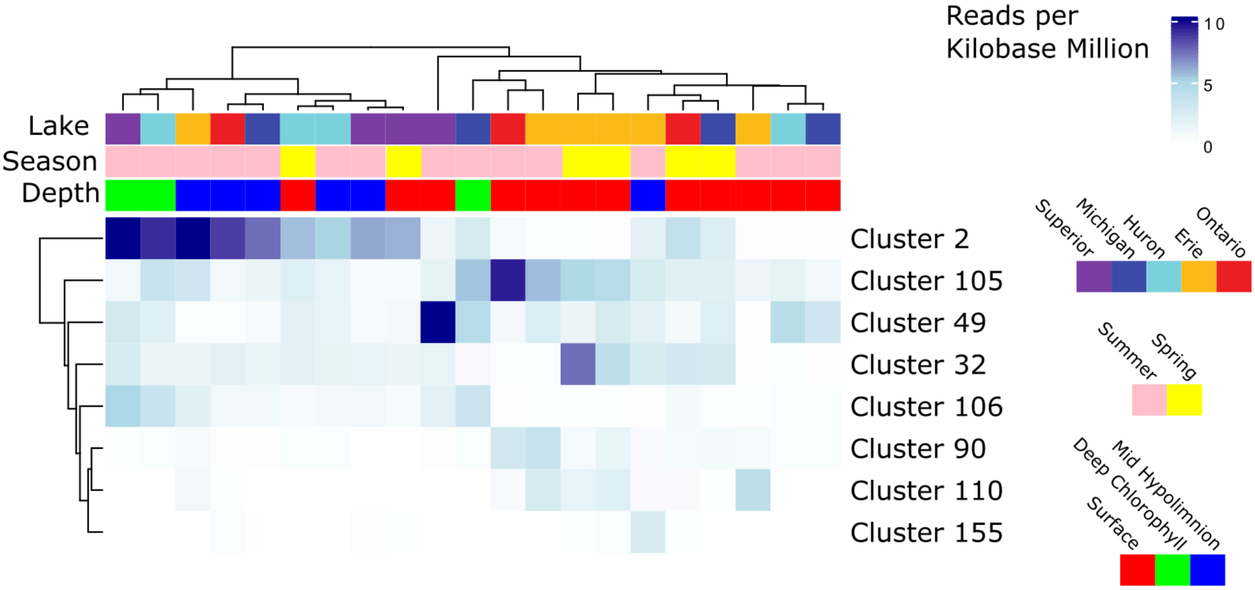


Figure 4.5 Relative abundance of MAG clusters that contain *rbcL* across lakes and depths in the Great Lakes. Relative abundance is presented as Reads per Kilobase Million to normalize for genome size and sequencing depth.

To understand the distribution of microbes in freshwater systems that contain *pufLM* and *rbcLS*, we searched genome databases and catalogs of freshwater genomes for microbes with this metabolic complement. MAGs that contain both *pufLM* and *rbcLS* are phylogenetically diverse and are found in a wide array of environments. Genomes that contain both AAP genes and *rbcL* were found in 6 different genera, for a total of 59 genomes (Table S4.3, Figure S4.1a, S4.1b, S4.1c, S4.1d). Overall, the wide phylogenetic and environmental breadth of genomes that contain *pufLM* genes and *rbcL* in these better sampled freshwater systems suggests a potential for global importance. This also suggests that studies that find bacteriochlorophyll (Masín et al. 2008) or *pufLM* (Salka et al. 2011) in diverse freshwater environments may also be detecting these AAPs that contain *rbcL*.

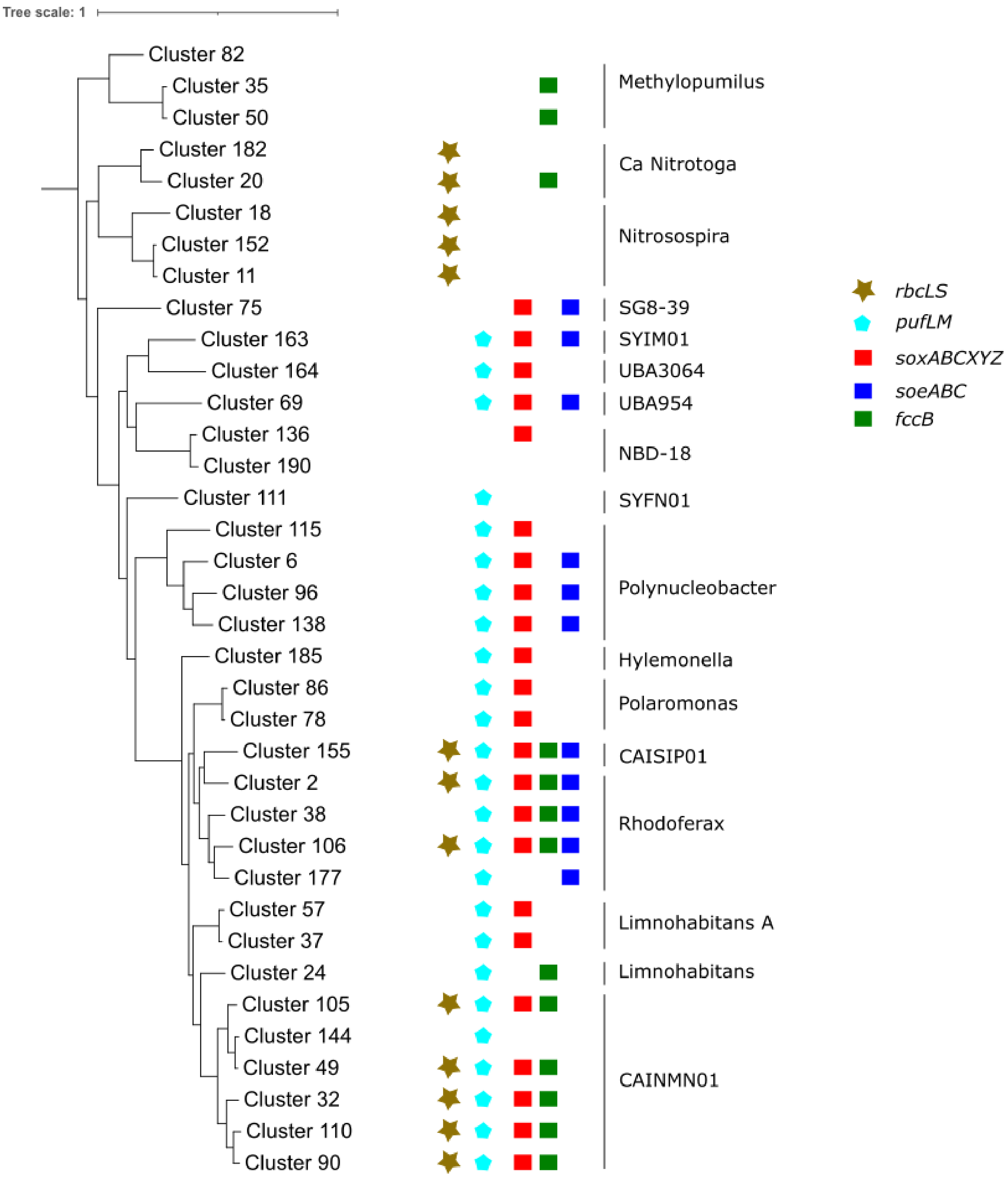


Figure 4.6 Phylogenomic tree of all Betaproteobacteria MAG clusters present from Great Lakes data presented. Gene content denoted is present in at least one of the MAGs present in the cluster.

4.5 Discussion

Dominance of eukaryotic phytoplankton in surface RuBisCO pools suggests high importance. Our results suggest that both at the gene and transcript level, *rbcL* from eukaryotic phytoplankton heavily outweighs *rbcL* from cyanobacterial picoplankton (Fig 4.1). The exception to this pattern appears to be in less common blooms at the deep chlorophyll layer (Fig 4.1, 4.2). Dominance of eukaryotic *rbcL* is further emphasized by what appears to be a record of the spring phytoplankton bloom in the relative abundance and taxonomic composition of *rbcL*. Focusing on Lake Michigan, we see a progression of April samples entirely dominated by eukaryotic *rbcL* (Fig 4.1), in June that dominance has waned but eukaryotic *rbcL* is still the majority of taxonomically assigned sequences, and by August, Burkholderiales *rbcL* is by far the majority of taxonomically assigned *rbcL*. But rather than a shift in relative taxonomic composition alone, it appears that an increase in the relative abundance of *rbcL* in the spring is driven by these eukaryotic *rbcL*, and as the eukaryotic *rbcL* recede in the summer, the relative abundance of *rbcL* altogether decreases two to four fold. Though eukaryotic associated *rbcL* did have more specific taxonomic assignment, a general lack of relevant freshwater phytoplankton genomes in RefSeq limits the utility of more specific taxonomic assignment. Chlorophyceae was the most common assignment of eukaryotic *rbcL* sequence, which is generally consistent with classical taxonomic identification of phytoplankton (Vollenweider, Munawar, and Stadelmann 1974). Members of these taxonomic groups are classically much larger than the 1.6 micron prefilter used in our sampling protocol, but a low similarity to reference organisms could explain this discrepancy. Inadvertent sequencing of chloroplast genomes from larger eukaryotic phytoplankton due to rupturing of eukaryotic cells during sampling could also explain this discrepancy

Increase in *rbcL* into the aphotic zone of the lower lakes associated with nitrifiers suggests potential for significant carbon fixation. The increase in the relative contribution of nitrifier taxa *Nitrosospira* and Ca. *Nitrotoga* in aphotic waters is unsurprising, given the well documented increase of nitrifiers with depth (Podowski et al. 2021; Small et al. 2013). While we do not see this dramatic increase in Lake Superior, this weaker signal is due to the fact that ammonia oxidizer *Nitrosarchaeum* and nitrite oxidizer *Nitrospira*, both of which are abundant in Lake Superior (Mukherjee et al. 2016; Podowski et al. 2021), fix carbon using the 3-hydroxypropionate/4-hydroxybutyrate (Konneke et al. 2014) and rTCA (Koch et al. 2015) pathways, respectively. An increase in *rbcL* relative abundance overall, rather than just an increase in the contribution of nitrifiers to *rbcL* is surprising. However, this could be an artifact of relative abundance, as absolute abundance of cells does decrease with depth in Lake Erie and Ontario (Paver, Newton, and Coleman 2020). Further, if most of the total *rbcL* in the surface water of these lower lakes was in phytoplankton cells larger than 1.6 microns, this may obscure the overall pattern as well. Given high relative abundance of nitrifiers (Podowski et al. 2021) and correspondingly high nitrification activity (Lean and Knowles 1987; Clevinger, Heath, and Bade 2014; Hoffman 2020) (Chapter 2), nitrifiers may make significant but unquantified contributions to carbon fixation budgets, especially as carbon fixation in the aphotic zone of some lakes has been shown to reach 87% of euphotic zone carbon fixation (Callieri et al. 2014).

Apparent metabolic flexibility of *rbcLS* containing Burkholderiales presents several distinct possible lifestyles. *rbcLS* containing MAGs presented here could be similar to well described anaerobic anoxygenic phototrophs (AnAPs), such as *Rhodoferax antarcticus* (Baker et al. 2017). This is especially the case for cluster 110 and 155 which contain terminal oxidases associated with low oxygen growth (Morris and Schmidt 2013). However presence of oxygen dependent ring cyclase *acsF* and lack of oxygen independent ring cyclase *bchE* suggests aerobic anoxygenic phototrophy (AAP) (Baker et al. 2017; G. E. Chen, Canniffe, and Hunter 2017). Aerobic chemoorganotrophic growth is also likely, given a complete Entner-Doudoroff in many clusters, which has been reported in photoheterotrophic AAPs (Tang, Tang, and Blankenship 2011). Aerobic chemoautotrophic growth is also possible, given thiosulfate, sulfide and sulfite oxidation genes present in many clusters, similar to those found in other sulfur oxidizing chemoautotrophs (Harrold et al. 2016; Larimer et al. 2004). Aerobic, rather than anaerobic chemoautotrophy is suggested by the oxygen niche of the form of *rbcL* found in these clusters (Badger and Bek 2008), the presence of low affinity terminal oxidases and the lack of high affinity terminal oxidases in some clusters (Morris and Schmidt 2013). While this also implies that these reduced sulfur compounds are available in oxygenated environments, stimulation of dark carbon fixation by thiosulfate has been demonstrated in oxygenated environments (Tuttle and Jannasch 1977). Careful metabolic study of isolates will be required to better detangle if aerobic anoxygenic phototrophy can power autotrophy, supplement heterotrophy in classic photoheterotrophic fashion, or supplement mixotrophic growth with light energy.

RuBisCO and sulfur oxidation genes in AAP bacteria challenge common understanding of the ecological role of *Limnohabitans* and *Rhodoferax* in freshwater ecosystems. These two genera are extremely common in freshwater systems as detected in amplicon sequencing studies, and are generally regarded as heterotrophs (Newton et al. 2011). While autotrophic isolates from both genera exist (Zeng et al. 2012; Baker et al. 2017) these were considered less likely metabolic modes for members of these genera in low nutrient environments. Photoheterotrophy as a metabolic strategy has been more recently presented as widespread in both *Rhodoferax* (Salka et al. 2011) and *Limnohabitans* (Caliz and Casamayor 2014; Kasalický et al. 2018) but even this would position all microbes in these genera as net sinks of organic carbon. Our findings expand the potential of these common freshwater genera as contributors to organic carbon production, whether by photoautotrophy or chemoautotrophy.

4.6 References

Alfreider, Albin, Victoria Grimus, Martin Luger, Anja Ekblad, Michaela M Salcher, and Monika Summerer. 2018. “Autotrophic Carbon Fixation Strategies Used by Nitrifying Prokaryotes in Freshwater Lakes.” *FEMS Microbiology Ecology* 94 (10). https://doi.org/10.1093/femsec/fiy163.

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz, Leo Lahti, Nicholas J Loman, Anders F Andersson, and Christopher Quince. 2014. “Binning Metagenomic Contigs by Coverage and Composition.” *Nature Methods* 11 (11): 1144–46. https://doi.org/10.1038/nmeth.3103.

Aramaki, Takuya, Romain Blanc-Mathieu, Hisashi Endo, Koichi Ohkubo, Minoru Kanehisa, Susumu Goto, and Hiroyuki Ogata. 2020. “KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold.” *Bioinformatics* 36 (7): 2251–52. https://doi.org/10.1093/bioinformatics/btz859.

Arora-Williams, Keith, Scott W. Olesen, Benjamin P. Scandella, Kyle Delwiche, Sarah J. Spencer, Elise M. Myers, Sonali Abraham, Alyssa Sooklal, and Sarah P. Preheim. 2018. “Dynamics of Microbial Populations Mediating Biogeochemical Cycling in a Freshwater Lake.” *Microbiome* 6 (1). https://doi.org/10.1186/s40168-018-0556-7.

Badger, Murray Ronald, and Emily Jane Bek. 2008. “Multiple Rubisco Forms in Proteobacteria: Their Functional Significance in Relation to CO2 Acquisition by the CBB Cycle.” *Journal of Experimental Botany* 59 (7): 1525–41. https://doi.org/10.1093/jxb/erm297.

Baker, Jennifer M., Carli J. Riester, Blair M. Skinner, Austin W. Newell, Wesley D. Swingley, Michael T. Madigan, Deborah O. Jung, et al. 2017. “Genome Sequence of Rhodoferax Antarcticus ANT.BRT; A Psychrophilic Purple Nonsulfur Bacterium from an Antarctic Microbial Mat.” *Microorganisms* 5 (1): 8. https://doi.org/10.3390/microorganisms5010008.

Baltar, Federico, and Gerhard J. Herndl. 2019. “Ideas and Perspectives: Is Dark Carbon Fixation Relevant for Oceanic Primary Production Estimates?” *Biogeosciences* 16 (19): 3793–99. https://doi.org/10.5194/bg-16-3793-2019.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Benedict, Matthew N, James R Henriksen, William W Metcalf, Rachel J Whitaker, and Nathan D Price. 2014. “ITEP: An Integrated Toolkit for Exploration of Microbial Pan-Genomes.” *BMC Genomics* 15 (1): 8. https://doi.org/10.1186/1471-2164-15-8.

Bianchi, Daniele, Thomas S. Weber, Rainer Kiko, and Curtis Deutsch. 2018. “Global Niche of Marine Anaerobic Metabolisms Expanded by Particle Microenvironments.” *Nature*

Geoscience 11 (4): 263–68. https://doi.org/10.1038/s41561-018-0081-0.

Biddanda, Bopaiah, Megan Ogdahl, and James Cotner. 2001. "Dominance of Bacterial Metabolism in Oligotrophic Relative to Eutrophic Waters." *Limnology and Oceanography* 46 (3): 730–39. https://doi.org/10.4319/lo.2001.46.3.0730.

Boddicker, Andrew M., and Annika C. Mosier. 2018. "Genomic Profiling of Four Cultivated Candidatus Nitrotoga Spp. Predicts Broad Metabolic Potential and Environmental Distribution." *The ISME Journal* 12 (12): 2864–82. https://doi.org/10.1038/s41396-018-0240-8.

Boughanemi, Souhela, Pascale Infossi, Marie-Thérèse Giudici-Orticoni, Barbara Schoepp-Cothenet, and Marianne Guiral. 2020. "Sulfite Oxidation by the Quinone-Reducing Molybdenum Sulfite Dehydrogenase SoeABC from the Bacterium Aquifex Aeolicus." *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1861 (11): 148279. https://doi.org/10.1016/j.bbabio.2020.148279.

Buchfink, Benjamin, Klaus Reuter, and Hajk-Georg Drost. 2021. "Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND." *Nature Methods* 18 (4): 366–68. https://doi.org/10.1038/s41592-021-01101-x.

Buchfink, Benjamin, Chao Xie, and Daniel H Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." *Nature Methods* 12 (1): 59–60. https://doi.org/10.1038/nmeth.3176.

Buck, Moritz, Sarahi L. Garcia, Leyden Fernandez, Gaëtan Martin, Gustavo A. Martinez-Rodriguez, Jatta Saarenheimo, Jakob Zopfi, Stefan Bertilsson, and Sari Peura. 2021. "Comprehensive Dataset of Shotgun Metagenomes from Oxygen Stratified Freshwater Lakes and Ponds." *Scientific Data* 8 (1): 131. https://doi.org/10.1038/s41597-021-00910-1.

Bushnell, Brian, Jonathan Rood, and Esther Singer. 2017. "BBMerge – Accurate Paired Shotgun Read Merging via Overlap." Edited by Patrick Jon Biggs. *PLOS ONE* 12 (10): e0185056. https://doi.org/10.1371/journal.pone.0185056.

Cabello-Yeves, Pedro J., Tamara I. Zemskaya, Riccardo Rosselli, Felipe H. Coutinho, Alexandra S. Zakharenko, Vadim V. Blinov, and Francisco Rodriguez-Valera. 2018. "Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal." Edited by Harold L. Drake. *Applied and Environmental Microbiology* 84 (1). https://doi.org/10.1128/AEM.02132-17.

Cabello-Yeves, Pedro J., Tamara I. Zemskaya, Alexandra S. Zakharenko, Mariya V. Sakirko, Vyacheslav G. Ivanov, Rohit Ghai, and Francisco Rodriguez-Valera. 2019. "Microbiome of the Deep Lake Baikal, a Unique Oxic Bathypelagic Habitat." *Limnology and Oceanography*.

Caliz, Joan, and Emilio O. Casamayor. 2014. “Environmental Controls and Composition of Anoxygenic Photoheterotrophs in Ultraoligotrophic High-Altitude Lakes (Central Pyrenees).” *Environmental Microbiology Reports* 6 (2): 145–51. https://doi.org/10.1111/1758-2229.12142.

Callbeck, Cameron M., Gaute Lavik, Timothy G. Ferdelman, Bernhard Fuchs, Harald R. Gruber-Vodicka, Philipp F. Hach, Sten Littmann, et al. 2018. “Oxygen Minimum Zone Cryptic Sulfur Cycling Sustained by Offshore Transport of Key Sulfur Oxidizing Bacteria.” *Nature Communications* 9 (1): 1729. https://doi.org/10.1038/s41467-018-04041-x.

Callieri, Cristiana, Manuela Coci, Ester M. Eckert, Michaela M. Salcher, and Roberto Bertoni. 2014. “Archaea and Bacteria in Deep Lake Hypolimnion: In Situ Dark Inorganic Carbon Uptake.” *Journal of Limnology* 73 (1). https://doi.org/10.4081/jlimnol.2014.937.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10 (1): 421. https://doi.org/10.1186/1471-2105-10-421.

Carr, Mary-Elena, Marjorie A. M. Friedrichs, Marjorie Schmeltz, Maki Noguchi Aita, David Antoine, Kevin R. Arrigo, Ichio Asanuma, et al. 2006. “A Comparison of Global Estimates of Marine Primary Production from Ocean Color.” *Deep Sea Research Part II: Topical Studies in Oceanography*, The US JGOFS Synthesis and Modeling Project: Phase III, 53 (5): 741–70. https://doi.org/10.1016/j.dsr2.2006.01.028.

Castresana, J. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution* 17 (4): 540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chaumeil, Pierre-Alain, Aaron J Mussig, Philip Hugenholtz, and Donovan H Parks. 2020. “GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database.” *Bioinformatics* 36 (6): 1925–27. https://doi.org/10.1093/bioinformatics/btz848.

Chen, Guangyu E., Daniel P. Canniffe, and C. Neil Hunter. 2017. “Three Classes of Oxygen-Dependent Cyclase Involved in Chlorophyll and Bacteriochlorophyll Biosynthesis.” *Proceedings of the National Academy of Sciences* 114 (24): 6280–85. https://doi.org/10.1073/pnas.1701687114.

Chen, Xi, Karoline Schreiber, Jens Appel, Alexander Makowka, Berit Fähnrich, Mayo Roettger, Mohammad R. Hajirezaei, et al. 2016. “The Entner–Doudoroff Pathway Is an Overlooked Glycolytic Route in Cyanobacteria and Plants.” *Proceedings of the National Academy of Sciences* 113 (19): 5441–46. https://doi.org/10.1073/pnas.1521916113.

Clevinger, Curtis C., Robert T. Heath, and Darren L. Bade. 2014. “Oxygen Use by Nitrification in the Hypolimnion and Sediments of Lake Erie.” *Journal of Great Lakes Research* 40 (1): 202–7. https://doi.org/10.1016/j.jglr.2013.09.015.

Delmont, Tom O., and A. Murat Eren. 2018. “Linking Pangenomes and Metagenomes: The Prochlorococcus Metapangenome.” *PeerJ* 6 (January): e4320. https://doi.org/10.7717/peerj.4320.

D’mello, Rita, Susan Hill, and Robert K. Poole. 1996. “The Cytochrome Bd Quinol Oxidase in Escherichia Coli Has an Extremely High Oxygen Affinity and Two Oxygen-Binding Haems: Implications for Regulation of Activity in Vivo by Oxygen Inhibition.” *Microbiology (Reading, England)* 142 (Pt 4) (April): 755–63. https://doi.org/10.1099/00221287-142-4-755.

Dolan, Stephen K., and Martin Welch. 2018. “The Glyoxylate Shunt, 60 Years On.” *Annual Review of Microbiology* 72 (1): 309–30. https://doi.org/10.1146/annurev-micro-090817-062257.

Dongen, Stijn van, and Cei Abreu-Goodger. 2012. “Using MCL to Extract Clusters from Networks.” *Methods in Molecular Biology (Clifton, N.J.)* 804: 281–95. https://doi.org/10.1007/978-1-61779-361-5_15.

Edgar, R. C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. 2015. “Anvi’o: An Advanced Analysis and Visualization Platform for ‘omics Data.” *PeerJ* 3 (October): e1319. https://doi.org/10.7717/peerj.1319.

Fahnenstiel, Gary L., Linda Sicko-Goad, Donald Scavia, and Eugene F. Stoermer. 1986. “Importance of Picoplankton in Lake Superior.” *Canadian Journal of Fisheries and Aquatic Sciences* 43 (1): 235–40. https://doi.org/10.1139/f86-028.

Finn, Robert D., Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, et al. 2014. “Pfam: The Protein Families Database.” *Nucleic Acids Research* 42 (Database issue): D222–30. https://doi.org/10.1093/nar/gkt1223.

Glooschenko, Walter A., James E. Moore, Mohiuddin Munawar, and R. A. Vollenweider. 1974. “Primary Production in Lakes Ontario and Erie: A Comparative Study.” *Journal of the Fisheries Research Board of Canada* 31 (3): 253–63. https://doi.org/10.1139/f74-045.

Graham, E. D., J. F. Heidelberg, and B. J. Tully. 2018. “Potential for Primary Productivity in a Globally-Distributed Bacterial Phototroph.” *The ISME Journal* 12 (7): 1861–66. https://doi.org/10.1038/s41396-018-0091-3.

Graham, Elaina D., John F. Heidelberg, and Benjamin J. Tully. 2017. “BinSanity: Unsupervised Clustering of Environmental Microbial Assemblies Using Coverage and Affinity Propagation.” *PeerJ* 5 (March): e3035. https://doi.org/10.7717/peerj.3035.

Harrold, Zoë R., Mark L. Skidmore, Trinity L. Hamilton, Libby Desch, Kirina Amada, Will van Gelder, Kevin Glover, Eric E. Roden, and Eric S. Boyd. 2016. "Aerobic and Anaerobic Thiosulfate Oxidation by a Cold-Adapted, Subglacial Chemoautotroph." Edited by G. Voordouw. *Applied and Environmental Microbiology* 82 (5): 1486–95. https://doi.org/10.1128/AEM.03398-15.

Heath, Robert, S-J Hwang, and M. Munawar. 2010. "A Hypothesis for the Assessment of the Importance of Microbial Food Web Linkages in Nearshore and Offshore Habitats of the Laurentian Great Lakes." *Aquatic Ecosystem Health & Management* 6 (November): 231–39. https://doi.org/10.1080/14634980301495.

Henson, Michael W., Megan E. Guidry, M. Katherine Carnes, and J. Cameron Thrash. 2020. "Draft Genome Sequence of the Novel Coastal Bacterium LSUCC0115 from the MWH-UniPo Clade, Order Burkholderiales, Class Betaproteobacteria." *Microbiology Resource Announcements* 9 (2). https://doi.org/10.1128/MRA.01492-19.

Hoffman, Daniel K. 2020. "The Fate of Anthropogenic Nitrogen Along Hydrologic Continuums: Patterns of Transformation and Recycling in a Eutrophic Lake and Coastal Marine Sediments." Wright State University. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=wright1598886275031918.

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (1). https://doi.org/10.1186/1471-2105-11-119.

Ivanikova, Natalia V., Audrey R. Matteson, Christopher J. Gobler, Linda C. Popels, Michael L. McKay, Steven W. Wilhelm, and George S. Bullerjahn. 2008. "Picoplanktonic Cyanobacteria in Lakes Superior and Erie: Phylogenies of Endemic Populations and Cultured Isolates." *SIL Proceedings, 1922-2010* 30 (3): 459–65. https://doi.org/10.1080/03680770.2008.11902166.

Ivanikova, Natalia V., Linda C. Popels, R. Michael L. McKay, and George S. Bullerjahn. 2007. "Lake Superior Supports Novel Clusters of Cyanobacterial Picoplankton." *Applied and Environmental Microbiology* 73 (12): 4055–65. https://doi.org/10.1128/AEM.00214-07.

Jain, Chirag, Luis M. Rodriguez-R, Adam M. Phillippy, Konstantinos T. Konstantinidis, and Srinivas Aluru. 2018. "High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries." *Nature Communications* 9 (1): 1–8. https://doi.org/10.1038/s41467-018-07641-9.

Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong Wang. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies." *PeerJ* 7 (July): e7359. https://doi.org/10.7717/peerj.7359.

Kasalický, Vojtěch, Yonghui Zeng, Kasia Piwosz, Karel Šimek, Hana Kratochvilová, and Michal Koblížek. 2018. "Aerobic Anoxygenic Photosynthesis Is Commonly Present within the Genus Limnohabitans." Edited by Claire Vieille. *Applied and Environmental Microbiology* 84 (1). https://doi.org/10.1128/AEM.02116-17.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30 (14): 3059–66. https://doi.org/10.1093/nar/gkf436.

Kitzinger, Katharina, Hanna Koch, Sebastian Lücker, Christopher J. Sedlacek, Craig Herbold, Jasmin Schwarz, Anne Daebeler, et al. 2018. "Characterization of the First 'Candidatus Nitrotoga' Isolate Reveals Metabolic Versatility and Separate Evolution of Widespread Nitrite-Oxidizing Bacteria." *MBio* 9 (4). https://doi.org/10.1128/mBio.01186-18.

Kitzinger, Katharina, Hannah K. Marchant, Laura A. Bristow, Craig W. Herbold, Cory C. Padilla, Abiel T. Kidane, Sten Littmann, et al. 2020. "Single Cell Analyses Reveal Contrasting Life Strategies of the Two Main Nitrifiers in the Ocean." *Nature Communications* 11 (1): 767. https://doi.org/10.1038/s41467-020-14542-3.

Koch, Hanna, Sebastian Lücker, Mads Albertsen, Katharina Kitzinger, Craig Herbold, Eva Spieck, Per Halkjaer Nielsen, Michael Wagner, and Holger Daims. 2015. "Expanded Metabolic Versatility of Ubiquitous Nitrite-Oxidizing Bacteria from the Genus *Nitrospira*." *Proceedings of the National Academy of Sciences* 112 (36): 11371–76. https://doi.org/10.1073/pnas.1506533112.

Konneke, M., D. M. Schubert, P. C. Brown, M. Hugler, S. Standfest, T. Schwander, L. Schada von Borzyskowski, T. J. Erb, D. A. Stahl, and I. A. Berg. 2014. "Ammonia-Oxidizing Archaea Use the Most Energy-Efficient Aerobic Pathway for CO2 Fixation." *Proceedings of the National Academy of Sciences* 111 (22): 8239–44. https://doi.org/10.1073/pnas.1402028111.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Larimer, Frank W., Patrick Chain, Loren Hauser, Jane Lamerdin, Stephanie Malfatti, Long Do, Miriam L. Land, et al. 2004. "Complete Genome Sequence of the Metabolically Versatile Photosynthetic Bacterium Rhodopseudomonas Palustris." *Nature Biotechnology* 22 (1): 55–61. https://doi.org/10.1038/nbt923.

Lean, D. R. S., and R. Knowles. 1987. "Nitrogen Transformations in Lake Ontario." *Canadian Journal of Fisheries and Aquatic Sciences* 44 (12): 2133–43. https://doi.org/10.1139/f87-262.

Linz, Alexandra M., Shaomei He, Sarah L. R. Stevens, Karthik Anantharaman, Robin R.

Rohwer, Rex R. Malmstrom, Stefan Bertilsson, and Katherine D. McMahon. 2018. “Freshwater Carbon and Nutrient Cycles Revealed through Reconstructed Population Genomes.” *PeerJ* 6 (December): e6075. https://doi.org/10.7717/peerj.6075.

Masín, Michal, Jirí Nedoma, Libor Pechar, and Michal Koblízek. 2008. “Distribution of Aerobic Anoxygenic Phototrophs in Temperate Freshwater Systems.” *Environmental Microbiology* 10 (8): 1988–96. https://doi.org/10.1111/j.1462-2920.2008.01615.x.

Mistry, Jaina, Robert D. Finn, Sean R. Eddy, Alex Bateman, and Marco Punta. 2013. “Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions.” *Nucleic Acids Research* 41 (12): e121–e121. https://doi.org/10.1093/nar/gkt263.

Morris, Rachel L., and Thomas M. Schmidt. 2013. “Shallow Breathing: Bacterial Life at Low O2.” *Nature Reviews Microbiology* 11 (3): 205–12. https://doi.org/10.1038/nrmicro2970.

Mukherjee, Maitreyee, Anirban Ray, Anton F. Post, Robert Michael McKay, and George S. Bullerjahn. 2016. “Identification, Enumeration and Diversity of Nitrifying Planktonic Archaea and Bacteria in Trophic End Members of the Laurentian Great Lakes.” *Journal of Great Lakes Research* 42 (1): 39–49. https://doi.org/10.1016/j.jglr.2015.11.007.

Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. “A Guide to the Natural History of Freshwater Lake Bacteria.” *Microbiology and Molecular Biology Reviews* 75 (1): 14–49. https://doi.org/10.1128/MMBR.00028-10.

Norton, Jeanette M., Martin G. Klotz, Lisa Y. Stein, Daniel J. Arp, Peter J. Bottomley, Patrick S. G. Chain, Loren J. Hauser, et al. 2008. “Complete Genome Sequence of Nitrosospira Multiformis, an Ammonia-Oxidizing Bacterium from the Soil Environment.” *Appl. Environ. Microbiol.* 74 (11): 3559–72. https://doi.org/10.1128/AEM.02722-07.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. “MetaSPAdes: A New Versatile Metagenomic Assembler.” *Genome Research* 27 (5): 824–34. https://doi.org/10.1101/gr.213959.116.

Okazaki, Yusuke, Yosuke Nishimura, Takashi Yoshida, Hiroyuki Ogata, and Shin-Ichi Nakano. 2019. “Genome-Resolved Viral and Cellular Metagenomes Revealed Potential Key Virus-Host Interactions in a Deep Freshwater Lake.” *Environmental Microbiology*, October. https://doi.org/10.1111/1462-2920.14816.

O’Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation.” *Nucleic Acids Research* 44 (D1): D733–45. https://doi.org/10.1093/nar/gkv1189.

Ondov, Brian D, Nicholas H Bergman, and Adam M Phillippy. 2011. “Interactive Metagenomic Visualization in a Web Browser.” *BMC Bioinformatics* 12 (1).

https://doi.org/10.1186/1471-2105-12-385.

Pachiadaki, Maria G., Eva Sintes, Kristin Bergauer, Julia M. Brown, Nicholas R. Record, Brandon K. Swan, Mary Elizabeth Mathyer, et al. 2017. “Major Role of Nitrite-Oxidizing Bacteria in Dark Ocean Carbon Fixation.” *Science* 358 (6366): 1046–51. https://doi.org/10.1126/science.aan8260.

Parks, Donovan H, Maria Chuvochina, Christian Rinke, Aaron J Mussig, Pierre-Alain Chaumeil, and Philip Hugenholtz. 2022. “GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy.” *Nucleic Acids Research* 50 (D1): D785–94. https://doi.org/10.1093/nar/gkab776.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55. https://doi.org/10.1101/gr.186072.114.

Paver, Sara F., Ryan J. Newton, and Maureen L. Coleman. 2020. “Microbial Communities of the Laurentian Great Lakes Reflect Connectivity and Local Biogeochemistry.” *Environmental Microbiology* 22 (1): 433–46. https://doi.org/10.1111/1462-2920.14862.

Peura, Sari, Moritz Buck, Sanni L Aalto, Sergio E Morales, Hannu Nykänen, and Alexander Eiler. 2018. “Novel Autotrophic Organisms Contribute Significantly to the Internal Carbon Cycling Potential of a Boreal Lake” 9 (4): 13.

Podowski, Justin C., Sara F. Paver, Ryan J. Newton, and Maureen L. Coleman. 2021. “Genome Streamlining, Proteorhodopsin, and Organic Nitrogen Metabolism in Freshwater Nitrifiers.” bioRxiv. https://doi.org/10.1101/2021.01.19.427344.

Preisig, O., R. Zufferey, L. Thöny-Meyer, C. A. Appleby, and H. Hennecke. 1996. “A High-Affinity Cbb3-Type Cytochrome Oxidase Terminates the Symbiosis-Specific Respiratory Chain of Bradyrhizobium Japonicum.” *Journal of Bacteriology* 178 (6): 1532–38. https://doi.org/10.1128/jb.178.6.1532-1538.1996.

Props, Ruben, and Vincent J. Denef. 2020. “Temperature and Nutrient Levels Correspond with Lineage-Specific Microdiversification in the Ubiquitous and Abundant Freshwater Genus Limnohabitans.” *Applied and Environmental Microbiology*, March. https://doi.org/10.1128/AEM.00140-20.

Reavie, Euan D., Meijun Cai, Carsten Meyer-Jacob, John P. Smol, and Josef P. Werne. 2021. “Long-Term Primary Production Trends in the Laurentian Great Lakes: A Comparison of Geochemical Methods.” *Journal of Paleolimnology* 65 (3): 299–314. https://doi.org/10.1007/s10933-020-00170-w.

Reinthaler, Thomas, Hendrik M. van Aken, and Gerhard J. Herndl. 2010. “Major Contribution of

Autotrophy to Microbial Carbon Cycling in the Deep North Atlantic's Interior." *Deep Sea Research Part II: Topical Studies in Oceanography* 57 (16): 1572–80. https://doi.org/10.1016/j.dsr2.2010.02.023.

Rodriguez-R, Luis M., and Konstantinos T. Konstantinidis. 2016. "The Enveomics Collection: A Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes." e1900v1. PeerJ Inc. https://doi.org/10.7287/peerj.preprints.1900v1.

Rozmarynowycz, Mark J., Benjamin F. N. Beall, George S. Bullerjahn, Gaston E. Small, Robert W. Sterner, Sandra S. Brovold, Nigel A. D'souza, Susan B. Watson, and Robert Michael L. McKay. 2019. "Transitions in Microbial Communities along a 1600 km Freshwater Trophic Gradient." *Journal of Great Lakes Research* 45 (2): 263–76. https://doi.org/10.1016/j.jglr.2019.01.004.

Salka, Ivette, Zuzana Čuperová, Michal Mašín, Michal Koblížek, and Hans-Peter Grossart. 2011. "Rhodoferax-Related PufM Gene Cluster Dominates the Aerobic Anoxygenic Phototrophic Communities in German Freshwater Lakes." *Environmental Microbiology* 13 (11): 2865–75. https://doi.org/10.1111/j.1462-2920.2011.02562.x.

Scavia, Donald, J. David Allan, Kristin K. Arend, Steven Bartell, Dmitry Beletsky, Nate S. Bosch, Stephen B. Brandt, et al. 2014. "Assessing and Addressing the Re-Eutrophication of Lake Erie: Central Basin Hypoxia." *Journal of Great Lakes Research* 40 (2): 226–46. https://doi.org/10.1016/j.jglr.2014.02.004.

Shah, Vega, Xiaowei Zhao, Rachel A. Lundeen, Anitra E. Ingalls, Daniela Nicastro, and Robert M. Morris. 2019. "Morphological Plasticity in a Sulfur-Oxidizing Marine Bacterium from the SUP05 Clade Enhances Dark Carbon Fixation." *MBio*, May. https://doi.org/10.1128/mBio.00216-19.

Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess, Susannah G. Tringe, and Jillian F. Banfield. 2018. "Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy." *Nature Microbiology* 3 (7): 836–43. https://doi.org/10.1038/s41564-018-0171-1.

Small, Gaston E., George S. Bullerjahn, Robert W. Sterner, Benjamin F. N. Beall, Sandra Brovold, Jacques C. Finlay, Robert M. L. McKay, and Maitreyee Mukherjee. 2013. "Rates and Controls of Nitrification in a Large Oligotrophic Lake." *Limnology and Oceanography* 58 (1): 276–86. https://doi.org/10.4319/lo.2013.58.1.0276.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Sterner, Robert W. 2010. "In Situ-Measured Primary Production in Lake Superior." *Journal of Great Lakes Research* 36 (1): 139–49. https://doi.org/10.1016/j.jglr.2009.12.007.

Swan, B. K., M. Martinez-Garcia, C. M. Preston, A. Sczyrba, T. Woyke, D. Lamy, T. Reinthaler, et al. 2011. “Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean.” *Science* 333 (6047): 1296–1300. https://doi.org/10.1126/science.1203690.

Tabita, F. R., S. Satagopan, T. E. Hanson, N. E. Kreel, and S. S. Scott. 2007. “Distinct Form I, II, III, and IV Rubisco Proteins from the Three Kingdoms of Life Provide Clues about Rubisco Evolution and Structure/Function Relationships.” *Journal of Experimental Botany* 59 (7): 1515–24. https://doi.org/10.1093/jxb/erm361.

Tabita, F. Robert, Thomas E. Hanson, Huiying Li, Sriram Satagopan, Jaya Singh, and Sum Chan. 2007. “Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs.” *Microbiology and Molecular Biology Reviews : MMBR* 71 (4): 576–99. https://doi.org/10.1128/MMBR.00015-07.

Tang, Kuo-Hsiang, Yinjie Tang, and Robert Blankenship. 2011. “Carbon Metabolic Pathways in Phototrophic Bacteria and Their Broader Evolutionary Implications.” *Frontiers in Microbiology* 2. https://www.frontiersin.org/article/10.3389/fmicb.2011.00165.

Tatusov, Roman L., Michael Y. Galperin, Darren A. Natale, and Eugene V. Koonin. 2000. “The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution.” *Nucleic Acids Research* 28 (1): 33–36.

Tellier, Joshua M., Nicholas I. Kalejs, Benjamin S. Leonhardt, David Cannon, Tomas O. Höök, and Paris D. Collingsworth. 2021. “Widespread Prevalence of Hypoxia and the Classification of Hypoxic Conditions in the Laurentian Great Lakes.” *Journal of Great Lakes Research*, November. https://doi.org/10.1016/j.jglr.2021.11.004.

Tran, Patricia Q., Samantha C. Bachand, Peter B. McIntyre, Benjamin M. Kraemer, Yvonne Vadeboncoeur, Ismael A. Kimirei, Rashid Tamatamah, Katherine D. McMahon, and Karthik Anantharaman. 2021. “Depth-Discrete Metagenomics Reveals the Roles of Microbes in Biogeochemical Cycling in the Tropical Freshwater Lake Tanganyika.” *The ISME Journal* 15 (7): 1971–86. https://doi.org/10.1038/s41396-021-00898-x.

Tuttle, Jon H., and Holger W. Jannasch. 1977. “Thiosulfate Stimulation of Microbial Dark Assimilation of Carbon Dioxide in Shallow Marine Waters.” *Microbial Ecology* 4 (1): 9–25. https://doi.org/10.1007/BF02010426.

Vargas Asensio, Juan Gabriel. 2020. “Molecular and Environmental Controls on Aerobic Anoxygenic Phototrophy.” Ph.D., United States -- Illinois: The University of Chicago. https://www.proquest.com/pqdtlocal1006268/docview/2447594181/abstract/5A6C94D636144EDBPQ/5.

Vollenweider, R. A., M. Munawar, and P. Stadelmann. 1974. “A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes.” *Journal of the Fisheries Research Board of Canada* 31 (5): 739–62. https://doi.org/10.1139/f74-100.

Wilson, Wayne A., Peter J. Roach, Manuel Montero, Edurne Baroja-Fernández, Francisco José Muñoz, Gustavo Eydallin, Alejandro M. Viale, and Javier Pozueta-Romero. 2010. “Regulation of Glycogen Metabolism in Yeast and Bacteria.” *FEMS Microbiology Reviews* 34 (6): 952–85. https://doi.org/10.1111/j.1574-6976.2010.00220.x.

Zeng, Yonghui, Vojtech Kasalický, Karel Šimek, and Michal Koblížeka. 2012. “Genome Sequences of Two Freshwater Betaproteobacterial Isolates, Limnohabitans Species Strains Rim28 and Rim47, Indicate Their Capabilities as Both Photoautotrophs and Ammonia Oxidizers.” *Journal of Bacteriology* 194 (22): 6302–3. https://doi.org/10.1128/JB.01481-12.

Zhou, Zhengzhen, Laodong Guo, and Elizabeth Minor. 2016. “Characterization of Bulk and Chromophoric Dissolved Organic Matter in the Laurentian Great Lakes during Summer 2013.” *Journal of Great Lakes Research* 42 (April). https://doi.org/10.1016/j.jglr.2016.04.006.

Chapter 5
Conclusion and Future Directions

This work has expanded our knowledge of the nitrifiers and nitrification in freshwater systems, and of the phylogenetic and metabolic diversity of autotrophs in freshwater systems. We have demonstrated that nitrifiers in the Great Lakes are streamlined and contain adaptations for use of organic sources of nitrogen, and in some cases novel adaptations for use of light energy. We showed that rates of nitrification are distinct across the Great Lakes, and that nitrification in Lake Erie is considerably higher than the other Great Lakes, and more closely resembles smaller, more productive lakes. Further, we demonstrated that nitrification is largely carried out by free living nitrifiers defined by our genomic survey. In cases where nitrification is carried out by larger or particle associated nitrifiers, the taxonomy of those nitrifiers was not distinct from those that are free living. Finally, we demonstrated that phylogenetically and metabolically diverse autotrophs in Burkholderiales are abundant in the Great Lakes and other freshwater systems, and that either reduced sulfur or sunlight energy may power carbon fixation in these microbes. Broadly, we investigated microbial communities in the Great Lakes and expand the diversity of known contributions to biogeochemical cycles, and demonstrated evolutionary adaptations to this unique ecosystem.

5.1 Conclusions

Using metagenomic sequencing of microbial communities in the Great Lakes, we established the dominant taxa of nitrifiers in the Great Lakes, and their distribution within the water column (Chapter 2). We find that in stratified water columns Through multiple years of sequencing, and through sequencing of both free living (0.2-1.6 micron) and whole water column (0.2-25 microns) size fractions, we establish the taxa of nitrifiers which reach high abundance in the Great Lakes. Ammonia oxidation was dominated by ammonia oxidizing bacteria in *Nitrosospira*, with a smaller contribution from ammonia oxidizing archaea in *Nitrosarchaeum* in Michigan and Superior. Nitrite oxidizers in *Nitrospira* and Ca. *Nitrotoga* were largely found to be numerically dominant in the Great Lakes, with *Nitrospira* in Lake Superior and some basins in Michigan and Ontario, and Ca. *Nitrotoga* in Huron and Erie, and some basins of Michigan and Ontario (Chapter 2, Chapter 3). This assessment agrees with past sequencing based studies (Paver, Newton, and Coleman 2020; Fujimoto et al. 2016), but disagrees in some key ways with past probe based studies in the Great Lakes (Small et al. 2013; Mukherjee et al. 2016). Disagreements with prior probe based studies can largely be attributed to general issues with targeted probe based approaches- those of designing DNA probes based on isolated microbes which may be dissimilar to microbes in the environment, and the issue that a targeted survey will not consider community members that are not specifically targeted. Overall, our more holistic approach has set the framework for understanding nitrifiers in the Great Lakes.

Metabolic reconstruction of Great Lakes nitrifiers demonstrated adaptations to the Great Lakes, and emphasized the importance of organic nitrogen. Our metabolic reconstruction of nitrifiers from the Great Lakes emphasized several key ways in which nitrifiers appear to have adapted to the Great Lakes. First, generalized adaptations to a low nutrient environment were apparent in all Great Lakes nitrifiers (Chapter 2) among them smaller genomes, a lack of genes involved in responding to changing environmental conditions (chemotaxis, flagella, sigma factors, etc), and a lack of some auxiliary metabolic functions. However, we saw that relative to references, many adaptations to light exist in nitrifiers in the Great Lakes (Chapter 2). We discovered proteorhodopsin, a light activated proton pump that may be involved in harvesting light energy, in an ammonia oxidizing bacteria. This is the first time this widely distributed gene has been found in a nitrifier. We also find an array of genes involved in DNA repair from light damage in both ammonia and nitrite oxidizers. The Great Lakes experience thermal mixing in which nitrifiers that are typically present in the deep water are transported to the surface, despite being largely located in aphotic waters when the water column is stratified. We believe these adaptations are related to this exposure to light, which is not typical for nitrifiers (Auguet et al. 2012; Smith, Chavez, and Francis 2014). Finally, we saw adaptations for use of organic nitrogen in the form of urea and cyanate in Great Lakes nitrifiers- though their distribution suggests these sources are not equally important across the Great Lakes (Chapter 2). Ammonia oxidizing archaea possed urease, as do NspGL1 ammonia oxidizing bacteria. However, these two groups are present in all the Great Lakes except Lake Erie, suggesting urea is less likely to be a source of ammonia for nitrification. We found in fact that this is the case (Chapter 3) that oxidation of urea derived nitrogen is much lower in Lake Erie than elsewhere in the Great Lakes. We also found cyanase in all Great Lakes nitrite oxidizers- though as nitrite oxidizers do not oxidize ammonium, this is likely for use of ammonia to generate amino acids. Together, these results suggested that nitrifiers have evolved to adapt to the unique conditions of the Great Lakes- and some of these adaptations suggest the importance of nitrogen sources not previously considered.

Nitrification rate measurement contextualized the Great Lakes against other aquatic ecosystems. We provided the first cohesive framework by which to compare the Great Lakes to one another in terms of rate of nitrification- likening Lake Superior to the open ocean (Chapter 3, (Santoro, Casciotti, and Francis 2010)), Michigan and Ontario to coastal regions (Tolar et al. 2017) and Lake Erie to high productivity lakes like Lake Taihu (Hampel et al. 2018) or Mendota (Small et al. 2013). While direct measurements of nitrification existed from Lake Superior (Small et al. 2013) and Erie (Hoffman 2020), lack of methodological consistency made direct comparison difficult. In terms of rates of nitrification (Chapter 3), genomic identity of nitrifiers (Chapter 2 and Chapter 3), size fractionation of nitrification (Chapter 3) and relative unimportance of urea to nitrification (Chapter 3), Lake Erie stands out as distinct with respect to nitrogen cycling. Differences in nitrifier taxonomy are likely only a small piece of why nitrogen biogeochemistry differs in Lake Erie- as nitrogen loss through denitrification in the sediment (Small et al. 2014) and likely the water column (Scavia et al. 2014) differentiates Erie, as does productivity (Vollenweider, Munawar, and Stadelmann 1974) and nitrogen fixation (Natwora and Sheik 2021). A better understanding of other components of the nitrogen cycle- both in terms of microbial taxonomy and rate measurement- will likely inform why nitrogen cycling is distinct in Lake Erie.

Discovery of diverse potential autotrophs in Burkholderiales suggests unexpected impact on carbon fixation, potential importance of reduced sulfur. Our discovery of abundant, widespread potential autotrophs in the surface and deep water of the Great Lakes suggests a potential for a source of organic carbon not previously accounted for (Chapter 4). We also find this metabolism distributed widely throughout other freshwater ecosystems (Chapter 4). The metabolic complement of photosystem II, RuBisCo and sulfur oxidation genes had been previously described in these lineages (Zeng et al. 2012, 27; M. W. Henson et al. 2020) and recognized as present in marine systems (Graham, Heidelberg, and Tully 2018). However, our contribution was the recognition of the abundance of microbes with this metabolic complement, the phylogenetic diversity of these microbes, and their widespread distribution throughout the Great Lakes and other freshwater ecosystems. As the isolation of an organism does not necessarily speak to its abundance in an ecosystem (M. Henson 2019), isolation of microbes with photosystem II, RuBisCo and sulfur oxidation genes did not necessarily indicate a meaningfully high abundance of those microbes. Further, while this metabolic complement had been recognized in marine systems (Graham, Heidelberg, and Tully 2018), in marine systems these microbes are of low abundance and are monophyletic. Our discovery of a high abundance of these microbes in the Great Lakes, as well as a broad phylogenetic diversity of microbes with this metabolic complement, suggests not only a potential to contribute to organic carbon availability, but also a complicated history of the acquisition of RuBisCo genes. A better understanding of under what conditions these microbes fix carbon, and whether it is entirely chemoautotrophic from reduced sulfur or potentially photoautotrophic, will inform the degree to which these microbes may impact organic carbon budgets.

Predominance of eukaryotic RuBisCo genes challenges the importance of Synechococcus in the Great Lakes. We find that RuBisCo genes associated with eukaryotic phytoplankton are by far more abundant in the surface and deep chlorophyll layer than RuBisCo genes associated with picocyanobacteria Synechococcus (Chapter 4). This is especially apparent at the level of gene expression, where RuBisCo transcripts associated with eukaryotic phytoplankton are more than 90 times more abundant than RuBisCo transcripts associated with Synechococcus (Chapter 4). A great deal of research has focused on Synechococcus and its contribution to carbon fixation (Ivanikova et al. 2007; 2008; Fahnenstiel and Carrick 2011; Caron, Pick, and Lean 2004).Our focus on ‘free living’ microbes less than 1.6 microns suggests that the source of eukaryotic RuBisCo genes and transcripts must be less than 1.6 microns, eliminating contributions of diatoms and other eukaryotic phytoplankton that are well studied in the Great Lakes (Vollenweider, Munawar, and Stadelmann 1974).

5.2 Future Directions

The future directions of these investigations will likely be concentrated around two methods: the isolation of key microbes identified through DNA sequencing, and the measurement of biogeochemical processes such as carbon fixation. Our results speak to in many cases the potential contributions of a microbe to contribute to biogeochemical cycling in the Great Lakes- and while we specifically measured the rates of nitrification, many other hypothetical interactions remain untested.

Isolation of NspGL1 *Nitrosospira* is key to understanding the role of proteorhodopsin. One of the most interesting discoveries from our genomic survey of nitrifiers was the discovery of proteorhodopsin in NspGL1 *Nitrosospira* (Chapter 2). Based on the sequence of the proteorhodopsin gene, it appears to be a light activated proton pump and not a light sensor (Chapter 2). However, verification of the role of proteorhodopsin in this nitrifier, and under what conditions it is expressed, requires isolation of this microbe. Careful growth experiments, combined with proteomic expression analysis targeted at proteorhodopsin, would help illuminate the role of proteorhodopsin in this nitrifier.

Measurement of carbon fixation by *Limnohabitans* and *Rhodoferax* can constrain contributions to organic carbon. Our identification of microbes that reside in the surface waters of the Great Lakes and appear to be autotrophic though use of either light energy or energy from reduced sulfur imply a potential for contribution to carbon fixation and organic carbon. However the actual contribution of these microbes to carbon fixation is unknown. Combination of radiolabeled bicarbonate uptake microscopy combined with fluorescent DNA probes (Pachiadaki et al. 2017) specific to these taxa would help better constrain the rates of carbon fixation possible by these identified autotrophs. Further, measurement of unamended bulk carbon fixation rates compared with carbon fixation rates after amendment with thiosulfate or other reduced sulfur species (Tuttle and Jannasch 1977) could help constrain the importance of reduced sulfur to carbon fixation in the Great Lakes.

5.3 References

Auguet, Jean-Christophe, Xavier Triadó-Margarit, Natalya Nomokonova, Lluís Camarero, and Emilio O Casamayor. 2012. "Vertical Segregation and Phylogenetic Characterization of Ammonia-Oxidizing Archaea in a Deep Oligotrophic Lake." *The ISME Journal* 6 (9): 1786–97. https://doi.org/10.1038/ismej.2012.33.

Caron, David, Frances Pick, and David Lean. 2004. "Chroococcoid Cyanobacteria in Lake Ontario: Vertical and Seasonal Distributions during 1982." *Journal of Phycology* 21 (October): 171–75. https://doi.org/10.1111/j.0022-3646.1985.00171.x.

Chrétiennot-Dinet, M-J., C. Courties, A. Vaquer, J. Neveux, H. Claustre, J. Lautier, and M. C. Machado. 1995. "A New Marine Picoeucaryote: Ostreococcus Tauri Gen. et Sp. Nov. (Chlorophyta, Prasinophyceae)." *Phycologia* 34 (4): 285–92. https://doi.org/10.2216/i0031-8884-34-4-285.1.

Derelle, Evelyne, Conchita Ferraz, Stephane Rombauts, Pierre Rouzé, Alexandra Z. Worden, Steven Robbens, Frédéric Partensky, et al. 2006. "Genome Analysis of the Smallest Free-Living Eukaryote Ostreococcus Tauri Unveils Many Unique Features." *Proceedings of the National Academy of Sciences* 103 (31): 11647–52. https://doi.org/10.1073/pnas.0604795103.

Fahnenstiel, Gary L., and Hunter J. Carrick. 2011. "Phototrophic Picoplankton in Lakes Huron and Michigan: Abundance, Distribution, Composition, and Contribution to Biomass and Production." *Canadian Journal of Fisheries and Aquatic Sciences*, April. https://doi.org/10.1139/f92-043.

Fujimoto, Masanori, Joann Cavaletto, James R. Liebig, Ann McCarthy, Henry A. Vanderploeg, and Vincent J. Denef. 2016. "Spatiotemporal Distribution of Bacterioplankton Functional Groups along a Freshwater Estuary to Pelagic Gradient in Lake Michigan." *Journal of Great Lakes Research* 42 (5): 1036–48. https://doi.org/10.1016/j.jglr.2016.07.029.

Graham, E. D., J. F. Heidelberg, and B. J. Tully. 2018. "Potential for Primary Productivity in a Globally-Distributed Bacterial Phototroph." *The ISME Journal* 12 (7): 1861–66. https://doi.org/10.1038/s41396-018-0091-3.

Hampel, Justyna J., Mark J. McCarthy, Wayne S. Gardner, Lu Zhang, Hai Xu, Guangwei Zhu, and Silvia E. Newell. 2018. "Nitrification and Ammonium Dynamics in Taihu Lake, China: Seasonal Competition for Ammonium between Nitrifiers and Cyanobacteria." *Biogeosciences* 15 (3): 733–48. https://doi.org/10.5194/bg-15-733-2018.

Henson, Michael. 2019. "High-Throughput Cultivation of Bacterioplankton from the Gulf of Mexico and Genomics of the First Cultured LD12 Representative." *LSU Doctoral Dissertations*, January. https://digitalcommons.lsu.edu/gradschool_dissertations/4796.

Henson, Michael W., Megan E. Guidry, M. Katherine Carnes, and J. Cameron Thrash. 2020. "Draft Genome Sequence of the Novel Coastal Bacterium LSUCC0115 from the MWH-

UniPo Clade, Order Burkholderiales, Class Betaproteobacteria." *Microbiology Resource Announcements* 9 (2). https://doi.org/10.1128/MRA.01492-19.

Hoffman, Daniel K. 2020. "The Fate of Anthropogenic Nitrogen Along Hydrologic Continuums: Patterns of Transformation and Recycling in a Eutrophic Lake and Coastal Marine Sediments." Wright State University. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=wright1598886275031918.

Ivanikova, Natalia V., Audrey R. Matteson, Christopher J. Gobler, Linda C. Popels, Michael L. McKay, Steven W. Wilhelm, and George S. Bullerjahn. 2008. "Picoplanktonic Cyanobacteria in Lakes Superior and Erie: Phylogenies of Endemic Populations and Cultured Isolates." *SIL Proceedings, 1922-2010* 30 (3): 459–65. https://doi.org/10.1080/03680770.2008.11902166.

Ivanikova, Natalia V., Linda C. Popels, R. Michael L. McKay, and George S. Bullerjahn. 2007. "Lake Superior Supports Novel Clusters of Cyanobacterial Picoplankton." *Applied and Environmental Microbiology* 73 (12): 4055–65. https://doi.org/10.1128/AEM.00214-07.

Mukherjee, Maitreyee, Anirban Ray, Anton F. Post, Robert Michael McKay, and George S. Bullerjahn. 2016. "Identification, Enumeration and Diversity of Nitrifying Planktonic Archaea and Bacteria in Trophic End Members of the Laurentian Great Lakes." *Journal of Great Lakes Research* 42 (1): 39–49. https://doi.org/10.1016/j.jglr.2015.11.007.

Natwora, Kaela E., and Cody S. Sheik. 2021. "Assessment of Nitrogen Fixation Rates in the Laurentian Great Lakes." *Journal of Great Lakes Research* 47 (5): 1288–95. https://doi.org/10.1016/j.jglr.2021.07.005.

Pachiadaki, Maria G., Eva Sintes, Kristin Bergauer, Julia M. Brown, Nicholas R. Record, Brandon K. Swan, Mary Elizabeth Mathyer, et al. 2017. "Major Role of Nitrite-Oxidizing Bacteria in Dark Ocean Carbon Fixation." *Science* 358 (6366): 1046–51. https://doi.org/10.1126/science.aan8260.

Paver, Sara F., Ryan J. Newton, and Maureen L. Coleman. 2020. "Microbial Communities of the Laurentian Great Lakes Reflect Connectivity and Local Biogeochemistry." *Environmental Microbiology* 22 (1): 433–46. https://doi.org/10.1111/1462-2920.14862.

Santoro, Alyson E., Karen L. Casciotti, and Christopher A. Francis. 2010. "Activity, Abundance and Diversity of Nitrifying Archaea and Bacteria in the Central California Current." *Environmental Microbiology* 12 (7): 1989–2006. https://doi.org/10.1111/j.1462-2920.2010.02205.x.

Scavia, Donald, J. David Allan, Kristin K. Arend, Steven Bartell, Dmitry Beletsky, Nate S. Bosch, Stephen B. Brandt, et al. 2014. "Assessing and Addressing the Re-Eutrophication of Lake Erie: Central Basin Hypoxia." *Journal of Great Lakes Research* 40 (2): 226–46. https://doi.org/10.1016/j.jglr.2014.02.004.

Small, Gaston E., George S. Bullerjahn, Robert W. Sterner, Benjamin F. N. Beall, Sandra Brovold, Jacques C. Finlay, Robert M. L. McKay, and Maitreyee Mukherjee. 2013. "Rates and Controls of Nitrification in a Large Oligotrophic Lake." *Limnology and Oceanography* 58 (1): 276–86. https://doi.org/10.4319/lo.2013.58.1.0276.

Small, Gaston E., James B. Cotner, Jacques C. Finlay, Rebecca A. Stark, and Robert W. Sterner. 2014. "Nitrogen Transformations at the Sediment–Water Interface across Redox Gradients in the Laurentian Great Lakes." *Hydrobiologia* 731 (1): 95–108. https://doi.org/10.1007/s10750-013-1569-7.

Smith, Jason M., Francisco P. Chavez, and Christopher A. Francis. 2014. "Ammonium Uptake by Phytoplankton Regulates Nitrification in the Sunlit Ocean." Edited by Connie Lovejoy. *PLoS ONE* 9 (9): e108173. https://doi.org/10.1371/journal.pone.0108173.

Tolar, Bradley B., Natalie J. Wallsgrove, Brian N. Popp, and James T. Hollibaugh. 2017. "Oxidation of Urea-Derived Nitrogen by Thaumarchaeota-Dominated Marine Nitrifying Communities." *Environmental Microbiology* 19 (12): 4838–50. https://doi.org/10.1111/1462-2920.13457.

Tuttle, Jon H., and Holger W. Jannasch. 1977. "Thiosulfate Stimulation of Microbial Dark Assimilation of Carbon Dioxide in Shallow Marine Waters." *Microbial Ecology* 4 (1): 9–25. https://doi.org/10.1007/BF02010426.

Vollenweider, R. A., M. Munawar, and P. Stadelmann. 1974. "A Comparative Review of Phytoplankton and Primary Production in the Laurentian Great Lakes." *Journal of the Fisheries Research Board of Canada* 31 (5): 739–62. https://doi.org/10.1139/f74-100.

Zeng, Yonghui, Vojtech Kasalický, Karel Šimek, and Michal Koblížeka. 2012. "Genome Sequences of Two Freshwater Betaproteobacterial Isolates, Limnohabitans Species Strains Rim28 and Rim47, Indicate Their Capabilities as Both Photoautotrophs and Ammonia Oxidizers." *Journal of Bacteriology* 194 (22): 6302–3. https://doi.org/10.1128/JB.01481-12.