Viral Community Dynamics and Implications for the Fate of cHABs

by

A.J Wing

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2024

Doctoral Committee:

Professor Melissa Duhaime, Chair
Professor Gregory Dick
Professor Meghan Duffy
Dr. Reagan Errera
Professor Luis Zaman

Anthony J Wing

ajwing@umich.edu

ORCID: 0009-0008-5056-2550

© Anthony J Wing 2024

Acknowledgements

To Special K: my rock, my partner and the ultimate teammate. You were there during the lowest lows and the highest highs. You loved me, fed me, and supported me in ways no one else could. This progression is yours as much as it is mine. Every second counts, and that's why we always stay until the last second.

Thank you to my father for instilling in me a work ethic that provided the foundations of my journey. Your daily 16-hour sacrifices guided my effort and my ambitions for a better tomorrow. You grew in a time where emotions were best left unshared, but I know I pulled them out of you and I hope we continue to be best friends. To my mother and our shared hardships: you did what you could when you could and always did it from a place of love. A constant reminder to both of us that happiness is homemade.

To my siblings and family members: life tried its best to keep us apart as time passed, but holidays always brought us together as if that time had never passed. It always feels like home when we are together.

To my one-of-a-kind advisor: my sensei. My coach. You took a chance on me as a scientist. You pushed me harder than anyone in my life. I thank you for all the things you did and did not do to get me to this point. You showed me that being a great scientist and a great human do not have to be mutually exclusive endeavors. I mean this from the bottom of my heart: None of this would be possible without your mind, your compassion, and your belief in me. I'll thank you forever.

To my old lifelong friends: Clay, P.J., T., Dan and T.J.; none of you scientists. You reminded me of life outside of science and also reminded me of its importance to the world. How about those lions? True brothers until the end.

To my lab mates and cohort members that struggled through the duration of this process with me, I thank you. In my opinion, shared struggles lead to a life well-lived.

Thank you to my research committee for their continued support through my scientific journey. Your feedback and knowledge kept me rooted in logic and made me a better scientist.

To the Richart conglomerate: From day one, you made me feel welcomed. I was fortunate enough to become one of the family and continue to be grateful that our paths crossed.

Thank you to my new friends made along the way. Sam and Meagan, you kept a roof over my head and an inflatable mattress under it. Mi familia de la Costa Este. Agent Dahal, we began our mission together and overcame so many obstacles early on in our journey. You have, without a doubt, proved to be worthy of the “Best Boy” award. John and Lorraine, you also kept a roof over my head, taught me how to farm, and most importantly, how to grow.

Finally, this work and its many stepping stones would not exist without the hardships I endured to reach this point in my life and career. In many ways, I view the pressure to overcome and succeed as a privilege. My journey serves as a reminder to anyone and everyone that to whom nothing is given, everything is earned. They say money doesn’t buy happiness, and while that may be true, money certainly affords opportunity. To everyone out there surviving the challenges of poverty, keep surviving. Your opportunity is coming.

Table of Contents

Acknowledgements ii

List of Figures v

Abstract x

Chapter 1: Introduction 1

Chapter 2:Tracking *Microcystis* Viruses and Infection Dynamics in a Multi-Peak *Microcystis*-Dominated Bloom 36

Chapter 3: Deciphering Total Viral Community Structure and Metabolic Potential in a Great Lakes Cyanobacterial Harmful Algal Bloom 75

Chapter 4: Evaluating the Relationship Between Viral Infection and Phylogeny of Lake Erie *Microcystis* Isolates 102

Chapter 5: Conclusions and the Road Ahead 133

List of Figures

Figure 1. A satellite image of Lake Erie overlaid on a map of the lake and its tributaries. Initially, the bloom forms in the western basin of Lake Erie before radiating out towards the central basin. Image from Michalek et al., 2013. 2

Figure 2. Lytic and lysogenic phage cycles. Lytic cycle: 1) Virus attaches to the host cell and injects DNA. 2) Phage DNA circularizes and enters the lytic cycle. 3a) New phage DNA and proteins are synthesized and assembled into virions. 4a) Cell lyses, releasing phage virions. Lysogenic cycle: 1) Virus attaches to the host cell and injects DNA. 2) Phage DNA circularizes and enters the lysogenic cycle. 3b) Phage DNA integrates within the bacterial chromosome. 4b) Virocell reproduces normally. 5) On occasion, the prophage excises from the bacterial chromosome and enters the lytic cycle. Image generated in BioRender. 6

Figure 3. Traditional example of rank-abundance curve demonstrating vOTU abundance in a community. The most abundant vOTU is ranked as 1, the next highest is 2, and so on and so forth. According to seed bank theory, only a few of the most abundant vOTUs are in the active fraction at any given point. As new hosts rise in abundance in response to changing environmental conditions, the viral predators of those hosts may also become abundant. The vOTUs that were previously in the active fraction become part of the bank fraction. Adapted from Breitbart, 2005. 9

Figure 4. Three stages of CRISPR-mediated phage immunity. New spacers from the phage genome are integrated into the host CRISPR array (acquisition), separated by palindromic repeats (white rectangles). The array is transcribed and processed by Cas proteins to generate small crRNAs (adaptation). crRNA leads Cas protein complex to matching target to induce cleavage of exogenous nucleic acid (interference). Adapted from Heler et al. 2014. 21

Figure 5. Lake Erie 2014 sampling overview and *Microcystis* MAG diversity and dynamics. (A) A map of sampling sites located in the western basin of Lake Erie. Three sites were sampled bi-monthly in June and weekly from July to October of 2014. (B) Phycocyanin, chlorophyll-a and microcystin measurements for the 2014 bloom. (C) Sampling filter size fractionation schematic of how samples were collected from the western basin of Lake Erie in 2014. (D) Heatmap of average nucleotide identity (% ANI) between 26 *Microcystis* MAGs reconstructed from samples July through August across three stations and four sample fractions. Bottom row and right column list the MAG identifiers. Sample fraction, station and date are identified by color of left panel. Dendrogram depicts the clustering of MAGs based on %ANI. (E) Relative abundances of *Microcystis* MAGs based on fraction of reads from each sample that mapped to each MAG in a competitive read mapping to all assembled MAGs from that sample. No data exist for 4 Aug 100 µm at WLE12 and 53 µm at WLE4. 41

Figure 6. Genome similarity and spatiotemporal distribution of four closely related *Microcystis* viruses found in Lake Erie. (A) Pairwise genome similarity between the Lake Erie Ma-LEF01 and three closely related known *Microcystis* viruses (Ma-LMM01 and MaMV-DC) and viral contig (MVGF-J19). Similarity reported as average nucleotide identity (%) and alignment fraction

(%) relative to the shortest of the two being compared. (B) Genome synteny plot of Ma-LMM01 and Lake Erie Ma-LEF01 and MVGF-J19 sequences. Gene color indicated DRAM gene function annotations for Ma-LEF01. 'Unknown' is assigned when there is a lack of information, which could arise from no sequence hits. 'Unknown functional' is assigned when a sequence has been identified, but its functional role is unidentified or uncharacterized. 'Hypothetical' is assigned when the function is not well-characterized or solely predicted computationally. Gray tracks linking portions of the genome represent nucleotide identity. Red box indicates the sequence portion unique to Lake Erie vOTUs. Red asterisk indicates nblA gene; purple asterisk indicates pentapeptide repeat proteins (C) Circos genome synteny and coverage plots of the four closely related *Microcystis* viruses/contig from panel A. Gray links in the inner circle represent the shared ANI between the viruses. Outer tracks represent read mapping coverage for different fractions on 4 August (left) and 29 September (right) collected at station WLE12. Colored stars on Sept 29 indicate genome regions with spikes in mapped reads. Colors indicate gene clusters with host homologues described in SI Table 3. (D) Strain-resolved population dynamics of the *Microcystis* viruses at two bloom peaks. Distribution of normalized read counts and percent identity when reads from each date and fraction were competitively mapped to Ma-LEF01 and Ma-LMM01 (all viruses mapped in SI Fig. 2-9). Histogram bar color represents the sampling fraction. 44

Figure 7. Networks of predicted infections between *Microcystis* viruses (>10 kb viral contigs) and bacterial host MAGs identified on the 4 Aug and 29 Sept bloom metagenomes. (A) Predicted infection network of 4 August toxic bloom peak. (B) Predicted infection network of 29 September non-toxic bloom peak. Circle nodes are host MAGs; circle size represents the number of viruses predicted to infect a given host. Triangle nodes are viruses. Node size represents TPM abundance. Node colors represent assigned taxonomy; for viruses, "Unknown" indicates the vOTU has no hit in the reference database, "Unassigned" indicates the vOTU has a hit to something unassigned in the reference database. Only predictions with >93% infection probability and viruses predicted to infect at least *Microcystis* are shown. The most abundant vOTUs are labeled in each network (described in SI Table 6 and Fig. 6) (C) Heat map of hosts with shared virus predictions on 4 August. (D) Heat map of hosts with shared virus predictions on 4 August. Axis colors represent host taxonomy (as in panels A-B) and heat map cell color represents the amount of shared phages between any two given hosts. 50

Figure 8. Proportion of vOTUs infecting bloom peak hosts and proportion of abundant viral genes at bloom peaks. (A) Comparison of host ranges. (B) Proportion of viral genes encoded by predicted *Microcystis* vOTUs at 4 August bloom peak (x-axis) and 29 September bloom peak (y-axis). Shape represents whether viral genes were identified in a single bloom peak or both. Point color reflects the assigned KEGG Ontology (KO) metabolic module. Points labels are derived from assigned protein family (Pfam) annotations for select functions of interest. 53

Figure 9. *Microcystis* virus diversity, turnover, and abundant vOTU temporal dynamics through the 2014 Lake Erie bloom season. (A) Diversity of predicted *Microcystis* vOTU assemblage across fractions. (B) NMDS ordination based on Bray-Curtis dissimilarities in distributions of predicted *Microcystis* vOTUs at the cHAB peaks. Point color represents the sampling fraction; shape represents sampling date. Dashed ovals indicate whether a point belongs to the free-living or colony-associated community. The ordinations are overlaid with the gradient of fit between measured environmental parameters and Bray-Curtis dissimilarities. Vector length of the environmental parameter represents the strength of the correlation with the data variation. (C) Temporal relative abundance dynamics of the ten most abundant predicted *Microcystis* vOTUs in terms of fraction of total metagenomic reads mapped to each vOTU per sample. Point

shape represents the sampling station and point color represents a specific *Microcystis* virus OTU. Dates of sampled bloom peaks are highlighted in bold with double asterisk. 56

Figure 10. Lake Erie viral community sampling overview and abundant vOTU taxonomic classification. (A) A map of sampling sites located in the western basin of Lake Erie. Three sites were sampled bi-monthly in June and weekly from July to October of 2014. (B) Sampling filter size fractionation breakdown. Percentage in pink box indicates fraction of total contigs derived from each fraction identified as viral. (C) Scatterplot showing the 148 vOTUs identified as abundant (mapping >0.5% of reads in a given sample) and the relationship between the number of samples where a vOTU was abundant and the number of sampling dates in which that vOTU was found. 79

Figure 11. Lake Erie viral community overview. (A) Shannon diversity of viral community through time and space. (B) Shannon diversity of bacterial community across time and station. Panel A-B color indicates sampling fraction. (C) Non-metric multidimensional scaling (NMDS) (stress-value = 0.14) ordination based on Bray-Curtis dissimilarities in vOTU abundances. (D) NMDS (stress-value = 0.09) ordination based on Bray-Curtis dissimilarities in viral metabolic gene abundances. Panels C-D color represents the sampling date; shape represents sample fraction; ordinations are overlaid with the gradient of the fit between the significant environmental parameters and the Bray-Curtis dissimilarities; vector lengths represent strength of the correlation with the data variation... 82

Figure 12. Lake Erie vOTU dynamics. (A) Tracking abundant vOTUs through time by fraction at site WLE12 (only site where viral fraction was sampled). Each line tracks a single vOTU. If >0.5% of the total reads that mapped to viruses map to a given vOTU on the specified date, it was assigned an abundance score of “1”; if <0.5% viral reads map to the vOTU it was assigned “0.” Sample sizes (*n*) on the y-axis in red font report the number of ‘not abundant’ and ‘abundant’ vOTUs, as well as percentage of abundant vOTUs, in each fraction. (B) Temporal relative abundance dynamics of the thirty most abundant vOTUs across the entire dataset separated by sampling fraction. (C) vOTUs depicted in panel B are listed to the right of the plots in which they rise above the zero line. Conceptual model depicting infection of bacteria in colonies by “Active” viruses that are colony-associated (53 and 100 μm), infection of bacteria by “Active” viruses *not* colony-associated (>0.22 and 3 μm), and free viruses of the Viral “Bank.” 85

Figure 13. Predicted virus-host interactions between bacterial metagenome assembled genomes (MAGs) and vOTUs of the 2014 Lake Erie bloom community. (A) Predicted virus-host network from 4 August bloom peak in the station WLE 12 sample, all fractions combined. Circle nodes are host MAGs; sizes represent the MAG relative abundance in the sample. Triangle nodes are viruses; sizes represent the average relative abundance of the vOTUs across fractions. Node colors represent assigned taxonomy. For vOTUs, taxonomy was assigned according to the Phage Taxonomy Tool (PTT; Kieft et al., 2020); for viruses, “Unknown” indicates the vOTU has no hit in the PTT reference database, “Unassigned” indicates the vOTU has a hit to something unassigned in the reference database, and “Ambiguous” refers to a situation in which a protein or virus had significant hits to multiple proteins within the database, but the program could not distinguish between taxonomic assignments. Only predictions with >93% infection probability are shown. Sept 29 virus-host network in SI Fig. 5. (B) Relationship between host relative abundance and the number of viruses predicted to infect a given host, separated by date. Line and point colors represent filter fractions. Grey shading represents 95% confidence interval of the linear regression... 88

Figure 14. Turnover of Lake Erie cHAB virus-host pairs. (A) Network of predicted virus-host pairs shared between different sampling dates and stations. Black nodes represent predicted virus-host networks of individual samples (individual networks are not visualized here; composite networks for Aug 4 and Sept 29 are in SI Fig. 5). Lines (edges) connecting nodes represent a virus-host pair shared between two samples. Numbers on nodes are the total number of virus-host pairs predicted in a sample. (B) Heatmap depicting the proportion of virus-host pairs shared between different samples scaled by the total number of virus-host pairs in the row sample. Dendrogram depicts hierarchical clustering of samples by proportion of shared virus-host pairs. Color blocks on the right indicate sample date and station. (C) Network of predicted virus-host pairs shared between different sampling dates and fractions. All else as described for panel A. (D) Heatmap depicting the proportion of virus-host pairs shared between different samples scaled by the total number of virus-host pairs in the row sample. Color blocks on the right indicate sample date and fraction. Dashed box indicates a cluster of colony-associated samples from the Lake Erie 2014 bloom peaks. 91

Figure 15. Spearman Rank Correlation Coefficient between *Microcystis* isolate phylogenetic distances and infection profile distances. Each point represents a comparison between two *Microcystis* isolates' phylogenetic distance and infection profile distance. The purple line represents a linear regression and the gray shading represents 95% confidence intervals..... 108

Figure 16. PCOA ordination based on Jaccard dissimilarities in distributions of *Microcystis* isolate infection profiles. Point color represents the collection date; shape represents station sampled... 110

Figure 17. Relative abundance of isolate vOTUs in Lake Erie samples between 2014-2021. Only relative abundance of vOTUs at stations WLE12, WLE2 and WLE4 for the >0.22 µm fraction are shown.. 112

Figure 18. Isolate infection networks containing *Microcystis* and abundant vOTU 17663 (vRhyme_175__447--samp_447_157055). These cultures include: ND-81, ND-85, ND-89, ND-95 and ND-102. Circular nodes represent isolate host MAGs and triangular nodes represent isolate vOTUs. Host nodes are colored by GTDB taxonomic classification. vOTU nodes are colored by the date the culture sample was originally collected. All nodes are sized by the number of interactions they have. vOTU 17663 is specified with a specific color and stroke in the legend. 114

Figure 19. Relative abundance of Lake Erie culture isolate MAGs between 2014-2021. Top row displays all isolate host MAGs. These were dereplicated with wild MAGs and competitively mapped to samples' reads with wild MAGs. Line color pertains to each individual isolate MAG's relative abundance. Bottom row displays only Lake Erie culture isolate *Microcystis* MAGs between 2014-2021. The original 21 *Microcystis* MAGs from isolates were dereplicated to only the 10 shown here. Note this only shows relative abundance of isolate MAGs at stations WLE12, WLE2 and WLE4 for the >0.22 µm fraction. 117

Figure 20. Bloom proxy for Lake Erie cHABs dynamics between 2014-2021. B) Chlorophyll-a, phycocyanin and microcystin measurements across bloom seasons for sampling stations WLE 12, WLE2 and WLE4 (refer to graphical abstract for sampling map). 119

Figure 21. Correlation between microcystin concentration, (A) vOTU 17663 and (B) LE19-12.2 relative abundance. Each point represents the community member at a given date, station and

fraction. Purple line depicts linear regression line and gray shadow represents 95% confidence intervals. 120

Abstract

Cyanobacterial harmful algal blooms (cHABs) pose significant ecological and public health concerns in freshwater ecosystems. Viruses, specifically phages (viruses that infect bacteria), are increasingly recognized as influential players in microbial community dynamics, yet their roles within cHABs remain poorly understood. This dissertation explores the intricate interactions between phages and their microbial hosts within the context of cHABs, revealing their potential impacts on bloom dynamics and genetic diversity. Chapter 2 delves into the temporal dynamics of phages infecting bloom-forming *Microcystis aeruginosa* populations in Lake Erie, a region susceptible to recurrent cHABs. Through extensive genomic analyses and a novel machine-learning model, we unveil the complex web of viral interactions within cHABs, highlighting the potential for cross-species exchange of genetic material and potential phage-driven alterations in key metabolic pathways crucial for *Microcystis* persistence in cHABs. Chapter 3 further explores the role of phages in cHABs more broadly by unraveling the viral community structure and its relationship with the bacterial host community beyond *Microcystis*. Using metagenomic data, we identify and characterize thousands of viral operational taxonomic units (vOTUs), decipher their metabolic functions, and predict their bacterial hosts. This chapter underscores the dynamic nature of viral communities within cHABs and emphasizes the impact of spatiotemporal variation on viruses and how community turnover affects virus-host interactions. Chapter 4 shifts focus to how *Microcystis* host evolutionary distance affects their infection profiles, by using a collection of Lake Erie *Microcystis* multispecies enrichments. We reveal a significant association between *Microcystis* strain phylogenetic relatedness and infection profiles, suggesting that hosts with similar phylogenies share comparable infection profiles. Furthermore, evidence of infection dynamics within the multispecies colonies formed by *Microcystis* and its associated bacteria assemblage emerges, as multiple phages are predicted to infect both *Microcystis* and non-*Microcystis* hosts within a culture.

Collectively, this dissertation advances our understanding of the intricate interplay between phages and their bacterial hosts within cHABs. As such, it provides valuable insights into viral ecology as it pertains to cHABs, paving the way for future research to bolster our understanding of viruses in the wild.

Chapter 1: Introduction

1.1 Toxic algal blooms and subsequent effects in Lake Erie

Causes of cHABs in Lake Erie and why these phenomenon matter

1.1.1 Cyanobacterial harmful algal blooms plague Lake Erie annually

"Lake Erie suffered immensely throughout the late 19th and 20th centuries as a receptacle for human, industrial and agricultural wastes. But nothing compares to what is happening today. Those millions of acres of destroyed wetlands, the overapplication of farm fertilizer, an increase in spring deluges and a lakebed smothered with invasive mussels have all conspired to create massive seasonal toxic algae blooms that are turning Erie's water into something that seems impossible for a sea of its size: poison."
Dan Egan, The Death and Life of the Great Lakes (2017)

The Lake Erie watershed, with approximately 12 million residents, contributes significantly to the economy, generating at least $7.4 billion annually from tourism and another $1 billion from seaports along the lake (French et al., 2011; Steffen et al., 2014; Wortman, 2014). Since the early twentieth century, Lake Erie has experienced annual toxic cyanobacterial harmful algal blooms (cHABs) during the summer months, primarily in its western basin (Fig. 1). These persistent blooms are driven by factors such as shallow waters, warm temperatures (>25 °C), anthropogenic impacts, and short water residence times (~2.1 years) (Harke et al., 2015). Approximately 64% of Lake Erie's shoreline is adjacent to agricultural lands, contributing to nutrient loading and availability (United States Department of Agriculture, 2014). Industrial runoff and agricultural inputs, especially from the Maumee River drainage basin, provide substantial amounts of phosphorus and nitrogen, promoting the proliferation of these blooms (Dolan and Chapra, 2012; Solomon et al., 2010). Despite international efforts to reduce nutrient

levels, particularly phosphorus from the Maumee River, cyanobacterial blooms continue to plague Lake Erie (Lewis et al., 2011; Schindler, 2012). In fact, over the past few decades, the intensity of cyanobacterial blooms in Lake Erie has significantly increased (Bridgeman et al., 2012; Stumpf et al., 2012).

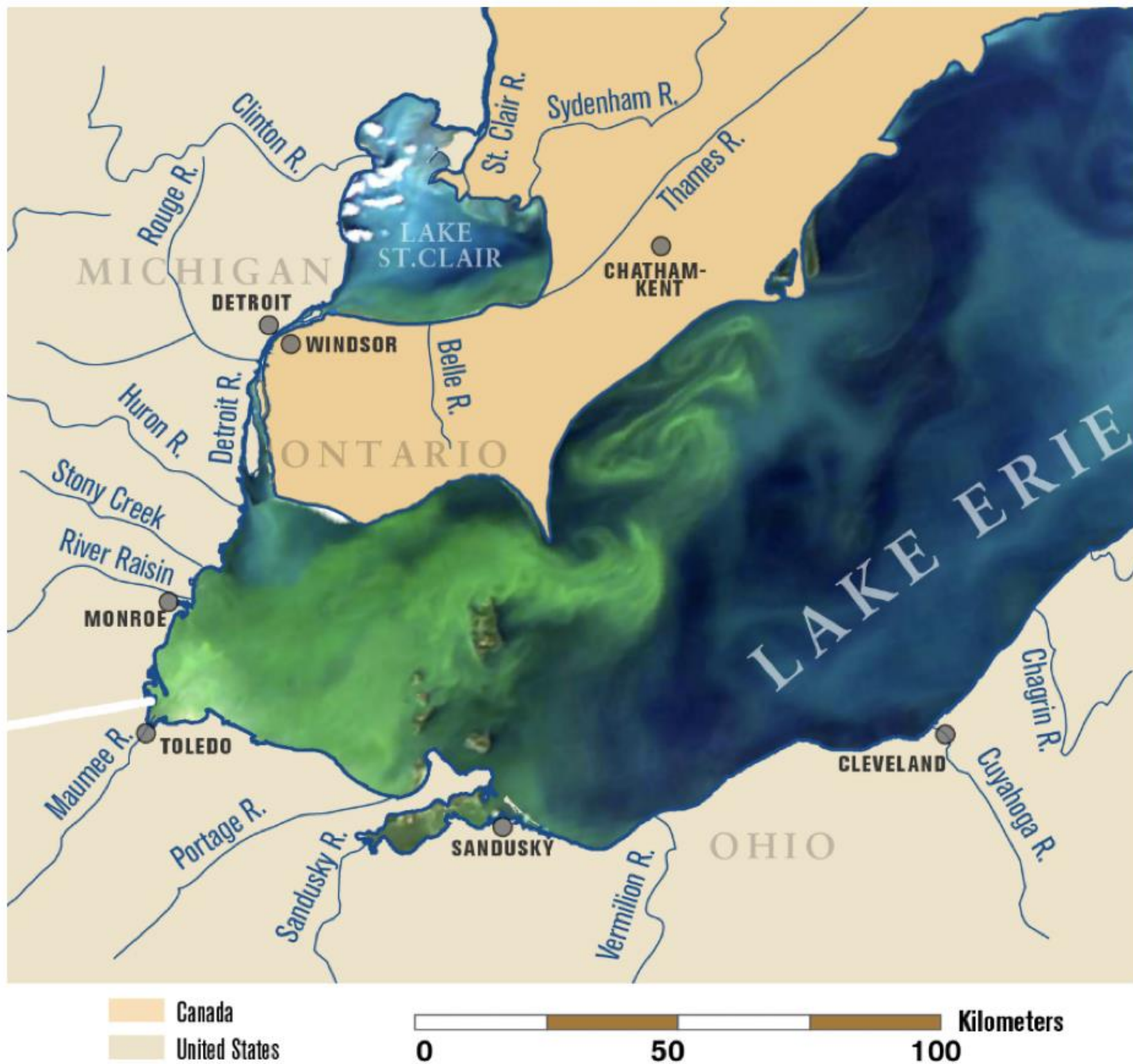


Figure 1. A satellite image of Lake Erie overlaid on a map of the lake and its tributaries. Initially, the bloom forms in the western basin of Lake Erie before radiating out towards the central basin. Image from Michalek et al., 2013.

The importance of Microcystis aeruginosa in cHABs

1.1.2 *Microcystis aeruginosa*: a key member of a complex cHAB community

Microcystis aeruginosa, a globally-distributed cyanobacterium known for forming cHABs, has the potential to produce microcystin, a hepatotoxin regulated by the World Health Organization for drinking and recreational water standards (Beasley et al., 1989; Jochimsen et al., 1998; Yoshida et al., 2008). In 2014, a toxic *Microcystis* blooom in Lake Erie led to a water crisis in Toledo, Ohio, affecting nearly half a million residents by disrupting the drinking water supply for more than three days. While the role of phosphorus and nitrogen in influencing *Microcystis* populations is well-documented (Harke et al., 2015; Harke & Gobler, 2013), the investigation of interactions between *Microcystis* and other bloom members remains ongoing.

While *Microcystis* may dominate these complex communities, recent studies have shown that cHABs maintain numerous forms of cyanobacteria and heterotrophic bacteria across seasons (Harke et al., 2016; Cook et al., 2020; Smith et al., 2021). Though the cooccurrence and interactions between *Microcystis* and various bloom members remains largely unknown, evidence shows *Microcystis* grows better in the presence of heterotrophic bacteria and that *Microcystis* growth affects environmental parameters such as pH and light availability (Kim et al., 2019; Cook et al., 2020; Smith et al., 2021), consequently impacting accompanying bacterial populations. Furthermore, these bacterial populations can live either freely (as unattached individuals) or attached to particles, in part created by *Microcystis* colonial growth, and these size fraction-specific assemblages have genomic and physiological differences (Rieck et al., 2015, Yung et al., 2016; Suzuki et al., 2017). These free-living and particle-associated bacterial populations are influenced by different ecological drivers. Even so, few studies have differentiated them, and most are relegated to how different size fractions are correlated with environmental conditions (Buchan et al., 2014; Rieck et al., 2015; Schmidt et al., 2015; Yung et al., 2016; Mestre et al., 2017).

The role of viruses in cHABs

1.1.3 Impacts of viral infection in cHAB community contexts

Viruses are omnipresent members of Earth's various environments and are known to influence the population dynamics and functional profiles of microbial communities (Fuhrman, 1999; Suttle, 2007; Weitz & Wilhelm, 2012; Koskella & Brockhurst, 2014). Bacteriophages, viruses that infect bacteria, manipulate microbial populations by way of infection and lysis, and by metabolic reprogramming via auxiliary metabolic genes (Breitbart, 2011; Hurwitz & Sullivan, 2013; Rosenwasser et al., 2016; Enav, 2018; Howard-Varona et al., 2020; Zimmerman et al, 2020) and mediating gene transfer between hosts (McDaniel et al., 2010; Soucy et al., 2015). As viral replication hinges upon successful infection of a microbial host, so too do the abundances of viral populations and the viral community structure in a given environment. Therefore, the abundances and community structure of viruses are unequivocally coupled with the coexisting microbial host populations upon which they prey (Srinivasiah et al., 2008; Flores et al., 2011, 2013; Mihara et al., 2016; Dion et al., 2020; Kauffman et al., 2022). Viruses can directly contribute to the collapse of algal blooms or indirectly influence their persistence by driving nutrient and organic matter turnover (Wilhelm and Suttle 1999). Cyanobacterial viruses, known as cyanophages, have been proposed to regulate the progression of cyanobacterial blooms, including *Microcystis* (Brussaard, 2004; Wommack and Colwell, 2000; Yoshida et al., 2008, Yoshida et al., 2012). These cyanophages require further study in terms of infection dynamics if we are to understand how phage predation influences the progression of *Microcystis* blooms, not only in Lake Erie, but globally.

The increased levels of both toxic and non-toxic *Microcystis* strains in Lake Erie are primarily attributed to agricultural runoff from the Maumee River (Davis et al., 2010; Han et al., 2012). Elevated nutrient loading can also lead to an increase in viral abundance, resulting in higher bacterial mortality rates within the ecosystem (Béchette et al., 2013; Hewson et al., 2001; Tapper and Hicks, 1998; Wilhelm and Smith, 2000). Although nitrogen and phosphorus loading can influence the relative abundance of

Microcystis and its phages, the combined effects of nutrient-driven (bottom-up) and viral-driven (top-down) controls on *Microcystis* populations, as well as the dynamics between toxic and non-toxic strains, remain poorly understood (Berry et al., 2017; Harke et al., 2016).

1.2 Viral Infection and Viral Community Dynamics

Strategies used by viruses to infect hosts

1.2.1 Phage infection strategies

"It is also reasonable to speculate that the capacity of prophages to be induced at a low frequency must in itself be advantageous to the phage genome. It is attractive to think of this as a hedging strategy, in which the genetically identical phage population can simultaneously exploit two different phenotypes - in this case, to optimize its probability of genetic success. Lysogeny can be considered as phage conservatism, a strategy suited to survival in adverse conditions, Lytic replication is high-stakes gambling that pays off with confident prediction of outcomes. A phage that never takes advantage of the rewards of the high-stakes game (except under dire and uncertain circumstances) will not be as evolutionary successful as the generally conservative phage with an occasionally successful flutter that rewards with a burst of more rapid amplification. It seems likely then that phages have evolved to spontaneously induce, in a stochastic manner, in order to take advantage of lytic replication while not jeopardizing the genetically identical population of prophages still languishing in the chromosomes of their slowly dividing hosts." Michael G Cordingley, Viruses: Agents of Evolutionary Invention (2017)

There exist two well-established modes of viral infection, known as lytic and lysogenic cycles (Fig. 2). In the lytic cycle, replication of bacterial viruses commences upon phage entry into the host cell, leading to the production of phage progeny and eventual host cell lysis. Conversely, phages may follow the lysogenic route, characterized by the integration of temperate phage genomes into the host genome and the vertical transmission of phages to daughter cells during host cell division. These phages are often referred to as prophages. Lysogeny can result in changes to the host phenotype and provide immunity against subsequent viral infections. Temperate phages have the ability to switch between lysogenic and lytic replication through a process called

induction, which can occur spontaneously or in response to biological or chemical cues (Ptashne, 2004).

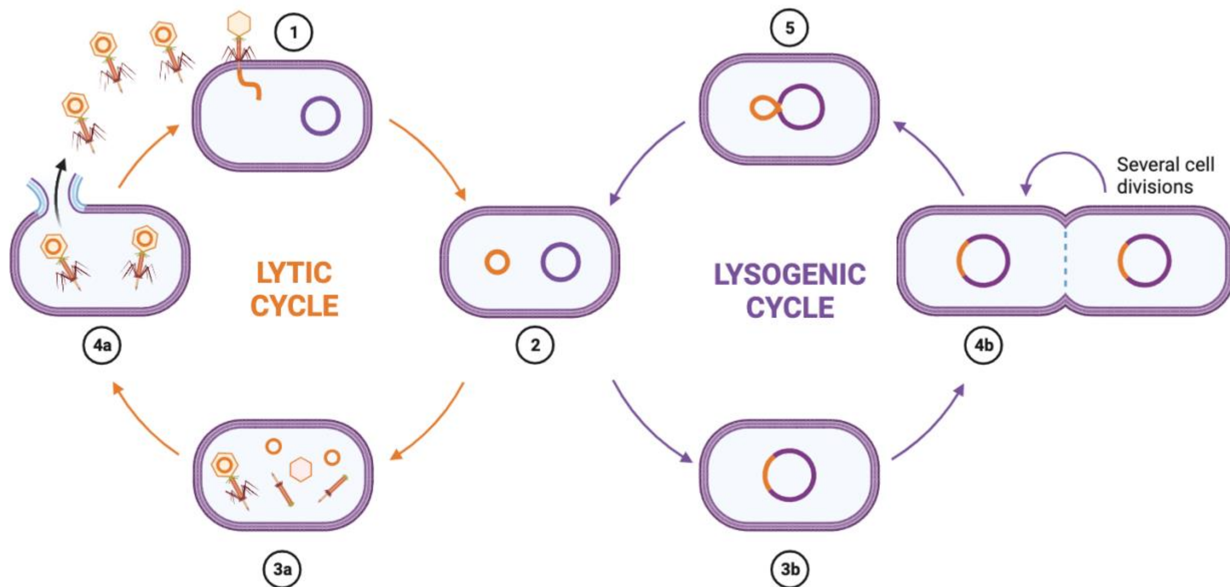


Figure 2. Lytic and lysogenic phage cycles. Lytic cycle: 1) Virus attaches to the host cell and injects DNA. 2) Phage DNA circularizes and enters the lytic cycle. 3a) New phage DNA and proteins are synthesized and assembled into virions. 4a) Cell lyses, releasing phage virions. Lysogenic cycle: 1) Virus attaches to the host cell and injects DNA. 2) Phage DNA circularizes and enters the lysogenic cycle. 3b) Phage DNA integrates within the bacterial chromosome. 4b) Virocell reproduces normally. 5) On occasion, the prophage excises from the bacterial chromosome and enters the lytic cycle. Image generated in BioRender.

This genetic switch of temperate phages has been extensively studied in systems involving *Escherichia coli* and *Escherichia* phage lambda (Ptashne, 2004). Upon phage attachment and genome insertion into the E. coli cell, a set of "early" genes are transcribed, including transcriptional regulators that promote either the lytic or lysogenic pathway (Weitz, 2015). Early genes associated with lysis typically encode structural components like phage tail sheaths or phage collars, while lysogeny-related genes include an integrase gene and a cI repressor gene. The balance of early gene expression determines whether the phage enters the lytic or lysogenic cycle. While we possess a detailed molecular understanding of this switch in well-defined systems, our comprehension of the environmental factors influencing this switch is still in its infancy.

Environmental impacts on the genetic switch

1.2.2 Environmental impacts on the switch between lysis and lysogeny

The switch between lytic and lysogenic replication strategies in phage has presumably evolved as a result of trade-offs between the relative costs of lysis and lysogeny (Goldhill and Turner, 2014). In general, lysogens are hypothesized to be more abundant as a result of low bacterial density, which may stem from low nutrient availability and low temperatures in the system (Ghosh et al., 2007, McDaniel and Paul, 2005; Middleboe, 2000; Shan et al., 2014). Bacterial cells are typically smaller under poor growth conditions and thus provide fewer nutrients for the generation of viral progeny (Akerlund et al., 1995; Volkmer and Heinemann, 2011). Unfavorable host growth conditions can also reduce concentrations of susceptible hosts, leading to suboptimal foraging and ultimately reducing the benefits of lysis for phage. Consequently, low temperatures and reduced nutrient availability may lead to gene acquisition and provide a selective advantage for host survival in these unfavorable conditions (Touchon et al., 2016). In recent studies however, conflicting dynamics between phage and host have been reported in what remains a contentious argument (Knowles and Rowher, 2017; Knowles et al., 2016; Weitz et al., 2017). Due to the direct effects of lysogeny on host behavior and fitness, discerning the role of prophages is a crucial step towards grasping a better understanding of phage-host interactions in freshwater ecosystems like the Laurentian Great Lakes.

Understanding phage-host interactions in the wild

1.2.3 Seed-bank theory in phage-host dynamics

"There is nothing so patient, in this world or any other, as a virus searching for a host." Mira Grant, Countdown: A Newsflash Novella (2011)

Factors driving temporal and spatial variations in viral communities are key to detangling complex host-virus interactions and the ecological implications of viruses.

Kill-the-winner (Winter et al., 2010), piggyback-the-winner (Knowles et al., 2016) and the seed bank theory (Breitbart et al., 2005) are all hypotheses that describe phage-host dynamics. Similar to traditional Lotka-Voltera equations, the kill-the-winner hypothesis posits that viruses target and lyse the more dominant and rapidly growing hosts in a system, while the piggyback-the-winner hypothesis suggests viruses delay host lysis to allow for host proliferation and avoid competition from other viruses capable of lysis. Seed bank theory (Fig. 3) suggests that the majority of viruses persist in an “inactive” state for extended periods, resulting in a reservoir of inactive individuals known as a "seed bank.” Seed banks increase the effective population size and reduce genetic drift, thereby buffering lineages from extinction and maintaining genetic diversity. Additionally, seed banks can alter species interactions by allowing competing species to coexist through the storage effect. Knowledge of which hypothesis is observed through Lake Erie’s cHABs would provide important insights into mechanistic underpinnings of viral community dynamics and their impacts on cHAB progression and functionality.

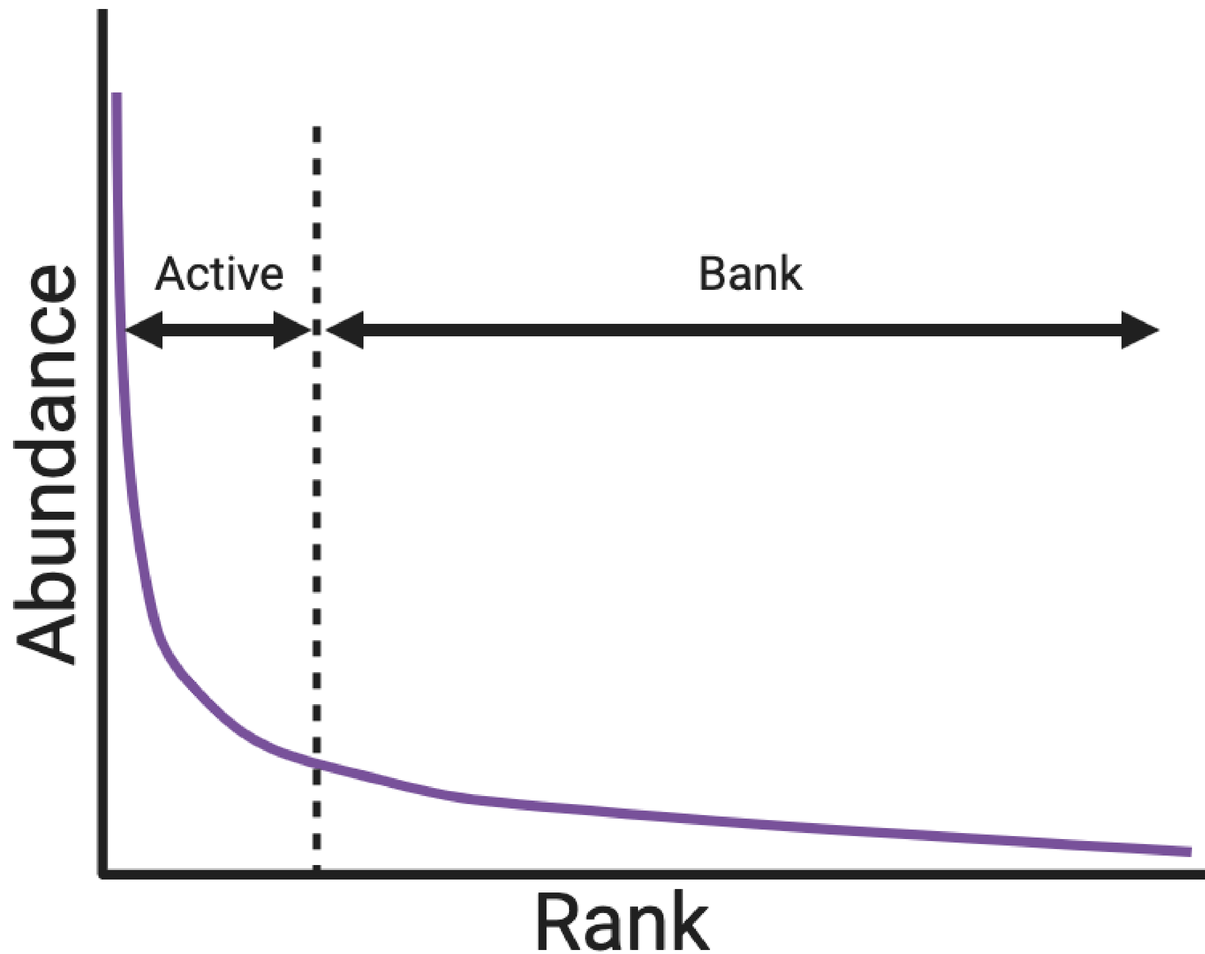


Figure 3. Traditional example of rank-abundance curve demonstrating vOTU abundance in a community. The most abundant vOTU is ranked as 1, the next highest is 2, and so on and so forth. According to seed bank theory, only a few of the most abundant vOTUs are in the active fraction at any given point. As new hosts rise in abundance in response to changing environmental conditions, the viral predators of those hosts may also become abundant. The vOTUs that were previously in the active fraction become part of the bank fraction. Adapted from Breitbart, 2005.

Defining coevolution and its importance in phage-host dynamics

1.2.4 Detangling the coevolutionary arms race between host and phage

"Because the thing about viruses is that they're easily manipulated. The DNA they inject doesn't have to be destructive. It can be replaced with almost any kind of DNA you want, and it can be programmed to

only replace certain parts of the host's genetic code. In other words, viruses are perfect vectors for genetic engineering." Christian Cantrell

Coevolution, characterized as the reciprocal adaptation and counter-adaptation between interacting species, plays a pivotal role in shaping the population dynamics of hosts and their infecting phages (Buckling and Rainey, 2003; Janzen, 1980; Laine and Tellier, 2008). Many ecosystems host diverse populations of hosts and phages engaged in persistent cycles of coevolution. In these cycles, phage-resistant hosts maintain bacterial lineages, while counter-resistant phages drive the evolution of surviving host strains (Labrie et al., 2010). However, studying coevolution can be challenging when organisms have long generation times or are difficult to culture. Furthermore, viruses and their hosts are not isolated entities but dynamic components woven into the fabric of all microbial communities.

Viruses and their hosts engage in a perpetual co-evolutionary arms race, marked by the continual development of new infection and defense strategies (Koskella and Brockhurst, 2014, van Houte et al., 2016). Over time, this relentless struggle between viruses and their hosts shapes the structure of microbial communities and profoundly influences host fitness. The influence of viruses extends beyond causing cell death; they also significantly shape microbial community structure and function by facilitating gene transfer between and across species, a process known as transduction (Jiang and Paul, 1998). Some viruses go a step further by encoding auxiliary metabolic genes (AMGs), which enhance host metabolic pathways to favor viral particle production (Warwick-Dugdale et al., 2019). Additionally, certain prophages establish mutualistic relationships by preventing other phages from successfully infecting the same cell, an occurrence referred to as "superinfection exclusion" (Rostol and Marraffini, 2019). Thus, virus-host interactions represent a pivotal aspect of comprehensive microbial ecology studies, impacting both hosts (Ogilvie and Jones, 2015) and their surrounding environments (Coutinho et al., 2018; Davenport et al., 2019).

1.3 **Tools for Diagnostics and Virus Discovery**

Approaches to study viruses in the wild

1.3.1 Virus Discovery Before Metagenomics

"...on opening the incubator I experienced one of those rare moments of intense emotion which rewarded the research worker for all his pains: at first glance I saw that the broth culture, which the night before had been very turbid, was perfectly clear: all the bacteria had vanished... As for my agar spread it was devoid of all growth and what caused my emotion was that in a flash I understood: what causes my spots was in fact an invisible microbe, a filterable virus, but a virus parasitic on bacteria." Felix D'Herelle, In Allan Chase, Magic Shots: A Human and Scientific Account of the Long and Continuing Struggle to Eradicated Infectious Diseases by Vaccination (1982)

Prior to the metagenomics era, the discovery of viruses heavily relied on classical experimental techniques, which required pure cultures of viruses and their potential hosts for spot and plaque assays. Other methods involved viral tagging that utilized fluorescent labeling and sorting of viruses (Edwards et al., 2016). These traditional approaches enabled researchers to delve into the morphology, host range, and replication cycles of cultured viruses, contributing significantly to the classification of viral lineages defined by the International Committee for the Taxonomy of Viruses (ICTV). However, these classical methodologies had limitations, including being low-throughput and requiring pure cultures, rendering them impractical for studying viruses in complex environmental samples where the isolation of both bacteria and viruses posed formidable challenges.

To address the need for detecting and categorizing microbes in environmental samples, scientists turned to universal prokaryotic marker genes, notably the small subunit rRNA gene (16S) (Woese et al., 1990), as well as domain-specific marker genes used in the Genome Taxonomy Database (GTDB) (Parks et al., 2018). In the quest to target specific viral groups, researchers employed marker genes associated with these groups for the purpose of virus detection and the assessment of diversity in environmental samples, primarily employing PCR-based fingerprinting methods

(Drosten et al., 2003). Such viral marker genes included major capsid proteins (for T4-like myoviruses), auxiliary metabolic genes like those encoding photosynthesis proteins in cyanophages, and DNA/RNA polymerases (Adriaenssens and Cowan, 2014).

Nonetheless, the use of marker genes for virus detection and classification presented challenges. Primer sets designed for marker genes were highly degenerate and required low annealing temperatures, indicating that even conserved group-specific genes exhibited diversity. This made them less suitable for quantitative PCR (Duhaime and Sullivan, 2012). Moreover, given the high variability of viral gene content, these primer sets were typically designed for specific viral groups, leaving a substantial portion of the virome unaccounted for. Additionally, PCR-based fingerprinting fell short when it came to identifying entirely novel viruses that lacked known marker genes.

Cultivating a diverse range of viruses was often hindered by the initial necessity of cultivating the host organisms in pure culture, presenting a significant bottleneck in virology research. This limitation stemmed from the fact that many microorganisms remained uncultured (Lloyd et al., 2018). The challenge of obtaining pure host cultures restricted the variety of viruses that could be successfully isolated and studied under controlled laboratory conditions. Even when host cultures were accessible, not all of them could be grown to the point of confluence on agar plates, which was a prerequisite for the formation of viral plaques (Chen and Novick, 2009; Willner and Hugenholtz, 2013).

For viruses lacking the ability to form plaques, alternative strategies were devised, particularly for detection purposes. These strategies included culture clearing methods, which involved inducing lysis of the host culture in a broth (Sullivan et al., 2003, Waterbury and Valois, 1993) and routine test dilution, where culture clearing on a plate occurred with near-confluent lysis (Thomas and Corbel, 1977). Notably, culture clearing could be automated and carried out in multi-well plates, enabling high-throughput monitoring of viral growth dynamics (Henry et al., 2012).

Furthermore, it is noteworthy that the host organism originally used for isolation, often obtained from a different sample than the virus isolate, might not always serve as the primary host for a virus. In some cases, it represented a suboptimal host, potentially leading to inaccuracies in estimating viral growth parameters (Howard-Varona et al., 2017, Enav et al., 2018). These challenges associated with cultivating virus hosts in the laboratory posed significant obstacles to the isolation, propagation, and ecological characterization of viral isolates under controlled conditions. In subsequent years, alternative methods to study viruses were not only sought, but considered inescapable in order to gain perspective on their roles in complex ecosystems.

Using modern approaches like metagenomics to study viruses

1.3.2 Capturing Viral Genomes Through Sequencing Technologies

In contrast to isolation procedures, which primarily target individual viruses or specific microbial clones, metagenomics offers a more comprehensive approach by involving the extraction of DNA from a given sample. The extracted DNA is subsequently fragmented into numerous smaller pieces and subjected to shotgun sequencing, resulting in a wealth of sequence data analyzed collectively to reconstruct the genomes of both bacteria and viruses present within a given environmental sample (Handelsman, 2004; Edwards and Rowher, 2005; Roux, 2019). Unlike virus isolation, metagenomics does not rely on the need for culturing, which provides an advantage when studying uncultivable microbes. This approach significantly broadened our understanding of microbial life in various environments and, notably, provided valuable insights into the presence and diversity of viruses within ecosystems (Handelsman, 2004; Daniel, 2005; Edwards and Rowher, 2005; Roux, 2019).

1.3.3 Resolving Virus Genomes amidst Abundant Sequence Data

In this section, we address several challenging aspects related to characterizing viruses in metagenomes. The foundation of these challenges lies in the fact that there are an estimated 10^{31} phages on Earth (Suttle, 2005), and there are thought to be hundreds of

thousands of viral genotypes in our oceans alone (Angly et al., 2006). The sheer magnitude of these numbers underscores the appeal of metagenomics as a method, as it is impossible to isolate all or even a majority of these biological entities. However, this abundance of data generated by metagenomic analyses can sometimes hinder our ability to precisely differentiate individual types of organisms, and this has been particularly problematic in resolving virus genomes. Regardless, two general approaches exist to enhance the resolution of virus genomes in metagenomic analyses: improved sequencing depth and enhanced sequence analysis. To be sure, these approaches include biases in terms of the DNA that is sequenced or analyzed.

Metagenomic analysis of random DNA samples can introduce biases stemming from various factors, including (1) the specific methodologies used for sample collection and storage, (2) the physical and chemical techniques employed for DNA extraction and subsequent amplification, and (3) the choice of bioinformatic tools used for metagenome reconstruction (Delmont et al., 2011). Many of these biases, however, can be mitigated through the implementation of standardized methodologies (Kunin et al., 2008; Roux et al., 2017; McLaren et al., 2019).

When collected, DNA sequences are often fragmented and diverse, making it difficult to assemble sequenced fragments into complete or nearly complete genomes. Less abundant genomes are often overlooked and result in a lack of fully sequenced genomes. Consequently, a constructed metagenome may not precisely reflect the actual collection of sequences in the original sample. To address the aforementioned issues, increasing the number of sequencing reads for a sample can enhance the recovery of metagenome-assembled genomes (MAGs) with lower error rates (Martinez-Hernandez et al., 2017,Sieradzki et al., 2019). Similar challenges are observed with viral genomes, despite their smaller size in comparison to bacterial genomes. Additionally, viral DNA is often less abundant in natural systems, making them less likely to be captured by sequencing efforts. Consequently, this results in far fewer virus sequences and genomes generated in metagenomes compared to their bacterial hosts.

Identifying viruses within metagenomes is further complicated by the immense diversity of viruses present, which can pose challenges in de novo assembly of viral contiguous sequences (contigs) (Sutton et al., 2019). Contig assembly algorithms rely on overlapping sequences (e.g., De Bruijn graph assembly) and are less effective when there are fewer copies of specific viral DNA sequences present in a sample. In particular, the scarcity of overlapping stretches of sequenced nucleotides can hinder the assembly of viral genomes without pre-existing templates (Sutton et al., 2019). As a result, virus sequences often constitute a small proportion of assembled sequences, and only partial virus genomes are typically obtained (Emerson et al., 2018).

1.3.4. The importance of sequence read depth and coverage

Sequencing efforts are often evaluated in two distinct ways: read coverage and read depth. Coverage pertains to the portion of a contig or genome that possesses aligned reads, indicating how complete an assembled genome is concerning a reference genome. In the case of bacterial and archaeal MAGs, researchers rely on the identification of universal marker genes to estimate completeness (Bowers et al., 2017). On the other hand, read depth represents the average number of reads aligning to each base in a contig or assembled genome. Read depth serves as a measure of the relative abundance of microbes or viruses within environments and is essential for assessing the reliability of certain analyses. In general, higher read depth is preferred as shallow read depth can result in the omission of less abundant viruses and their hosts. This limitation can impact estimates of metagenomic diversity. Nonetheless, interpreting the ecological implications solely based on read depth of sequencing reads can be challenging and misleading. While higher relative abundance in a metagenome suggests a potentially greater impact of certain viruses on an ecosystem due to a higher read depth, abundance alone does not provide qualitative insights into their impact. Metagenomes represent a snapshot of a community and lack information about community dynamics over time unless generated as part of a time series.

1.3.5 Needles in a haystack: Identifying viruses in large metagenomic datasets

Distinguishing viral genomic sequences from cellular genomic sequences presents a fundamental challenge in metagenomic analyses of large quantities of environmental DNA used to study virus ecology. Significant progress has been made in addressing this challenge,as tools have been developed to identify viral hallmark genes (e.g., VirSorter (Roux et al., 2015) or virus-specific motifs (e.g., VirFinder/DeepVirFinder (Ren et al., 2017, Ren et al., 2020)) to identify likely viral contigs in metagenomes. VirSorter relies on a database of known viral genes for category prediction and is particularly effective for marine viruses, given their better genomic characterization. DeepVirFinder also relies on a virus reference database and utilizes a machine learning approach for robust detection of virus fragments ≥3 kb. It offers conservative and sensitive approaches for selecting contigs based on scores and p-values. VirSorter and DeepVirFinder can be used in parallel to optimize viral identification from metagenomic data. Since the introduction of these tools, a cascade of viral identification tools have been developed to address challenges in identifying viral contigs in metagenomes and combinations of these tools have been analyzed to provide recommendations for users based on their sampling environment and research goals (Hegarty et al., 2023; Wu et al., 2023).

1.3.6 Benefits and Limitations of Viromics

Viromes, or metagenomes primarily consisting of sequence data obtained from the viral fraction (<0.22 µm) of environments, involves the separation of viruses from cells, followed by the lysis of these particles and subsequent sequencing of the liberated nucleic acid. Essentially, a virome can be viewed as a "targeted metagenome," focusing on a specific aspect of a metagenome to provide a more detailed description of the taxonomic content and related characteristics of that fraction. The first virome, published in 2002 (Breitbart et al., 2002), originated from marine water samples, and since then, this approach has become one of the predominant methods for characterizing viruses across diverse environments (Pratama and van Elsas, 2018; Breitbart et al., 2018).

Viromes offer a notable advantage over the extraction of viral signals from less specific metagenomes in that they provide enhanced coverage of viral genomes. This increased coverage is made possible by the prior removal of both prokaryotic and eukaryotic DNA, as these entities possess larger genomes that account for a significant portion of sequencing reads. By specifically targeting the viral fraction for sequencing, viromes can yield a greater recovery of viral contigs than larger, less-targeted fractions (Wing et al., 2024a in prep). As a result, greater read depth is achieved, leading to a more encompassing snapshot of viral diversity and the revelation of micro-diversity within viral populations (Gregory et al., 2019). These benefits translate into the acquisition of complete or nearly complete viral genomes that can subsequently serve as reference genomes. Reference genomes are valuable for identifying new viruses from metagenomic or viromic data, determining viral taxonomic affiliations, and facilitating the prediction of viral gene functions.

However, viromics shares several drawbacks with metagenomic studies These include biases introduced during sample preparation, high costs due to the extensive number of required sequencing reads (although costs are continually decreasing), computational intensity (Roux, 2019), and the challenge of annotating most predicted genes. Unlike untargeted metagenomic approaches where DNA is collectively

extracted, viromics necessitates additional wet lab procedures to separate viruses from various forms of environmental DNA before viral DNA itself can be extracted.

1.3.7 Short- and Long-Read Sequencing: Different Sized Pieces of the Same Puzzle

The choice between short- and long-read sequencing in metagenomic studies carries both advantages and drawbacks concerning explorations into virus diversity and distribution. High-throughput sequencing of short reads, typically spanning 100 to 250 base pairs, represents the most prevalent and cost-effective sequencing platform. It also benefits from extensively refined computational tools for the detection and characterization of viral genomes. In recent years, significant endeavors have been invested in enhancing long-read sequencing in terms of quality and throughput, with certain platforms now generating reads exceeding 2,000,000 base pairs in length (Payne et al., 2018).

The merits of employing long reads encompass several aspects including enhanced sensitivity, identification of hypervariable regions, detection of recombinants and perhaps most importantly, improved genome assemblies. The enhanced sensitivity of long-read sequencing can capture taxa that might elude detection by short reads, including less common viral Operational Taxonomic Units (vOTUs). Long reads also facilitate the identification of hypervariable regions within viral genomes, shedding light on their genetic diversity. Furthermore, the extended read lengths provided by long-read sequencing aid in the identification of recombinants within viral populations, uncovering genetic exchanges that may impact virus evolution. Finally, long reads contribute to more robust assemblies of viral genomes, increasing opportunities to recover complete viral genomes (Zablocki et al., 2020).

To unlock the full potential of metagenomic studies and comprehensively explore virus diversity and distribution, a compelling strategy emerges: the integration of both short and long-read sequencing approaches. This combined approach harnesses the strengths of each technology, providing a more comprehensive view of the virome.

Short reads excel at efficiently identifying abundant vOTUs, while long reads enhance sensitivity to less common vOTUs, uncover hypervariable regions, detect genetic exchanges, and enable more robust genome assemblies. By synergizing these capabilities, researchers can delve deeper into the intricate world of viruses, ultimately advancing our understanding of viral ecology and evolution in complex microbial communities.

Applying network-based approaches to viral ecology

1.3.8 Constructing Networks in Viral Ecology

1.3.8.1 Who is Who? Viral Taxonomy in the Age of Networks

While viral phylogenomics, the study of reconstruction and analysis of evolutionary relationships and diversification among viruses, is valuable for understanding viral relatedness and taxonomy, this methodology faces limitations in representing the mosaic and highly diverse nature of viral genomes (Lima-Mendez et al., 2008). Additionally, hierarchical tree structures often inadequately portray the actual evolutionary trajectories of viruses, and different viral lineages may not fit onto the same phylogram if they lack common genes (Low et al., 2019, Iranzo et al., 2017, Corel et al., 2016). To address these limitations, the field of virology has increasingly turned to network-based approaches (Lima-Mendez et al., 2008). These networks use shared proteins to establish links between viral genomes. In some networks, known as monopartite networks, nodes, which represent viral genomes, are connected by edges, which are weighted based on the total protein sequence similarity between two given nodes (Lima-Mendez et al., 2008). VConTACT2 is an example of a tool that employs monopartite networks of reference viral genomes to classify the taxonomy of user-provided viral sequences (Bin Jang et al., 2019). Sequences that share proteins with reference viruses cluster together within the network, allowing for the identification of novel viral sequences as outliers (Bolduc et al., 2017). On the other hand, bipartite networks not only show relatedness between viral genomes but also indicate which proteins are shared between groups of viruses (Bolduc et al., 2017). These networks

have facilitated the identification of hallmark genes commonly shared by double-stranded DNA (dsDNA) viruses, providing valuable insights into the evolutionary history of viruses (Iranzo et al., 2016a, Iranzo et al., 2016b).

1.3.8.2 Who Infects Whom? Virus-Host Interactions as Networks

In viral ecology, disentangling virus-host interactions is paramount (Mihara et al., 2016). The effects of viruses are observed through interactions with their hosts. Virus-host interactions are often considered as part of vast ecological networks capable of varying across environmental gradients (Tylianakis et al., 2017). These networks are invaluable as they provide mathematical foundations to quantify interactions between members of ecosystems, where members are represented by nodes and interactions between members are represented by edges (Bascompte, 2009).

When applying network thinking to viral ecology, nodes become hosts and viruses and the edges become potential interactions between hosts and viruses based on a metagenomic data (Mihara et al., 2016). The Virus-Host Database is a valuable resource for virus-host information, compiling data from sources like RefSeq, GenBank, UniProt, ViralZone, and manual literature surveys (Brister et al., 2015; Benson et al., 2013; Consortium TU, 2019; Hulo et al., 2011). The remarkable diversity of viruses, encompassing variations in structure, genetic material, host ranges, and habitats, presents challenges for both traditional molecular methods and computational techniques. To address these challenges, viral ecologists currently use three genomic features to identify interactions between hosts and their viral predators. Those features are described in the following sections.

1.3.8.2.1 CRISPR-Cas System

In recent years, a bacterial defense mechanism based on a region of DNA known as clustered regularly interspaced short palindromic repeats (CRISPR) has emerged as a powerful tool for studying coevolution through DNA sequence analysis. The CRISPR-Cas system, a form of adaptive immunity present in over 40% of bacteria and over 90%

of archaea, targets and degrades recognized phage DNA, preventing infection, and simultaneously keeps a record of phage infections (Fig. 4) (Dutilh et al., 2014; Koskella and Brockhurst, 2014; Staals and Brouns, 2012). Several sequenced *Microcystis* genomes contain arrays of CRISPR spacers, some with over 70 spacers, indicating persistent infections and providing a valuable tool for studying the rapid evolution of this cyanobacterium (Grissa et al., 2007; Makarova and Koonin, 2012; Makarova et al., 2013). Additionally, analyzing spacers within the CRISPR-Cas systems of community genomes, referred to as metagenomes, helps establish links between *Microcystis* and its phages and sheds light on their infection dynamics by identifying patterns in CRISPR spacers that persist within populations and across communities (Berg-Miller et al., 2011).

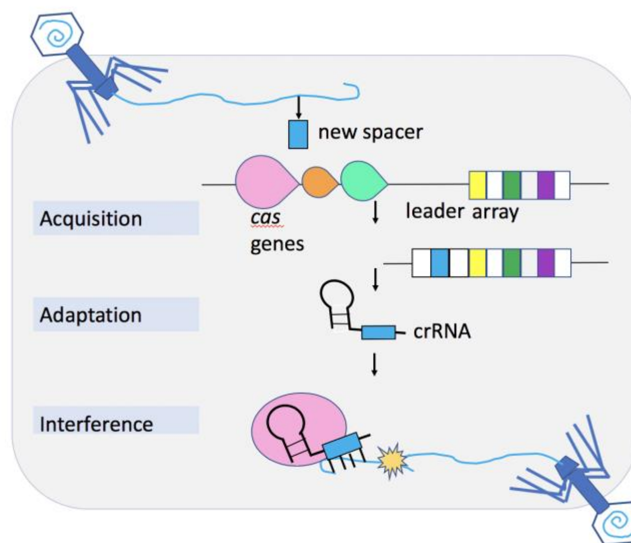


Figure 4. Three stages of CRISPR-mediated phage immunity. New spacers from the phage genome are integrated into the host CRISPR array (acquisition), separated by palindromic repeats (white rectangles). The array is transcribed and processed by Cas proteins to generate small crRNAs (adaptation). crRNA leads Cas protein complex to matching target to induce cleavage of exogenous nucleic acid (interference). Adapted from Heler et al. 2014.

Although spacer acquisition depends on the host microorganism, numerous studies have used CRISPR arrays to study the chronological order of phage infections. Newly incorporated spacers, acting as proxies for the most recent infections, are typically found at the leader end of the CRISPR array (Andersson and Banfield, 2008; Barrangou and Dudley, 2016; Stern and Sorek, 2010). Researchers have also utilized

the CRISPR-Cas system to differentiate between closely related strains (Heidelberg et al., 2009; Held et al., 2013; Tyson and Banfield, 2007). Coevolutionary dynamics between phages and hosts, including selective sweeps and bottlenecks that can shape species composition within bacterial populations, have been investigated using this system (Paez-Espino, 2013; Pride et al., 2010; Touchon and Rocha, 2010). For example, an analysis of spacer sequences within a *Microcystis* population in a small pond during a cyanobacterial bloom revealed multiple coexisting CRISPR types, defined as groups of CRISPR sequences sharing at least two trailer-end spacers, suggesting that the population did not experience a complete selective sweep but rather an incomplete one, allowing specific genotypes to persist over two years (Kuno et al., 2014). While the spacer-repeat arrays of the CRISPR-Cas system have been used to identify sweeps in various bacterial populations in other studies, the seasonal variation experienced by many *Microcystis* populations during Lake Erie blooms presents a unique challenge (Paerl and Otten, 2013). These blooms may undergo annual bottleneck effects due to harsh winter conditions, preventing a typical selective sweep from occurring (Kimura et al., 2012; Kimura et al., 2018; Yoshida et al., 2010).

1.3.8.2.2 Hi-C: The Power of Physical Linkage

High-Throughput Chromosome Conformation Capture (Hi-C) has been applied to investigate virus-host interactions by capturing the 3D architecture of chromosomes through proximity-based fixation and high-throughput sequencing (Marbouty et al., 2017; Marbouty et al., 2021). Although it characterizes prophages and slow-growing lytic phages, it may not capture highly virulent phages. Sequencing of viromes remains necessary to generate a comprehensive inventory of viral genomes. Single-cell Hi-C enables studies of chromosome-viral genome interactions at a single-cell level, facilitating the identification of virus-host physical linkages (Kim et al., 2020, Nagano et al., 2015).

1.3.8.2.3 Sequence Homology, AMG homology and k-mer frequency

Recombination sites with recognition sequences and the identification of auxiliary metabolic genes (AMGs) in viral sequences have also been used for host prediction (Edwards et al., 2016; Roux et al., 2016). Abundance profiles, which reflect the sequencing coverage of viral or host sequences across multiple samples, offer another approach for host prediction, especially effective with time-series metagenomic data (Thingstad, 2000; Van Goethem et al., 2019; Arkhipova, 2018). K-mer (oligonucleotide) frequency profiles have been employed to predict virus-host relationships, as viruses often exhibit similar profiles to their hosts (Ahlgren et al., 2017; Villarroel et al., 2016; Galan et al., 2019). Tools like VirHostMatcher, HostPhinder, Host Taxon Predictor and VHIP leverage k-mer frequency distributions for host prediction (Ahlgren et al., 2017; Villarroel et al., 2016; Galan et al., 2019; Bastien et al., 2023). Tetranucleotide (4-mer) frequency profiles have been particularly useful for alignment-free host prediction (Coutinho, 2018; Emerson et al., 2018; Roux et al., 2015; Roux et al., 2016).

In practice, combining multiple approaches, both homology-based and non-homology-based, and considering consensus results is common to achieve comprehensive and accurate predictions of virus-host associations (Edwards et al., 2016). In summary, advancements in virology have been driven by computational approaches that leverage genomic data, including sequence homology, abundance profiles, and k-mer frequency profiles, to predict virus-host interactions and enhance our understanding of viral dynamics within microbial communities (Mihara et al., 2016).

1.4 Dissertation Overview

In the realm of freshwater ecosystems, Cyanobacterial Harmful Algal Blooms (cHABs) present considerable ecological and public health concerns. While their significance is well-recognized, the roles of viruses, particularly phages, in shaping microbial communities within cHABs remain enigmatic. This dissertation embarks on a journey to unravel the intricate interactions between phages and their bacterial hosts within the context of cHABs, unveiling their potentially profound impacts on bloom dynamics.

Chapter 2 of this dissertation reports on the temporal dynamics of phages infecting *Microcystis aeruginosa*, a predominant bloom-forming cyanobacterium in Lake Erie—an area prone to recurring cHABs. Through comprehensive metagenomic analyses and the innovative application of a machine-learning model, this chapter unravels the complex web of viral interactions within cHABs. It illuminates the potential for cross-species exchange of genetic material and unveils phage-driven alterations in crucial metabolic pathways essential for *Microcystis* adaptation.

Chapter 3 takes us further into the world of phages in cHABs by dissecting the viral community structure and its relationships with the host community. Leveraging metagenomic data, this chapter identifies and characterizes thousands of viral populations. It deciphers their metabolic functions and predicts their microbial hosts. The focus here lies on highlighting the dynamic nature of viral communities within cHABs at the smallest of spatial scales and underlying factors that impact viral community structure and function, and how this variation can lead to changes in virus-host interactions.

Chapter 4 shifts our attention to exploring the correlation between host genome evolution and how it affects their infection profiles, using a collection of Lake Erie *Microcystis* isolates. This chapter uncovers a significant association between phylogenetic relatedness and infection profiles. This suggests that more evolutionary related hosts share comparable infection dynamics. Additionally, intriguing evidence emerges regarding intra-colony infection dynamics, as multiple phages are predicted to infect both *Microcystis* and non-*Microcystis* hosts within a culture.

In summation, this dissertation advances our comprehension of the intricate interplay between bacterial hosts and their viral predators within cHABs. These revelations provide invaluable insights into the viral ecology of cHABs, planting seeds for future research aimed at a more complete understanding of viruses in all natural systems.

1.5 References

Adriaenssens, E. M., & Cowan, D. A. (2014). Using Signature Genes as Tools To Assess Environmental Viral Ecology and Diversity. *Applied and Environmental Microbiology*, *80*(15), 4470–4480. https://doi.org/10.1128/AEM.00878-14

Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Research*, *45*(1), 39–53. https://doi.org/10.1093/nar/gkw1002

Akerlund, T., Nordström, K., & Bernander, R. (1995). Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of Escherichia coli. *Journal of Bacteriology*, *177*(23), 6791–6797.

Andersson, A. F., & Banfield, J. F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*. https://doi.org/10.1126/science.1157358

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., Chan, A. M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J. M., Mueller, J. E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C. A., & Rohwer, F. (2006). The Marine Viromes of Four Oceanic Regions. *PLOS Biology*, *4*(11), e368. https://doi.org/10.1371/journal.pbio.0040368

Arkhipova, I. R. (2018). Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. *Molecular Biology and Evolution*, *35*(6), 1332–1337. https://doi.org/10.1093/molbev/msy083

Barrangou, R., & Dudley, E. G. (2016). CRISPR-Based Typing and Next-Generation Tracking Technologies. *Annual Review of Food Science and Technology*, *7*, 395–411. https://doi.org/10.1146/annurev-food-022814-015729

Bascompte, J. (2009). Disentangling the web of life. *Science (New York, N.Y.)*, *325*(5939), 416–419. https://doi.org/10.1126/science.1170749

Bastien, E. G., Cable, R. N., Zaman, L., Batterbee, C., Wing, A. J., & Duhaime, M. B. (2023). *Virus-Host Interactions Predictor (VHIP): Machine learning approach to resolve microbial virus-host interaction networks* (p. 2023.11.03.565433). bioRxiv. https://doi.org/10.1101/2023.11.03.565433

Beasley, V. R., Dahlem, A. M., Cook, W. O., Valentine, W. M., Lovell, R. A., Hooser, S. B., Harada, K.-I., Suzuki, M., & Carmichael, W. W. (1989). Diagnostic and Clinically Important Aspects of Cyanobacterial (Blue-Green Algae) Toxicoses. *Journal of Veterinary Diagnostic Investigation*, *1*(4), 359–365. https://doi.org/10.1177/104063878900100417

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(Database issue), D36–D42. https://doi.org/10.1093/nar/gks1195

Berg Miller, M. E., Yeoman, C. J., Chia, N., Tringe, S. G., Angly, F. E., Edwards, R. A., Flint, H. J., Lamed, R., Bayer, E. A., & White, B. A. (2012). Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environmental Microbiology*, *14*(1), 207–227. https://doi.org/10.1111/j.1462-2920.2011.02593.x

Berry, M. A., Davis, T. W., Cory, R. M., Duhaime, M. B., Johengen, T. H., Kling, G. W., Marino, J. A., Den Uyl, P. A., Gossiaux, D., Dick, G. J., & Denef, V. J. (n.d.). *Cyanobacterial harmful*

algal blooms are a biological disturbance to Western Lake Erie bacterial communities. https://doi.org/10.1111/1462-2920.13640

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., Brister, J. R., Kropinski, A. M., Krupovic, M., Lavigne, R., Turner, D., & Sullivan, M. B. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology*, *37*(6), 632–639. https://doi.org/10.1038/s41587-019-0100-8

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B. (2017). vConTACT: An iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, e3243. https://doi.org/10.7717/peerj.3243

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., … Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, *35*(8), Article 8. https://doi.org/10.1038/nbt.3893

Breitbart, M. (2012). Marine viruses: Truth or dare. *Annual Review of Marine Science*, *4*, 425–448. https://doi.org/10.1146/annurev-marine-120709-142805

Breitbart, M., Bonnain, C., Malki, K., & Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0166-y

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. https://doi.org/10.1016/j.tim.2005.04.003

BREITBART, M., THOMPSON, L. R., SUTTLE, C. A., & SULLIVAN, M. B. (2007). Exploring the Vast Diversity of Marine Viruses. *Oceanography*, *20*(2), 135–139.

Bridgeman, T. B., Chaffin, J. D., & Filbrun, J. E. (2013). A novel method for tracking western Lake Erie Microcystis blooms, 2002–2011. *Journal of Great Lakes Research*, *39*(1), 83–89. https://doi.org/10.1016/j.jglr.2012.11.004

Brister, J. R., Ako-Adjei, D., Bao, Y., & Blinkova, O. (2015). NCBI viral genomes resource. *Nucleic Acids Research*, *43*(Database issue), D571-577. https://doi.org/10.1093/nar/gku1207

Brussaard, C. P. D., Wilhelm, S. W., Thingstad, F., Weinbauer, M. G., Bratbak, G., Heldal, M., Kimmance, S. A., Middelboe, M., Nagasaki, K., Paul, J. H., Schroeder, D. C., Suttle, C. A., Vaqué, D., & Wommack, K. E. (2008). Global-scale processes with a nanoscale drive: The role of marine viruses. *The ISME Journal*, *2*(6), 575–578. https://doi.org/10.1038/ismej.2008.31

Buchan, A., LeCleir, G. R., Gulvik, C. A., & González, J. M. (2014). Master recyclers: Features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, *12*(10), Article 10. https://doi.org/10.1038/nrmicro3326

Buckling, A., & Rainey, P. B. (2002). Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings. Biological Sciences*, *269*(1494), 931–936. https://doi.org/10.1098/rspb.2001.1945

Chen, J., & Novick, R. P. (2009). Phage-mediated intergeneric transfer of toxin genes. *Science (New York, N.Y.)*, *323*(5910), 139–141. https://doi.org/10.1126/science.1164783

Cook, K. V., Li, C., Cai, H., Krumholz, L. R., Hambright, K. D., Paerl, H. W., Steffen, M. M., Wilson, A. E., Burford, M. A., Grossart, H.-P., Hamilton, D. P., Jiang, H., Sukenik, A., Latour, D., Meyer, E. I., Padisák, J., Qin, B., Zamor, R. M., & Zhu, G. (2020). The global Microcystis interactome. *Limnology and Oceanography*, *65*(S1), S194–S207. https://doi.org/10.1002/lno.11361

Corel, E., Lopez, P., Méheust, R., & Bapteste, E. (2016). Network-Thinking: Graphs to Analyze Microbial Complexity and Evolution. *Trends in Microbiology*, *24*(3), 224–237. https://doi.org/10.1016/j.tim.2015.12.003

Coutinho, F. H., Gregoracci, G. B., Walter, J. M., Thompson, C. C., & Thompson, F. L. (2018). Metagenomics Sheds Light on the Ecology of Marine Microbes and Their Viruses. *Trends in Microbiology*, *26*(11), 955–965. https://doi.org/10.1016/j.tim.2018.05.015
Daniel, R. (2005). The metagenomics of soil. *Nature Reviews. Microbiology*, *3*(6), 470–478. https://doi.org/10.1038/nrmicro1160
Davenport, E. J., Neudeck, M. J., Matson, P. G., Bullerjahn, G. S., Davis, T. W., Wilhelm, S. W., Denney, M. K., Krausfeldt, L. E., Stough, J. M. A., Meyer, K. A., Dick, G. J., Johengen, T. H., Lindquist, E., Tringe, S. G., & McKay, R. M. L. (2019). Metatranscriptomic Analyses of Diel Metabolic Functions During a Microcystis Bloom in Western Lake Erie (United States). *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/articles/10.3389/fmicb.2019.02081
Davis, T. W., Harke, M. J., Marcoval, M. A., Goleski, J., Orano-Dawson, C., Berry, D. L., & Gobler, C. J. (2010). Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of Microcystis during cyanobacterial blooms. *Aquatic Microbial Ecology*, *61*(2), 149–162. https://doi.org/10.3354/ame01445
Delmont, T. O., Malandain, C., Prestat, E., Larose, C., Monier, J.-M., Simonet, P., & Vogel, T. M. (2011). Metagenomic mining for microbiologists. *The ISME Journal*, *5*(12), Article 12. https://doi.org/10.1038/ismej.2011.61
Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews. Microbiology*, *18*(3), 125–138. https://doi.org/10.1038/s41579-019-0311-5
Dolan, D. M., & Chapra, S. C. (2012). Great Lakes total phosphorus revisited: 1. Loading analysis and update (1994–2008). *Journal of Great Lakes Research*, *38*(4), 730–740. https://doi.org/10.1016/j.jglr.2012.10.001
Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A. M., Berger, A., Burguière, A.-M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.-C., Müller, S., … Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *The New England Journal of Medicine*, *348*(20), 1967–1976. https://doi.org/10.1056/NEJMoa030747
Duhaime, M. B., & Sullivan, M. B. (2012). Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology*, *434*(2), 181–186. https://doi.org/10.1016/j.virol.2012.09.036
Dutilh, B. E. (2014). Metagenomic ventures into outer sequence space. *Bacteriophage*, *4*(4), e979664. https://doi.org/10.4161/21597081.2014.979664
Edwards, R. A., McNair, K., Faust, K., Raes, J., & Dutilh, B. E. (2016). Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews*. https://doi.org/10.1093/femsre/fuv048
Edwards, R. A., & Rohwer, F. (2005). Opinion: Viral metagenomics. *Nature Reviews Microbiology*, *3*(6), 504–510. https://doi.org/10.1038/nrmicro1163
Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. B., Singleton, C. M., Solden, L. M., Naas, A. E., Boyd, J. A., Hodgkins, S. B., Wilson, R. M., Trubl, G., Li, C., Frolking, S., Pope, P. B., Wrighton, K. C., Crill, P. M., Chanton, J. P., … Sullivan, M. B. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology*, *3*(8), 870–880. https://doi.org/10.1038/s41564-018-0190-y
Enav, H., Kirzner, S., Lindell, D., Mandel-Gutfreund, Y., & Béjà, O. (2018). Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. *Nature Communications*, *9*, 4698. https://doi.org/10.1038/s41467-018-07164-3
Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., & Weitz, J. S. (2011). Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences*, *108*(28). https://doi.org/10.1073/pnas.1101595108

Flores, C. O., Valverde, S., & Weitz, J. S. (2013). Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal*, *7*(3), Article 3. https://doi.org/10.1038/ismej.2012.135

French, R., Cargnelli, L., & Doyle, M.-C. (n.d.). *Binational nutrient management*.

Fuhrman. (1999). *Marine viruses and their biogeochemical and ecological effects | Nature*. https://www.nature.com/articles/21119

Gałan, W., Bąk, M., & Jakubowska, M. (2019). Host Taxon Predictor—A Tool for Predicting Taxon of the Host of a Newly Discovered Virus. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-39847-2

Ghosh, D., Roy, K., Williamson, K. E., White, D. C., Wommack, K. E., Sublette, K. L., & Radosevich, M. (2008). Prevalence of Lysogeny among Soil Bacteria and Presence of 16S rRNA and trzN Genes in Viral-Community DNA. *Applied and Environmental Microbiology*, *74*(2), 495–502. https://doi.org/10.1128/AEM.01435-07

Goldhill, D. H., & Turner, P. E. (2014). The evolution of life history trade-offs in viruses. *Current Opinion in Virology*, *8*, 79–84. https://doi.org/10.1016/j.coviro.2014.07.005

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., … Sullivan, M. B. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, *177*(5), 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, *35*(Web Server issue), W52-57. https://doi.org/10.1093/nar/gkm360

Han, J., Jeon, B., Futatsugi, N., & Park, H.-D. (2013). The effect of alum coagulation for in-lake treatment of toxic Microcystis and other cyanobacteria related organisms in microcosm experiments. *Ecotoxicology and Environmental Safety*, *96*, 17–23. https://doi.org/10.1016/j.ecoenv.2013.06.008

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry & Biology*, *5*(10), R245-9. https://doi.org/10.1016/S1074-5521(98)90108-9

Harke, M. J., & Gobler, C. J. (2013). Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. *PLOS ONE*, *8*(7), e69834. https://doi.org/10.1371/journal.pone.0069834

Harke, M. J., & Gobler, C. J. (2015). Daily transcriptome changes reveal the role of nitrogen in controlling microcystin synthesis and nutrient transport in the toxic cyanobacterium, Microcystis aeruginosa. *BMC Genomics*, *16*(1), 1068. https://doi.org/10.1186/s12864-015-2275-9

Harke, M. J., Steffen, M. M., Gobler, C. J., Otten, T. G., Wilhelm, S. W., Wood, S. A., & Paerl, H. W. (2016). A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae*, *54*, 4–20. https://doi.org/10.1016/j.hal.2015.12.007

Hegarty, B., V, J. R., Bastien, E., Langenfeld, K., Lindback, M., Saini, J. S., Wing, A., Zhang, J., & Duhaime, M. (2023). *Benchmarking informatics approaches for virus discovery: Caution is needed when combining in silico identification methods* (p. 2023.08.07.552334). bioRxiv. https://doi.org/10.1101/2023.08.07.552334

Heidelberg, J. F., Nelson, W. C., Schoenfeld, T., & Bhaya, D. (2009). Germ Warfare in a Microbial Mat Community: CRISPRs Provide Insights into the Co-Evolution of Host and Viral Genomes. *PLOS ONE*, *4*(1), e4169. https://doi.org/10.1371/journal.pone.0004169

Held, N. L., Herrera, A., & Whitaker, R. J. (2013). Reassortment of CRISPR repeat-spacer loci in Sulfolobus islandicus. *Environmental Microbiology*, *15*(11), 3065–3076. https://doi.org/10.1111/1462-2920.12146

Henry, M., & Debarbieux, L. (2012). Tools from viruses: Bacteriophage successes and beyond. *Virology*, *434*(2), 151–161. https://doi.org/10.1016/j.virol.2012.09.017

Hewson, I., O'Neil, J. M., Fuhrman, J. A., & Dennison, W. C. (2001). Virus-like particle distribution and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. *Limnology and Oceanography*, *46*(7), 1734–1746. https://doi.org/10.4319/lo.2001.46.7.1734

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME Journal*, *11*(7). https://doi.org/10.1038/ismej.2017.16

Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H., Andreopoulos, B., Brewer, H. M., Glavina del Rio, T., Adkins, J. N., Paul, S., Sullivan, M. B., & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *The ISME Journal*, *14*(4), Article 4. https://doi.org/10.1038/s41396-019-0580-z

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., & Le Mercier, P. (2011). ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Research*, *39*(Database issue), D576-582. https://doi.org/10.1093/nar/gkq901

Hurwitz, B. L., Hallam, S. J., & Sullivan, M. B. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology*, *14*(11), R123. https://doi.org/10.1186/gb-2013-14-11-r123

Iranzo, J., Koonin, E. V., Prangishvili, D., & Krupovic, M. (2016). Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements. *Journal of Virology*, *90*(24), 11043–11055. https://doi.org/10.1128/JVI.01622-16

Iranzo, J., Krupovic, M., & Koonin, E. V. (2016). The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. *mBio*, *7*(4), e00978-16. https://doi.org/10.1128/mBio.00978-16

Iranzo, J., Krupovic, M., & Koonin, E. V. (2017). A network perspective on the virus world. *Communicative & Integrative Biology*, *10*(2), e1296614. https://doi.org/10.1080/19420889.2017.1296614

Janzen, D. H. (1980). WHEN IS IT COEVOLUTION? *Evolution; International Journal of Organic Evolution*, *34*(3), 611–612. https://doi.org/10.1111/j.1558-5646.1980.tb04849.x

Jiang, S. C., & Paul, J. H. (1998). Gene transfer by transduction in the marine environment. *Applied and Environmental Microbiology*, *64*(8), 2780–2787. https://doi.org/10.1128/AEM.64.8.2780-2787.1998

Jochimsen, E. M., Carmichael, W. W., An, J. S., Cardo, D. M., Cookson, S. T., Holmes, C. E., Antunes, M. B., de Melo Filho, D. A., Lyra, T. M., Barreto, V. S., Azevedo, S. M., & Jarvis, W. R. (1998). Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *The New England Journal of Medicine*, *338*(13), 873–878. https://doi.org/10.1056/NEJM199803263381304

Kauffman, K. M., Chang, W. K., Brown, J. M., Hussain, F. A., Yang, J., Polz, M. F., & Kelly, L. (2022). Resolving the structure of phage–bacteria interactions in the context of natural diversity. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-021-27583-z

Kim, M., Shin, B., Lee, J., Park, H. Y., & Park, W. (2019). Culture-independent and culture-dependent analyses of the bacterial community in the phycosphere of cyanobloom-forming Microcystis aeruginosa. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-56882-1

Kimura, S., Sako, Y., & Yoshida, T. (2013). Rapid microcystis cyanophage gene diversification revealed by long- and short-term genetic analyses of the tail sheath gene in a natural pond. *Applied and Environmental Microbiology*, *79*(8), 2789–2795. https://doi.org/10.1128/AEM.03751-12

Kimura-Sakai, S., Sako, Y., & Yoshida, T. (2015). Development of a real-time PCR assay for the quantification of Ma-LMM01-type Microcystis cyanophages in a natural pond. *Letters in Applied Microbiology*, *60*(4), 400–408. https://doi.org/10.1111/lam.12387

Knowles, B., Bailey, B., Boling, L., Breitbart, M., Cobián-Güemes, A., Del Campo, J., Edwards, R., Felts, B., Grasis, J., Haas, A. F., Katira, P., Kelly, L. W., Luque, A., Nulton, J., Paul, L., Peters, G., Robinett, N., Sandin, S., Segall, A., … Rohwer, F. (2017). Variability and host density independence in inductions-based estimates of environmental lysogeny. *Nature Microbiology*. https://doi.org/10.1038/nmicrobiol.2017.64

Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobián-Güemes, A. G., Coutinho, F. H., Dinsdale, E. A., Felts, B., Furby, K. A., George, E. E., Green, K. T., Gregoracci, G. B., Haas, A. F., Haggerty, J. M., Hester, E. R., Hisakawa, N., Kelly, L. W., Lim, Y. W., … Rohwer, F. (2016). Lytic to temperate switching of viral communities. *Nature*, *531*(7595), 466–470. https://doi.org/10.1038/nature17193

Koskella, B., & Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, *38*(5). https://doi.org/10.1111/1574-6976.12072

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews: MMBR*, *72*(4), 557–578, Table of Contents. https://doi.org/10.1128/MMBR.00009-08

Kuno, S., Sako, Y., & Yoshida, T. (2014). Diversification of CRISPR within coexisting genotypes in a natural population of the bloom-forming cyanobacterium Microcystis aeruginosa. *Microbiology (Reading, England)*, *160*(5), 903–916. https://doi.org/10.1099/mic.0.073494-0

Kuno, S., Yoshida, T., Kaneko, T., & Sako, Y. (2012). Intricate interactions between the bloom-forming cyanobacterium Microcystis aeruginosa and foreign genetic elements, revealed by diversified clustered regularly interspaced short palindromic repeat(CRISPR) signatures. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.00626-12

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews. Microbiology*, *8*(5), 317–327. https://doi.org/10.1038/nrmicro2315

Laine, A., & Tellier, A. (2008). Heterogeneous selection promotes maintenance of polymorphism in host–parasite interactions. *Oikos*, *117*(9), 1281–1288. https://doi.org/10.1111/j.0030-1299.2008.16563.x

Lewis, C. F., Cameron, G. D. M., Anderson, T., Jr, C., & Gareau, P. (2012). Lake levels in Erie Basin of the Laurentian Great Lakes. *Journal of Paleolimnology*, *47*. https://doi.org/10.1007/s10933-012-9578-5

Lima-Mendez, G., Van Helden, J., Toussaint, A., & Leplae, R. (2008). Reticulate Representation of Evolutionary and Functional Relationships between Phage Genomes. *Molecular Biology and Evolution*, *25*(4), 762–777. https://doi.org/10.1093/molbev/msn023

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., & Crosby, L. (2018). Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems*, *3*(5), 10.1128/msystems.00055-18. https://doi.org/10.1128/msystems.00055-18

Low, S. J., Džunková, M., Chaumeil, P.-A., Parks, D. H., & Hugenholtz, P. (2019). Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nature Microbiology*, *4*(8), Article 8. https://doi.org/10.1038/s41564-019-0448-z

Makarova, K. S., Anantharaman, V., Aravind, L., & Koonin, E. V. (2012). Live virus-free or die: Coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biology Direct*, *7*(1), 40. https://doi.org/10.1186/1745-6150-7-40

Makarova, K. S., & Koonin, E. V. (2013). Evolution and Classification of CRISPR-Cas Systems and Cas Protein Families. In R. Barrangou & J. van der Oost (Eds.), *CRISPR-Cas Systems: RNA-mediated Adaptive Immunity in Bacteria and Archaea* (pp. 61–91). Springer. https://doi.org/10.1007/978-3-642-34657-6_3
Marbouty, M., Baudry, L., Cournac, A., & Koszul, R. (2017). Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Science Advances*, *3*(2), e1602105. https://doi.org/10.1126/sciadv.1602105
Marbouty, M., Thierry, A., Millot, G. A., & Koszul, R. (2021). MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. *eLife*, *10*, e60608. https://doi.org/10.7554/eLife.60608
Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M. J., Martínez, J. M., Anton, J., Gasol, J. M., Rosselli, R., Rodriguez-Valera, F., Sullivan, M. B., Acinas, S. G., & Martinez-Garcia, M. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms15892
McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B., & Paul, J. H. (2010). High frequency of horizontal gene transfer in the oceans. *Science (New York, N.Y.)*, *330*(6000), 50. https://doi.org/10.1126/science.1192243
McDaniel, L., & Paul, J. H. (2005). Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine synechococcus species. *Applied and Environmental Microbiology*, *71*(2), 842–850. https://doi.org/10.1128/AEM.71.2.842-850.2005
McLaren, M. R., Willis, A. D., & Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *eLife*, *8*, e46923. https://doi.org/10.7554/eLife.46923
Mestre, M., Borrull, E., Sala, M. M., & Gasol, J. M. (2017). Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *The ISME Journal*, *11*(4), Article 4. https://doi.org/10.1038/ismej.2016.166
Middelboe, M., Jorgensen, N., & Kroer, N. (1996). Effects of viruses on nutrient turnover and growth efficiency of noninfected marine bacterioplankton. *Applied and Environmental Microbiology*, *62*(6), 1991–1997. https://doi.org/10.1128/aem.62.6.1991-1997.1996
Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., & Ogata, H. (2016). Linking Virus Genomes with Host Taxonomy. *Viruses*, *8*(3), Article 3. https://doi.org/10.3390/v8030066
Nagano, T., Lubling, Y., Yaffe, E., Wingett, S. W., Dean, W., Tanay, A., & Fraser, P. (2015). Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nature Protocols*, *10*(12), Article 12. https://doi.org/10.1038/nprot.2015.127
Ogilvie, L. A., & Jones, B. V. (2015). The human gut virome: A multifaceted majority. *Frontiers in Microbiology*, *6*, 918. https://doi.org/10.3389/fmicb.2015.00918
Paerl, H. W., & Otten, T. G. (2013). Harmful cyanobacterial blooms: Causes, consequences, and controls. *Microbial Ecology*, *65*(4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y
Paez-Espino, D., Morovic, W., Sun, C. L., Thomas, B. C., Ueda, K., Stahl, B., Barrangou, R., & Banfield, J. F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature Communications*, *4*(1), Article 1. https://doi.org/10.1038/ncomms2440
Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), Article 10. https://doi.org/10.1038/nbt.4229

Payne, P., Geyrhofer, L., Barton, N. H., & Bollback, J. P. (2018). CRISPR-based herd immunity can limit phage epidemics in bacterial populations. *eLife*, *7*, e32035. https://doi.org/10.7554/eLife.32035
Pratama, A. A., & van Elsas, J. D. (2018). The "Neglected" Soil Virome—Potential Role and Impact. *Trends in Microbiology*, *26*(8), 649–662. https://doi.org/10.1016/j.tim.2017.12.004
Pride, D. T., Salzman, J., & Relman, D. A. (2012). Comparisons of clustered regularly interspaced short palindromic repeats and viromes in human saliva reveal bacterial adaptations to salivary viruses. *Environmental Microbiology*. https://doi.org/10.1111/j.1462-2920.2012.02775.x
Ptashne. (n.d.). *A Genetic Switch, Third Edition, Phage Lambda—ProQuest*. Retrieved January 10, 2024, from https://www.proquest.com/docview/207668229?sourcetype=Scholarly%20Journals
Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, *5*(1), 69. https://doi.org/10.1186/s40168-017-0283-5
Ren, J., Song, K., Deng, C., Ahlgren, N. A., Fuhrman, J. A., Li, Y., Xie, X., Poplin, R., & Sun, F. (2020). Identifying viruses from metagenomic data using deep learning. *Quantitative Biology (Beijing, China)*, *8*(1), 64–77. https://doi.org/10.1007/s40484-019-0187-4
Rieck, A., Herlemann, D. P. R., Jürgens, K., & Grossart, H.-P. (2015). Particle-Associated Differ from Free-Living Bacteria in Surface Waters of the Baltic Sea. *Frontiers in Microbiology*, *6*. https://www.frontiersin.org/articles/10.3389/fmicb.2015.01297
Rosenwasser, S., Ziv, C., Creveld, S. G. van, & Vardi, A. (2016). Virocell Metabolism: Metabolic Innovations During Host-Virus Interactions in the Ocean. *Trends in Microbiology*, *24*(10), 821–832. https://doi.org/10.1016/j.tim.2016.06.006
Rostøl, J. T., & Marraffini, L. (2019). (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host & Microbe*, *25*(2), 184–194. https://doi.org/10.1016/j.chom.2019.01.009
Roux and Bolduc. (2016). *MAVERICLab / stampede-clustergenomes—Bitbucket*. https://bitbucket.org/MAVERICLab/stampede-clustergenomes/src/master/
Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Lavigne, R., Brister, J. R., Varsani, A., Amid, C., Aziz, R. K., Bordenstein, S. R., Bork, P., Breitbart, M., Cochrane, G. R., Daly, R. A., Desnues, C., Duhaime, M. B., … Eloe-Fadrosh, E. A. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology*, *37*(1), Article 1. https://doi.org/10.1038/nbt.4306
Roux, S., & Brum, J. R. (2019). A viral reckoning: Viruses emerge as essential manipulators of global ecosystems. *Environmental Microbiology Reports*, *11*(1), 3–8. https://doi.org/10.1111/1758-2229.12700
Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., & Sullivan, M. B. (2017). Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*, *5*. https://doi.org/10.7717/peerj.3817
Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: Mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. https://doi.org/10.7717/peerj.985
Schindler, D. W. (2012). The dilemma of controlling cultural eutrophication of lakes. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1746), 4322–4333. https://doi.org/10.1098/rspb.2012.1032
Schmidt, R., Cordovez, V., de Boer, W., Raaijmakers, J., & Garbeva, P. (2015). Volatile affairs in microbial interactions. *The ISME Journal*, *9*(11), Article 11. https://doi.org/10.1038/ismej.2015.42
Shan, J., Korbsrisate, S., Withatanung, P., Adler, N. L., Clokie, M. R. J., & Galyov, E. E. (2014). Temperature dependent bacteriophages of a tropical bacterial pathogen. *Frontiers in Microbiology*, *5*, 599. https://doi.org/10.3389/fmicb.2014.00599

Sieradzki, E. T., Ignacio-Espinoza, J. C., Needham, D. M., Fichot, E. B., & Fuhrman, J. A. (2019). Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09106-z

Smith, D. J., Tan, J. Y., Powers, M. A., Lin, X. N., Davis, T. W., & Dick, G. J. (2021). Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. *Environmental Microbiology*, *23*(6), 3020–3036. https://doi.org/10.1111/1462-2920.15514

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

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews. Genetics*, *16*(8), 472–482. https://doi.org/10.1038/nrg3962

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., & Wommack, K. E. (2008). Phages across the biosphere: Contrasts of viruses in soil and aquatic environments. *Research in Microbiology*, *159*(5), 349–357. https://doi.org/10.1016/j.resmic.2008.04.010

Staals, R. H. J., & Brouns, S. J. J. (2013). Distribution and Mechanism of the Type I CRISPR-Cas Systems. In R. Barrangou & J. van der Oost (Eds.), *CRISPR-Cas Systems: RNA-mediated Adaptive Immunity in Bacteria and Archaea* (pp. 145–169). Springer. https://doi.org/10.1007/978-3-642-34657-6_6

Steffen, M. M., Dearth, S. P., Dill, B. D., Li, Z., Larsen, K. M., Campagna, S. R., & Wilhelm, S. W. (2014). Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis. *The ISME Journal*, *8*(10), 2080–2092. https://doi.org/10.1038/ismej.2014.78

Stern, A., & Sorek, R. (2010). The phage-host arms race: Shaping the evolution of microbes. *BioEssays*, *33*(1), Article LBNL-7066E. https://doi.org/10.1002/bies.201000071

Stumpf, R. P., Wynne, T. T., Baker, D. B., & Fahnenstiel, G. L. (2012). Interannual Variability of Cyanobacterial Blooms in Lake Erie. *PLOS ONE*, *7*(8), e42444. https://doi.org/10.1371/journal.pone.0042444

Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., & Chisholm, S. W. (2005). Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLOS Biology*, *3*(5), e144. https://doi.org/10.1371/journal.pbio.0030144

Sullivan, M. B., Waterbury, J. B., & Chisholm, S. W. (2003). Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature*, *424*(6952), Article 6952. https://doi.org/10.1038/nature01929

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), Article 7057. https://doi.org/10.1038/nature04160

Suttle, C. A. (2007). Marine viruses—Major players in the global ecosystem. *Nature Reviews Microbiology*, *5*(10), 801–812. https://doi.org/10.1038/nrmicro1750

Sutton, T. D. S., Clooney, A. G., Ryan, F. J., Ross, R. P., & Hill, C. (2019). Choice of assembly software has a critical impact on virome characterisation. *Microbiome*, *7*(1), 12. https://doi.org/10.1186/s40168-019-0626-5

Suzuki, S., Kaneko, R., Kodama, T., Hashihama, F., Suwa, S., Tanita, I., Furuya, K., & Hamasaki, K. (2017). Comparison of community structures between particle-associated and free-living prokaryotes in tropical and subtropical Pacific Ocean surface waters. *Journal of Oceanography*, *73*(3), 383–395. https://doi.org/10.1007/s10872-016-0410-0

Tapper, M. A., & Hicks, R. E. (1998). Temperate viruses and lysogeny in Lake Superior bacterioplankton. *Limnology and Oceanography*, *43*(1), 95–103. https://doi.org/10.4319/lo.1998.43.1.0095

Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, *45*(6), 1320–1328. https://doi.org/10.4319/lo.2000.45.6.1320
Thomas, E. L., & Corbel, M. J. (1977). Isolation of a phage lytic for several Brucella species following propagation of Tbilisi phage in the presence of mitomycin C. *Archives of Virology*, *54*(3), 259–261. https://doi.org/10.1007/BF01314793
Touchon, M., Bernheim, A., & Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal*, *10*(11), 2744–2754. https://doi.org/10.1038/ismej.2016.47
Touchon, M., & Rocha, E. P. C. (2010). The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. *PloS One*, *5*(6), e11126. https://doi.org/10.1371/journal.pone.0011126
Tylianakis, J. M., & Morris, R. J. (2017). Ecological Networks Across Environmental Gradients. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 25–48. https://doi.org/10.1146/annurev-ecolsys-110316-022821
Tyson, G. W., & Banfield, J. F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*, *10*(1), 200–207. https://doi.org/10.1111/j.1462-2920.2007.01444.x
UniProt Consortium. (2019a). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049
UniProt Consortium. (2019b). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049
Van Goethem, N., Descamps, T., Devleesschauwer, B., Roosens, N. H. C., Boon, N. A. M., Van Oyen, H., & Robert, A. (2019). Status and potential of bacterial genomics for public health practice: A scoping review. *Implementation Science*, *14*(1), 79. https://doi.org/10.1186/s13012-019-0930-2
van Houte, S., Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A., & Westra, E. R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature*, *532*(7599), 385–388. https://doi.org/10.1038/nature17436
Villarroel, J., Kleinheinz, K. A., Jurtz, V. I., Zschach, H., Lund, O., Nielsen, M., & Larsen, M. V. (2016). HostPhinder: A Phage Host Prediction Tool. *Viruses*, *8*(5), Article 5. https://doi.org/10.3390/v8050116
Volkmer, B., & Heinemann, M. (2011). Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling. *PloS One*, *6*(7), e23126. https://doi.org/10.1371/journal.pone.0023126
Warwick-Dugdale, J., Buchholz, H. H., Allen, M. J., & Temperton, B. (2019). Host-hijacking and planktonic piracy: How phages command the microbial high seas. *Virology Journal*, *16*(1), 15. https://doi.org/10.1186/s12985-019-1120-1
Waterbury, J. B., & Valois, F. W. (1993). Resistance to co-occurring phages enables marine synechococcus communities to coexist with cyanophages abundant in seawater. *Applied and Environmental Microbiology*, *59*(10), 3393–3399. https://doi.org/10.1128/aem.59.10.3393-3399.1993
Weitz, J. S., Beckett, S. J., Brum, J. R., Cael, B. B., & Dushoff, J. (2017). Lysis, lysogeny and virus–microbe ratios. *Nature 2017 549:7672*, *549*(7672), E1–E1. https://doi.org/10.1038/nature23295
Weitz, J. S., Stock, C. A., Wilhelm, S. W., Bourouiba, L., Coleman, M. L., Buchan, A., Follows, M. J., Fuhrman, J. A., Jover, L. F., Lennon, J. T., Middelboe, M., Sonderegger, D. L., Suttle, C. A., Taylor, B. P., Frede Thingstad, T., Wilson, W. H., & Eric Wommack, K. (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and

ecosystem processes. *The ISME Journal*, *9*(6), Article 6. https://doi.org/10.1038/ismej.2014.220
Weitz, J. S., & Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports*, *4*, 17. https://doi.org/10.3410/B4-17
Wilhelm, S. W., & Smith, R. E. (2000). Bacterial carbon production in Lake Erie is influenced by viruses and solar radiation. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*(2), 317–326. https://doi.org/10.1139/f99-202
Wilhelm, S. W., & Suttle, C. A. (1999). Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. *BioScience*, *49*(10), 781–788. https://doi.org/10.2307/1313569
Willner, D., & Hugenholtz, P. (2013). From deep sequencing to viral tagging: Recent advances in viral metagenomics. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *35*(5), 436–442. https://doi.org/10.1002/bies.201200174
Wing, A.J, Hegarty, B., Bastien, G.E., Evans, J., Cable, R., Smith, D., NOAA, Denef, V., Dick, G., & Duhaime, M.B. "Tracking Microcystis Viruses and Infection Dynamics in a Multi-Peak Microcystis-Dominated Bloom." [Submitted to PLoS Biology]
Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., Roth, J. R., & Bäumler, A. J. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, *467*(7314), Article 7314. https://doi.org/10.1038/nature09415
Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(12), 4576–4579.
Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: Viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews: MMBR*, *64*(1), 69–114. https://doi.org/10.1128/MMBR.64.1.69-114.2000
Wortman, S. (2018). *U.S. Action Plan for Lake Erie*.
Wu, L.-Y., Pappas, N., Wijesekara, Y., Piedade, G. J., Brussaard, C. P. D., & Dutilh, B. E. (2023). *Benchmarking Bioinformatic Virus Identification Tools Using Real-World Metagenomic Data across Biomes* [Preprint]. Bioinformatics. https://doi.org/10.1101/2023.04.26.538077
Yoshida, M., Yoshida, T., Kashima, A., Takashima, Y., Hosoda, N., Nagasaki, K., & Hiroishi, S. (2008). Ecological dynamics of the toxic bloom-forming cyanobacterium Microcystis aeruginosa and its cyanophages in freshwater. *Applied and Environmental Microbiology*, *74*(10), 3269–3273. https://doi.org/10.1128/AEM.02240-07
Yoshida-Takashima, Y., Yoshida, M., Ogata, H., Nagasaki, K., Hiroishi, S., & Yoshida, T. (2012). Cyanophage infection in the bloom-forming cyanobacteria Microcystis aeruginosa in surface freshwater. *Microbes and Environments*, *27*(4), 350–355. https://doi.org/10.1264/JSME2.ME12037
Zablocki, O., Michelsen, M., Burris, M., Solonenko, N., Warwick-Dugdale, J., Ghosh, R., Pett-Ridge, J., Sullivan, M. B., & Temperton, B. (2021). VirION2: A short- and long-read sequencing and informatics workflow to study the genomic diversity of viruses in nature. *PeerJ*, *9*, e11088. https://doi.org/10.7717/peerj.11088
Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., Waldbauer, J. R., & Coleman, M. L. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews. Microbiology*, *18*(1), 21–34. https://doi.org/10.1038/s41579-019-0270-x

Chapter 2:Tracking *Microcystis* Viruses and Infection Dynamics in a Multi-Peak *Microcystis*-Dominated Bloom

A.J Wing[1], Bridget Hegarty[1,2], Eric Bastien[1], Vincent Denef[1], Jacob Evans[1], Derek Smith[3], Gregory Dick[1,3], Melissa Duhaime[1]

[1] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI

[2] Department of Civil and Environmental Engineering, Case Western Reserve University, Cleveland, OH

[3] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI

Abstract

Given the impact of viruses on microbial community composition and function, viruses have the potential to play a significant role in the fate of freshwater cyanobacterial harmful algal blooms (cHABs). Yet the role of viruses in complex bloom communities remains poorly understood. As the frequency and intensity of cHABs are increasing globally, we sought to address this knowledge gap by tracking viruses of bloom-forming *Microcystis aeruginosa* through a cHAB in the western basin of Lake Erie. We identified *Microcystis* virus Ma-LEF01, a relative of the well-studied *Microcystis* virus Ma-LMM01, and tracked the temporal succession of its population variants through the Lake Erie bloom, highlighting the local provenance and persistence of Ma-LEF01-like viruses in the lake over a five year period. Size-fractionation of the water allowed us to identify significant fraction-specific trends in viral diversity, which corresponded with *Microcystis* genetic diversity. Using a new machine-learning model, we predicted infections between viral and microbial host populations. We found hundreds of viral populations shared between *Microcystis* and non-*Microcystis* hosts, suggesting extensive interconnectivity and the potential for virus-mediated cross-species exchange of genetic material within

cHABs communities. Abundant viral genes belonging to predicted *Microcystis* viruses revealed their potential role in key metabolic pathways involved in carbohydrate biosynthesis, photosynthesis, nitrogen metabolism, and adaptation to environmental changes. These findings advance our understanding of uncultivated *Microcystis* virus diversity, their potential effects on host metabolism, and their potential influence on the complex microbial communities associated with *Microcystis*-dominated cHABs.

Importance

Understanding interactions between viruses, their hosts, and environmental parameters is central to identifying the triggers and mechanisms underlying the onset, persistence, and demise of cyanobacterial harmful algal blooms. In this study we describe the viral diversity, metabolic potential, and host ranges of viruses predicted to infect *Microcystis*, describing the distribution of these properties across time, space, and different bloom-associated size fractions. These findings contribute to a better understanding of the interplay between viruses, *Microcystis*, and their accompanying bacterial communities, shedding light on the mechanisms driving bloom dynamics, species interactions, and coevolutionary processes.

2.1 Introduction

Microcystis aeruginosa is a cyanobacterium that can form toxic blooms in freshwater and estuarine systems worldwide (Yoshida et al., 2008; Preece et al., 2017). Microcystins, the most prolific in a suite of toxins produced by *Microcystis aeruginosa* (Perez-Carrascal et al., 2019), have toxic effects on humans and diminish drinking water quality and overall aquatic ecosystem health (Paerl et al., 2013b; Steffen et al., 2017; Huisman et al., 2018). The frequency and intensity of *M. aeruginosa* blooms are increasing (Paerl and Huisman, 2009; Harke et al., 2016), largely due to climate change and eutrophication of aquatic habitats (Paerl and Huisman, 2008; O'Neil et al., 2012; Michalek et al., 2013; Pearl and Otten; 2013b; Visser et al., 2016). While the role of such abiotic controls on *Microcystis* bloom progression has been the focus of study for years, elucidating the role of biotic controls, such as predator-prey relationships, has

been less explored, partly owing to the challenges of studying microbial interactions in complex community contexts.

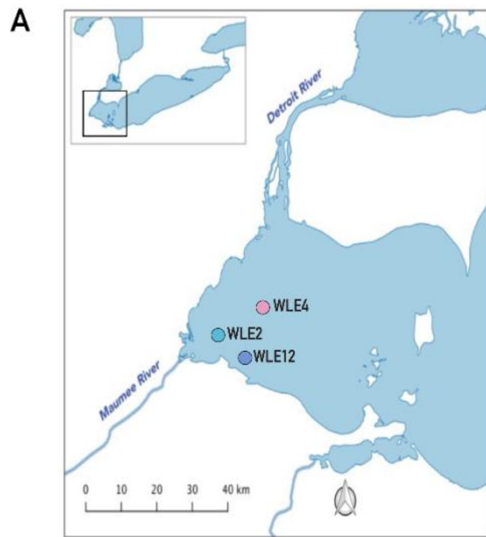
Viruses profoundly influence microbial communities by infecting and lysing microbial host populations (Fuhrman, 1999; Suttle, 2007; Weitz & Wilhelm, 2012; Koskella & Brockhurst, 2014), reprogramming host metabolisms (Breitbart, 2011; Hurwitz & Sullivan, 2013; Rosenwasser et al., 2016; Enav, 2018; Howard-Varona et al., 2020; Zimmerman et al, 2020), and facilitating gene transfer (McDaniel et al., 2010; Soucy et al., 2015). Evidence suggests that *Microcystis* in the annual Lake Erie cHABs of the lake's western basin is susceptible to viral infection and subsequent lysis (Steffen et al., 2017; Jiang et al., 2019; McKindles et al., 2020), and release of microcystins from cells due to viral lysis contributed to the Toledo drinking water crisis in 2014 (Steffen et al., 2017). Yet, much of what is known of *Microcystis* viruses is through tracking the abundance and distribution of viral marker genes of a single *Microcystis* virus isolate, Ma-LMM01, and its close relatives (e.g. Yoshida et al., 2007; Mankiewicz-Boczek et al., 2016; Steffen et al., 2017; McKindles et al., 2020; Rozon and Short; 2013; Yoshida-Takashima et al., 2012). In contrast, a community genomic approach captures the complex system of microbial populations interacting with one another, their viral predators, and their environment.

In this study, we sought to address (1) How do the diversity and distribution of viruses infecting *Microcystis* vary across time, space, and different size fractions during a cHAB in Lake Erie, and (2) What are the possible implications of *Microcystis* virus infection in terms of host metabolisms and gene flow with other host taxa? With a combination of metagenomic analyses of viral and cellular communities, we set out to detail the viral diversity and potential host range of viruses predicted to infect *Microcystis*, the distribution of these viruses across time, space, and different size fractions, and explored their metabolic potential. These findings will contribute to a better understanding of the mechanisms driving cHAB dynamics, species interactions, and potential coevolutionary processes.

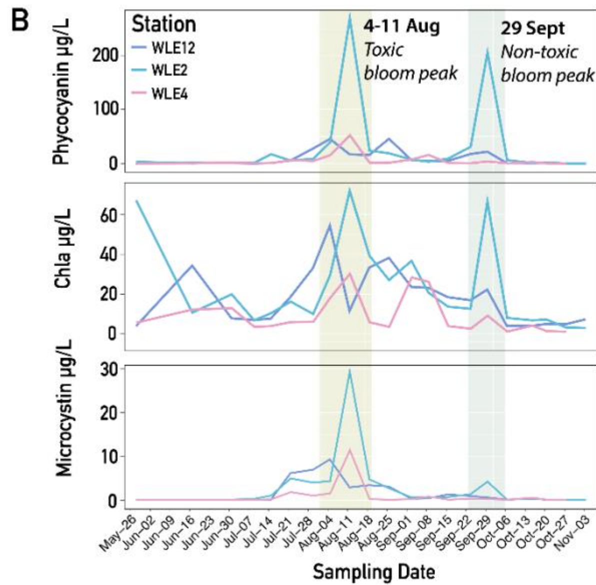
2.2 Results and Discussion

2.2.1 Tracking *Microcystis* during the 2014 Lake Erie cHAB

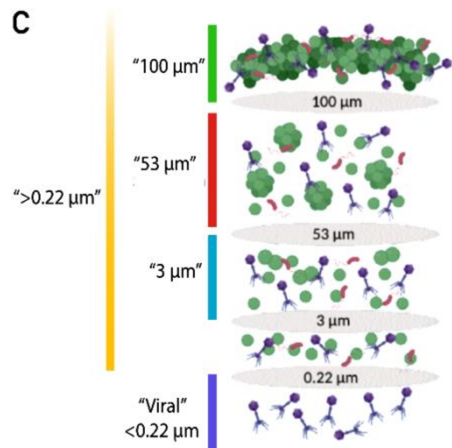
We analyzed metagenomic data obtained from cyanobacterial harmful algal blooms (cHABs) in the western region of Lake Erie (Fig. 5A) that persisted from July to October in 2014 (Cory et al., 2016; Berry et al., 2017; Smith et al., 2021). Concentrations of particulate phycocyanin (used as a proxy for cyanobacteria) and microcystin (indicative of bloom toxicity) revealed a toxic *Microcystis* bloom at all three sampling locations (WLE12, WLE2, WLE4) in early August (Fig. 5A-B). The cyanobacterial bloom, in late September occurred primarily at the nearshore stations (WLE12 and WLE2), showed lower microcystin concentrations than August 4, and was dominated by *Microcystis* genotypes that contained a partial operon of *mcy* genes or lacked *mcy* genes altogether (Yancey et al. 2022). To investigate the viruses and virus-host interactions associated with these blooms, we analyzed metagenomic data from each station at the two bloom peaks and across five size fractions chosen to target *Microcystis* in both free-living and colony forms and their viruses (Fig. 5C). From these data, we reconstructed 17 *Microcystis* metagenome assembled genomes ('MAGs'; Fig. 5D-E; SI Table 1) and 27,086 viral contigs >3 kb. Relative abundances of *Microcystis* ranged from 0-29%, based on proportion of the total reads in each sample that mapped to the *Microcystis* MAGs in the cellular fraction metagenomes (0.22-100 µm; Fig. 5E). Highest relative abundances of *Microcystis* MAGs were observed in the 100 µm fraction where large *Microcystis* colonies that typify blooms are expected. The greatest variability in abundances between *Microcystis* MAGs was also observed in this fraction (Fig. 5E), which we suspect is due to the sporadic rise and fall of different colony-forming *Microcystis* populations captured in the 100 µm fraction. Viral contigs identified across all size fractions were clustered into 15,461 viral operational taxonomic units (vOTUs), which approximate viral species, based on broadly accepted thresholds (95% ANI across 85% of the contig length (Roux, 2019).



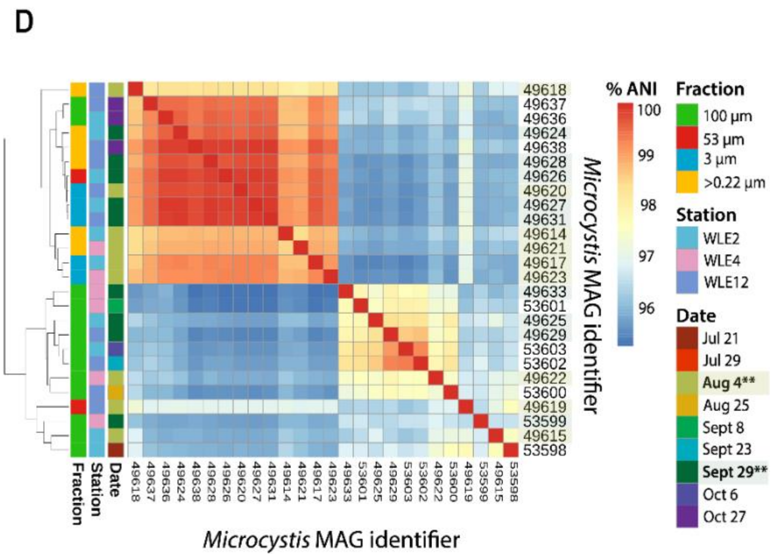
A
Detroit River
WLE4
WLE2
WLE12
Maumee River
0 10 20 30 40 km



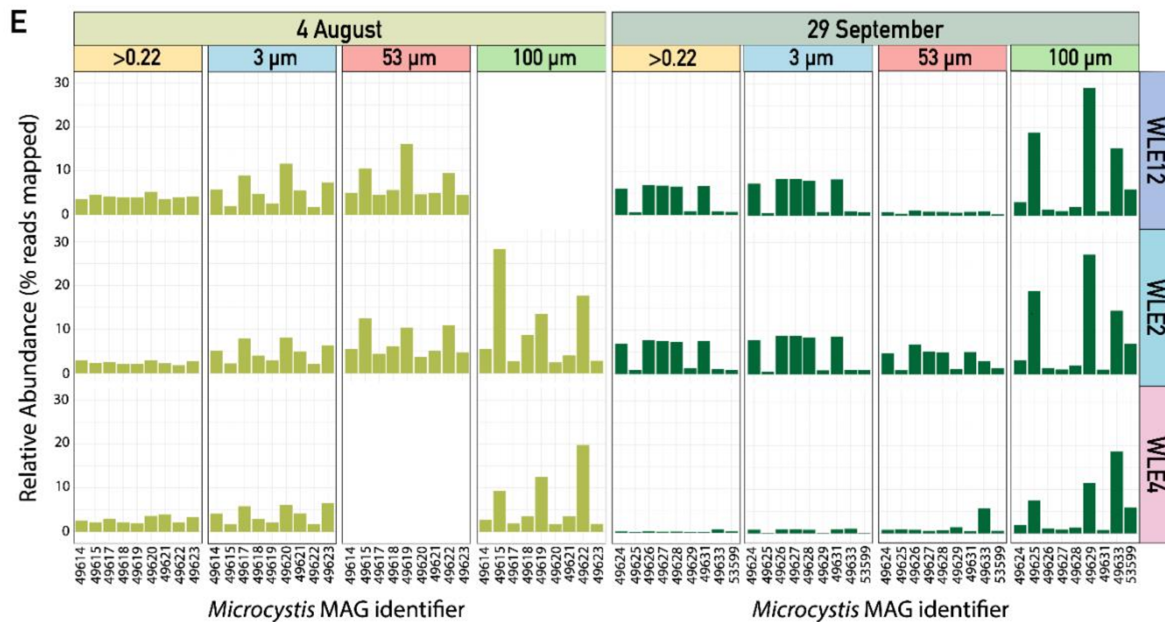
B
Station
WLE12
WLE2
WLE4
4-11 Aug
Toxic bloom peak
29 Sept
Non-toxic bloom peak
Phycocyanin µg/L
Chla µg/L
Microcystin µg/L
Sampling Date



C
"100 µm"
"53 µm"
"3 µm"
">0.22 µm"
"Viral" <0.22 µm
100 µm
53 µm
3 µm
0.22 µm



D
% ANI
Fraction
100 µm
53 µm
3 µm
>0.22 µm
Station
WLE2
WLE4
WLE12
Date
Jul 21
Jul 29
Aug 4**
Aug 25
Sept 8
Sept 23
Sept 29**
Oct 6
Oct 27
Microcystis MAG identifier
Fraction
Station
Date



E
4 August
29 September
>0.22
3 µm
53 µm
100 µm
WLE12
WLE2
WLE4
Relative Abundance (% reads mapped)
Microcystis MAG identifier

Figure 5. Lake Erie 2014 sampling overview and *Microcystis* MAG diversity and dynamics. (A) A map of sampling sites located in the western basin of Lake Erie. Three sites were sampled bi-monthly in June and weekly from July to October of 2014. (B) Phycocyanin, chlorophyll-a and microcystin measurements for the 2014 bloom. (C) Sampling filter size fractionation schematic of how samples were collected from the western basin of Lake Erie in 2014. (D) Heatmap of average nucleotide identity (% ANI) between 26 *Microcystis* MAGs reconstructed from samples July through August across three stations and four sample fractions. Bottom row and right column list the MAG identifiers. Sample fraction, station and date are identified by color of left panel. Dendrogram depicts the clustering of MAGs based on %ANI. (E) Relative abundances of *Microcystis* MAGs based on fraction of reads from each sample that mapped to each MAG in a competitive read mapping to all assembled MAGs from that sample. No data exist for 4 Aug 100 µm at WLE12 and 53 µm at WLE4.

2.2.2 Lake Erie-specific populations of known globally distributed *Microcystis* viruses

2.2.2.1 Four previously isolated *Microcystis* viruses identified in Lake Erie cHAB bloom peaks

We first sought to identify and track known *Microcystis* virus OTUs (vOTUs) in Lake Erie. To date, 10 freshwater *Microcystis* viral lab isolates have been sequenced and described (Tucker and Pollard, 2005; Yoshida et al., 2008; Ou et al., 2015; Lin et al., 2020; Yang et al., 2020; Naknaen et al., 2021; Cai et al., 2022; Qian et al., 2022; Wang et al., 2022; Zhang et al., 2022). We identified four Lake Erie vOTUs (vOTU_4, vOTU_1398, vOTU_4148, vOTU_6227) with a high degree of similarity to four *Microcystis* virus isolates (SI Fig. 1; SI Table 2): *Microcystis aeruginosa* viruses Ma-LMM01 (Lake Mikata, Japan, 2006), MaMV-DC (Lake Dianchi, China, 2012), and Mic1 (freshwater estuary, China) and *Microcystis weisenbergii* virus vB-MweS-Yong2.

Most stretches of homology between Lake Erie vOTUs and known viruses were short, ranging from 3-10 kb (SI Fig. 1). A notable exception was the representative virus of vOTU4, which we renamed as Ma-LEF01 (*Microcystis aeruginosa* Lake Erie Fukuivirus-01). Ma-LEF01 has high similarity along the full length of its genome to a viral contig, MVGF-J-19, that was previously assembled from a 2019 Lake Erie cHAB metagenome (McKindles et al., 2020; Fig. 6A-C). While not members of vOTU4, MVGF-J-19 and Ma-LEF01 were also highly similar to *Microcystis* isolates MaMV-DC and Ma-LMM01 (Fig. 6A-C). That Ma-LEF01 and MVGF-J19 have nearly identical genomes, yet were detected 5 years apart, aligns with similar observations in a marine system, where nearly identical viral genomes (>99% shared ANI) were found over the course of a

decade (Marston and Martiny, 2016). Overall, the detection of these vOTUs highly similar to *Microcystis* viruses isolated from around the world and spanning decades, suggests the study of Lake Erie *Microcystis* virus-host dynamics may provide insights into persistent predator-prey relationships relevant for cHAB dynamics in other regions as well.

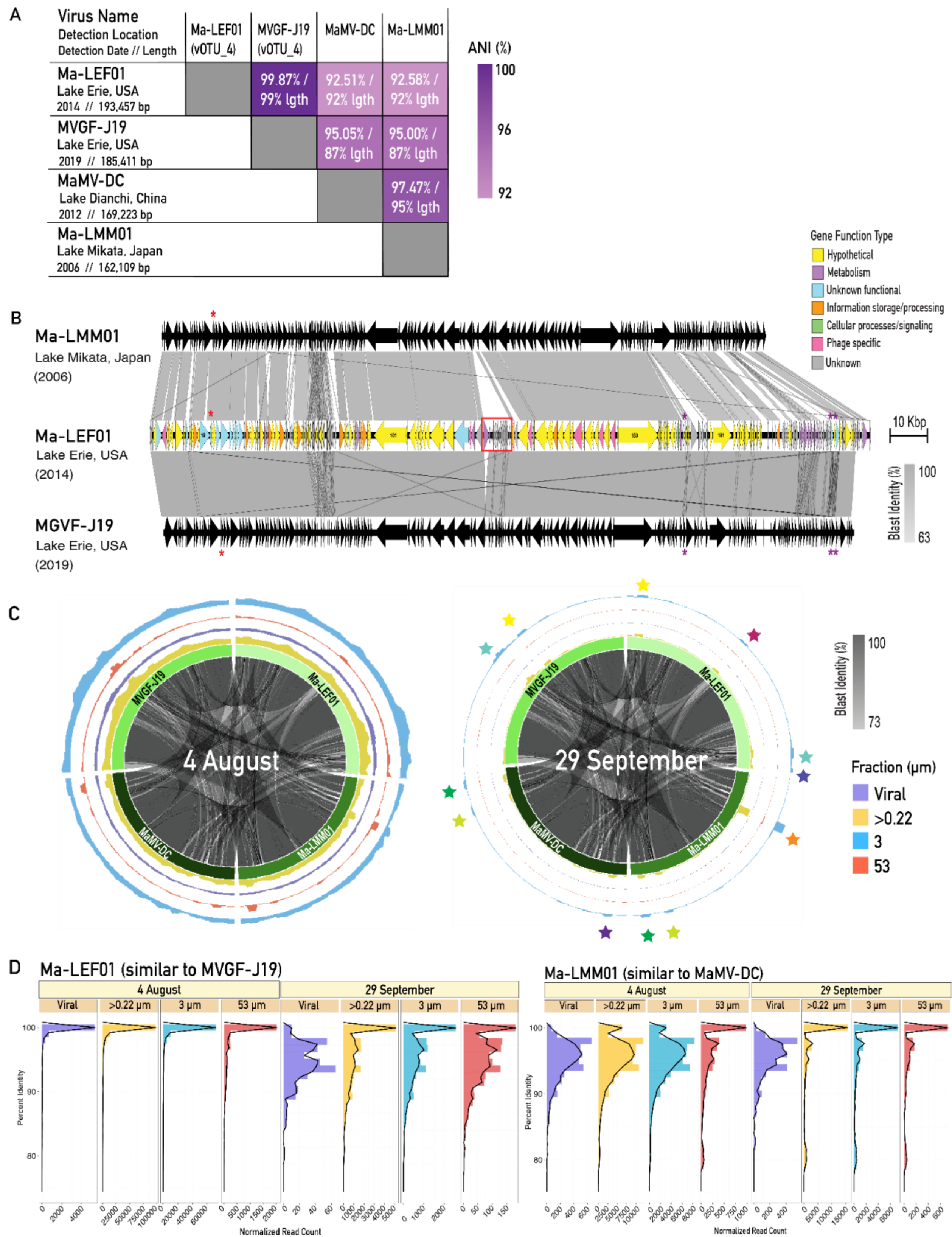

A
Virus Name
Detection Location
Detection Date // Length
Ma-LEF01 (vOTU_4)
MVGF-J19 (vOTU_4)
MaMV-DC
Ma-LMM01
Ma-LEF01
Lake Erie, USA
2014 // 193,457 bp
99.87% / 99% lgth
92.51% / 92% lgth
92.58% / 92% lgth
MVGF-J19
Lake Erie, USA
2019 // 185,411 bp
95.05% / 87% lgth
95.00% / 87% lgth
MaMV-DC
Lake Dianchi, China
2012 // 169,223 bp
97.47% / 95% lgth
Ma-LMM01
Lake Mikata, Japan
2006 // 162,109 bp
ANI (%)
100
96
92
B
Ma-LMM01
Lake Mikata, Japan
(2006)
Ma-LEF01
Lake Erie, USA
(2014)
MGVF-J19
Lake Erie, USA
(2019)
Gene Function Type
Hypothetical
Metabolism
Unknown functional
Information storage/processing
Cellular processes/signaling
Phage specific
Unknown
10 Kbp
Blast Identity (%)
100
63
C
4 August
29 September
MVGF-J19
Ma-LEF01
MaMV-DC
Ma-LMM01
Blast Identity (%)
100
73
Fraction (µm)
Viral
>0.22
3
53
D
Ma-LEF01 (similar to MVGF-J19)
Ma-LMM01 (similar to MaMV-DC)
4 August
29 September
Viral
>0.22 µm
3 µm
53 µm
Percent Identity
Normalized Read Count

Figure 6. Genome similarity and spatiotemporal distribution of four closely related *Microcystis* viruses found in Lake Erie. (A) Pairwise genome similarity between the Lake Erie Ma-LEF01 and three closely related known *Microcystis* viruses (Ma-LMM01 and MaMV-DC) and viral contig (MVGF-J19). Similarity reported as average nucleotide identity (%) and alignment fraction (%) relative to the shortest of the two being compared. (B) Genome synteny plot of Ma-LMM01 and Lake Erie Ma-LEF01 and MVGF-J19 sequences. Gene color indicated DRAM gene function annotations for Ma-LEF01. 'Unknown' is assigned when there is a lack of information, which could arise from no sequence hits. 'Unknown functional' is assigned when a sequence has been identified, but its functional role is unidentified or uncharacterized. 'Hypothetical' is assigned when the function is not well-characterized or solely predicted computationally. Gray tracks linking portions of the genome represent nucleotide identity. Red box indicates the sequence portion unique to Lake Erie vOTUs. Red asterisk indicates nblA gene; purple asterisk indicates pentapeptide repeat proteins (C) Circos genome synteny and coverage plots of the four closely related *Microcystis* viruses/contig from panel A. Gray links in the inner circle represent the shared ANI between the viruses. Outer tracks represent read mapping coverage for different fractions on 4 August (left) and 29 September (right) collected at station WLE12. Colored stars on Sept 29 indicate genome regions with spikes in mapped reads. Colors indicate gene clusters with host homologues described in SI Table 3. (D) Strain-resolved population dynamics of the *Microcystis* viruses at two bloom peaks. Distribution of normalized read counts and percent identity when reads from each date and fraction were competitively mapped to Ma-LEF01 and Ma-LMM01 (all viruses mapped in SI Fig. 2-9). Histogram bar color represents the sampling fraction.

2.2.2.2 Genome characterization of Lake Erie *Microcystis* virus Ma-LEF01

Given the length, high degree of synteny with *Microcystis* virus isolates, and high coverage of the vOTU_4 in the Lake Erie metagenomes at bloom peaks, we further characterized this population. We renamed vOTU_4 as Ma-LEF01 for *Microcystis aeruginosa* Lake Erie (candidate) *Fukuivirus* number 01. Ma-LEF01 has a 193,457 bp long genome with 243 predicted genes (SI Table 4), 223 of which were genes shared with MVGF-J-19, and 168 and 173 shared with the less closely related Ma-LMM01 and MaMV-DC viruses, respectively (SI Table 3). Seven genes are unique to Ma-LEF01 and missing from MVGF-J-19, Ma-LMM01 and MaMV-DC (genes 58, 59, 128, 199, 238, 239 and 241), all of which are of unknown function (SI Table 4). Ma-LEF01 has genes characteristic of both lytic (viral tail sheath) and lysogenic (putative phage anti-repressors, site-specific recombinase, resolvase, lysis inhibition proteins rIIA and B) replication strategies. Prior study of its Ma-LMM01 relative in Lake Tai (China) used patterns in the transcription levels of these genes to make inferences about the population-wide infection status (i.e., lytic vs. lysogenic) in the sampled community (Stough et al., 2017). Neither Ma-LEF01 nor its relatives were identified as integrated prophages in the 32 *Microcystis* MAGs reconstructed in this study, but we cannot rule out that integration may be a strategy the viruses used in undetected *Microcystis* populations or at other times of the year. As with its relatives, Ma-LEF01 has a

homologue of *nblA* that encodes a phycobilisome degradation protein (red asterisk, Fig. 6B). If this enzyme is active during infection, phage-mediated degradation of the light-harvesting complex may benefit viral fitness by recycling biomolecules needed for replication (especially N) or, as others have proposed, by reducing absorption of light energy and thus photodamage to the new phage particles (Yoshida et al., 2008). The Ma-LEF01 genome also encodes three pentapeptide repeat proteins (purple asterisks, Fig. 6B). PRPs are found in bacteria and with high frequencies in cyanobacteria, though their biochemical functions are unknown (Zhang et al., 2020). Notably, the PRPs are specific to the Lake Erie viral strains Ma-LEF01 and MVGF-J-19, but are not found in Ma-LMM01 and MaMV-DC isolated in Asia. Also unique to the Lake Erie strains is a gene cluster of hypothetical genes not seen in existing sequence databases (red box, Fig. 6B), which is preceded by an adenylyltransferase-encoding gene found only in Ma-LEF-01 (purple gene in red box), but *not* MVGF-J-19. The lack of annotations regarding many of these strain-specific loci make it impossible to infer fitness consequences, if any, but their presence is evidence of either viral strain diversity (multiple variants arising to detection in different years and locations) or strain diversification at different spatial (Lake Erie compared to lakes in China and Japan) and temporal (Lake Erie population in 2014 compared to 2019) scales.

2.2.2.3 Local spatiotemporal patterns of *Microcystis* virus Ma-LEF01

When sequence reads were mapped back to the genomes of Ma-LEF01 and its close relatives, distinct date- and fraction-specific coverage patterns were observed that offer insights into their population ecology. Average genome coverage of Ma-LEF01, normalized by per sample sequencing effort, was the highest in the >0.22 μm and 3 μm fractions on August 4 (Fig. 6C) and an order of magnitude lower in the viral and 53 μm fractions on that date. Ma-LEF01 was not present on Sept 29 (Fig. 6C; SI Fig. 3; SI Table 5). The high standard deviation relative to read depth on Sept 29 was attributed to a few high coverage spikes across the genome (Fig. 6C). All of these regions contained viral genes with homologues in the Lake Erie *Microcystis* MAGs also assembled from those dates (SI Table 3). We attributed these spikes to narrow bands of non-specific recruitment of reads that originated from *Microcystis* populations, rather than from the

viruses themselves. When these spikes occur, they are found exclusively in the cellular fractions (not 'viral'), supporting the cellular origins of their reads. These regions of *Microcystis* virus-host homology included pentapeptide repeats (a family of motifs found at high frequencies in cyanobacterial genomes) and genes involved in sulfatase modification, DNA replication and repair, and energy metabolism (SI Table 3).

The disappearance of the Ma-LEF01 population from August 4 to September 29, despite the persistent detection of *Microcystis* MAGs from July through October (Fig. 5D), is consistent with other observations of dynamic rise and fall of *Microcystis* viral populations. For example, the frequency of Ma-LMM01-infected *Microcystis* cells throughout a bloom in Japan varied between 0.002 to 1.5% and usually remained below 0.3% during the year long study (Kimura-Sakai et al., 2015). The authors proposed the perpetual replacement of phage-sensitive populations with phage-resistant populations, whereby viruses promoted host diversification (Kimura et al., 2013). We also observed spatial patterns emerge in the relative abundances of the Ma-LEF01 population. On August 4, Ma-LEF01 was nearly two orders of magnitude more abundant at station WLE12 than WLE4, despite those stations being only seven nautical miles apart (SI Fig. 7). This spatial variability could be related to varying abundance of the specific *Microcystis* population infected by Ma-LEF01, however the distribution of the *Microcystis* MAGs across the stations for which sequence data exists was remarkably consistent (Fig. 5D). This discordance could be explained by (i) Ma-LEF01 may have infected a host population at WLE12 not captured in our reconstructed MAGs, (ii) Ma-LEF01 infection is occurring at WLE4, but is below the detection limit of metagenomics, (iii) *Microcystis* populations at WLE4 acquired resistance encoded at finer scales that MAGs (e.g., CRISPR), or (iv) Ma-LEF01 may engage in a relationship with its host that differs from the canonical lytic infection system where a host boom is expected to be followed by (or depending on the infection timing, coincident with) a viral boom and host bust.

2.2.2.4 *Microcystis* viruses show strain-level population dynamics

The current standard for vOTU definition (95% ANI across 85% alignment fraction relative to the shorter sequence (Roux, 2019)) would place Ma-LEF01 and MVGF-J19

in a group and Ma-LMM01 and MaMV-DC in a group (Fig. 6A). We sought to evaluate these groupings through the lens of population genomics and thus inform future applications of the vOTU definition in wild communities.

While read alignment to the four close-relatives of Lake Erie Ma-LEF01 (Fig. 6A) showed similar patterns through the bloom (Fig. 6C; SI Fig. 2-9), tracking each strain revealed temporal and spatial patterns conserved within two groups. On August 4, the read alignment identity was heavily right-skewed in a broad peak ranging from 88-98% ANI for the Ma-LMM01/MaMV-DC group, whereas the Ma-LEF01/MVGF-J19 group showed only a prominent narrow peak near 100% identity (Fig. 5D; SI Fig. 10). We interpret this as the presence of the Ma-LEF01/MVGF-J19 at this date, but not the Ma-LMM01/MaMV-DC group. By the second bloom on September 29 the Ma-LEF01/MVGF-J19 group disappeared. The wide low identity peak during this second bloom suggested that a close-relative emerged (Fig. 5D), but we were not able to reconstruct its genome from these data. We attributed the narrow high identity peak that appeared *with* the wide low identity peak in the cellular fractions (but missing from the viral) to the non-specific mapping to cellular homologues (gene spikes from Fig. 6C).

Considering the sequence similarity thresholds proposed for distinguishing viral species (Roux et al., 2016) supported by infection and fitness profiles of cultured virus-host systems (Duhaime et al., 2017), we proposed that Ma-LEF01 and MVGF-J19 are different strains of the same viral OTU, whereas MaMV-DC and Ma-LMM01 belong to a different vOTU. Further, these findings indicated that the *Microcystis* Ma-LEF01/MVGF-J19 viral OTU persisted over a five-year period in Lake Erie's western basin cHABs, suggesting either sustained infection of local host populations or persistence without infection in the local viral 'bank' (Breitbart et al., 2005). The sporadic detection of the Ma-LEF01/MVGF-J19 vOTU (present in the August 4 bloom, but not the Sept 29 bloom) provides evidence for the latter. Overall, this analysis demonstrated how tracking read abundances and identities offers insights into population cohesion among wild virus populations and can be used to delineate ecologically meaningful boundaries between uncultured vOTUs reconstructed from metagenomic datasets.

2.2.3 Microbial virus-host interaction networks at toxic and nontoxic bloom peaks.

2.2.3.1 Linking uncultivated viral and host populations using genomic signals of coevolution

Much of what we know of the ecology and diversity of *Microcystis* viruses in natural systems has been limited to marker gene analyses. These studies have been performed near exclusively using the gp91 tail sheath gene of only one viral population Ma-LMM01/MaMV-DC (described in detail the prior section) (Takashima et al., 2007; Yoshida et al., 2008; Kimura et al., 2012; Mankiewicz-Boczek et al., 2016; McKindles et al., 2020; Pound 2020). This approach cannot account for the existence, diversity, host range, and ecology of all viruses infecting *Microcystis*, especially those yet to be discovered. This can lead to a fragmentary and biased view of *Microcystis* virus-host dynamics in cHABs. Some studies have leveraged metagenomic data to study uncultivated *Microcystis* viruses and hosts in cHABs by identifying virus-host pairs linked by CRISPR spacers (Morimoto et al., 2019; Morimoto et al., 2023) or by relying on coassociations between viral and *Microcystis* marker genes (Pound 2020). However, these approaches are limited. Relying on CRISPR has limited utility in a whole community context, as roughly 50% of bacterial genomes do not encode detectable CRISPR systems (Burstein et al., 2016). When coassociation studies are used to link viruses with hosts they oversimplify ecological complexities, e.g., assuming interactions between species are binary (presence/absence) and that positive linear correlations of species abundances imply predator-prey relationships, when in reality these ecological relationships are varied and rarely linear (Hevroni et al., 2020; Correa et al., 2021**).**

We sought to identify likely infection linkages between uncultivated viruses and *Microcystis* population genomes (MAGs) reconstructed from the two 2014 Lake Erie cHAB bloom peaks using a method not restricted by the limitations of marker gene, CRISPR, or coassociation analyses. For this we used Virus-Host Interaction Predictor (VHIP) (Bastien et al., 2023), a machine learning-based infection prediction tool that leverages genome-encoded signals of coevolution (e.g., nucleotide frequencies, %G+C patterns, shared nucleotide and protein sequences, etc.) in a model trained and tested on 8,849 lab-verified infection/no-infection data. We expanded the VHIP analysis to also include non-*Microcystis* MAGs that were predicted to be infected by identified

Microcystis viruses. While interpreting these networks, it is helpful to consider them as a superposition of all past infection networks, rather than as a snapshot of active infections captured at the time of sampling. Because VHIP relies on signals of coevolution detected between viral and putative host genomes and because different coevolutionary signals establish and degrade at different rates, there will be remnant signals linking populations that may not be able to carry out infections. Nonetheless, when tested on lab-verified infection pairs, the model predicts species-level virus-host linkages with 87.8% accuracy. We applied VHIP to gain a better understanding of the diversity, turnover, and metabolic impacts of *Microcystis* viruses at the bloom peaks.

2.2.3.2 Hundreds of viral OTUs predicted to infect *Microcystis* at bloom peaks.

At the August 4 toxic bloom peak, a total of 2,026 virus-host pairs were predicted between 454 vOTUs and 17 bacterial MAGs (9 of which were *Microcystis* MAGs; SI Table 1) (Fig. 7A). On the September 29 non-toxic bloom peak, 1,995 virus-host pairs were predicted between 339 vOTUs and 24 bacterial MAGs (8 *Microcystis* MAGs; SI Table 1) (Fig. 7B). The majority of viruses predicted to infect *Microcystis* were present at low abundances in the bloom peak samples, particularly on August 4. Abundant vOTUs (vOTUs that recruited >0.1% of the total reads mapped to vOTUs) represented 6.6% and 13.9% of the total *Microcystis* vOTUs on 4 Aug and 29 Sept, respectively. This trend of majority low abundance and few high abundance populations is common in microbial virus assemblages, which tend to have long tailed rank-abundance curves like their bacterial and archaeal hosts (Breitbart, 2002; Luque et al., 2020; Dart et al., 2023; Cai et al., 2023). The most numerically abundant *Microcystis* vOTUs that emerged on the bloom peaks (labeled in Fig. 7A-B) are discussed in the context of assemblage evenness and turnover in later sections.

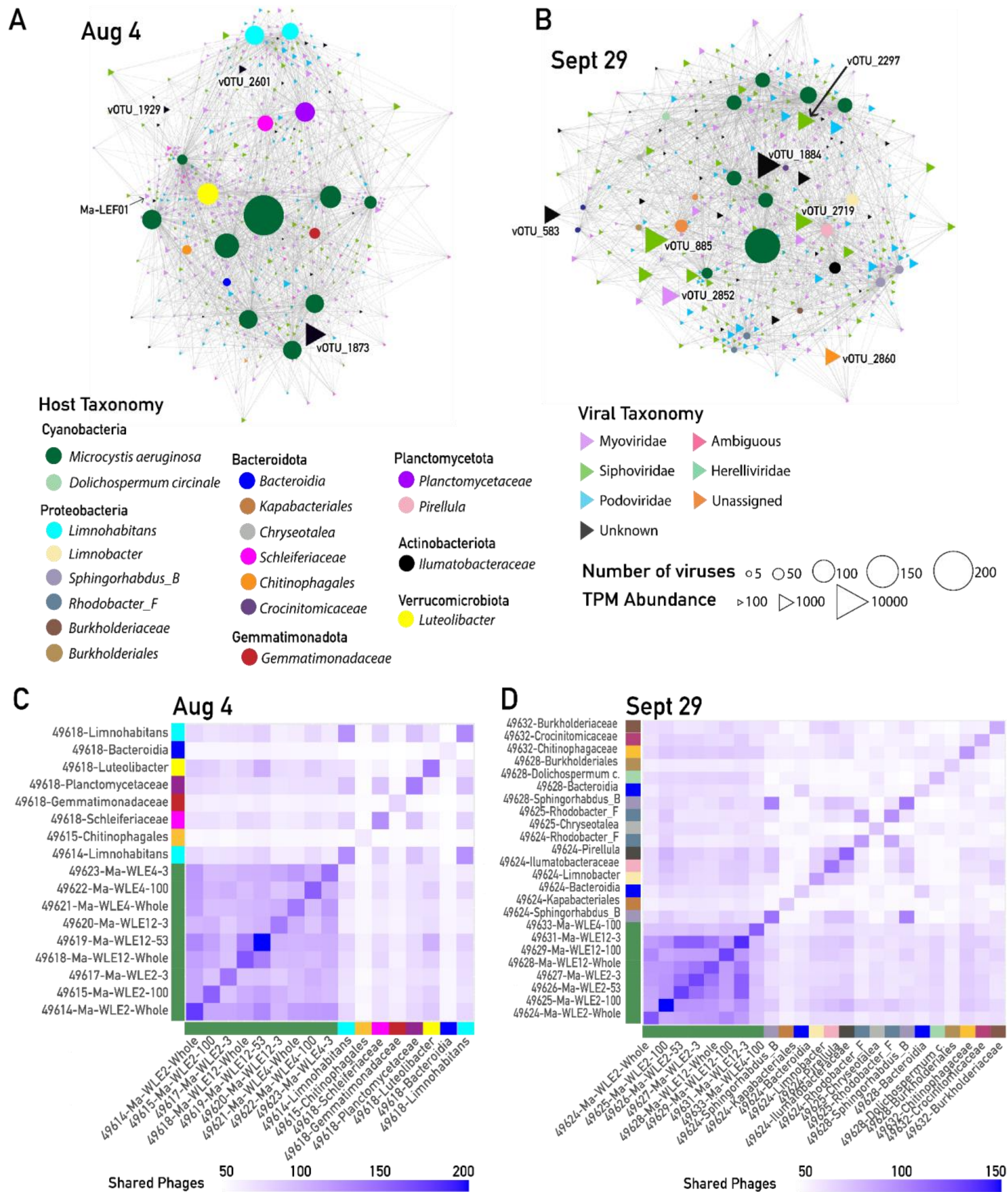


Figure 7. Networks of predicted infections between *Microcystis* viruses (>10 kb viral contigs) and bacterial host MAGs identified on the 4 Aug and 29 Sept bloom metagenomes. (A) Predicted infection network of 4 August toxic bloom peak. (B) Predicted infection network of 29 September non-toxic bloom peak. Circle nodes are host MAGs; circle size represents the number of viruses predicted to infect a given host. Triangle nodes are viruses. Node size represents TPM abundance. Node colors represent assigned

taxonomy; for viruses, "Unknown" indicates the vOTU has no hit in the reference database, "Unassigned" indicates the vOTU has a hit to something unassigned in the reference database. Only predictions with >93% infection probability and viruses predicted to infect at least *Microcystis* are shown. The most abundant vOTUs are labeled in each network (described in SI Table 6 and Fig. 6) (C) Heat map of hosts with shared virus predictions on 4 August. (D) Heat map of hosts with shared virus predictions on 4 August. Axis colors represent host taxonomy (as in panels A-B) and heat map cell color represents the amount of shared phages between any two given hosts.

While we have no available methods to determine the true number of uncultivated viruses infecting a given uncultivated host, these numbers of predicted *Microcystis* viruses are higher than previously reported from metagenomic studies that used other approaches to identify putative *Microcystis*-infecting viruses (McKindles et al., 2020, Morimoto et al., 2023, Pound et al., 2020). Given the 88% accuracy of VHIP in predicting infections (Bastien et al., 2023), we are confident that a substantial portion of the co-evolutionary associations identified in these data reflect true virus-host interactions. In addition, the elevated number of predicted *Microcystis* vOTUS may be a consequence of the incomplete nature of metagenome-reconstructed population genomes. We presume there is a substantial number of genomic "shrapnel", i.e., genome fragments that could not be linked with their corresponding parts, owing to the challenges of binning viral genomes (Roux, 2016; Kieft et al., 2022), which would elevate the number of predicted infections (e.g., one virus could be counted as many). So while the numbers of predicted viruses are not likely to reflect the true number of viruses in a given sample, the overall data structures are informative. These linkages provide important and novel perspectives regarding the structure of the predator-prey networks (e.g., narrow versus broad host ranges), the potential for gene flow between host and virus populations (e.g., *"which host populations are or have been evolutionarily connected via viral infection?"*), and insights into how viral diversity fluctuates through time and space over the course of a bloom.

2.2.3.3 Most *Microcystis* vOTUs host ranges are within-genus, some span phyla

To evaluate *Microcystis* vOTU host range and potential for virus-mediated cross-host gene transfer, we identified the predicted *Microcystis* vOTUs that were also predicted to infect non-*Microcystis* MAGs. The vast majority of shared viruses were between *Microcystis* populations (Fig. 7C-D). In addition, eight and 16 non-*Microcystis* hosts

were predicted on Aug 4 and Sept 29, respectively (Fig. 7). Most notably, on August 4, four prominent host nodes representing four non-Cyanobacteria phyla emerged with a relatively high number of *Microcystis* viruses predicted to infect them (Fig. 7A). While *Microcystis* viruses have been shown to infect multiple cyanobacterial genera in lab studies (Watkins et al., 2014), tests across higher phylogenetic levels have not been reported. However, viral isolates of other host taxa *are* known to infect across multiple phyla (Malki et al., 2015); we suspect the dearth of cross-phyla reports is not only due to evolutionary constraints that underlie host range, but also because of how extremely rare it is for cross-phyla infections to be tested (Bastien et al., 2023).

A previous study relying on qPCR to track metagenome-identified viral groups showed how narrow host range *Microcystis* viruses were observed at markedly lower abundances than broad host range viruses, which tended to be more dominant (Morimoto et al., 2023). We tested whether similar observations were observed in the Lake Erie bloom. A range of host range breadths was identified that depended on whether only *Microcystis* hosts were considered (as in the Morimoto et al., 2023 study) or all hosts were considered (Fig. 9A). When only *Microcystis* hosts were considered, most vOTUs were predicted to infect only one host at the bloom peak and 60% of the viruses were predicted to infect only one or two other Microcystis MAGs (Fig. 8A); these would be classified as “narrow” host range viruses by Morimoto et al., 2023. When all hosts were considered, the mode shifted to three predicted hosts per vOTU with 76% of the vOTU belonging to Morimoto’s “broad” host range category (Fig. 8A). However, unlike Morimoto et al., 2023, we found no or only extremely weak correlations between vOTU host range breadth (i.e., number of hosts) and abundance (Pearson R^2 <0.05 for all tests; SI Table 7).

In discussions of host range, it is important to consider that the terms “broad” and “narrow” are operational and often defined relative to a given study. In the Morimoto study, despite delineating narrow and broad host range groups, all host diversity was constrained to a single host genus. Here, we report *Microcystis* virus-host pairs with genomic evidence that suggests past infections that span phyla. While rampant recombination can spread these signal across the reticulate phylogeny of viruses, there is also culture-based evidence to support these signals could reflect true cross-phyla

infections (Malki et al., 2015) and community genomic evidence to support cross-domain infections (Hwang et al., 2023).

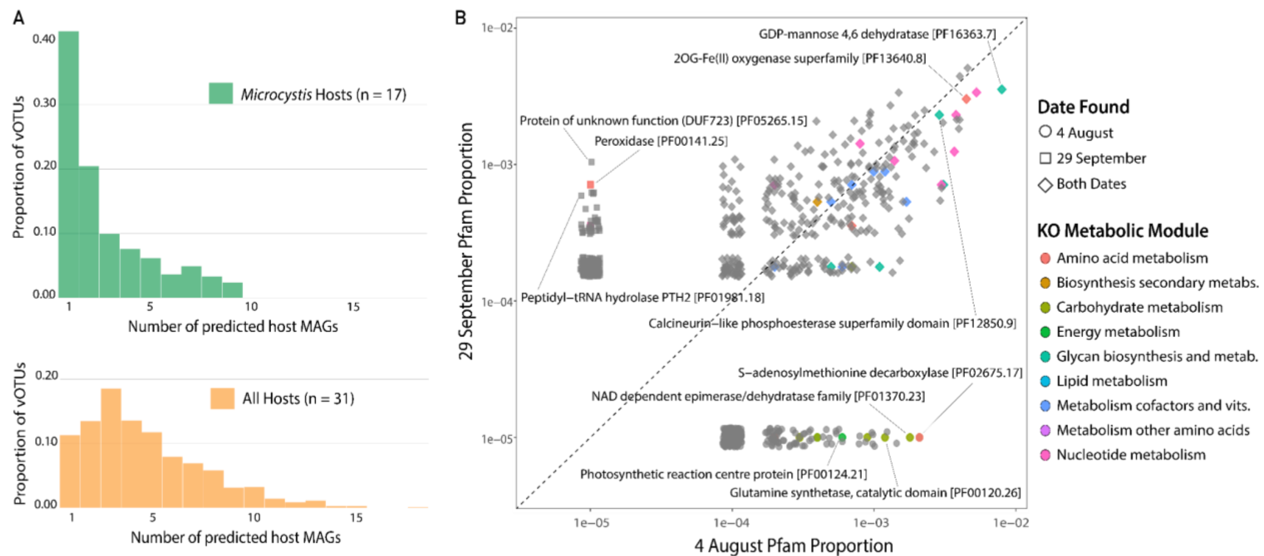


Figure 8. Proportion of vOTUs infecting bloom peak hosts and proportion of abundant viral genes at bloom peaks. (A) Comparison of host ranges. (B) Proportion of viral genes encoded by predicted *Microcystis* vOTUs at 4 August bloom peak (x-axis) and 29 September bloom peak (y-axis). Shape represents whether viral genes were identified in a single bloom peak or both. Point color reflects the assigned KEGG Ontology (KO) metabolic module. Points labels are derived from assigned protein family (Pfam) annotations for select functions of interest.

2.2.3.4 Metabolic genes encoded by predicted *Microcystis* vOTUs can be bloom peak-specific

The coevolutionary signals that underlie the predicted virus-host interaction networks represent paths of potential virus-mediated gene flow between bacterial populations. Viruses are well known to encode genes involved in myriad metabolic processes, e.g., photosynthesis, nitrogen (Roux et al., 2016), and sulfur metabolism (Anantharaman et al., 2014; Kieft and Zhou, 2020). When viruses facilitate the cross-taxa (virus and host) transfer of genes central to bacterial metabolism, there are potential consequences for the evolutionary trajectory of proteins with important biogeochemical functions (Lindell et al., 2004). Further, during infection, virus-encoded auxiliary metabolic genes (AMGs) rewire host metabolisms in ways that influence the flow of matter and energy during infection (Zimmerman et al., 2020; Howard-Varona et al., 2020). To evaluate this

potential, we identified the metabolic genes carried by the viruses predicted to interact with *Microcystis.*

Most identified *Microcystis* vOTU genes encode proteins with unknown metabolic functions (Fig. 8B; SI Table 8). Of the AMGs that could be annotated, some were shared at the bloom peaks and some were specific to the peaks on either August 4th or September 29th (Fig 8B). Virus-encoded AMGs shared at the bloom peaks included the biosynthesis of complex carbohydrates important for cell wall biosynthesis, cell-to-cell communication, and biofilm formation (GDP-Mannose 4,6 dehydratase; Kehr et al., 2015), the production of phycobiliprotein light-harvesting pigments and secondary metabolites (2OG-Fe(II) oxygenase superfamily; Jia et al., 2017; Herr and Hausinger, 2018) and the regulation of phosphate metabolism (2OG-Fe(II) oxygenase superfamily; Wanner, 1993; Morohoshi et al., 2002). Virus-encoded functions abundant on 4 August included those central to photosynthesis (photosynthetic reaction center proteins), nitrogen, amino acids, and energy metabolism (Glutamine synthetase; Bolay et al., 2018; NAD-dependent epimerase/dehydratase), and the cellular responses to nutrient, light and temperature fluctuations (S-adenosylmethionine decarboxylase; Jantaro et al., 2003; Zhu et al., 2016). Virus-encoded metabolic functions on 29 September included the viral takeover of host protein synthesis (peptidyl-tRNA hydrolase PTH2 enzymes; Garcia-Villegas et al., 1991) and cellular responses to oxidative stress, such as can arise from exposure to high light intensity and reactive oxygen species generated during photosynthesis (peroxidase enzymes; Zinser, 2018). Of the latter, a recent study of the 2014 Lake Erie bloom found anomalously high expression of the peroxidase subunit *ahpC* in *Microcystis* on 29 September in the >100 µm fraction at stations WLE12 and WLE2 (Smith et al., 2022). Our observation of high peroxidase gene abundance among vOTUs on 29 September suggests a potential role of viral AMGs in supplementing *Microcystis* ROS defense during bloom peaks. Overall, the functional gene analysis of the *Microcystis* vOTUs indicated that there are date-dependent consequences for how viruses may be supplementing or rewiring host metabolisms. These findings are motivation for future work to better resolve the degree to which viruses shape the genotypic and phenotypic changes that manifest in the cyanobacterial populations that dominate Lake Erie’s cHABs.

2.2.4 Turnover of predicted *Microcystis* vOTUs depends on colony formation

2.2.4.1 Diversity of predicted *Microcystis* vOTUs is highest in the viral fraction, lowest in the colony-associated fraction

vOTU abundances estimated based on sequence read recruitment were used to evaluate trends in diversity within the subset of vOTUs predicted to infect *Microcystis*. Within this *Microcystis* vOTU assemblage, evenness was the highest in the virus fraction and decreased with each consecutively larger pore size filter fraction (Fig. 9A). This trend is consistent with observed host strain diversity; the *Microcystis* MAGs were less evenly distributed in the larger colony-associated fraction (100 μm) than the smaller size fractions (>0.22 and 3 μm) (Fig. 5E). We considered 53 and 100 μm to be 'colony-associated' fractions, while those 'not colony-associated' were viral, 3 μm and >0.22 μm, the latter of which has been shown in other lake systems to be numerically dominated by the free-living cells (Schmidt et al., 2020). The genomic variation observed in the *Microcystis* MAGs from Lake Erie showed a distinct partitioning by size fraction. Such within-species genotypic differences that associate with different filter fractions have been observed in ocean taxa as well, such as *Vibrio splendidus* (Hunt et al., 2008). Notably, a significant clustering of genotypes was evident in the 100 μm fraction (Fig. 5D). This aligns with previous reports by Yancey et al., 2023, who also documented similar patterns of *Microcystis* strain diversity within the 100 μm fraction during the 2014 bloom in the western basin of Lake Erie.

The observed fraction-specific genotypic variation in *Microcystis* combined with patchy representation of the *Microcystis* MAGs in the colony-associated fractions can be explained by the proliferation of single phylotypes leading to colony formation during bloom development. Such episodes may occur frequently through the Lake Erie bloom, with distinctive phylotypes emerging at each sampling date among 100 μm fraction samples (Fig. 5C). In line with a 'Kill the Winner' dynamic (Winter et al., 2010), each such episode may select for a subset of viruses able to infect the dominant phylotype causing a bottleneck in viral diversity in the colony-associated fractions, which is supported by a corresponding drop in vOTU diversity in the 100 μm fraction (Fig. 9A). In contrast, the non-colony associated hosts may sustain infection relationships with more

members of the viral “seed bank” (Breitbart, 2005; Hevroni et al., 2020; Dart et al., 2023), thus maintaining a higher overall *Microcystis* vOTU assemblage diversity.

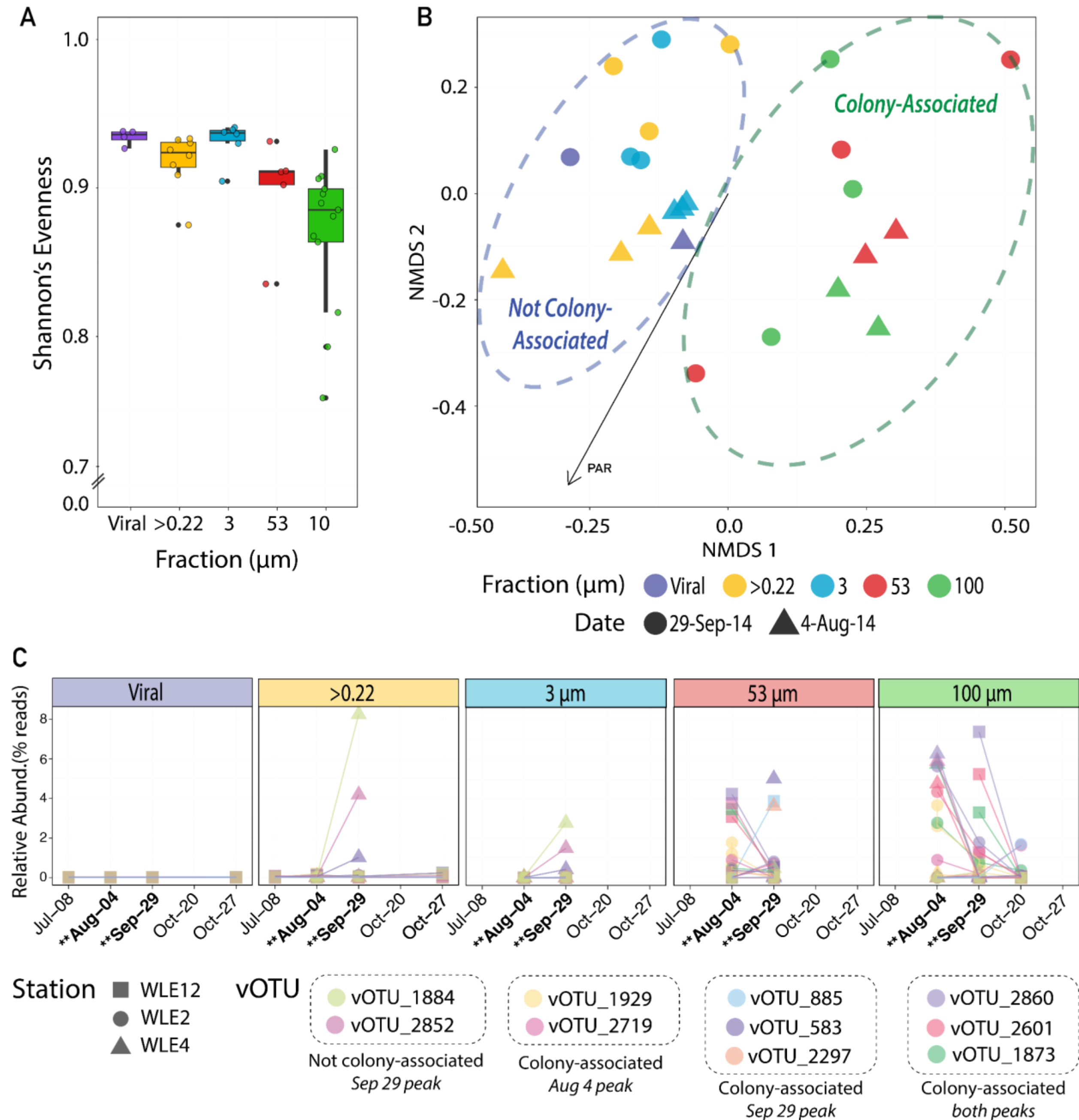


Figure 9. *Microcystis* virus diversity, turnover, and abundant vOTU temporal dynamics through the 2014 Lake Erie bloom season. (A) Diversity of predicted *Microcystis* vOTU assemblage across fractions. (B) NMDS ordination based on Bray-Curtis dissimilarities in distributions of predicted *Microcystis* vOTUs at the cHAB peaks. Point color represents the sampling fraction; shape represents sampling date. Dashed ovals indicate whether a point belongs to the free-living or colony-associated community. The ordinations are overlaid with the gradient of fit between measured environmental parameters and Bray-Curtis dissimilarities. Vector length of the environmental parameter represents the strength of the correlation with the data variation. (C) Temporal relative abundance dynamics of the ten most abundant predicted

Microcystis vOTUs in terms of fraction of total metagenomic reads mapped to each vOTU per sample. Point shape represents the sampling station and point color represents a specific *Microcystis* virus OTU. Dates of sampled bloom peaks are highlighted in bold with double asterisk.

2.2.4.2 Sampling fraction correlates with turnover of vOTUs predicted to infect *Microcystis*

Between-sample variation in the *Microcystis* vOTU assemblage structure significantly correlated with filter fraction (Fig. 9B; PERMANOVA R^2=0.21, p-value = 0.0001; SI Table 9) and to a lesser extent sampling date (R^2=0.06, p-value = 0.001). There was colony-dependent partitioning of the variation between the *Microcystis* vOTU assemblages (Fig. 9B). This pattern may be explained by the discussed fraction-specific differences in relative abundances of *Microcystis* genotypes (Fig. 5E). Even though free-living and colony-associated communities exist within the same aquatic matrix, they represent distinct ecological contexts allowing for distinct dynamics in host populations (Rosenberg et al., 2021), which we posit also extends to distinct virus-host interactions. These fraction-associated *Microcystis* genotypic differences may have implications for viral host range; viral mediated top down control on *Microcystis* may depend on whether the *Microcystis* host is colonial or single cellular.

Sample location was not significantly correlated with turnover in the *Microcystis* vOTU assemblage, (R^2=0.08, p-value = 0.30). This was surprising, given the station specificity of Ma-LEF01 (SI Fig. 2-3). This insignificance of station in explaining variation in *Microcystis* vOTU assemblage turnover, despite station impacting the presence of Ma-LEF01 when sampled on the same date, points to the importance of disentangling dynamics at multiple scales. In understanding top down controls on *Microcystis*, factors that underlie the dynamics of a single *Microcystis* virus population (let alone a single gene) will not reflect the dynamics of the total assemblage of *Microcystis* viruses. Of the environmental variables tested, only photoactive radiation (PAR) was significantly correlated with variation in *Microcystis* vOTU turnover (R^2= 0.06, p-value = 0.0014; SI Table 10). Given that *Microcystis* growth and physiology are influenced by variations in light intensity and duration (Wilson et al., 2006), it is understandable that a parameter that affects host population fitness may also indirectly affect the viruses of that host

population. This trend may also suggest light availability as a driver of *Microcystis* infection during high bloom densities.

2.2.4.3 High turnover of abundant *Microcystis* virus populations

We next sought to better understand the seasonal fluctuation and host ranges of the numerically important *Microcystis* vOTUs in the 2014 Lake Erie cHAB. We identified the ten *Microcystis* vOTUs most abundant across the entire sample set (SI Table 6). Notably, Ma-LEF01 (Fig. 6), relative of Ma-LMM01 whose gp91 gene has been used as the sole proxy for quantifying *Microcystis* viruses in some studies, was not among these abundant vOTUs. The numerically dominant *Microcystis* vOTUs were never abundant in the viral fraction (Fig. 5C). When they peaked, they were most abundant at the bloom peaks (Fig. 9C). The dominant *Microcystis* vOTUs peaked in abundance in either the colony-associated or not-colony associated fractions, but not both (Fig. 9C). In both fractions, most vOTUs peaked on either Aug 4 or Sept 29, with only three of the 10 vOTUs peaking at both dates (Fig. 9C).

The sporadic peaks of these abundant *Microcystis* vOTUs offers more support for Bank dynamics in this system, while only a select few are abundant at any given moment throughout the seasons (Breitbart, 2005; Hevroni et al., 2020). These changes in viral community composition over time are likely influenced by factors such as shifts in host availability, environmental conditions, and the viral populations' specific interactions with *Microcystis* and other host populations. Indeed, even at the coarse level of oligotypes, different *Microcystis* genotypes were present at the two bloom peaks (Berry et al., 2016). Of the 10 most abundant *Microcystis* viral populations identified, 80% were primarily found in either the 53 μm or 100 μm fraction, which represents the *Microcystis* colony-associated viral community. As host cells within colonies are tightly-packed, the opportunity to infect and increase in abundance is more likely to occur here than in those viral populations that persist in the free-living microbial community, where host interactions may be scarce. Alternatively, colonies may also provide defense against infection (Wucher, 2023). Understanding the turnover and dynamics of these viral populations is essential for comprehending their roles in regulating harmful algal blooms and the broader dynamics of aquatic ecosystems in Lake Erie.

2.3 Conclusions and outlook

Previous research has established that the two 2014 bloom peaks represented different *Microcystis* genotypes (Yancey et al., 2022), including the transition from toxin-producing to non-toxin producing genotypes. This was hypothesized to be in part due to the low ammonium and nitrate availability in September relative to August coupled with the nitrogen-rich nature of microcystin metabolites. Our work suggests viral activity as an additional control on *Microcystis* strain succession in the 2014 bloom. Distinct assemblages of *Microcystis* vOTUs were predicted to infect the *Microcystis* genotypes at the two bloom peaks, suggesting strain specificity among predicted *Microcystis* viruses. Strain-specific viral predation and the diversity and ecological dynamics of *Microcystis* populations are inextricably connected, likely influencing one another over the course of the bloom.

The patterns uncovered in this work describing the spatial and temporal dynamics of *Microcystis* viruses in the 2014 Lake Erie cHAB help to reveal the intricate dynamics of cyanobacterial blooms and their ecological implications. Additional molecular techniques, such as metatranscriptomics, metaproteomics, metabolomics, will help to identify viral gene and protein expression patterns and to identify relationships between viral activity and community-level metabolite (including toxin) production through the blooms. These omics approaches can also shed light on viral strategies for manipulating host metabolism, toxin regulation, and avoiding host antiviral strategies within cHABs. Moreover, longitudinal studies encompassing multiple bloom seasons and locations will contribute valuable insights into the temporal and spatial dynamics of *Microcystis* vOTUs. Understanding connections between the *Microcystis* viruses, the total virus community, environmental parameters, and overall bloom progression is essential for unraveling the complex dynamics of these cHABs.

2.4 Materials and Methods

2.4.1 Field Sampling and Collection

Field sites were sampled with the joint NOAA Great Lakes Environmental Research Laboratory / University of Michigan Cooperative Institute for Great Lakes Research weekly sampling program for Lake Erie. In 2014, three sites were sampled bi-monthly in June then weekly from July through October. Bloom stages were determined by phycocyanin fluorescence and relative abundance. Metagenomic data were generated from samples collected from three regularly sampled stations (WE2, near the mouth of Maumee River, 41° 45.743' N, 83°19.874' W; WE4, offshore towards the center of the western basin, 41°49.595' N, 83°11.698' W; and WE12, adjacent to the water intake crib for the city of Toledo, 41°42.535' N, 83°14.989' W). All samples were collected upon arriving on-station using a peristaltic pump to obtain 20 L of water from 0.1 m below the surface. Water was filtered onto 100 µm polycarbonate filters. This size was used to concentrate *Microcystis* colonies retained on the filter while excluding smaller particles. Previous work has shown that in Lake Erie the >100 µm assemblage comprised over 90% of all *Microcystis* cells in the water column (Chaffin et al., 2011). The filtered water was subsequently passed through a 53 µm and 3 µm polycarbonate filter to collect smaller colonies and large single-celled microbes, including *Microcystis*, whose cell sizes range from 1.7 to 7 µm in diameter (Xiao, 2018). Whole water was passed through a 0.22 µm filter to collect the total cellular microbial community; community structure of whole community fractions have been shown to be dominated by single cells (Schmidt et al., 2020). To enrich for viruses 10 g/L iron chloride stock solution (0.966 g $FeCl_3$-$6H_2O$ in 20 mL 0.02 µm-filtered autoclaved MilliQ-H_2O) was added to the <0.22 µm ("viral") fraction (Poulos, 2017). The flocculant incubated overnight to maximize virus recovery before being passed through a 0.45 µm 142 mm Millipore Express Plus filter and stored at 4°C.

2.4.2 DNA Extraction and Sequencing of Hosts and Viruses

DNA was extracted from samples using the DNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. Shotgun sequencing of DNA was performed on the Illumina® HiSeq™ platform (2000 PE 100, Illumina, Inc., San Diego, CA, USA) at the University of Michigan DNA Sequencing Core.

2.4.3 Host Assembly and Binning

BBDuk (https://sourceforge.net/projects/bbduk/) was used to identify and remove contaminated sequences and denoised reads were evaluated using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). The reads were dereplicated using BBnorm (https://sourceforge.net/projects/bbnorm/). Reads of all 36 samples were assembled on a per-sample basis into contigs using Megahit (Li et al., 2015). Following contig assembly, Centrifuge (Kim et al., 2016) was used to taxonomically classify contigs to the species level, or lowest resolution available. Automated binning was performed on the contigs using Concoct on default settings (Alneberg et al., 2013) to generate metagenome assembled genomes (MAGs). Quality trimmed raw reads were mapped to each individual assembly using bwa (https://sourceforge.net/projects/bio-bwa/) with bwa-mem on default settings. SAMtools (Li et al., 2009) was used to convert, sort, and index compressed BAM files. Quality trimmed reads were mapped to MAGs using a pileup shell script provided by BBtools (https://sourceforge.net/projects/bbtools/). The Anvi'o platform v2.3.0 (Eren et al., 2021) was used to manually refine the unique MAGs identified through Vizbin (Laczny et al., 2015) by evaluating differential coverage patterns across the samples. Bins with >50% completeness (completeness statistics inferred from a CheckM (Parks et al., 2015)), >10% contamination, and <75% strain heterogeneity were manually refined in Anvi'o based on differential coverage and contamination. Multiple rounds of Anvi'o refinement were performed to curate bins until they passed the aforementioned thresholds (SI Table 11).

2.4.4 Viral Population Identification and Taxonomic Assignment

CheckV v0.7.0 (Nayfech et al., 2021), VIBRANT v1.2.1 (Kieft et al., 2020), VirFinder v1.1 (Ren et al., 2017), VirSorter v1.0.6 (Roux et al., 2015) and VirSorter2 v2.1 (Guo et al., 2021) were used to predict presumed viral contigs. Contigs from each tool were preserved according to recently established rules for viral contig identification (Hegarty et al., 2022). Additionally, only contigs >3kb were kept from the viral prediction tools and used to identify viral populations. Viruses were then binned using vRhyme default settings (Kieft et al., 2022) to create a collection of viral bins and high-quality unbinned contigs for population clustering. Viral bins and unbinned contigs were clustered (Roux and Bolduc, 2016; stampede-clustergenomes) according to previously established standards defining viral populations (Roux et al. 2019). Contigs sharing an average nucleotide identity (ANI) of 95% across 85% of the contig length were clustered and the longest sequence of each cluster was considered the representative for a cluster, referred to as a viral operational taxonomic unit (vOTU). Taxonomy of viral populations from the two Lake Erie bloom peaks was estimated using the Phage Taxonomy Tool approach (PTT; Kieft et al., 2021).

2.4.5 Viral Community Read Mapping, Quantification and Alpha/Beta Diversity

Filtered and trimmed reads were assembled from the same sample and quantified using Samtools v1.11 (Li et al., 2009). These reads were then competitively mapped to all contigs using Bowtie2 (Langmead and Salzberg, 2012). Reads mapped to trimmed viral sequences were quantified using BLAST v2.9.0 (NCBI, 2019) to align trimmed viral sequences to the contigs and then using FeatureCounts (Liao et al., 2014) from the Subread package (Liao et al., 2013) to quantify reads overlapping this region. The viral reads for each sample were downsampled by 1,000,000 reads for alpha diversity analyses using seqtk v1-3 (http://github.com/lh3/seqtk). Only viral contigs with reads covering at least 3 kb of their length were extracted and included in diversity analyses. Alpha diversity measures were calculated using the vegan v2.5-2 package in R v4.0.2 based on downsampled reads. Transcripts per million (TPM) was determined for each viral population and used to calculate the Bray-Curtis distance between samples in R using the vegan package and then NMDS ordination was performed. PERMANOVA

using the adonis function in vegan was used to test the effects of sampling location, sampling date, sampling fraction as well as effects of environmental parameters on the full viral community structure and metabolic potential. Viral contig information and NCBI accession numbers can be found in SI Table 12.

2.4.6 Viral Metabolic Potential Analyses

KEGG and Pfam databases (Shaffer et al., 2020) were accessed to assign viral contig gene metabolic annotations using Distilled and Refined Annotation of Metabolism (DRAM; Shaffer et al., 2020) following the generation of open reading frames (ORFs) using Prodigal (Hyatt et al., 2010). FeatureCounts (Liao et al., 2014) from the Subread package (Liao et al., 2013) were then used to calculate read coverages of the ORFs generated by DRAM. Bray Curtis distances between samples were calculated using the vegdist function followed by an NMDS ordination with the vegan package in R. Water quality parameters were then applied to a PERMANOVA model to evaluate their effects on the abundance of protein families (Pfams). Parameters were considered significantly correlated with a p-value ≤ 0.05.

2.4.7 Virus-Host Infection Prediction Network

To better understand potential phage infections associated with bacterial hosts in the bloom, we applied VHIP (v.1.), a gradient-boosted machine learning model informed by signals of coevolution embedded in sequences between viruses and hosts. (Bastien et al, 2023). All possible virus-host combinations were considered initially. For network visualization, only viral sequences larger than 10kb and prediction scores higher than 0.93 were considered. (viral sequences >10kb and a >93% probability of infection) were plotted using Gephi 0.9.0 software (Bastian et al., 2009).

2.4.8 *Microcystis* phage comparisons

Normalized read coverage values of *Microcystis* viruses from 4 August and 29 September were obtained and formatted using circos-0.69-9 (Krzywinski et al., 2009). FastANI (Jain et al., 2018) was performed on *Microcystis* viruses to determine shared

ANI between genomes. Configuration files specifying the layout, colors, labels and data tracks for the circos plot were generated prior to creating the circos plot (Fig. 3). EasyFig (Sullivan et al., 2011) was used to generate synteny plots between *Microcystis* viruses and NCBI Blast hits (Fig. 2B; SI Fig. S1).

For a complete list of SI Tables, visit:
https://docs.google.com/spreadsheets/d/1aQf2AOnYSI5cXsjzXj8ub3xF2Q8m625ITLLMhY96nr8/edit?usp=sharing

For a complete list of SI Figures, visit:
https://docs.google.com/document/d/1ItKAGaSW3aKXZe2jcEv7Q51qgZt0xC-WnGIlfgvdJ3I/edit?usp=sharing

2.5 Acknowledgements

We would like to thank Michelle Berry for early assistance with viral sample collection and processing and Lizy Michaelson for assistance verifying and publishing protocols in protocols.io. We acknowledge the members of the Duhaime Lab for years of constructive feedback on the research and writing that went into this work. We also thank the field crew at CIGLR/GLERL for allowing us to sample with them and assisting in field measurements. This research was also funded in part by the University of Michigan Water Center, NOAA Michigan Sea Grant Core Program awards NA22OAR4170084, and NA18OAR4170102, the National Science Foundation grant 2055455, and the NOAA Omics program awarded via the Cooperative Institute of Great Lake Research (CIGLR) through the NOAA cooperative agreements with the University of Michigan (NA17OAR4320152 and NA22OAR4320150).

2.6 References

Aguirre de Cárcer, D., López-Bueno, A., Pearce, D. A., & Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Science Advances*, *1*(5), e1400127. https://doi.org/10.1126/sciadv.1400127

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Loman, N. J., Andersson, A. F., & Quince, C. (2013). *CONCOCT: Clustering cONtigs on COverage and ComposiTion* (arXiv:1312.4038). arXiv. https://doi.org/10.48550/arXiv.1312.4038

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M., & Dick, G. J. (2014). Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. *Science*, *344*(6185), 757–760. https://doi.org/10.1126/science.1252229

Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., Romano, A. E., & DeLong, E. F. (2017). Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), 11446–11451. https://doi.org/10.1073/pnas.1714821114

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Bastien, E. G., Cable, R. N., Zaman, L., Batterbee, C., Wing, A. J., & Duhaime, M. B. (2023). *Virus-Host Interactions Predictor (VHIP): Machine learning approach to resolve microbial virus-host interaction networks* (p. 2023.11.03.565433). bioRxiv. https://doi.org/10.1101/2023.11.03.565433

Berry, M. A., Davis, T. W., Cory, R. M., Duhaime, M. B., Johengen, T. H., Kling, G. W., Marino, J. A., Den Uyl, P. A., Gossiaux, D., Dick, G. J., & Denef, V. J. (n.d.). *Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities*. https://doi.org/10.1111/1462-2920.13640

Berry, M. A., White, J. D., Davis, T. W., Jain, S., Johengen, T. H., Dick, G. J., Sarnelle, O., & Denef, V. J. (2017). Are Oligotypes Meaningful Ecological and Phylogenetic Units? A Case Study of Microcystis in Freshwater Lakes. *Frontiers in Microbiology*, *08*, 365–365. https://doi.org/10.3389/fmicb.2017.00365

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., Brister, J. R., Kropinski, A. M., Krupovic, M., Lavigne, R., Turner, D., & Sullivan, M. B. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology*, *37*(6), 632–639. https://doi.org/10.1038/s41587-019-0100-8

Bolay, P., Muro-Pastor, M. I., Florencio, F. J., & Klähn, S. (2018). The Distinctive Regulation of Cyanobacterial Glutamine Synthetase. *Life (Basel, Switzerland)*, *8*(4), 52. https://doi.org/10.3390/life8040052

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B. (2017). vConTACT: An iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, e3243. https://doi.org/10.7717/peerj.3243

Breitbart, M. (2012). Marine viruses: Truth or dare. *Annual Review of Marine Science*, *4*, 425–448. https://doi.org/10.1146/annurev-marine-120709-142805

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. https://doi.org/10.1016/j.tim.2005.04.003

Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W., & Sullivan, M. B. (2016). Seasonal time bombs: Dominant temperate viruses affect Southern Ocean microbial dynamics. *The ISME Journal*, *10*(2), 437–449. https://doi.org/10.1038/ismej.2015.125

Brum, J. R., & Sullivan, M. B. (2015). Rising to the challenge: Accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology*, *13*(3), 147–159. https://doi.org/10.1038/nrmicro3404

Buchfink et al. (2015). *Fast and sensitive protein alignment using DIAMOND | Nature Methods*. https://www.nature.com/articles/nmeth.3176

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., Thomas, B. C., & Banfield, J. F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms10613

Cai, R., Li, D., Lin, W., Qin, W., Pan, L., Wang, F., Qian, M., Liu, W., Zhou, Q., Zhou, C., & Tong, Y. (2022). Genome sequence of the novel freshwater Microcystis cyanophage Mwe-Yong1112-1. *Archives of Virology*, *167*, 1–6. https://doi.org/10.1007/s00705-022-05542-3

Chaffin, J. D., Bridgeman, T. B., Heckathorn, S. A., & Mishra, S. (2011). Assessment of Microcystis growth rate potential and nutrient status across a trophic gradient in western Lake Erie. *Journal of Great Lakes Research*, *37*(1), 92–100. https://doi.org/10.1016/j.jglr.2010.11.016

Chow, C.-E. T., Sachdeva, R., Cram, J. A., Steele, J. A., Needham, D. M., Patel, A., Parada, A. E., & Fuhrman, J. A. (2013). Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *The ISME Journal*, *7*(12), Article 12. https://doi.org/10.1038/ismej.2013.122

Correa, A. M. S., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., & Weitz, J. S. (2021). Revisiting the rules of life for viruses of microorganisms. *Nature Reviews Microbiology*, *19*(8), Article 8. https://doi.org/10.1038/s41579-021-00530-x

Cory, R. M., Davis, T. W., Dick, G. J., Johengen, T., Denef, V. J., Berry, M. A., Page, S. E., Watson, S. B., Yuhas, K., & Kling, G. W. (2016). Seasonal Dynamics in Dissolved Organic Matter, Hydrogen Peroxide, and Cyanobacterial Blooms in Lake Erie. *Frontiers in Marine Science*, *3*. https://www.frontiersin.org/articles/10.3389/fmars.2016.00054

Dart, E., Fuhrman, J. A., & Ahlgren, N. A. (2023). Diverse Marine T4-like Cyanophage Communities Are Primarily Comprised of Low-Abundance Species Including Species with Distinct Seasonal, Persistent, Occasional, or Sporadic Dynamics. *Viruses*, *15*(2), Article 2. https://doi.org/10.3390/v15020581

Deboutte, W., Beller, L., Yinda, C. K., Maes, P., de Graaf, D. C., & Matthijnssens, J. (2020). Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(19), 10511–10519. https://doi.org/10.1073/pnas.1921859117

Duhaime et al. (2017). *Frontiers | Comparative Omics and Trait Analyses of Marine Pseudoalteromonas Phages Advance the Phage OTU Concept*. https://www.frontiersin.org/articles/10.3389/fmicb.2017.01241/full

Enav, H., Kirzner, S., Lindell, D., Mandel-Gutfreund, Y., & Béjà, O. (2018). Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. *Nature Communications*, *9*, 4698. https://doi.org/10.1038/s41467-018-07164-3

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., Fink, I., Pan, J. N., Yousef, M., Fogarty, E. C., Trigodet, F., Watson, A. R., Esen, Ö. C., Moore, R. M., Clayssen, Q., Lee, M. D., Kivenson, V., Graham, E. D., Merrill, B. D., … Willis, A. D. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology*, *6*(1), Article 1. https://doi.org/10.1038/s41564-020-00834-3

Fuhrman. (1999). *Marine viruses and their biogeochemical and ecological effects | Nature*. https://www.nature.com/articles/21119

Garcia-Villegas et al. (n.d.). *Peptidyl-tRNA hydrolase is involved in lambda inhibition of host protein synthesis.* Retrieved January 10, 2024, from https://www.embopress.org/doi/epdf/10.1002/j.1460-2075.1991.tb04919.x

Gu, X., Tay, Q. X. M., Te, S. H., Saeidi, N., Goh, S. G., Kushmaro, A., Thompson, J. R., & Gin, K. Y.-H. (2018). Geospatial distribution of viromes in tropical freshwater ecosystems. *Water Research*, *137*, 220–232. https://doi.org/10.1016/j.watres.2018.03.017

Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, *9*(1), 37. https://doi.org/10.1186/s40168-020-00990-y

Harke, M. J., & Gobler, C. J. (2013). Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. *PLOS ONE*, *8*(7), e69834. https://doi.org/10.1371/journal.pone.0069834

Harke, M. J., & Gobler, C. J. (2015). Daily transcriptome changes reveal the role of nitrogen in controlling microcystin synthesis and nutrient transport in the toxic cyanobacterium, Microcystis aeruginosa. *BMC Genomics*, *16*(1), 1068. https://doi.org/10.1186/s12864-015-2275-9

Harke, M. J., Steffen, M. M., Gobler, C. J., Otten, T. G., Wilhelm, S. W., Wood, S. A., & Paerl, H. W. (2016). A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae*, *54*, 4–20. https://doi.org/10.1016/j.hal.2015.12.007

Hegarty, B., Dai, Z., Raskin, L., Pinto, A., Wigginton, K., & Duhaime, M. (2022). A snapshot of the global drinking water virome: Diversity and metabolic potential vary with residual disinfectant use. *Water Research*, *218*, 118484. https://doi.org/10.1016/j.watres.2022.118484

Herr, C. Q., & Hausinger, R. P. (2018). Amazing Diversity in Biochemical Roles of Fe(II)/2-Oxoglutarate Oxygenases. *Trends in Biochemical Sciences*, *43*(7), 517–532. https://doi.org/10.1016/j.tibs.2018.04.002

Hevroni, G., Flores-Uribe, J., Béjà, O., & Philosof, A. (2020). Rising through the Ranks: Seasonal and Diel Patterns of Marine Viruses. *bioRxiv*, 2020.05.07.082883. https://doi.org/10.1101/2020.05.07.082883

Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H., Andreopoulos, B., Brewer, H. M., Glavina del Rio, T., Adkins, J. N., Paul, S., Sullivan, M. B., & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *The ISME Journal*, *14*(4), Article 4. https://doi.org/10.1038/s41396-019-0580-z

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471–483. https://doi.org/10.1038/s41579-018-0040-1

Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., & Polz, M. F. (2008). Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science*, *320*(5879), 1081–1085. https://doi.org/10.1126/science.1157890

Hurwitz, B. L., Hallam, S. J., & Sullivan, M. B. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology*, *14*(11), R123. https://doi.org/10.1186/gb-2013-14-11-r123

Hwang, Y., Roux, S., Coclet, C., Krause, S. J. E., & Girguis, P. R. (2023). Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. *Nature Microbiology*, *8*(5), Article 5. https://doi.org/10.1038/s41564-023-01347-5

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), 119. https://doi.org/10.1186/1471-2105-11-119

Ignacio-Espinoza, J. C., Ahlgren, N. A., & Fuhrman, J. A. (2020). Long-term stability and Red Queen-like strain dynamics in marine viruses. *Nature Microbiology*, *5*(2), 265–271. https://doi.org/10.1038/s41564-019-0628-x

Jain. (2018). *High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries | Nature Communications*. https://www.nature.com/articles/s41467-018-07641-9

Jantaro, S., Mäenpää, P., Mulo, P., & Incharoensakdi, A. (2003). Content and biosynthesis of polyamines in salt and osmotically stressed cells of Synechocystis sp. PCC 6803. *FEMS Microbiology Letters*, *228*(1), 129–135. https://doi.org/10.1016/S0378-1097(03)00747-X

Jia, B., Jia, X., Kim, K. H., & Jeon, C. O. (2017). Integrative view of 2-oxoglutarate/Fe(II)-dependent oxygenase diversity and functions in bacteria. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1861*(2), 323–334. https://doi.org/10.1016/j.bbagen.2016.12.001

Jiang, X., Ha, C., Lee, S., Kwon, J., Cho, H., Gorham, T., & Lee, J. (2019). Characterization of Cyanophages in Lake Erie: Interaction Mechanisms and Structural Damage of Toxic Cyanobacteria. *Toxins*, *11*(8), 444. https://doi.org/10.3390/toxins11080444

Kehr, J.-C., & Dittmann, E. (2015). Biosynthesis and function of extracellular glycans in cyanobacteria. *Life (Basel, Switzerland)*, *5*(1), 164–180. https://doi.org/10.3390/life5010164

Kieft et al. (2022). *vRhyme enables binning of viral genomes from metagenomes | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/50/14/e83/6584432

Kieft, K., Zhou, Z., & Anantharaman, K. (2019). VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of virome function from genomic sequences. *bioRxiv*, 855387. https://doi.org/10.1101/855387

Kieft, K., Zhou, Z., Anderson, R. E., Buchan, A., Campbell, B. J., Hallam, S. J., Hess, M., Sullivan, M. B., Walsh, D. A., Roux, S., & Anantharaman, K. (2021). Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-23698-5

Kim, J.-G., Park, S.-J., Sinninghe Damsté, J. S., Schouten, S., Rijpstra, W. I. C., Jung, M.-Y., Kim, S.-J., Gwak, J.-H., Hong, H., Si, O.-J., Lee, S., Madsen, E. L., & Rhee, S.-K. (2016). Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7888–7893. https://doi.org/10.1073/pnas.1605501113

Kimura, S., Sako, Y., & Yoshida, T. (2013). Rapid microcystis cyanophage gene diversification revealed by long- and short-term genetic analyses of the tail sheath gene in a natural pond. *Applied and Environmental Microbiology*, *79*(8), 2789–2795. https://doi.org/10.1128/AEM.03751-12

Kimura-Sakai, S., Sako, Y., & Yoshida, T. (2015). Development of a real-time PCR assay for the quantification of Ma-LMM01-type Microcystis cyanophages in a natural pond. *Letters in Applied Microbiology*, *60*(4), 400–408. https://doi.org/10.1111/lam.12387

Koskella, B., & Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, *38*(5), 916–931. https://doi.org/10.1111/1574-6976.12072

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10.1101/gr.092759.109

Laczny, C. C., Sternal, T., Plugaru, V., Gawron, P., Atashpendar, A., Margossian, H. H., Coronado, S., der Maaten, L. van, Vlassis, N., & Wilmes, P. (2015). VizBin—An application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome*, *3*(1), 1. https://doi.org/10.1186/s40168-014-0066-1

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

Lennon, J. T., den Hollander, F., Wilke-Berenguer, M., & Blath, J. (2021). Principles of seed banks and the emergence of complexity from dormancy. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-24733-1

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Gu, L., Bai, S., Wang, J., Su, L., Wei, B., Zhang, L., & Fang, J. (2021). Characterization of particle-associated and free-living bacterial and archaeal communities along the water columns of the South China Sea. *Biogeosciences*, *18*(1), 113–133. https://doi.org/10.5194/bg-18-113-2021

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108. https://doi.org/10.1093/nar/gkt214

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, S., Boegman, L., & Rao, Y. R. (2021). Characterizing spatial and temporal distributions of turbulent mixing and dissipation in Lake Erie. *Journal of Great Lakes Research*, *47*(1), 168–179. https://doi.org/10.1016/j.jglr.2020.11.014

Lin, W., Li, D., Sun, Z., Tong, Y., Yan, X., Wang, C., Zhang, X., & Pei, G. (2020). A novel freshwater cyanophage vB_MelS-Me-ZS1 infecting bloom-forming cyanobacterium Microcystis elabens. *Molecular Biology Reports*, *47*(10), 7979–7989. https://doi.org/10.1007/s11033-020-05876-8

Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., & Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(30), 11013–11018. https://doi.org/10.1073/pnas.0401526101

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., Ackermann, M., Hahn, A. S., Srivastava, D. S., Crowe, S. A., Doebeli, M., & Parfrey, L. W. (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, *2*(6), Article 6. https://doi.org/10.1038/s41559-018-0519-1

Luque, A., & Silveira, C. B. (2020). Quantification of Lysogeny Caused by Phage Coinfections in Microbial Communities from Biophysical Principles. *mSystems*, *5*(5), 10.1128/msystems.00353-20. https://doi.org/10.1128/msystems.00353-20

Malki, K., Kula, A., Bruder, K., Sible, E., Hatzopoulos, T., Steidel, S., Watkins, S. C., & Putonti, C. (2015). Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. *Virology Journal*, *12*(1), 164. https://doi.org/10.1186/s12985-015-0395-0

Malki, K., Sawaya, N. A., Tisza, M. J., Coutinho, F. H., Rosario, K., Székely, A. J., & Breitbart, M. (2021). Spatial and Temporal Dynamics of Prokaryotic and Viral Community Assemblages in a Lotic System (Manatee Springs, Florida). *Applied and Environmental Microbiology*, *87*(18), e0064621. https://doi.org/10.1128/AEM.00646-21

Mankiewicz-Boczek, J., Jaskulska, A., Pawełczyk, J., Gągała, I., Serwecińska, L., & Dziadek, J. (2016). Cyanophages Infection of Microcystis Bloom in Lowland Dam Reservoir of Sulejów, Poland. *Microbial Ecology*, *71*(2), 315–325. https://doi.org/10.1007/s00248-015-0677-5

Marston, M. F., & Martiny, J. B. H. (2016). Genomic diversification of marine cyanophages into stable ecotypes. *Environmental Microbiology*, *18*(11), 4240–4253. https://doi.org/10.1111/1462-2920.13556

McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B., & Paul, J. H. (2010). High frequency of horizontal gene transfer in the oceans. *Science (New York, N.Y.)*, *330*(6000), 50. https://doi.org/10.1126/science.1192243

McKindles, K. M., Manes, M. A., DeMarco, J. R., McClure, A., McKay, R. M., Davis, T. W., & Bullerjahn, G. S. (2020). Dissolved Microcystin Release Coincident with Lysis of a Bloom Dominated by Microcystis spp. In Western Lake Erie Attributed to a Novel Cyanophage. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01397-20

Meyer, K. A., Davis, T. W., Watson, S. B., Denef, V. J., Berry, M. A., & Dick, G. J. (2017). Genome sequences of lower Great Lakes Microcystis sp. Reveal strain-specific genes that are present and expressed in western Lake Erie blooms. *PLOS ONE*, *12*(10), e0183859. https://doi.org/10.1371/journal.pone.0183859

Michalek et al. (2013). *Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions | PNAS*. https://www-pnas-org.proxy.lib.umich.edu/doi/10.1073/pnas.1216006110

Morimoto, D., Tominaga, K., Nishimura, Y., Yoshida, N., Kimura, S., Sako, Y., & Yoshida, T. (2019). Co-occurrence of broad and narrow host-range viruses infecting the toxic bloom-forming cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, AEM.01170-19. https://doi.org/10.1128/AEM.01170-19

Morimoto, D., Yoshida, N., Sasaki, A., Nakagawa, S., Sako, Y., & Yoshida, T. (2023). Ecological Dynamics of Broad- and Narrow-Host-Range Viruses Infecting the Bloom-Forming Toxic Cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, *89*(2), e02111-22. https://doi.org/10.1128/aem.02111-22

Morohoshi, T., Maruo, T., Shirai, Y., Kato, J., Ikeda, T., Takiguchi, N., Ohtake, H., & Kuroda, A. (2002). Accumulation of inorganic polyphosphate in phoU mutants of Escherichia coli and Synechocystis sp. Strain PCC6803. *Applied and Environmental Microbiology*, *68*(8), 4107–4110. https://doi.org/10.1128/AEM.68.8.4107-4110.2002

Naknaen, A., Suttinun, O., Surachat, K., Khan, E., & Pomwised, R. (2021). A Novel Jumbo Phage PhiMa05 Inhibits Harmful Microcystis sp. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.660351

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, *39*(5), 578–585. https://doi.org/10.1038/s41587-020-00774-7

NCBI. (2019, April 2). *BLAST+ v2.9.0*. NCBI Insights. https://ncbiinsights.ncbi.nlm.nih.gov/2019/04/02/blast-2-9-0-now-available-with-enhanced-support-for-new-database-format-and-improved-performance/

O'neil. (2012). *The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change—ScienceDirect*. https://www-sciencedirect-com.proxy.lib.umich.edu/science/article/pii/S1568988311001557

Ou, T., Gao, X.-C., Li, S.-H., & Zhang, Q.-Y. (2015). Genome analysis and gene nblA identification of Microcystis aeruginosa myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *Journal of General Virology*, *96*(12), 3681–3697. https://doi.org/10.1099/jgv.0.000290

Ouellette, A. J. A., Handy, S. M., & Wilhelm, S. W. (2006). Toxic Microcystis is widespread in Lake Erie: PCR detection of toxin genes and molecular characterization of associated cyanobacterial communities. *Microbial Ecology*, *51*(2), 154–165. https://doi.org/10.1007/s00248-004-0146-z

Paerl and Otten. (2013b). *Blooms Like It Hot | Science*. https://www.science.org/doi/10.1126/science.1155398

Paerl, H. W., & Huisman, J. (2009). Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports*, *1*(1), 27–37. https://doi.org/10.1111/j.1758-2229.2008.00004.x

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology*, *65*(4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y

Palermo, C. N., Shea, D. W., & Short, S. M. (2021). Analysis of Different Size Fractions Provides a More Complete Perspective of Viral Diversity in a Freshwater Embayment. *Applied and Environmental Microbiology*, *87*(11), e00197-21. https://doi.org/10.1128/AEM.00197-21

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

Pérez-Carrascal, O. M., Terrat, Y., Giani, A., Fortin, N., Greer, C. W., Tromas, N., & Shapiro, B. J. (2019). Coherence of Microcystis species revealed through population genomics. *The ISME Journal*, *13*(12), 2887–2900. https://doi.org/10.1038/s41396-019-0481-1

Poulos, B. T., John, S. G., & Sullivan, M. B. (2018). *Iron Chloride Flocculation of Bacteriophages from Seawater* (pp. 49–57). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7343-9_4

Pound, H. L., & Wilhelm, S. W. (2020). Tracing the active genetic diversity of Microcystis and Microcystis phage through a temporal survey of Taihu. *PLoS ONE*, *15*(12), e0244482. https://doi.org/10.1371/journal.pone.0244482

Preece, E. P., Hardy, F. J., Moore, B. C., & Bryan, M. (2017). A review of microcystin detections in Estuarine and Marine waters: Environmental implications and human health risk. *Harmful Algae*, *61*, 31–45. https://doi.org/10.1016/j.hal.2016.11.006

Qian et al. (2022). *Viruses | Free Full-Text | A Novel Freshwater Cyanophage, Mae-Yong924-1, Reveals a New Family*. https://www-mdpi-com.proxy.lib.umich.edu/1999-4915/14/2/283

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, *5*(1), 69. https://doi.org/10.1186/s40168-017-0283-5

Rosenwasser, S., Ziv, C., Creveld, S. G. van, & Vardi, A. (2016). Virocell Metabolism: Metabolic Innovations During Host-Virus Interactions in the Ocean. *Trends in Microbiology*, *24*(10), 821–832. https://doi.org/10.1016/j.tim.2016.06.006

Roth Rosenberg, D., Haber, M., Goldford, J., Lalzar, M., Aharonovich, D., Al-Ashhab, A., Lehahn, Y., Segrè, D., Steindler, L., & Sher, D. (2021). Particle-associated and free-living bacterial communities in an oligotrophic sea are affected by different environmental factors. *Environmental Microbiology*, *23*(8), 4295–4308. https://doi.org/10.1111/1462-2920.15611

Roux and Bolduc. (2016). *MAVERICLab / stampede-clustergenomes—Bitbucket*. https://bitbucket.org/MAVERICLab/stampede-clustergenomes/src/master/

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Lavigne, R., Brister, J. R., Varsani, A., Amid, C., Aziz, R. K., Bordenstein, S. R., Bork, P., Breitbart, M., Cochrane, G. R., Daly, R. A., Desnues, C., Duhaime, M. B., … Eloe-Fadrosh, E. A. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology*, *37*(1), Article 1. https://doi.org/10.1038/nbt.4306

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: Mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. https://doi.org/10.7717/peerj.985

Rozon and Short. (2013). *Complex seasonality observed amongst diverse phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario—Rozon—2013—Freshwater Biology—Wiley Online Library*. https://onlinelibrary-wiley-com.proxy.lib.umich.edu/doi/full/10.1111/fwb.12241

Salazar, G., Cornejo-Castillo, F. M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D., Arrieta, J. M., Duarte, C. M., Gasol, J. M., & Acinas, S. G. (2015). Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Molecular Ecology*, *24*(22), 5692–5706. https://doi.org/10.1111/mec.13419

Santos-Medellin, C., Zinke, L. A., ter Horst, A. M., Gelardi, D. L., Parikh, S. J., & Emerson, J. B. (2021). Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *The ISME Journal*, *15*(7), Article 7. https://doi.org/10.1038/s41396-021-00897-y

Schmidt, K. C., Jackrel, S. L., Smith, D. J., Dick, G. J., & Denef, V. J. (2020). Genotype and host microbiome alter competitive interactions between Microcystis aeruginosa and Chlorella sorokiniana. *Harmful Algae*, *99*, 101939. https://doi.org/10.1016/j.hal.2020.101939

Schmidt, M. L., White, J. D., & Denef, V. J. (2016). Phylogenetic conservation of freshwater lake habitat preference varies between abundant bacterioplankton phyla. *Environmental Microbiology*, *18*(4), 1212–1226. https://doi.org/10.1111/1462-2920.13143

Schroeder, D. C., Oke, J., Hall, M., Malin, G., & Wilson, W. H. (2003). Virus Succession Observed during an Emiliania huxleyi Bloom. *Applied and Environmental Microbiology*, *69*(5), 2484–2490. https://doi.org/10.1128/AEM.69.5.2484-2490.2003

Shaffer et al. (2020). *DRAM for distilling microbial metabolism to automate the curation of microbiome function | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/48/16/8883/5884738

Smith, D. J., Berry, M. A., Cory, R. M., Johengen, T. H., Kling, G. W., Davis, T. W., & Dick, G. J. (2022). Heterotrophic Bacteria Dominate Catalase Expression during Microcystis Blooms. *Applied and Environmental Microbiology*, *88*(14), e0254421. https://doi.org/10.1128/aem.02544-21

Smith, D. J., Tan, J. Y., Powers, M. A., Lin, X. N., Davis, T. W., & Dick, G. J. (2021). Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. *Environmental Microbiology*, *23*(6), 3020–3036. https://doi.org/10.1111/1462-2920.15514

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, *16*(8), Article 8. https://doi.org/10.1038/nrg3962

Steffen, M. M., Davis, T. W., McKay, R. M., Bullerjahn, G. S., Krausfeldt, L. E., Stough, J. M. A., Neitzey, M. L., Gilbert, N. E., Boyer, G. L., Johengen, T. H., Gossiaux, D. C., Burtner, A. M., Rowe, M., Dick, G. J., Meyer, K., Levy, S., Boone, B., Wynne, T., Zimba, P. V., … Wilhelm, S. W. (n.d.). Ecophysiological examination of the Lake Erie Microcystis bloom in 2014: Linkages between biology and the water supply shutdown of Toledo, Ohio. *Environmental Science*.

Steffen, M. M., Dearth, S. P., Dill, B. D., Li, Z., Larsen, K. M., Campagna, S. R., & Wilhelm, S. W. (2014). Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis. *The ISME Journal*, *8*(10), 2080–2092. https://doi.org/10.1038/ismej.2014.78

Stough, J. M. A., Tang, X., Krausfeldt, L. E., Steffen, M. M., Gao, G., Boyer, G. L., & Wilhelm, S. W. (2017). Molecular prediction of lytic vs lysogenic states for Microcystis phage: Metatranscriptomic evidence of lysogeny during large bloom events. *PLoS ONE*, *12*(9), e0184146. https://doi.org/10.1371/journal.pone.0184146

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics (Oxford, England)*, *27*(7), 1009–1010. https://doi.org/10.1093/bioinformatics/btr039

Suttle, C. A. (2007). Marine viruses—Major players in the global ecosystem. *Nature Reviews. Microbiology*, *5*(10), 801–812. https://doi.org/10.1038/nrmicro1750

Takashima, Y., Yoshida, T., Yoshida, M., Shirai, Y., Tomaru, Y., Takao, Y., Hiroishi, S., & Nagasaki, K. (2007). Development and Application of Quantitative Detection of

Cyanophages Phylogenetically Related to Cyanophage Ma-LMM01 Infecting *Microcystis aeruginosa* in Fresh Water. *Microbes and Environments*, *22*(3), 207–213. https://doi.org/10.1264/jsme2.22.207
Tucker, S., & Pollard, P. (2005). Identification of Cyanophage Ma-LBP and Infection of the Cyanobacterium Microcystis aeruginosa from an Australian Subtropical Lake by the Virus. *Applied and Environmental Microbiology*, *71*(2), 629–635. https://doi.org/10.1128/AEM.71.2.629-635.2005
Visser, P. M., Verspagen, J. M. H., Sandrini, G., Stal, L. J., Matthijs, H. C. P., Davis, T. W., Paerl, H. W., & Huisman, J. (2016). How rising CO2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae*, *54*, 145–159. https://doi.org/10.1016/j.hal.2015.12.006
Wang, F., Li, D., Cai, R., Pan, L., Zhou, Q., Liu, W., Qian, M., & Tong, Y. (2022). A Novel Freshwater Cyanophage Mae-Yong1326-1 Infecting Bloom-Forming Cyanobacterium Microcystis aeruginosa. *Viruses*, *14*(9), 2051. https://doi.org/10.3390/v14092051
Wanner, B. L. (1993). Gene regulation by phosphate in enteric bacteria. *Journal of Cellular Biochemistry*, *51*(1), 47–54. https://doi.org/10.1002/jcb.240510110
Watkins, S. C., Smith, J. R., Hayes, P. K., & Watts, J. E. M. (2014). Characterisation of Host Growth after Infection with a Broad-Range Freshwater Cyanopodophage. *PLOS ONE*, *9*(1), e87339. https://doi.org/10.1371/journal.pone.0087339
Weitz, J. S., & Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports*, *4*, 17. https://doi.org/10.3410/B4-17
Williamson, S. J., Allen, L. Z., Lorenzi, H. A., Fadrosh, D. W., Brami, D., Thiagarajan, M., McCrow, J. P., Tovchigrechko, A., Yooseph, S., & Venter, J. C. (2012). Metagenomic Exploration of Viruses throughout the Indian Ocean. *PLOS ONE*, *7*(10), e42047. https://doi.org/10.1371/journal.pone.0042047
Wilson, A. E., Wilson, W. A., & Hay, M. E. (2006). Intraspecific variation in growth and morphology of the bloom-forming cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, *72*(11), 7386–7389. https://doi.org/10.1128/AEM.00834-06
Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., Roth, J. R., & Bäumler, A. J. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, *467*(7314), Article 7314. https://doi.org/10.1038/nature09415
Wucher, B. R., Winans, J. B., Elsayed, M., Kadouri, D. E., & Nadell, C. D. (2023). Breakdown of clonal cooperative architecture in multispecies biofilms and the spatial ecology of predation. *Proceedings of the National Academy of Sciences*, *120*(6), e2212650120. https://doi.org/10.1073/pnas.2212650120
Xiao, M., Willis, A., & Burford, M. A. (2017). Differences in cyanobacterial strain responses to light and temperature reflect species plasticity. *Harmful Algae*, *62*, 84–93. https://doi.org/10.1016/j.hal.2016.12.008
Yancey, C. E., Smith, D. J., Den Uyl, P. A., Mohamed, O. G., Yu, F., Ruberg, S. A., Chaffin, J. D., Goodwin, K. D., Tripathi, A., Sherman, D. H., & Dick, G. J. (2022). Metagenomic and Metatranscriptomic Insights into Population Diversity of Microcystis Blooms: Spatial and Temporal Dynamics of mcy Genotypes, Including a Partial Operon That Can Be Abundant and Expressed. *Applied and Environmental Microbiology*, *88*(9), e0246421. https://doi.org/10.1128/aem.02464-21
Yang, F., Jin, H., Wang, X.-Q., Li, Q., Zhang, J.-T., Cui, N., Jiang, Y.-L., Chen, Y., Wu, Q.-F., Zhou, C.-Z., & Li, W.-F. (2020). Genomic Analysis of Mic1 Reveals a Novel Freshwater Long-Tailed Cyanophage. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.00484
Ye, W., Tan, J., Liu, X., Lin, S., Pan, J., Li, D., & Yang, H. (2011). Temporal variability of cyanobacterial populations in the water and sediment samples of Lake Taihu as determined

by DGGE and real-time PCR. *Harmful Algae*, *10*(5), 472–479. https://doi.org/10.1016/j.hal.2011.03.002
Yoshida, M., Yoshida, T., Kashima, A., Takashima, Y., Hosoda, N., Nagasaki, K., & Hiroishi, S. (2008). Ecological dynamics of the toxic bloom-forming cyanobacterium Microcystis aeruginosa and its cyanophages in freshwater. *Applied and Environmental Microbiology*, *74*(10), 3269–3273. https://doi.org/10.1128/AEM.02240-07
Yoshida, T., Takashima, Y., Tomaru, Y., Shirai, Y., Takao, Y., Hiroishi, S., & Nagasaki, K. (2006). Isolation and Characterization of a Cyanophage Infecting the Toxic Cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, *72*(2), 1239–1247. https://doi.org/10.1128/AEM.72.2.1239-1247.2006
Yoshida-Takashima, Y., Yoshida, M., Ogata, H., Nagasaki, K., Hiroishi, S., & Yoshida, T. (2012). Cyanophage infection in the bloom-forming cyanobacteria Microcystis aeruginosa in surface freshwater. *Microbes and Environments*, *27*(4), 350–355. https://doi.org/10.1264/JSME2.ME12037
Zhang, D., He, Y., & Gin, K. Y.-H. (2022). Genomic Characterization of a Novel Freshwater Cyanophage Reveals a New Lineage of Cyanopodovirus. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.768868
Zhang, S., He, X., Cao, L., Tong, Y., Zhao, B., & An, W. (2022). A Novel Wide-Range Freshwater Cyanophage MinS1 Infecting the Harmful Cyanobacterium Microcystis aeruginosa. *Viruses*, *14*(2), 433. https://doi.org/10.3390/v14020433
Zhu, L., Zancarini, A., Louati, I., De Cesare, S., Duval, C., Tambosco, K., Bernard, C., Debroas, D., Song, L., Leloup, J., & Humbert, J.-F. (2016). Bacterial Communities Associated with Four Cyanobacterial Genera Display Structural and Functional Differences: Evidence from an Experimental Approach. *Frontiers in Microbiology*, *7*. https://www.frontiersin.org/articles/10.3389/fmicb.2016.01662
Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., Waldbauer, J. R., & Coleman, M. L. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology*, *18*(1), Article 1. https://doi.org/10.1038/s41579-019-0270-x

Chapter 3: Deciphering Total Viral Community Structure and Metabolic Potential in a Great Lakes Cyanobacterial Harmful Algal Bloom

A.J Wing[1], Eric Bastien[1], Vincent Denef[1], Bridget Hegarty[1,2], Jacob Evans[1], Gregory Dick[1,3], Melissa Duhaime[1]

[1] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI

[2] Department of Civil and Environmental Engineering, Case Western Reserve University, Cleveland, OH

[3] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI

Abstract

Viruses play a vital role in microbial communities and cyanobacterial harmful algal blooms (cHABs) are no exception. However, the viral community structure and microbial predator-prey dynamics in freshwater cHABs are largely undescribed. We used a community genomics approach to better understand the role of viruses in the 2014 Lake Erie cHAB event, a bloom often dominated by toxigenic, colony-forming cyanobacterium, *Microcystis aeruginosa*. We sequenced four virus-enriched metagenomes (<0.22 μm) and 32 cellular metagenomes from four filter fractions that were either colony-associated (53 and 100 μm) or not (>0.22 and 3 μm). We identified 15,461 viral operational taxonomic units (vOTUs), tracked viral community ecology, elucidated their metabolic functions, and predicted potential microbial hosts from the same samples. We identified a significant correlation between viral community structure and sampling date and filter size fraction, but not station. Our study offered a novel perspective by tracking viral abundances and virus-host interactions across five size fractions. The observations are consistent with fraction-specific Kill-the-Winner dynamics that give rise to trends that support the 'viral bank' model. We observed a

diverse viral fraction containing the viral bank, with rare instances of ‘active’ viruses becoming abundant in the cellular, and especially colony-associated, fractions. Notably, virus-host interactions predominantly varied by both date and size fraction. Despite the presumed dominance of *Microcystis*, most virus-host interactions occur independently of this cyanobacterium. This study sheds new light on the ecological and evolutionary aspects of cHABs emphasizing the significance of genomic novelty and predator-prey interactions contributed by viruses.

3.1 Introduction

Climate change and anthropogenic nutrient inputs have led to greater occurrence and intensity of cyanobacterial harmful algal blooms (cHABs; Paerl and Huisman, 2009; Michalek et al., 2013; Visser et al., 2016). A cHAB occurs in the western basin of Lake Erie each year that is often attributed to *Microcystis*, a toxin-producing cyanobacteria that can comprise a substantial portion of the bloom community (Berry et al., 2017). While phosphorus and nitrogen inputs are thought to be primary drivers of bloom dynamics (Harke and Gobler, 2013; Harke et al., 2015), top-down controls on *Microcystis*, such as viral infection known to control aquatic phytoplankton blooms (Schroeder et al., 2003; Sorensen et al., 2009; Trainic et al., 2018), have been less explored. Evidence suggests that Lake Erie *Microcystis* is infected by viruses (Jiang et al., 2019; McKindles et al. 2020; Meyer et al., 2017; Steffen et al., 2015), but the impact of infection on the bloom dynamics is not known. We lack a comprehensive understanding of the viral community dynamics tied to cHAB progression, specifically the prevalence and diversity of viral populations that comprise these complex viral communities.

While *Microcystis* can dominate the Lake Erie cHABs, they do not comprise a majority of the community (Berry et al., 2017). In the Great Lakes and elsewhere, these blooms are far from homogeneous (Cook et al., 2020; Eiler and Bertilsson, 2004; Smith et al., 2021). The specific interactions between *Microcystis* and these co-existing community members remain enigmatic, though lab and field-based evidence suggests *Microcystis* tends to proliferate in the presence of heterotrophic bacteria (Kim et al., 2019). It is believed that *Microcystis*-induced changes in the environmental conditions

can play a role in shaping the ecology of its mutualistic partners (Bullerjahn et al., 2016, Paerl and Otten, 2013, Smith et al., 2021). The unknown effects of ecological interactions, mutualistic and otherwise, among members of the bloom community limits our understanding of cHAB dynamics.

In addition to infecting *Microcystis*, viruses can also exert an indirect "top-down" influence by infecting other community members. Research efforts have predominantly focused on *Microcystis* and its viruses, with less attention given to the broader viral community. The potential for viruses to suppress the growth of competitive species is explained through the "kill the winner" (KtW) hypothesis (Winter et al., 2010), a framework developed to understand the role viruses may play in equilibrium dynamics in plankton communities. In a KtW scenario, viral infections can lead to population reductions in competitive species, thereby creating opportunities for the proliferation of new competitors (Thingstad and Lignell, 1997). In this way, expanding our knowledge of viruses at the community level, not just viruses of *Microcystis*, holds promise for improving our ability to understand and predict overall cHAB dynamics in Lake Erie.

In this study, we described the viral community across time, space and size fractions through the 2014 Lake Erie cHAB season. We observed that abundant vOTUs are rare and sporadically observed, suggesting that cHAB viral activity is fraction-specific. We demonstrate virus-host network turnover through cHAB progression, where we observe few shared virus-host pairs between colony and non-colony associated fractions on the same dates. This study provides novel insights into the Lake Erie cHAB viral community, and in a broader context, contributes to a more comprehensive understanding of the ecological and evolutionary impact of vOTUs on complex microbial communities in cHABs.

3.2 Results and Discussion

3.2.1 Lake Erie cHAB viral community ecology

3.2.1.1 Thousands of undescribed viral OTUs highlight the novelty of Lake Erie viral communities

In the western basin of Lake Erie (Fig. 10A), a multispecies cHAB persisted between July and October of 2014. Particulate phycocyanin (proxy for cyanobacteria) and

microcystin (proxy for bloom toxicity) concentrations indicated a toxic *Microcystis* bloom occurred at all three sampling stations (WLE12, WLE2 and WLE4) in early August, followed by a non-toxic cyanobacterial bloom that occurred primarily at the nearshore stations late September (WLE12 and WLE2) (SI Fig. 1). The dual peak 2014 Lake Erie bloom pattern, whereby only the first peak was toxin-producing, has been described (Cory et al., 2016; Berry et al., 2017; Smith et al., 2021; Yancey et al., 2022; Wing et al., 2024a). To study the viruses associated with the bloom, we sampled the WLE12, WLE2, and WLE4 stations across time and different fractions (Fig 10B), resulting in 36 metagenomes that were individually assembled to form 155,025 contigs longer than 3 kb, 27,086 (20%) of which were identified as viral (Fig. 10B; SI Fig. 2). The fraction of contigs identified as viral from the cell-enriched fractions ranged from 1-12% (Fig. 10B).

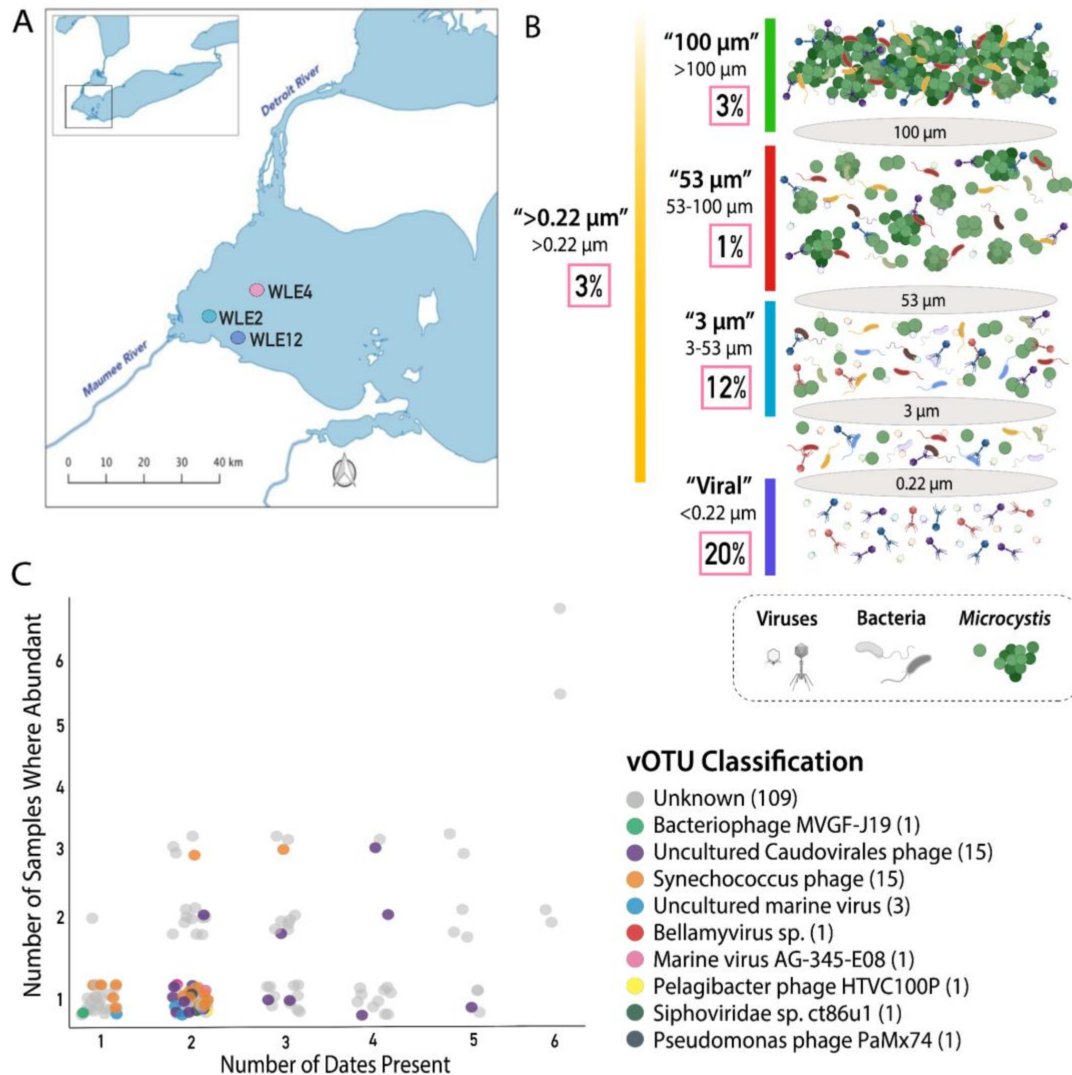


Figure 10. Lake Erie viral community sampling overview and abundant vOTU taxonomic classification. (A) A map of sampling sites located in the western basin of Lake Erie. Three sites were sampled bi-monthly in June and weekly from July to October of 2014. (B) Sampling filter size fractionation breakdown. Percentage in pink box indicates fraction of total contigs derived from each fraction identified as viral. (C) Scatterplot showing the 148 vOTUs identified as abundant (mapping >0.5% of reads in a given sample) and the relationship between the number of samples where a vOTU was abundant and the number of sampling dates in which that vOTU was found.

Based on sequence similarity, 14,744 (54%) of the viral contigs formed 3,527 viral clusters (Bin Jang et al., 2019) that approximate genus-level groups (80% ANI across 80% of the contig; Bolduc et al., 2017) (SI Fig. 3) and 15,461 viral operational taxonomic units (vOTUs) (Table S1) that approximate species-level groups (95% ANI across 85% of the contig; Roux et al., 2019). Of the 3,527 viral clusters that approximated genus level groups, 3,447 of these genus level groups were novel. Only 80 clustered with known reference viruses from the database used for vConTACT2.

Only 26% of the 148 abundant and ubiquitous vOTUs were identified as known based on sequence similarity to viruses, both isolated and not, in the NCBI viral sequence database, 15 of which were *Synechococcus* phages (Fig. 10C). The prevalence of *Synechococcus* phages suggests active infection of *Synechococcus* hosts. This is plausible given that *Synechococcus* was the genera detected in greatest relative abundance in a study of the 2014 bloom's bacterial community (Berry et al., 2016); its cooccurrence with *Microcystis* has been reported previously in this (Ouellette et al., 2006; Smith et al., 2021) other lakes (Ye et al., 2011). We next sought to describe spatial and temporal diversity of Lake Erie cHAB vOTUs through the bloom season.

3.2.1.2 Virome diversity greatest in the viral fraction, lowest in colony-associated fractions during bloom peaks

We first examined trends in within-sample diversity (alpha-diversity) across dates, locations and sample filter fractions. The diversity (Shannon's H) of the viral fraction did not display major shifts from July to October (Fig. 11A). The bacterial community diversity, assessed using metagenomic data, showed a decreasing trend for the entire bacterial community (>0.22 µm fraction). This is consistent with the overall seasonal trend in bacterial community evenness, though opposite to increasing richness measured using 16S rRNA gene sequencing (Berry 2017). However, the viral and metagenomic bacterial sampling used here lacked the same temporal resolution of Berry et al., 2017. As could be expected, the bacterial diversity was highest in the full community fraction (>0.22 µm) and lower in other fractions enriched for portions of the entire community (Fig. 11B; SI Table S2).

The viral diversity patterns were more complex: diversity observed in the colony-associated fractions (53 and 100 µm; Fig. 10B) was lower than that of the non-colony-associated fractions (viral, >0.22 and 3 µm). Significant differences in alpha diversity were most common between fractions that were colony associated and not colony associated (Fig. 11A; Table S3). Size-fractionation virome studies are rare, but one other such report of ocean viruses similarly observed greater viral diversity in the viral fraction than cellular fractions (Dart et al., 2023). This trend, which was supported by our Lake Erie data, was attributed to seed bank theory, whereby a highly diverse 'bank'

of viruses persists in the submicron size fraction from which infections are drawn when a virus meets a sensitive host.

Notably, the lowest viral diversity observed at nearly every station was in the colony-associated fractions on the dates of the bloom peaks (Aug 4, Sept 29; Fig. 11A). We attribute this to active infection of colony-associated host cells captured on those filter fractions; a reduction in host diversity caused by the bloom led to a corresponding reduction in viral diversity not seen in the lake at other points in the bloom. The low diversity of the colony-associated fraction on the bloom peaks was similar to that of bathypelagic and mesopelagic polar ocean viral communities (Gregory et al., 2019), which are among the lowest reported using the same community genomic approach. The high diversity of the Lake Erie viral fraction is on par with those of viromes from freshwater reservoirs in China (Shannon H' mean 10.4; richness >20,000; Gu et al., 2018). These Shannon H' measures are higher than those reported in polar and ocean studies where they range ~3-8 for arctic lakes (Aguirre De Carcer et al., 2015), ~8 for the global ocean (Brum, 2016), ~6-8.5 in the South China Sea (Liang et al., 2019). Though comparing Shannon H' measures using virome data can be challenging given constraints on data acquisition, quality, and quantity and the myriad bioinformatic approaches used to generate these estimates, our Lake Erie whole viral community data support a trend of temperate lake systems having among the most diverse viral communities described thus far.

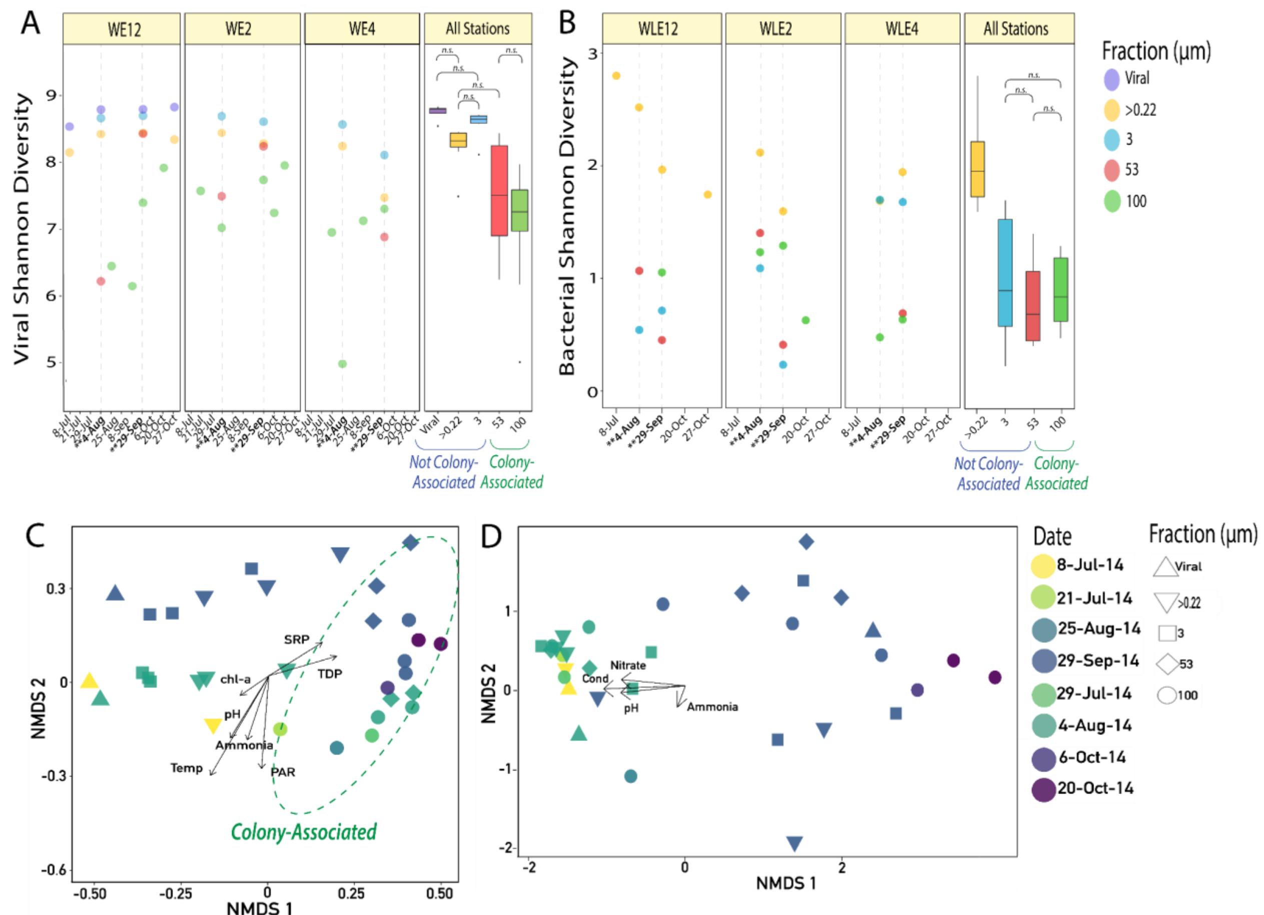


Figure 11. Lake Erie viral community overview. (A) Shannon diversity of viral community through time and space. (B) Shannon diversity of bacterial community across time and station. Panel A-B color indicates sampling fraction. (C) Non-metric multidimensional scaling (NMDS) (stress-value = 0.14) ordination based on Bray-Curtis dissimilarities in vOTU abundances. (D) NMDS (stress-value = 0.09) ordination based on Bray-Curtis dissimilarities in viral metabolic gene abundances. Panels C-D color represents the sampling date; shape represents sample fraction; ordinations are overlaid with the gradient of the fit between the significant environmental parameters and the Bray-Curtis dissimilarities; vector lengths represent strength of the correlation with the data variation.

3.2.1.3 Viral community turnover correlates with with date and fraction, not station

We next evaluated trends in vOTU turnover between sample dates, stations, and fractions, i.e., viral community beta diversity. Both date and fraction significantly correlated with viral community turnover, correlating with 25% and 18% of the total between-sample variation, respectively, thereby together explaining nearly half of the total variability (Fig. 11C; Table S4). Time has been identified as a significant factor in explaining shifts in viral community structure in the ocean (Chow & Fuhrman, 2012; Brum, 2016) and freshwater springs (Malki et al., 2021), likely due to shifts in host communities as microbes respond to seasonal physiochemical changes. Of the

environmental parameters tested, temperature, chlorophyll *a*, ammonia, soluble reactive phosphate, and total dissolved phosphate measurements were significantly correlated with variation in viral community structure but had little explanatory power (each with R^2 <0.05; Table S5). Variation in the taxonomic representation of viruses across different filter fractions has been reported in ocean (Williamson et al., 2012), soil (Santos-Medellin et al., 2021; Palermo et al., 2021, Dart et al., 2023), though the difference is not always significant (Hegarty et al. 2022). Such partitioning of viral community variability by filter fractions can be explained by changes in host microbial populations, as aquatic bacterial communities collected across different filter pore size fractions in freshwater and marine systems are reported to partition by fraction (Schmidt et al., 2016; Salazar et al., 2015).

Sampling location was not significantly correlated with variation in whole viral community structure (Table S4). In a 2014 Lake Erie study of viruses predicted to infect *Microcystis* (rather than total viral community reported here) it was similarly observed that despite being only seven miles apart, the station did not significantly influence variability of *Microcystis* viral assemblage structure in Lake Erie (Wing et al., 2024a). These trends in viral community composition align with previous analyses of broader Lake Erie bloom community dynamics, whereby bacterial community structure was found to vary more seasonally than spatially (Berry et al., 2017a; Smith et al., 2021). This may be due to the fact that Lake Erie, especially the western basin, is known for its high turbulent kinetic energy that can contribute to mixing, especially at the lake surface (Lin et al., 2021). Combined with the influx of water to the western basin from the Maumee and Detroit Rivers (Fig. 10A), turbulent eddies can move water on short length, time, and velocity scales that are significant for dispersal of plankton (Lin et al., 2021), including bacterioplankton and blooming biomass. Mixing could also explain why there is only weak correlation between viral community structure and environmental parameters. While temporal trends and fraction-specific patterns in viral diversity were observed at the micron scale, the western basin appears to have been mixed well enough to dissipate station-driven influences on viral community structure.

3.2.1.4 Viral metabolic diversity correlates with date only, suggesting viral functional redundancy across fractions

In addition to vOTUs, the distribution of viral-encoded auxiliary metabolic genes (AMGs) was tracked through space (stations and fractions) and time. As with viral community structure, viral genes were found to cluster by sampling date (Fig. 11D; Table S6) and to not significantly vary with sampling location (Table S6). Conductivity, ammonia, and pH significantly correlated with variation in viral AMG distribution but had little explanatory power (each with R^2 <0.04; Table S7). Unlike for vOTUs distributions, sampling fraction did *not* significantly correlate with viral metabolic gene distributions (Table S6). While size fractionation has previously been found in early pyrosequencing virome studies to distinguish viruses with different functional potentials (Williamson et al., 2012), our findings are consistent with those of a drinking water study that found that size fractionation had no significant effect on viral metabolic gene distribution (Hegarty et al., 2022). Notably, the drinking water study and our work here used a shared Illumina sequencing approach; we suspect that the ability to more deeply sequence the functional virome in each fraction (rather than only the dominant viruses detected in Williamson et al.) can explain why we saw similar trends as Hegarty et al. These patterns could be explained by functional redundancy across fractions that arises even when taxonomic differences are found. Functional redundancy is known to exist in microbial communities and thought to emerge as a consequence of biotic, environmental, and spatial processes (Louca et al., 2021), though is not often considered in viral communities. Alternatively, this trend may be a consequence of undersampling functions, given that most viral genes identified in this study shared little to no homology with reference genes in public databases; a common finding for environmental metagenomes and metaviromes (Deboutte et al., 2020; Gregory et al., 2019).

3.2.2 Viral 'Bank' framework suggests cHAB viral activity is fraction-specific

3.2.2.1 Abundant Lake Erie vOTUs were rare and sporadically observed

Only 5.7% (n=889) of the vOTUs were ever considered highly abundant (defined here as >0.5% of viral reads) in any sample (Table S8). The vOTUs identified as abundant

on a given date were rarely abundant in following dates, as indicated by the prominent diagonal lines as vOTUs rise and fall based on their abundance score (Fig. 12A). The abundant viruses ranged from 0.1-5.7% relative abundance of viral reads, with the highest reached by vOTU_596 on 4 August in the 100 μm fraction (Fig. 12B; Table S9). This virus had reads mapped across the full length of its genome (SI Fig. 4) and is a novel virus with no sequence homology to known viruses in NCBI.

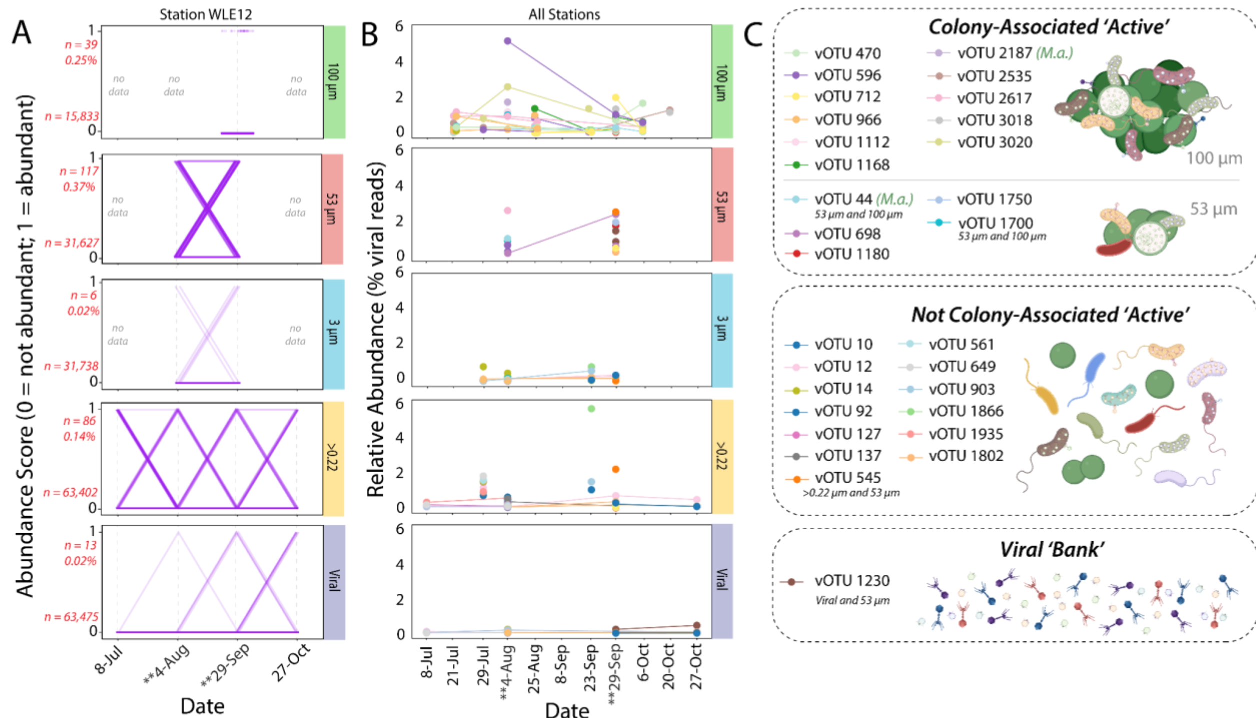


Figure 12. Lake Erie vOTU dynamics. (A) Tracking abundant vOTUs through time by fraction at site WLE12 (only site where viral fraction was sampled). Each line tracks a single vOTU. If >0.5% of the total reads that mapped to viruses map to a given vOTU on the specified date, it was assigned an abundance score of "1"; if <0.5% viral reads map to the vOTU it was assigned "0." Sample sizes (*n*) on the y-axis in red font report the number of 'not abundant' and 'abundant' vOTUs, as well as percentage of abundant vOTUs, in each fraction. (B) Temporal relative abundance dynamics of the thirty most abundant vOTUs across the entire dataset separated by sampling fraction. (C) vOTUs depicted in panel B are listed to the right of the plots in which they rise above the zero line. Conceptual model depicting infection of bacteria in colonies by "Active" viruses that are colony-associated (53 and 100 μm), infection of bacteria by "Active" viruses *not* colony-associated (>0.22 and 3 μm), and free viruses of the Viral "Bank."

3.2.2.2 Succession in Lake Erie vOTUs explained by Viral Bank model

We hypothesized that if the Lake Erie viral community adheres to principles of viral Bank theory, seasonal succession of vOTUs would arise through the progression of the bloom and abundant vOTUs would sporadically rise to dominance in the colony-associated size fractions. The virus 'Bank' model was proposed by Breitbart and Rohwer (2005) to reconcile the observed global dispersal and high local diversity of

viruses. It is premised on the assumption that in a given system the majority of viruses exist in a non-active state, while only a small subset of this 'bank' are active at any given time (Breitbart et al., 2005), similar to seed bank theory (Lennon et al., 2021). Indeed, among the Lake Erie cHAB viruses, few vOTUs (0.02-0.37% vOTUs in a fraction) ever reach an 'abundant' level (Fig. 12A-B), at which point they represented the active fraction of the community (Fig. 12B-C). Meanwhile, the majority of the viruses existed at low abundance in the viral bank (Fig. 12B-C). Not all active vOTUs were observed in the viral bank across sampling points, as Bank model would predict, but that is likely due to the inability of metagenomics to capture all rare members of the viral community.

Previous studies have used the Bank theory framework to understand marine viral community dynamics (Breitbart et al., 2005; Aylward et al., 2017; Hevroni et al., 2020; Dart et al., 2023). Given our study design, our work expanded on these previous findings by offering a unique opportunity to distinctly describe the viruses of the colony-associated and not-colony-associated cellular fractions. We found that each fraction contained different sets of active viruses (Fig. 12B-C). After infection, we posit that the newly released free virions were released to the viral bank, contributing to its high local diversity (Fig. 11A-B). As for similar results described by Dart et al., 2023, this observation can be explained through the 'Kill the Winner' framework (Winter et al., 2010), whereby the viruses that dominated in the different cellular fractions were infecting the distinct colony-associated/not colony-associated "winner" hosts (Table S10). This is supported by the observation that the highest vOTU relative abundances were observed near the bloom peaks (Fig. 12B; SI Fig. 1). We propose that the dense colonies, which are composed of cyanobacteria and their associated heterotrophic bacteria (Smith et al., 2021), may be hotspots of viral activity. Colonies may serve as multi-species islands that increase the opportunities for viruses to encounter new hosts in close proximity without needing to "forage" as planktonic viruses in the bank. Further work to examine virus-host dynamics within individual colonies can evaluate the potential for cross-species infection and how microbial dynamics at the microscale may impact cHAB seasonal progression.

3.2.3 Predicted virus-host networks underscore the complexity of interactions in the diverse Lake Erie cHAB community

3.2.3.1 The majority of virus-host interactions are among non-*Microcystis* community members

Virus-Host Infection Predictor (VHIP) (Bastien et al., 2023), a machine-learning approach, was used to predict infection-related interactions between bacterial metagenome assembled genomes (MAGs; Table S11-Table S12) and the vOTUs from the 2014 Lake Erie bloom community). 29,976 infection interactions were predicted between 4,090 vOTUs and 31 bacterial MAGs reconstructed from samples collected on Aug 4 at station WLE12 (Fig. 13A). Similarly, 26,615 infection interactions were predicted between 4,147 vOTUs and 32 bacterial MAGs on Sept 29 at WLE 12 (SI Fig. 5). Hundreds of vOTUs were predicted to infect the two *Microcystis* MAGs (Ma_MAG_1 and Ma_MAG_2), yet over 9,000 vOTUs were predicted to infect other community members and formed a dense cluster far from *Microcystis* in the network (Fig. 13A). This supports the growing recognition that though *Microcystis* can at times be observed as the most abundant genus (Berry et al., 2017), it exists in a diverse community context. Our observations further highlight the diversity of virus-host interactions that exist as well, an appreciation of which may be critical to understanding the underpinning of cHAB progression.

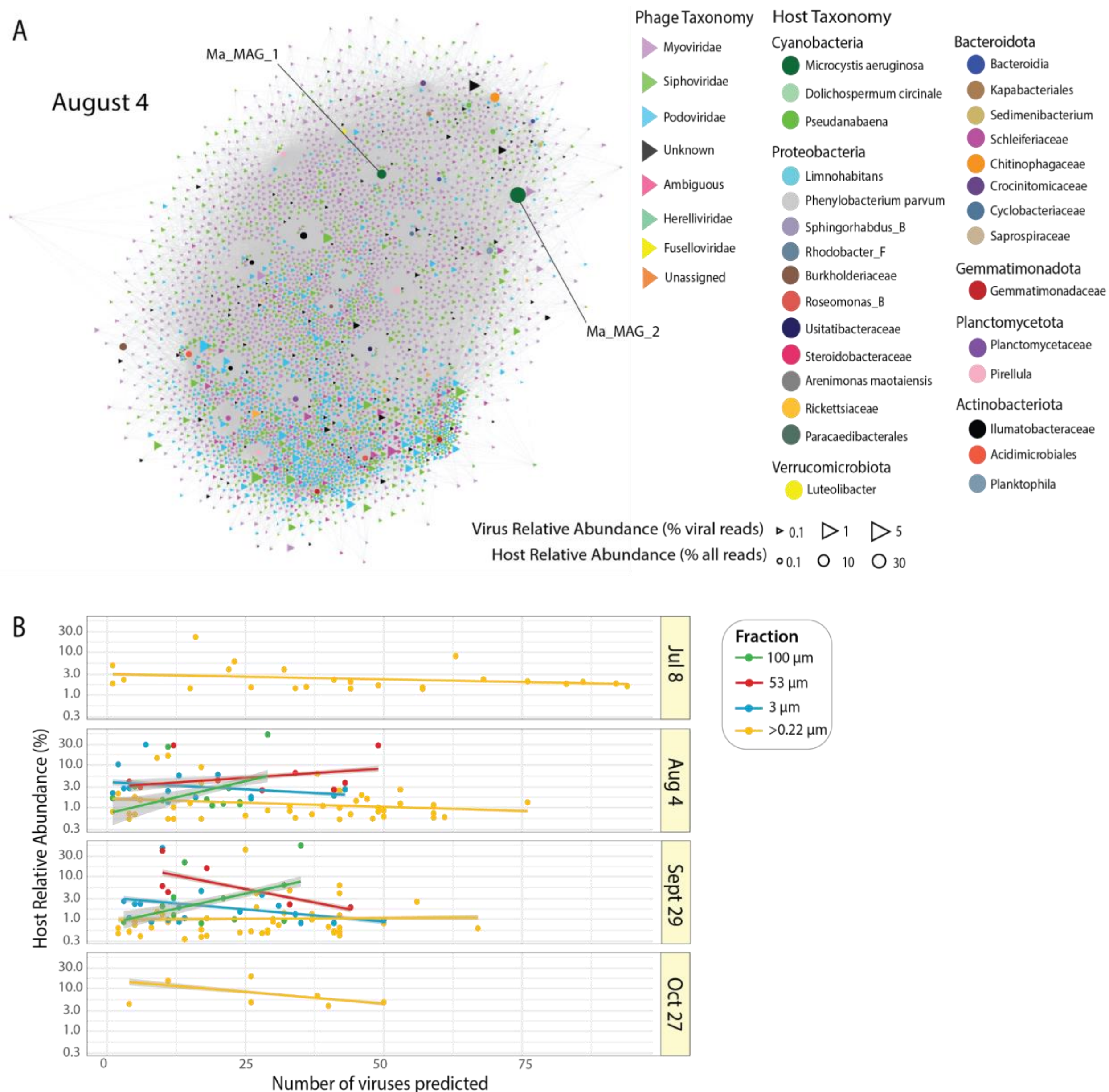


Figure 13. Predicted virus-host interactions between bacterial metagenome assembled genomes (MAGs) and vOTUs of the 2014 Lake Erie bloom community. (A) Predicted virus-host network from 4 August bloom peak in the station WLE 12 sample, all fractions combined. Circle nodes are host MAGs; sizes represent the MAG relative abundance in the sample. Triangle nodes are viruses; sizes represent the average relative abundance of the vOTUs across fractions. Node colors represent assigned taxonomy. For vOTUs, taxonomy was assigned according to the Phage Taxonomy Tool (PTT; Kieft et al., 2020); for viruses, “Unknown” indicates the vOTU has no hit in the PTT reference database, “Unassigned” indicates the vOTU has a hit to something unassigned in the reference database, and “Ambiguous” refers to a situation in which a protein or virus had significant hits to multiple proteins within the database, but the program could not distinguish between taxonomic assignments. Only predictions with >93% infection probability are shown. Sept 29 virus-host network in SI Fig. 5. (B) Relationship between host relative abundance and the number of viruses predicted to infect a given host, separated by date. Line and point colors represent filter fractions. Grey shading represents 95% confidence interval of the linear regression.

3.2.3.2 The majority of predicted *Microcystis* viruses are present at low abundance through the bloom

Though hundreds of vOTUs were predicted to infect each *Microcystis* MAG, most remained at low abundance throughout the bloom. Two exceptions were vOTU_2187 and vOTU_44, which rose to abundance in colony-associated fractions on Aug 4 (Fig. 12B-C). These were only two of the top 30 most abundant vOTUs across the entire dataset (Fig. 12), further supporting the importance of non-*Microcystis* virus-host dynamics in understanding the community ecology. Notably, Ma_MAG_2, the more abundant *Microcystis* MAG at each bloom peak, had fewer predicted viruses than Ma_MAG_1, suggesting that the VHIP viral predictions for *Microcystis* are strain-specific and that host relative abundance may not be predictive of the number of viruses of a given host. To expand upon this observation, we evaluated the relationship between all hosts and their predicted viruses.

3.2.3.3 Fraction-specific trends in predicted viral host ranges

The number of viruses predicted to infect host MAGs depended on the fraction in which the viruses were observed. The greatest number of predicted viruses per host was observed among the >0.22 µm fraction, followed by the 3, 53, and 100 µm fractions (Fig. 13B). In most cases, the number of viruses predicted per host showed no trend or was slightly negatively correlated with the host relative abundance (Fig. 13B; Table S13). A notable exception was the case of the 100 µm samples from the Aug 4 and Sept 29 bloom peaks, which showed a positive correlation between host relative abundance and the number of viruses predicted to infect the host MAG (Fig. 13B; Table S13). We propose that this trend can be explained through the lens of a fraction-dependent Kill-the-Winner and Bank frameworks. When a given host population is more abundant (a ‘winner’), it draws more viruses from the viral ‘bank’, thereby moving them to the active pool. Here they are detected in the colony-associated fractions where the “winner” host taxa are thriving. The colony may provide new hosts in close proximity that can lead to an increase in the number of viruses infecting a given host, thereby

representing a unique fraction-specific dynamic that occurs in the context of the cHAB colonies.

3.2.3.4 Virus-host interaction dynamics reflect vOTUs dynamics through the cHAB season

We next looked at the turnover of networks through time at spatial scales of miles (cross-station comparison) and microns (cross-filter comparison) to better understand constraints on virus-microbe interactions. To do this, we tracked the occurrence of each predicted virus-host pair across dates, stations, and sample fractions. Examining these trends between stations, the networks collected from different stations shared more virus-host pairs than those collected on different dates but at the same station (Fig. 14A). A notable exception is the limited overlap between WLE4 compared to WLE2 and WLE12 (Fig. 14A), in line with the limited bloom at WLE4 at that time (Berry et al., 2017). The importance of date in structuring these interactions is consistent with the significant role of date in explaining variation in viral community structure (Fig. 11C). This observation underscores the dynamic nature of these interactions, where ecological (assembly and turnover of host populations based on environmental conditions, competitive and consumer-prey interactions) and coevolutionary processes continually shape the relationships between viruses and their bacterial hosts.

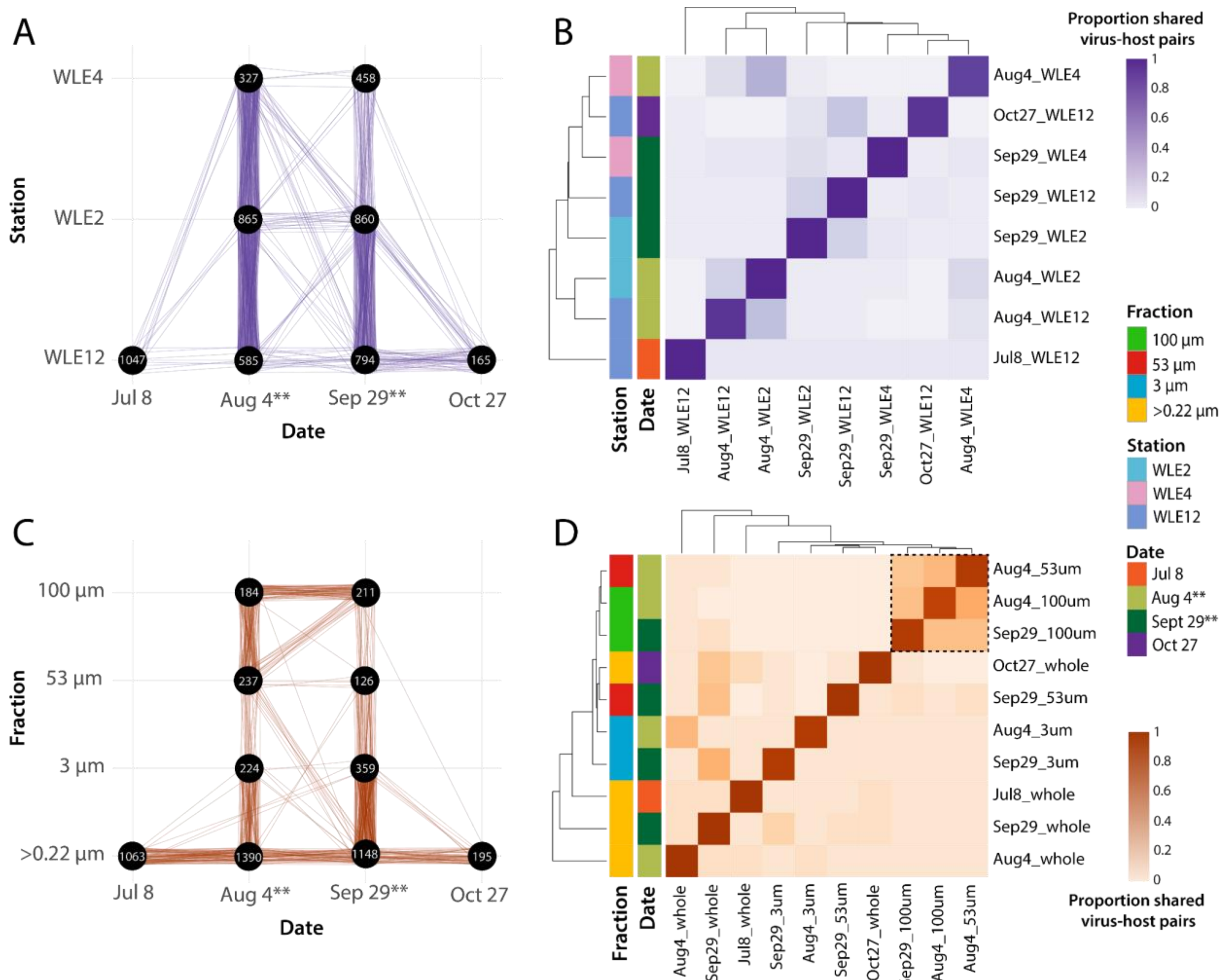


Figure 14. Turnover of Lake Erie cHAB virus-host pairs. (A) Network of predicted virus-host pairs shared between different sampling dates and stations. Black nodes represent predicted virus-host networks of individual samples (individual networks are not visualized here; composite networks for Aug 4 and Sept 29 are in SI Fig. 5). Lines (edges) connecting nodes represent a virus-host pair shared between two samples. Numbers on nodes are the total number of virus-host pairs predicted in a sample. (B) Heatmap depicting the proportion of virus-host pairs shared between different samples scaled by the total number of virus-host pairs in the row sample. Dendrogram depicts hierarchical clustering of samples by proportion of shared virus-host pairs. Color blocks on the right indicate sample date and station. (C) Network of predicted virus-host pairs shared between different sampling dates and fractions. All else as described for panel A. (D) Heatmap depicting the proportion of virus-host pairs shared between different samples scaled by the total number of virus-host pairs in the row sample. Color blocks on the right indicate sample date and fraction. Dashed box indicates a cluster of colony-associated samples from the Lake Erie 2014 bloom peaks.

3.2.3.5 Distinct virus-host interactions emerge in colony- and not colony-associated assemblages

In contrast, there were marked differences in how these virus-host interactions were distributed across fractions, even when collected on the same date (Fig. 14C-D). A greater percentage of pairs were shared *within* the colony-associated fractions (17% on average) and within the non-colony-associated fractions (8% on average), than between

colony- and not-colony-associated fractions (2.4% on average). Our previous work (Wing et al., 2024a) described colony-dependent partitioning among *Microcystis* vOTU assemblages and how this pattern may be explained by observed fraction-specific differences in relative abundances of *Microcystis* genotypes. The observed partitioning of variation between host-virus assemblages based on the presence of distinct ecological contexts (Rosenberg et al., 2021), such as free-living and host-associated communities, is a pattern that can be generalized beyond just *Microcystis* to various host-virus systems. This phenomenon underscores the importance of considering different ecological niches within a larger ecosystem when studying host-virus interactions. Overall, the network of networks representation (Fig. 14A,C) provides a novel approach to capture the high-dimensional data in a framework suitable to study the dynamism of viral-host interactions across spatial scales (microhabitats to lake regions) and time.

Across ecosystems, microbial hosts exist in various forms and associations, such as single-cell organisms and multicellular colonies or biofilms (Schmidt et al., 2020; Li et al., 2021; Rosenberg et al., 2021). These different ecological contexts offer unique dynamics that impact the interactions between hosts and their viral predators. The variation in host populations within these ecological niches may, in turn, influence the range of viruses that can infect and influence these hosts. This partitioning suggests that the dynamics of host-virus interactions can be context-dependent, where the same host species may exhibit different susceptibility or resistance to viruses depending on whether it exists as a free-living individual or forms more complex structures (Palermo et al., 2021). In the case of cHABs, the presence of colonies versus single cells may result in distinct virus-host interactions. We recommend future considerations of micro and macro spatial contexts when studying virus-host interactions, as these unique ecological constraints likely influence the dynamics and outcomes of virus-host interactions.

3.3 Outlook/Future Directions

The characterization of Lake Erie's vOTUs and their metabolic potential presented here is an important step towards understanding viral community dynamics in a freshwater

cHAB. Our work detailed thousands of undescribed vOTUs, revealing the novelty of Great Lakes viral communities. Our findings support the application of the Bank model to describe the seasonal succession of vOTUs, with a small portion of abundant viruses representing the “active” fraction while the majority of populations remain in a dormant “bank” state. This viral bank likely contributes to the high level of viral richness and evenness observed in the cHAB viral community. We identified abundant vOTUs in the cellular, especially colony-associated, fractions that appeared on multiple dates, suggesting their active role in shaping bloom community dynamics. Overall, sampling date and size fractionation primarily shaped the spatiotemporal dynamics of the viral community, highlighting the necessity to incorporate different sampling fractions through time to better understand the interplay between local and global processes that shape viral community dynamics. Further analysis of the ecological roles of virally encoded metabolic functions will provide a deeper understanding of host-virus interactions and their implications for the rise and demise of harmful algal blooms. Our findings suggest that the standard approach of studying viruses in only one filter fraction results in the underestimation of viral diversity and a skewed representation of their ecology, as viral community analysis is restricted to only those viruses that are physically captured. Together, our results highlight the importance of incorporating different sampling fractions into future studies aimed at describing the ecological importance of viruses in ecosystems.

3.4 Materials and Methods

3.4.1 Field Sampling and Collection

For full field sampling methods, refer to Wing et al., 2024a. Briefly, The field sampling was conducted in cooperation with the NOAA Great Lakes Environmental Research Laboratory sampling program for Lake Erie in 2014. Three sampling sites were selected, and samples were collected bi-monthly in June and weekly from July through October. Metagenomic data were generated from samples collected at three regularly sampled stations (WE2, WE4, and WE12) located at different positions in Lake Erie. Water samples were collected from 0.1m below the water surface using a peristaltic pump, yielding 20 liters of water. Samples were filtered using 100μm polycarbonate

filters to concentrate Microcystis colonies while excluding smaller particles. This >100µm Microcystis community constituted over 90% of all Microcystis cells in Lake Erie. Additional filtration steps were employed to collect single-celled microbes, the bulk bacterial community, and viruses associated with the bloom stages. Iron chloride stock solution was added to the <0.22µm fraction to create a flocculant containing iron chloride and viruses, which was allowed to sit overnight to maximize virus recovery. Finally, the flocculant was filtered through a 0.45µm filter and stored at 4°C.

3.4.2 DNA Extraction and Sequencing of Hosts and Viruses

Host DNA was extracted from samples using the DNeasy Mini Kit (QIAGEN) according to manufacturer’s instructions. All sequencing was performed at the University of Michigan Sequencing Core. Paired-end DNA sequencing (2 125) was conducted on Illumina HiSeq 2000 with V4 chemistry reagents with “low-input prep” using the Rubicon ThruPlex kit. For viral DNA, refer to our process described in protocols.io (https://www.protocols.io/view/iron-chloride-flocculation-resuspension-and-dna-ex-c4ygyxtw). Metagenome information can be found in Table S14.

3.4.3 Host Assembly and Binning

For full host assembly and binning instructions, refer to Wing et al., 2024a. Metagenomic data from 36 samples were processed to create high-quality Metagenome Assembled Genomes (MAGs). Contaminated sequences were removed using BBDuk, and denoised reads were assessed for quality with FastQC. Dereplication was performed with BBnorm, followed by contig assembly using Megahit. MAGs were generated using Concoct (Alneberg, 2013). Host bins were dereplicated using dRep default settings (Olm, 2017). Manual refinement of MAGs was carried out using Anvi'o (Eren, 2021). MAGs meeting specific criteria for completeness, contamination, and strain heterogeneity were retained as the final results. Multiple rounds of refinement were performed, and CheckM was used to estimate MAG quality (Table S15). Read mapping was conducted to determine host relative abundance, where bins with 70% coverage at 1x read depth were considered present in a given sample.

3.4.4 Viral vOTU Generation, mapping and community diversity (alpha and beta)

For full viral vOTU methods, mapping and diversity metrics, refer to Wing et al., 2024a. Briefly, a combination of viral identification tools were employed to predict potential viral contigs from metagenomic data obtained from 2014 Lake Erie samples. Only contigs exceeding 3 kilobases in length were retained and used to identify viral operational taxonomic units (vOTUs). These vOTUs were defined by clustering contigs with an average nucleotide identity (ANI) of 95% across 85% of their length. The longest sequence within each vOTU was designated as the representative sequence for subsequent analyses. Reads from each sample were mapped to these contigs, and quantification was performed using Bowtie2 and Samtools. Alpha diversity measures were calculated based on downsampled reads, and Bray-Curtis distances were used to assess the viral community structure among samples. PERMANOVA analysis was conducted to evaluate the impact of various factors, including sampling location, date, fraction, and environmental parameters, on the viral community structure. Taxonomy of vOTUs from Lake Erie was estimated using an approach previously described in Kieft et al. (2021). Briefly, Prodigal v2.6.3 (Hyatt et al., 2010) was used to generate open reading frames (ORFs) for viral contigs and DIAMOND v0.9.14.115 (Buchfink et al., 2015) was used to generate a custom NCBI GenBank reference database. DIAMOND BLASTp (Altschul et al., 1990) was then used to match proteins from viral contigs to the custom database, providing hits that were then filtered and assigned a taxonomic lineage according to DIAMOND BLASTp hit counts. Assignments were provided beginning at the Order level and subsequently assigned down to the Subfamily when available.

3.4.5 Viral Metabolic Potential Analyses

For full viral metabolic potential analyses methods, refer to Wing et al., 2024a. Viral contig gene metabolic annotations were assigned using the KEGG and Pfam databases through the DRAM (Distilled and Refined Annotation of Metabolism) tool (Shaffer et al., 2020), which was preceded by the generation of open reading frames (ORFs) with

Prodigal (Hyatt et al., 2010). Read coverages of these ORFs were calculated using FeatureCounts (Liao et al., 2014). Bray-Curtis distances were computed between samples, followed by NMDS ordination using the vegan package in R. The impact of water quality parameters on the abundance of protein families (Pfams) was assessed using a PERMANOVA model, with parameters showing significant correlations at a p-value of ≤ 0.05.

3.4.6 Virus-Host Predictions

Virus-Host Infection Predictor VHIP (v.1.0) was used to predict virus-host pairs from the 2014 Lake Erie cHAB metagenomic data. VHIP is a gradient-boosted machine learning model that relies on sequence-based signals of coevolution detected in viral and putative host genomes (Bastien et al., 2023). vOTUs were binned using vRhyme (Kieft et al., 2022). Viral bin representative sequences >10 kb were and virus-host pairs with VHIP prediction score >0.93 were used for interaction network analysis. Networks were visualized with Gephi 0.9.0 (Bastian et al., 2009).

For a full list of SI Tables, visit:
https://docs.google.com/spreadsheets/d/1Vn4FnikvIpvI40X-DpH4S3vJYZ5JJ4qcfFjYk1HjmCQ/edit?usp=sharing

For a full list of SI Figures, visit:
https://docs.google.com/document/d/1dZqx40UssIb_n5lqqZonhAZ8e4h6JQcU1uId7EG87tE/edit?usp=sharing

3.5 Acknowledgements

We would like to thank Derek Smith for providing the high-quality, manually refined non-Microcystis host bins used in this study. We would also like to thank all Duhaime Lab members for their efforts in research question generation and evaluation as well as discussions on best approaches to assess the viral community. We would also like to

thank Michelle Berry for early assistance with viral sample collection and processing and Lizy Michaelson for assistance verifying and publishing protocols in protocols.io. We also thank the field crew at CIGLR/GLERL for allowing us to sample with them and assisting in field measurements. This research was also funded in part by the University of Michigan Water Center, NOAA Michigan Sea Grant Core Program awards NA22OAR4170084, and NA18OAR4170102, the National Science Foundation grant 2055455, and the NOAA Omics program awarded via the Cooperative Institute of Great Lake Research (CIGLR) through the NOAA cooperative agreements with the University of Michigan (NA17OAR4320152 and NA22OAR4320150).

3.6 References

Aguirre de Cárcer, D., López-Bueno, A., Pearce, D. A., & Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Science Advances*, *1*(5), e1400127. https://doi.org/10.1126/sciadv.1400127

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., Romano, A. E., & DeLong, E. F. (2017). Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), 11446–11451. https://doi.org/10.1073/pnas.1714821114

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Bastien, E. G., Cable, R. N., Zaman, L., Batterbee, C., Wing, A. J., & Duhaime, M. B. (2023). *Virus-Host Interactions Predictor (VHIP): Machine learning approach to resolve microbial virus-host interaction networks* (p. 2023.11.03.565433). bioRxiv. https://doi.org/10.1101/2023.11.03.565433

Berry, M. A., Davis, T. W., Cory, R. M., Duhaime, M. B., Johengen, T. H., Kling, G. W., Marino, J. A., Den Uyl, P. A., Gossiaux, D., Dick, G. J., & Denef, V. J. (n.d.). *Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities*. https://doi.org/10.1111/1462-2920.13640

Berry, M. A., White, J. D., Davis, T. W., Jain, S., Johengen, T. H., Dick, G. J., Sarnelle, O., & Denef, V. J. (2017). Are Oligotypes Meaningful Ecological and Phylogenetic Units? A Case Study of Microcystis in Freshwater Lakes. *Frontiers in Microbiology*, *08*, 365–365. https://doi.org/10.3389/fmicb.2017.00365

Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., Brister, J. R., Kropinski, A. M., Krupovic, M., Lavigne, R., Turner, D., & Sullivan, M. B. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology*, *37*(6). https://doi.org/10.1038/s41587-019-0100-8

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B. (2017). vConTACT: An iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, e3243. https://doi.org/10.7717/peerj.3243

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. https://doi.org/10.1016/j.tim.2005.04.003

Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W., & Sullivan, M. B. (2016). Seasonal time bombs: Dominant temperate viruses affect Southern Ocean microbial dynamics. *The ISME Journal*, *10*(2), 437–449. https://doi.org/10.1038/ismej.2015.125

Buchfink et al. (2015). *Fast and sensitive protein alignment using DIAMOND | Nature Methods*. https://www.nature.com/articles/nmeth.3176

Chow, C.-E. T., Sachdeva, R., Cram, J. A., Steele, J. A., Needham, D. M., Patel, A., Parada, A. E., & Fuhrman, J. A. (2013). Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *The ISME Journal*, *7*(12), Article 12. https://doi.org/10.1038/ismej.2013.122

Cook, K. V., Li, C., Cai, H., Krumholz, L. R., Hambright, K. D., Paerl, H. W., Steffen, M. M., Wilson, A. E., Burford, M. A., Grossart, H.-P., Hamilton, D. P., Jiang, H., Sukenik, A., Latour, D., Meyer, E. I., Padisák, J., Qin, B., Zamor, R. M., & Zhu, G. (2020). The global Microcystis interactome. *Limnology and Oceanography*, *65*(S1), S194–S207. https://doi.org/10.1002/lno.11361

Cory, R. M., Davis, T. W., Dick, G. J., Johengen, T., Denef, V. J., Berry, M. A., Page, S. E., Watson, S. B., Yuhas, K., & Kling, G. W. (2016). Seasonal Dynamics in Dissolved Organic Matter, Hydrogen Peroxide, and Cyanobacterial Blooms in Lake Erie. *Frontiers in Marine Science*, *3*. https://www.frontiersin.org/articles/10.3389/fmars.2016.00054

Dart, E., Fuhrman, J. A., & Ahlgren, N. A. (2023). Diverse Marine T4-like Cyanophage Communities Are Primarily Comprised of Low-Abundance Species Including Species with Distinct Seasonal, Persistent, Occasional, or Sporadic Dynamics. *Viruses*, *15*(2), Article 2. https://doi.org/10.3390/v15020581

Deboutte, W., Beller, L., Yinda, C. K., Maes, P., de Graaf, D. C., & Matthijnssens, J. (2020). Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(19), 10511–10519. https://doi.org/10.1073/pnas.1921859117

Eiler, A., & Bertilsson, S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environmental Microbiology*, *6*(12), 1228–1243. https://doi.org/10.1111/j.1462-2920.2004.00657.x

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., Fink, I., Pan, J. N., Yousef, M., Fogarty, E. C., Trigodet, F., Watson, A. R., Esen, Ö. C., Moore, R. M., Clayssen, Q., Lee, M. D., Kivenson, V., Graham, E. D., Merrill, B. D., … Willis, A. D. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology*, *6*(1), Article 1. https://doi.org/10.1038/s41564-020-00834-3

Gregory, A. C., Solonenko, S. A., Ignacio-Espinoza, J. C., LaButti, K., Copeland, A., Sudek, S., Maitland, A., Chittick, L., dos Santos, F., Weitz, J. S., Worden, A. Z., Woyke, T., & Sullivan, M. B. (2016). Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics*, *17*(1), 930. https://doi.org/10.1186/s12864-016-3286-x

Gu, X., Tay, Q. X. M., Te, S. H., Saeidi, N., Goh, S. G., Kushmaro, A., Thompson, J. R., & Gin, K. Y.-H. (2018). Geospatial distribution of viromes in tropical freshwater ecosystems. *Water Research*, *137*, 220–232. https://doi.org/10.1016/j.watres.2018.03.017

Harke, M. J., & Gobler, C. J. (2013). Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and

Growth on Organic Matter. *PLOS ONE*, *8*(7), e69834. https://doi.org/10.1371/journal.pone.0069834

Harke, M. J., Steffen, M. M., Gobler, C. J., Otten, T. G., Wilhelm, S. W., Wood, S. A., & Paerl, H. W. (2016). A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae*, *54*, 4–20. https://doi.org/10.1016/j.hal.2015.12.007

Hegarty, B., Dai, Z., Raskin, L., Pinto, A., Wigginton, K., & Duhaime, M. (2022). A snapshot of the global drinking water virome: Diversity and metabolic potential vary with residual disinfectant use. *Water Research*, *218*, 118484. https://doi.org/10.1016/j.watres.2022.118484

Hevroni, G., Flores-Uribe, J., Béjà, O., & Philosof, A. (2020). Rising through the Ranks: Seasonal and Diel Patterns of Marine Viruses. *bioRxiv*, 2020.05.07.082883. https://doi.org/10.1101/2020.05.07.082883

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), 119. https://doi.org/10.1186/1471-2105-11-119

Jiang, X., Ha, C., Lee, S., Kwon, J., Cho, H., Gorham, T., & Lee, J. (2019). Characterization of Cyanophages in Lake Erie: Interaction Mechanisms and Structural Damage of Toxic Cyanobacteria. *Toxins*, *11*(8), 444. https://doi.org/10.3390/toxins11080444

Kieft et al. (2022). *vRhyme enables binning of viral genomes from metagenomes | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/50/14/e83/6584432

Kieft, K., Zhou, Z., Anderson, R. E., Buchan, A., Campbell, B. J., Hallam, S. J., Hess, M., Sullivan, M. B., Walsh, D. A., Roux, S., & Anantharaman, K. (2021). Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-23698-5

Kim, J.-G., Park, S.-J., Sinninghe Damsté, J. S., Schouten, S., Rijpstra, W. I. C., Jung, M.-Y., Kim, S.-J., Gwak, J.-H., Hong, H., Si, O.-J., Lee, S., Madsen, E. L., & Rhee, S.-K. (2016). Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7888–7893. https://doi.org/10.1073/pnas.1605501113

Lennon, J. T., den Hollander, F., Wilke-Berenguer, M., & Blath, J. (2021). Principles of seed banks and the emergence of complexity from dormancy. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-24733-1

Li, J., Gu, L., Bai, S., Wang, J., Su, L., Wei, B., Zhang, L., & Fang, J. (2021). Characterization of particle-associated and free-living bacterial and archaeal communities along the water columns of the South China Sea. *Biogeosciences*, *18*(1), 113–133. https://doi.org/10.5194/bg-18-113-2021

Liang, X., Wang, Y., Zhang, Y., Zhuang, J., & Radosevich, M. (2021). Viral abundance, community structure and correlation with bacterial community in soils of different cover plants. *Applied Soil Ecology*, *168*, 104138. https://doi.org/10.1016/j.apsoil.2021.104138

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, W., Li, D., Sun, Z., Tong, Y., Yan, X., Wang, C., Zhang, X., & Pei, G. (2020). A novel freshwater cyanophage vB_MelS-Me-ZS1 infecting bloom-forming cyanobacterium Microcystis elabens. *Molecular Biology Reports*, *47*(10), 7979–7989. https://doi.org/10.1007/s11033-020-05876-8

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., Ackermann, M., Hahn, A. S., Srivastava, D. S., Crowe, S. A., Doebeli, M., & Parfrey, L. W. (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, *2*(6), Article 6. https://doi.org/10.1038/s41559-018-0519-1

Malki, K., Sawaya, N. A., Tisza, M. J., Coutinho, F. H., Rosario, K., Székely, A. J., & Breitbart, M. (2021). Spatial and Temporal Dynamics of Prokaryotic and Viral Community Assemblages in a Lotic System (Manatee Springs, Florida). *Applied and Environmental Microbiology*, *87*(18), e0064621. https://doi.org/10.1128/AEM.00646-21

McKindles, K. M., Manes, M. A., DeMarco, J. R., McClure, A., McKay, R. M., Davis, T. W., & Bullerjahn, G. S. (2020). Dissolved Microcystin Release Coincident with Lysis of a Bloom Dominated by Microcystis spp. In Western Lake Erie Attributed to a Novel Cyanophage. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01397-20

Meyer, K. A., Davis, T. W., Watson, S. B., Denef, V. J., Berry, M. A., & Dick, G. J. (2017). Genome sequences of lower Great Lakes Microcystis sp. Reveal strain-specific genes that are present and expressed in western Lake Erie blooms. *PLOS ONE*, *12*(10), e0183859. https://doi.org/10.1371/journal.pone.0183859

Michalek et al. (2013). *Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions | PNAS*. https://www-pnas-org.proxy.lib.umich.edu/doi/10.1073/pnas.1216006110

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), Article 12. https://doi.org/10.1038/ismej.2017.126

Ouellette, A. J. A., Handy, S. M., & Wilhelm, S. W. (2006). Toxic Microcystis is widespread in Lake Erie: PCR detection of toxin genes and molecular characterization of associated cyanobacterial communities. *Microbial Ecology*, *51*(2), 154–165. https://doi.org/10.1007/s00248-004-0146-z

Paerl, H. W., & Huisman, J. (2009). Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports*, *1*(1), 27–37. https://doi.org/10.1111/j.1758-2229.2008.00004.x

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology*, *65*(4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y

Palermo, C. N., Shea, D. W., & Short, S. M. (2021). Analysis of Different Size Fractions Provides a More Complete Perspective of Viral Diversity in a Freshwater Embayment. *Applied and Environmental Microbiology*, *87*(11), e00197-21. https://doi.org/10.1128/AEM.00197-21

Roth Rosenberg, D., Haber, M., Goldford, J., Lalzar, M., Aharonovich, D., Al-Ashhab, A., Lehahn, Y., Segrè, D., Steindler, L., & Sher, D. (2021). Particle-associated and free-living bacterial communities in an oligotrophic sea are affected by different environmental factors. *Environmental Microbiology*, *23*(8), 4295–4308. https://doi.org/10.1111/1462-2920.15611

Roux, S., & Brum, J. R. (2019). A viral reckoning: Viruses emerge as essential manipulators of global ecosystems. *Environmental Microbiology Reports*, *11*(1), 3–8. https://doi.org/10.1111/1758-2229.12700

Salazar, G., Cornejo-Castillo, F. M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D., Arrieta, J. M., Duarte, C. M., Gasol, J. M., & Acinas, S. G. (2015). Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Molecular Ecology*, *24*(22), 5692–5706. https://doi.org/10.1111/mec.13419

Santos-Medellin, C., Zinke, L. A., ter Horst, A. M., Gelardi, D. L., Parikh, S. J., & Emerson, J. B. (2021). Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *The ISME Journal*, *15*(7), Article 7. https://doi.org/10.1038/s41396-021-00897-y

Schmidt, M. L., White, J. D., & Denef, V. J. (2016). Phylogenetic conservation of freshwater lake habitat preference varies between abundant bacterioplankton phyla. *Environmental Microbiology*, *18*(4), 1212–1226. https://doi.org/10.1111/1462-2920.13143

Schroeder, D. C., Oke, J., Hall, M., Malin, G., & Wilson, W. H. (2003). Virus Succession Observed during an Emiliania huxleyi Bloom. *Applied and Environmental Microbiology*, *69*(5), 2484–2490. https://doi.org/10.1128/AEM.69.5.2484-2490.2003

Shaffer et al. (2020). *DRAM for distilling microbial metabolism to automate the curation of microbiome function | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/48/16/8883/5884738

Smith, D. J., Tan, J. Y., Powers, M. A., Lin, X. N., Davis, T. W., & Dick, G. J. (2021). Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. *Environmental Microbiology*, *23*(6), 3020–3036. https://doi.org/10.1111/1462-2920.15514

Sorensen, G., Baker, A. C., Hall, M. J., Munn, C. B., & Schroeder, D. C. (2009). Novel virus dynamics in an Emiliania huxleyi bloom. *Journal of Plankton Research*, *31*(7), 787–791. https://doi.org/10.1093/plankt/fbp027

Steffen, M. M., Dearth, S. P., Dill, B. D., Li, Z., Larsen, K. M., Campagna, S. R., & Wilhelm, S. W. (2014). Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis. *The ISME Journal*, *8*(10), 2080–2092. https://doi.org/10.1038/ismej.2014.78

Thingstad, T., & Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, *13*, 19–27. https://doi.org/10.3354/ame013019

Trainic, M., Koren, I., Sharoni, S., Frada, M., Segev, L., Rudich, Y., & Vardi, A. (2018). Infection Dynamics of a Bloom-Forming Alga and Its Virus Determine Airborne Coccolith Emission from Seawater. *iScience*, *6*, 327–335. https://doi.org/10.1016/j.isci.2018.07.017

Visser, P. M., Verspagen, J. M. H., Sandrini, G., Stal, L. J., Matthijs, H. C. P., Davis, T. W., Paerl, H. W., & Huisman, J. (2016). How rising CO2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae*, *54*, 145–159. https://doi.org/10.1016/j.hal.2015.12.006

Williamson, S. J., Allen, L. Z., Lorenzi, H. A., Fadrosh, D. W., Brami, D., Thiagarajan, M., McCrow, J. P., Tovchigrechko, A., Yooseph, S., & Venter, J. C. (2012). Metagenomic Exploration of Viruses throughout the Indian Ocean. *PLOS ONE*, *7*(10), e42047. https://doi.org/10.1371/journal.pone.0042047

Wing, A.J, Hegarty, B., Bastien, G.E., Evans, J., Cable, R., Smith, D., NOAA, Denef, V., Dick, G., & Duhaime, M.B. “Tracking Microcystis Viruses and Infection Dynamics in a Multi-Peak Microcystis-Dominated Bloom.” [Submitted to PLoS Biology]

Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., Roth, J. R., & Bäumler, A. J. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, *467*(7314), Article 7314. https://doi.org/10.1038/nature09415

Yancey, C. E., Smith, D. J., Den Uyl, P. A., Mohamed, O. G., Yu, F., Ruberg, S. A., Chaffin, J. D., Goodwin, K. D., Tripathi, A., Sherman, D. H., & Dick, G. J. (2022). Metagenomic and Metatranscriptomic Insights into Population Diversity of Microcystis Blooms: Spatial and Temporal Dynamics of mcy Genotypes, Including a Partial Operon That Can Be Abundant and Expressed. *Applied and Environmental Microbiology*, *88*(9), e0246421. https://doi.org/10.1128/aem.02464-21

Ye, W., Tan, J., Liu, X., Lin, S., Pan, J., Li, D., & Yang, H. (2011). Temporal variability of cyanobacterial populations in the water and sediment samples of Lake Taihu as determined by DGGE and real-time PCR. *Harmful Algae*, *10*(5), 472–479. https://doi.org/10.1016/j.hal.2011.03.002

Chapter 4: Evaluating the Relationship Between Viral Infection and Phylogeny of Lake Erie *Microcystis* Isolates

A.J Wing[1], Anders Kiledal[1], Eric Bastien[1], Lizy Michaelson[1], Vincent Denef[1], Bridget Hegarty[1,2], Gregory Dick[1,3], Melissa Duhaime[1]

[1] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI

[2] Department of Civil and Environmental Engineering, Case Western Reserve University, Cleveland, OH

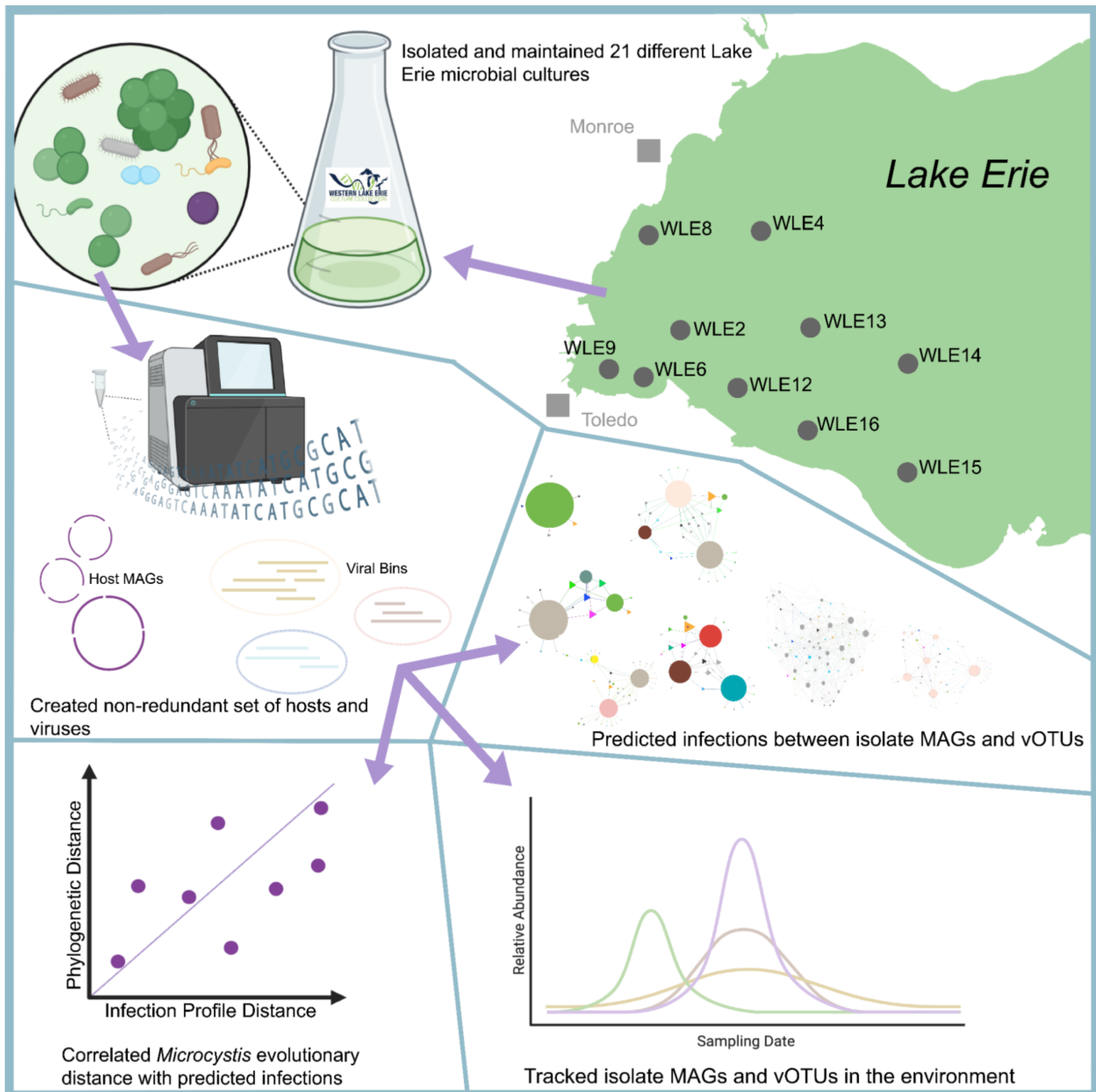
[3] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI

Abstract

As the importance of viruses continues to be underscored in the context of cyanobacterial harmful algal blooms (cHABs), it is imperative to gain a better perspective on the role of phage-host interactions in these complex communities. Despite ongoing research efforts, our understanding of phage-host dynamics in cHABs remains limited. While cHABs are increasing in intensity and frequency across the globe, our study aims to address the lack of knowledge in cHAB phage-host dynamics across eight separate blooms in the western basin of Lake Erie. We identified a significantly positive correlation between the phylogenetic distances of Western Lake Erie Culture Collection (WLECC) *Microcystis* strains and their infection profiles. Underscoring the role of shared genetic and physiological traits among closely related *Microcystis* strains in influencing their susceptibility or resistance to specific phages. Intra-colony infection dynamics within WLECC cultures are examined, providing evidence for intra-colony infections between *Microcystis* and other consortia members.

Viral OTU (vOTU) 17663 and *Microcystis* strain LE19-12.2 emerge as a meaningful phage-host pair as their predator-prey relationship is tracked across multiple blooms, linking phage, host and bloom dynamics in Lake Erie cHABs. This work advances our understanding of *Microcystis*-phage dynamics during cHABs and provides a powerful stepping stone to further explore the importance of phage-host dynamics in the rise and demise of cHABs.

Graphical Abstract



4.1 Introduction

Microcystis aeruginosa, a globally distributed cyanobacterium, forms toxic blooms in freshwater systems, estuaries, and coastal systems worldwide (Yoshida et al., 2008; Preece et al., 2017). As *M. aeruginosa* can produce microcystins, which are potent hepatotoxins, the ability to predict when and where toxic blooms occur is critical to ensuring public health in at-risk habitats. While the importance of nutrient availability for *M. aeruginosa* bloom proliferation has been established (Harke et al., 2016; Dolan and Chapra, 2012; Pearl and Huisman 2009), these bottom-up influences are not able to explain finer resolution changes within a bloom, such as shifts from toxin- to non-toxin producing blooms (Bozarth et al., 2010; Davis et al., 2010). These shifts may be better explained by top-down controls that function at strain-level resolution (Yoshida et al., 2006, Yoshida et al. 2020), such as viral predation. Cyanophage, phage that specifically infect cyanobacteria, are capable of regulating bloom diversity and community structure via lysis-induced mortality and the reprogramming of host metabolisms during infection (Gao et al., 2016; Howard-Varona et al., 2020, Thompson et al., 2011, Wang et al., 2022). While the role of viruses as top-down controls of marine phytoplankton blooms is well-established (Schroeder et al., 2003; Sorensen et al., 2009; Trainic et al., 2018), the impact of viral infection on freshwater algal blooms and bloom toxicity is less well understood.

In 2014, proliferation of a toxic *Microcystis* bloom in Lake Erie led to the drinking water shut-off of Toledo, Ohio. More than 400,000 residents were left without access to potable water for over three days. Broad-scale viral infection of *Microcystis* has been suggested to have contributed to the high concentration of dissolved microcystin during this period, presumably by releasing the toxin following cell lysis (Steffen et al., 2015; McKindles et al., 2020). Previous studies have revealed that *Microcystis* is targeted by a wide array of cyanophage (Kuno et al., 2012; Yang et al., 2015), suggesting *Microcystis*-infecting phage are likely to play a role in determining cyanobacterial bloom dynamics. Though paramount to understanding bloom progression, the interactions between viral infection and *Microcystis* strain-type remain poorly resolved.

Due to the clear public health implications associated with *Microcystis*-dominated cyanobacterial harmful algal blooms (cHABs), numerous studies on the diversity of

Microcystis have focused on the presence of toxigenic and non-toxigenic strains, which often cohabitate in these blooms. Often, these blooms display a temporal strain succession, as toxigenic strains dominate early bloom peaks while non-toxigenic strains persist in the later stages (Davis et al., 2010; Bozart et al., 2010; Singh et al., 2015; Gobler et al.; 2016), though this succession cannot always explain bloom toxicity dynamics (Kardinaal et al., 2007; Rinta-Kanto et al., 2009). While this strain succession remains ambiguous, it is evident that ecologically distinct strains of *Microcystis* exist and delineate from one another based on adaptations to temperature conditions (Xiao et al., 2017), light availability (Fontana et al., 2019), oxidative stressors (Dziallis and Grossart, 2011; Zilleges et al., 2011; Schuurmans et al., 2018) and nitrogen requirements (Alexova et al., 2011; Alexova et al., 2016; Monchamp et al., 2014). Yet, previous work has failed to adequately explain these delineations in terms of top-down controls such as viral predation. While prior research has proposed viral resistance in one of two distinct *Microcystis* populations identified in separate cHABs peaks based on presence/absence of viral marker genes (Yoshida et al., 2007), marker gene analyses and qPCR targets alone are not enough to establish virus-host interactions. While virus-driven intraspecific diversification could impact bloom toxicity, the addition of 'omics-based approaches like community level genomic sequencing (Morimoto et al., 2019) are required to better elucidate the susceptibility of *Microcystis* strains to viral infection as well as the host range of cultured and uncultured *Microcystis* viruses.

In this study, we described the relationship between viral predation and *Microcystis* population diversification as a means to explore the potential impacts of viral infection on toxin production. We sought to answer the following research questions: 1) Are Microcystis isolates' phylogenetic distances correlated with infection profiles? If so, what host traits explain the variation in infection profiles between isolates? 2) Do we see intra-colony infection dynamics? Using high quality toxic and non-toxic xenic *Microcystis* isolates from the Western Lake Erie Culture Collection (WLECC) (Yancey et al., 2023), we establish a relationship between *Microcystis* strains and wild viral populations from the same season and location. Furthermore, we assess whether wild viral populations predicted to infect *Microcystis* are also predicted to infect other consortia members associated with these isolates. To date, these linkages have only been made between

Microcystis viruses and reference genomes isolated from different temporal and geographic contexts. By leveraging *Microcystis* isolates, their well-described genome content, their characterized toxigenic potential, and their inferred infection profiles, we link host, toxin production, and viral predation for the first time.

4.2 Results/Discussion

4.2.1 Phylogenetic distance and predicted viral infection profiles of *Microcystis* isolates from Lake Erie are correlated

Understanding the factors that shape the genetic diversity of *Microcystis* and interactions with members of the phycosphere community are necessary to better understand microbial community dynamics in cHABs. In this context, we examined the relationship between *Microcystis* strain phylogenetic distances and infection profile distances, with the hypothesis that: *If phylogenetic distances are correlated with infection profiles, then Microcystis isolates with greater phylogenetic similarity will share similar infection profiles.* This similarity in infection profiles may be explained by shared host traits as isolates belonging to the same phylogenetic cluster or lineage are expected to display more similar infection profiles due to their shared genetic and physiological characteristics, ultimately influencing their susceptibility or resistance to specific viruses.

To test our hypothesis, we calculated phylogenetic distances among 21 *Microcystis* isolates (Table 1) and also calculated their infection profile distances resulting from infection predictions generated with VHIP, a novel infection prediction model with 87% accuracy in predicting phage-host pairs (Bastien et al., 2023), using 23,347 virus OTUs (vOTUs) identified in Lake Erie samples and the 21 *Microcystis* cultures. We then performed a Spearman rank correlation analysis to assess the relationship between phylogenetic distances and infection profile distances. Our analysis revealed a statistically significant correlation (r-squared = 0.1955, p-value = 0.0223) between *Microcystis* isolate phylogenetic distances and infection profile distances (Fig. 15).

Table 1. Summary of Lake Erie *Microcystis* isolate culture collection.

***Microcystis* Isolate**	**Culture Key**	**Mcy genotype**	**Associated Bacteria Catalase**	**Collection Date**	**Station**	**Latitude**	**Longitude**
LE19-41.2	ND-100	complete_C1	yes	7/15/2019	WE2	41.7621	-83.33
LE19-10.1C	ND-101	partial_C1	yes	7/8/2019	WE8	41.8357	-83.359
LE19-12.2C	ND-102	absent	yes	7/8/2019	WE8	41.8357	-83.359
LE17-10A	ND-78	absent	yes	9/18/2017	WE8	41.8357	-83.359
LE18-22.4A	ND-79	complete_C1	yes	6/25/2018	WE12	41.7035	-83.2537
LE19-114.1A	ND-80	absent	yes	7/22/2019	WE8	41.8354	-83.3584
LE18-13.4	ND-81	absent	no	6/12/2018	WE2	41.7621	-83.33
LE19-196.1	ND-82	absent	yes	8/5/2019	WE6	41.7057	-83.3831
LE19-338.1	ND-84	absent	yes	8/20/2019	WE6	41.7109	-83.3668
LE19-55.1	ND-85	absent	yes	7/15/2019	WE2	41.7621	-83.33
LE19-8.1	ND-86	absent	yes	7/8/2019	WE8	41.8357	-83.359
LE19-84.1	ND-87	complete_B1	no	7/29/2019	WE4	41.8261	-83.1946
LE19-131.1A	ND-89	absent	yes	7/29/2019	WE8	41.8328	-83.3625
LE19-195.1B	ND-90	complete_B1	yes	8/5/2019	WE6	41.7057	-83.3831
LE19-197.1	ND-91	absent	no	8/5/2019	WE8	41.8347	-83.3587
LE19-4.1	ND-93	absent	yes	7/8/2019	WE8	41.8357	-83.359
LE19-59.1	ND-94	complete_C1	yes	7/15/2019	WE12	41.7035	-83.2537
LE19-98.1	ND-95	absent	yes	7/29/2019	WE4	41.8261	-83.1946
LE17-20C	ND-97	absent	yes	9/18/2017	WE8	41.8357	-83.359
LE19-251.1	ND-98	partial_C1	yes	8/12/2019	WE8	41.8325	-83.3598
LE19-388.1	ND-99	absent	yes	7/29/2019	WE6	41.7049	-83.3869

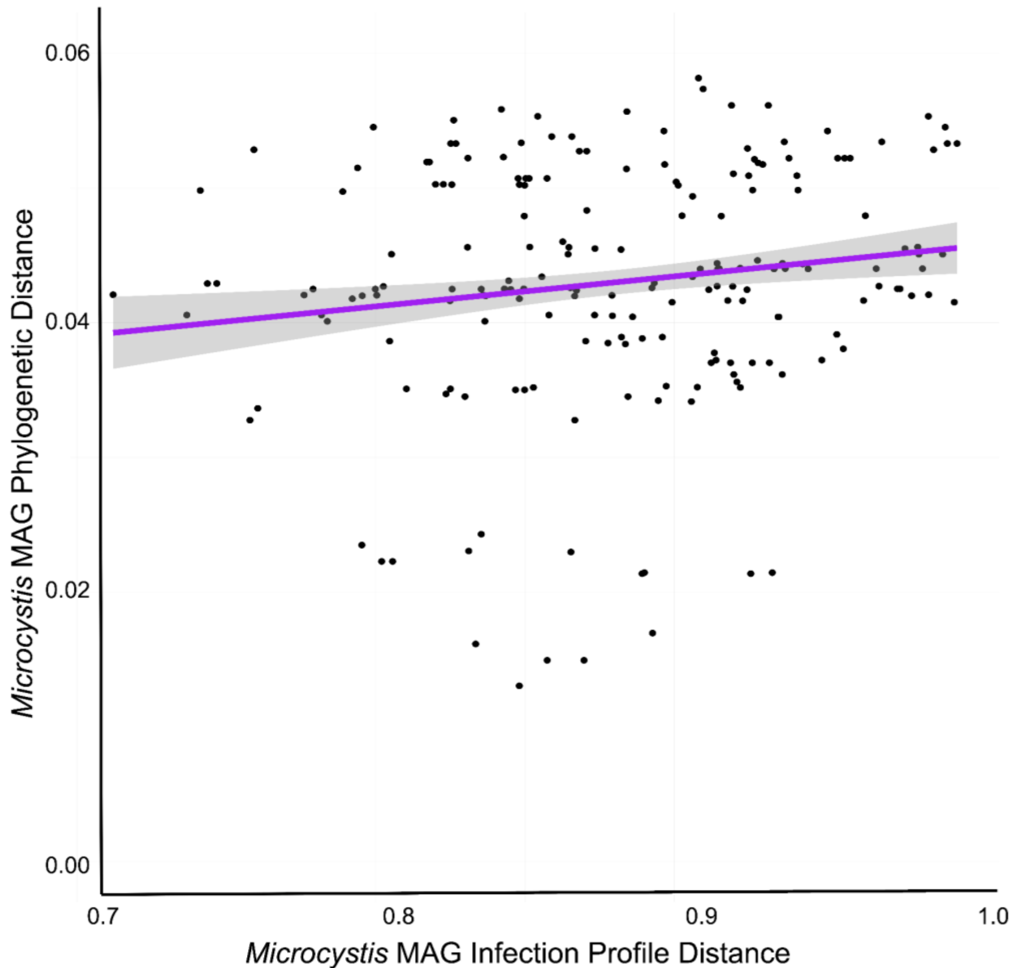


Figure 15. Spearman Rank Correlation Coefficient between *Microcystis* isolate phylogenetic distances and infection profile distances. Each point represents a comparison between two *Microcystis* isolates' phylogenetic distance and infection profile distance. The purple line represents a linear regression and the gray shading represents 95% confidence intervals.

This finding suggests a moderately strong positive correlation between these variables. While phylogenetic distances and infection profile distances are not directly proportional, they are linked, and the significance of the correlation suggests that underlying biological or ecological factors may drive this association. Our findings build on previous work that temperate phages are more likely to infect bacteria from the same clade as their original host, relative to bacteria from distantly related clades (Wendling et al., 2018). Our study delves even deeper into the intricacies of this phenomenon. We

discovered correlations between phylogenetic distance and infecting viruses within sub-species of bacteria, highlighting a finer level of specificity in the relationships between phages and their bacterial hosts. This nuanced understanding suggests that phage-host interactions are not solely dictated by broad taxonomic affiliations but can be influenced by more subtle genetic and ecological factors within bacterial populations. We next sought to identify and determine which host traits and environmental parameters might explain the variation observed between *Microcystis* isolate phylogenetic distances and infection profiles.

4.2.1.1 Collection date is significant predictor of predicted viral infection profiles for *Microcystis* isolates

We conducted a comprehensive ANOVA analysis, focusing on several environmental and spatiotemporal variables. Our results, as summarized in SI Table 2, reveal insights into the significance of these variables in shaping the *Microcystis* community. Among the variables examined, the collection date (month_year) emerged as the sole statistically significant predictor of *Microcystis* infection profiles. This finding highlights the crucial role of temporal dynamics in *Microcystis*-phage interactions within the western basin of Lake Erie. The significant F-statistic ($F = 1.9357$, $p = 0.0438$*) associated with the collection date variable underscores its importance in explaining this variation and the positive R-squared value ($R^2 = 0.28271$) suggests that approximately 28.27% of the variability in the infection profiles between *Microcystis* strains can be attributed to changes over time.

Contrastingly, the remaining environmental variables, including sampling site (Station), mcy cassette genotype (mcy_genotype), host-associated catalase activity, temperature (temp_c), dissolved oxygen concentration (do_mg_L), specific conductivity (SpCond_uS_cm), photosynthetically active radiation (PAR_uE_cm2_s), total phosphorus (TP_ug_L), and total dissolved phosphorus (TDP_ug_L), did not exhibit statistically significant relationships with the *Microcystis* strain infection profiles within our dataset. While our PCOA ordination based on Jaccard dissimilarity in *Microcystis* isolate infection profiles captured 48% of the variation observed in infection profiles

between *Microcystis* isolates (Fig. 16), it is essential to acknowledge that there are likely other unexplored factors contributing to the unexplained variation.

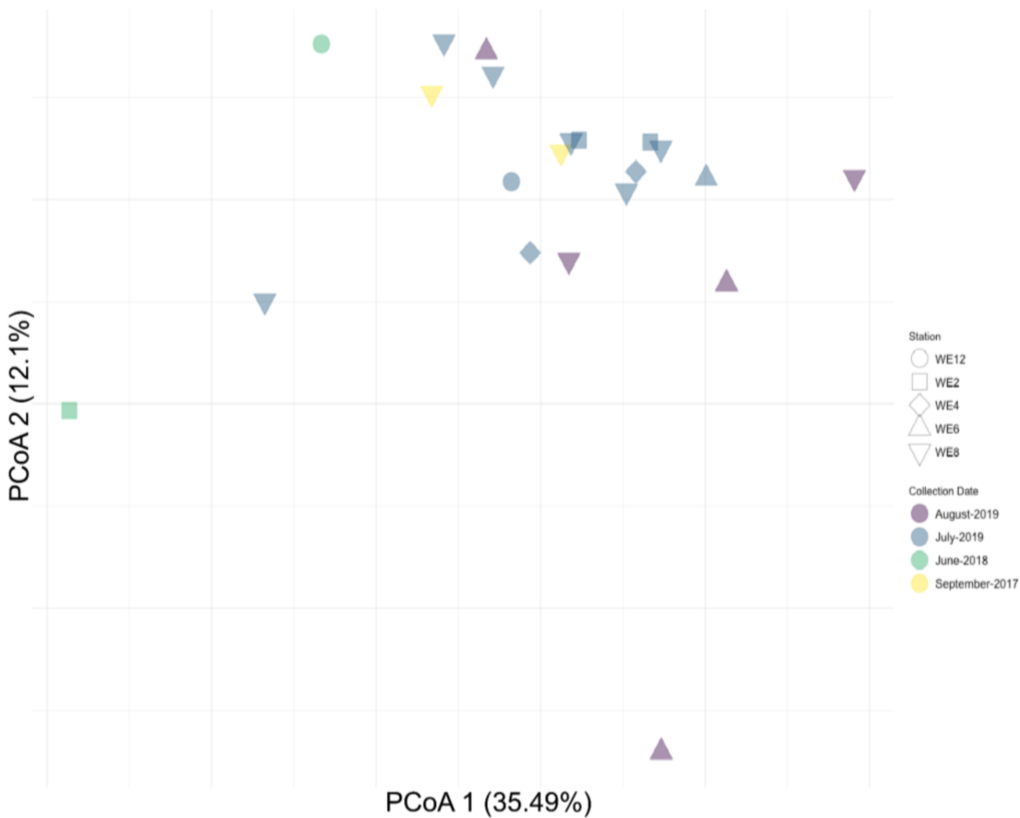


Figure 16. PCOA ordination based on Jaccard dissimilarities in distributions of *Microcystis* isolate infection profiles. Point color represents the collection date; shape represents station sampled.

These findings underscore the importance of specific factors, including the temporal relationship between *Microcystis* isolate infection profiles, in shaping the coevolutionary arms race between *Microcystis* and its predicted phages, while highlighting the limited influence of tested environmental parameters in the context of our study. As we consider the dynamic nature of microbial communities in Lake Erie, it is crucial to recognize that multiple factors could be at play, influencing *Microcystis* strain diversity over time. Thus, further investigations are warranted to explore the nuanced interplay between these variables, as well as the potential impacts of unexamined factors on *Microcystis* ecology and virus-host interactions. This holistic approach will contribute to a more comprehensive understanding of the temporal

dynamics of *Microcystis* strain diversity in Lake Erie and its broader ecological implications.

4.2.2 WLECC offers foundations for phage-host dynamics in the wild

4.2.2.1 Analysis of Western Lake Erie Culture Collection reveals 41 isolate vOTUs

After exploring the relationship between *Microcystis* isolate phylogenetic distances and infection profiles, we investigated the Western Lake Erie Culture Collection for vOTUs. Using the high MCC ruleset recommended in Hegarty et al., 2023, we identified 57 viral contigs ranging from 3,000 to 350,000 base pairs in length. Clustering these viral contigs with one another and wild viral contigs identified in Lake Erie samples between 2014-2021 resulted in 41 isolate vOTUs. Of the 57 viral contigs identified in culture, only 6 (82--ND_82_k141_3083, 93--ND_93_k141_2167, 94--ND_94_k141_2606, 97--ND_97_k141_135, 86--ND_86_k141_3379 and 92--ND_92_k141_17537) clustered with wild vOTUs. Of the 41 isolate vOTUs, 3 sequences from wild vOTUs (vRhyme_175__447--samp_447_157055, vRhyme_259__449--samp_449_49203, and vRhyme_37__448--samp_448_44905) from their respective clusters were used as isolate vOTU representatives given these were the longest sequence within the cluster. To contextualize these isolate vOTUs within the confines of Lake Erie, we performed competitive mapping of metagenomic reads from Lake Erie samples to these isolate vOTUs and 24,218 vOTUs identified in the wild to gain insight into the relative abundances of these isolates within the viral community (Fig. 17).

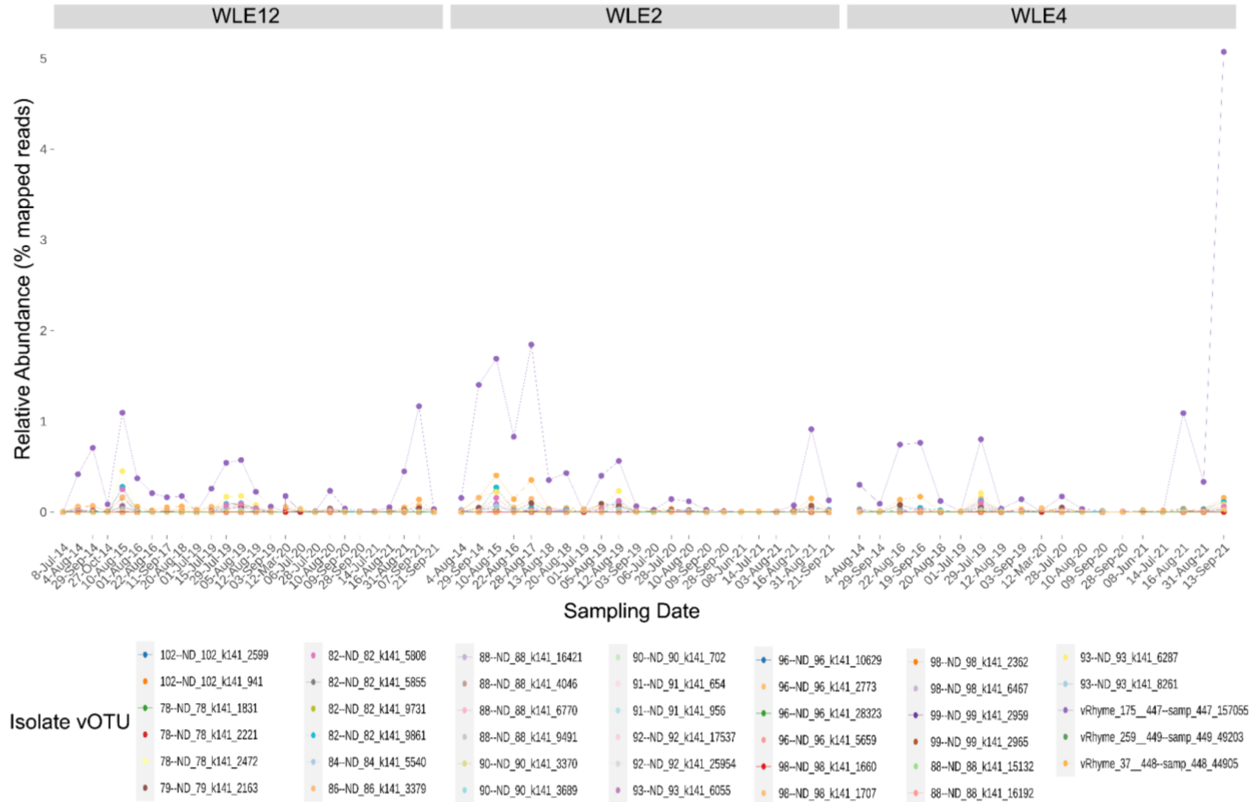


Figure 17. Relative abundance of isolate vOTUs in Lake Erie samples between 2014-2021. Only relative abundance of vOTUs at stations WLE12, WLE2 and WLE4 for the >0.22 µm fraction are shown.

4.2.2.2 Within the Lake Erie viral community, isolate vOTU 17663 registers the highest relative abundance through eight cHABs

Among the relative abundances generated for vOTUs in the >0.22 µm fraction, only one isolate vOTU was found to eclipse 1% relative abundance within the viral community at the primary sampling stations WLE12 (nearshore site by Toledo water intake crib), WLE2 (nearshore site by outlet of Maumee River) and WLE4 (offshore site closest to the Detroit River outlet). This vOTU, vRhyme_175__447--samp_447_157055, is 13,280bp in length and is referred to hereon as vOTU 17663. Derived from its cluster number within the Lake Erie viral community, vOTU 17663 was detected at all sampling stations except WLE6. When investigating vOTU 17663 in other sampling fractions collected during the 2014 Lake Erie cHAB, we found this vOTU peaked at 22% relative

abundance in the 100 μm fraction collected September 29th at WLE2 (data not shown). vOTU 17663 reached 18% and 9.6% relative abundance in the 100 μm fraction at WLE12 on September 29th and WLE2 on August 4th, respectively.

This was not unexpected as previous work (Wing et al., 2024a, Wing et al., 2024b) has shown the most abundant members of the viral community often reside in the larger colony associated fractions. That said, vOTU 17663 registered a relative abundance of <0.5% in all 3 and 53 μm fractions, likely indicating the virus-host dynamics of this vOTU are best captured in the 100 μm fraction. Furthermore, vOTU 17663 had a relative abundance <0.0001% in all viral fractions (<0.22 μm) collected demonstrating this vOTU likely reaches peak abundance when infecting hosts that are part of dense, colony-associated communities. To better understand why vOTU 17663 registered the highest relative abundance of all isolate vOTUs in Lake Erie, we next determined whom this vOTU was predicted to infect.

4.2.2.3 Isolate vOTU 17663 predicted to infect 5 *Microcystis* isolates, dozens of other consortia members

Using VHIP, vOTU 17663 was predicted to infect a host MAG in 18 of 21 WLECC cultures. Across 18 cultures, vOTU was found to infect 96 isolate host metagenome assembled genomes (MAGs) with a 93% predicted probability of infection. Of the 96 predicted hosts in culture, 5 of these were *Microcystis* MAGs (LE18-13.4, LE19-12.2, LE19-131.1, LE19-55.1 and LE19-98.1). Intriguingly, each of these *Microcystis* MAGs lacked an mcy operon, preventing the production of microcystin. To visualize these interactions between vOTU 17663 and these 5 *Microcystis* MAGs, in addition to other isolate vOTUs and isolate MAGs in these cultures, we generated 5 separate interaction networks (Fig. 18) complete with assigned host taxonomy and the date in which isolate vOTUs' cultures were collected.

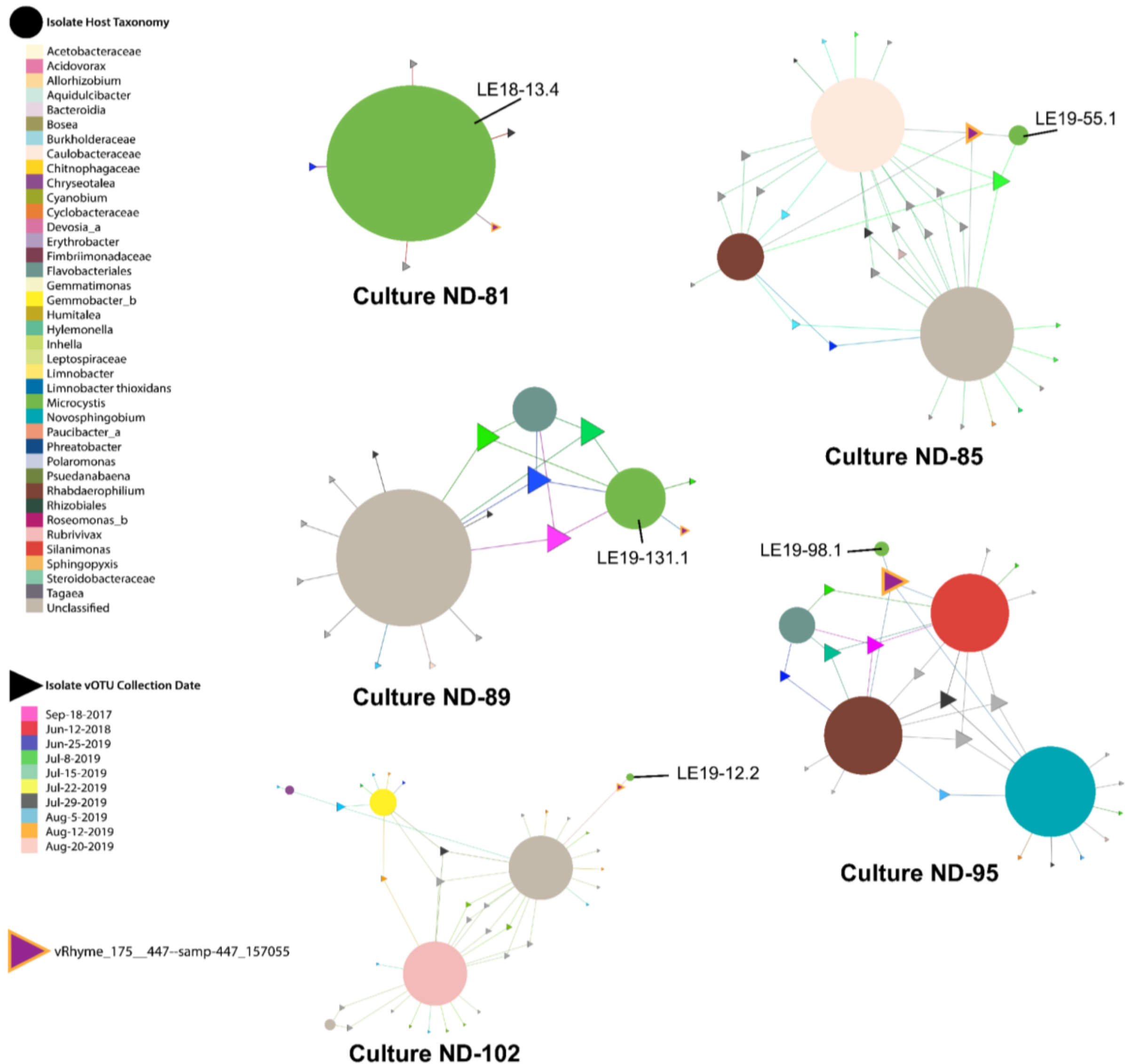


Figure 18. Isolate infection networks containing *Microcystis* and abundant vOTU 17663 (vRhyme_175__447--samp_447_157055). These cultures include: ND-81, ND-85, ND-89, ND-95 and ND-102. Circular nodes represent isolate host MAGs and triangular nodes represent isolate vOTUs. Host nodes are colored by GTDB taxonomic classification. vOTU nodes are colored by the date the culture sample was originally collected. All nodes are sized by the number of interactions they have. vOTU 17663 is specified with a specific color and stroke in the legend.

In addition to a *Microcystis* host, vOTU 17663 was predicted to infect at least one other host in 3 of the 5 cultures in which it was predicted to infect *Microcystis*. It is worth noting that Culture ND-81 had no high-confidence (93% infection probability) predicted infections outside of *Microcystis*, resulting in only one host being plotted from this culture. Nevertheless, these observations begin to address the research question: *Do*

we see intra-colony infection dynamics? Even when we introduce the bottleneck of visualizing only those cultures with a *Microcystis* MAG that vOTU 17663 was predicted to infect, we observe intra-colony infection dynamics. In fact, we observe that multiple vOTUs within these 5 networks are predicted to infect multiple host MAGs.

With VHIP demonstrating an 87% accuracy in predicting infections (Bastien et al., 2023), we hold confidence that a substantial portion of the identified co-evolutionary associations genuinely represent virus-host interactions, whether they be past or current. We suspect numerous genomic fragments, often referred to as genomic "shrapnel," could not be effectively matched with their corresponding sequences due to the challenges associated with binning viral genomes (Roux, 2016; Kieft et al., 2022). This circumstance might lead to an inflation in the number of predicted infections, as a single virus could be counted multiple times. Therefore, although the predicted virus counts may not precisely reflect the actual number of infections in each culture, the overall data structures offer valuable insights. These networks offer novel perspectives on virus-host network structures, including host range characteristics (e.g., narrow versus broad host ranges), the potential for gene exchange between host and vOTUs (e.g., identifying which host populations have been evolutionarily connected through viral infections), and insights into how viral diversity fluctuates over time and space throughout a bloom.

These culture-derived networks highlight the diversity of isolate vOTUs predicted to infect high quality *Microcystis* MAGs, and predictions of cross-infection of diverse hosts by the same vOTU. As *Microcystis* occurs in densely packed colonies with a closely associated and even physically attached, complex microbiome mainly composed of heterotrophic bacteria (Smith et al., 2021; Yancey et al., 2023), there is high potential for interactions with other bacterial taxa. Unlike specialist viruses that infect only a few host strains, generalist viruses can infect a broad range of hosts (de Jonge et al., 2019). This includes infections of different bacterial genera (Cazares et al., 2021) and phyla (Malki et al., 2015) and even cross-domain infections (Hwang et al., 2023). A virus' host range has been discovered to be a highly evolvable trait (Heineman et al., 2008; Meyer et al., 2012; Meyer et al., 2016; Holtzman et al., 2020; Sant et al., 2021; De Sordi et al., 2017; Cornuault et al., 2020); a trait that can either narrow or

expand (Heineman et al., 2008; Meyer et al., 2012). Prior meta-analyses have demonstrated nested patterns that allow for the coexistence of generalist and specialist phages (Flores et al., 2011; Flores et al., 2013; Weitz et al., 2013). To provide additional context to the dynamics of these isolate vOTUs and their predicted isolate host MAGs in the environment, we next tracked the relative abundances of isolate MAGs across eight bloom years.

4.2.2.4 Isolate vOTU 17663 tracks abundance of *Microcystis* MAG LE19-12.2 through blooms

To gain insights into the dynamics of WLECC isolate MAGs in the wild, we first dereplicated 210 isolate MAGs with 424 wild MAGs from environmental samples using Galah (99% ANI across 50% of the genome), resulting in a non-redundant set of 579 hosts to track moving forward. We competitively mapped this non-redundant set of host MAGs to samples' reads across eight different bloom seasons. The relative abundances of all isolate MAGs (SI Fig. 1) within the non-redundant host set were collected and tracked across seasons. When we restricted tracking relative abundance to only the *Microcystis* isolate MAGs (Fig. 19), a clear pattern emerged between vOTU 17663 and *Microcystis* MAG LE19-12.2. Like vOTU 17663, LE19-12.2 was by far the most abundant member of its respective community, reaching a relative abundance of 16.8% in the 100 μm fraction at station WLE2 during October 20th of the 2014 bloom. In the >0.22 μm fraction, LE19-12.2 had relative abundance peaks surpassing 15% on August 10th of the 2015 bloom. When comparing relative abundance peaks between vOTU 17663 (Fig. 17) and LE19-12.2, we see a clear tracking pattern at nearly every peak outside of vOTU 17663's final relative abundance peak during September 13th of the 2021 bloom. While uncertain, this boom peak may have been the result of a host switching event or the detection of vOTU in the >0.22 μm fraction following a final lysis event of LE19-12.2.

Given our results of 96 predicted isolate hosts for vOTU 17663, it is intriguing to see this vOTU follow the dynamics of one specific host, LE19-12.2, so closely. While

our work continues to support that these abundant community members likely rise to the peak of their relative abundances in the 100 μm fraction, one might expect an increased likelihood of host switching in a densely-populated colonial host community space. Yet, vOTU 17663 continued to track the dynamic abundance of LE19-12.2 year after year, bloom after bloom. Our findings suggest a fascinating level of consistency in the interactions between vOTU 17663 and its specific host, LE19-12.2, despite the potentially dynamic and competitive environment of a densely-populated colonial host community. This persistent tracking of host dynamics over multiple years and blooms raises intriguing questions about the mechanisms and selective pressures that govern such associations. It beckons further exploration into the coevolutionary processes, ecological advantages, and potential adaptations that allow vOTU 17663 to maintain a close relationship with LE19-12.2. Understanding these dynamics could provide valuable insights into the stability and intricacies of virus-host interactions within cHABs. We next explored whether these phage-host dynamics coincided with cHAB dynamics across eight seasons.

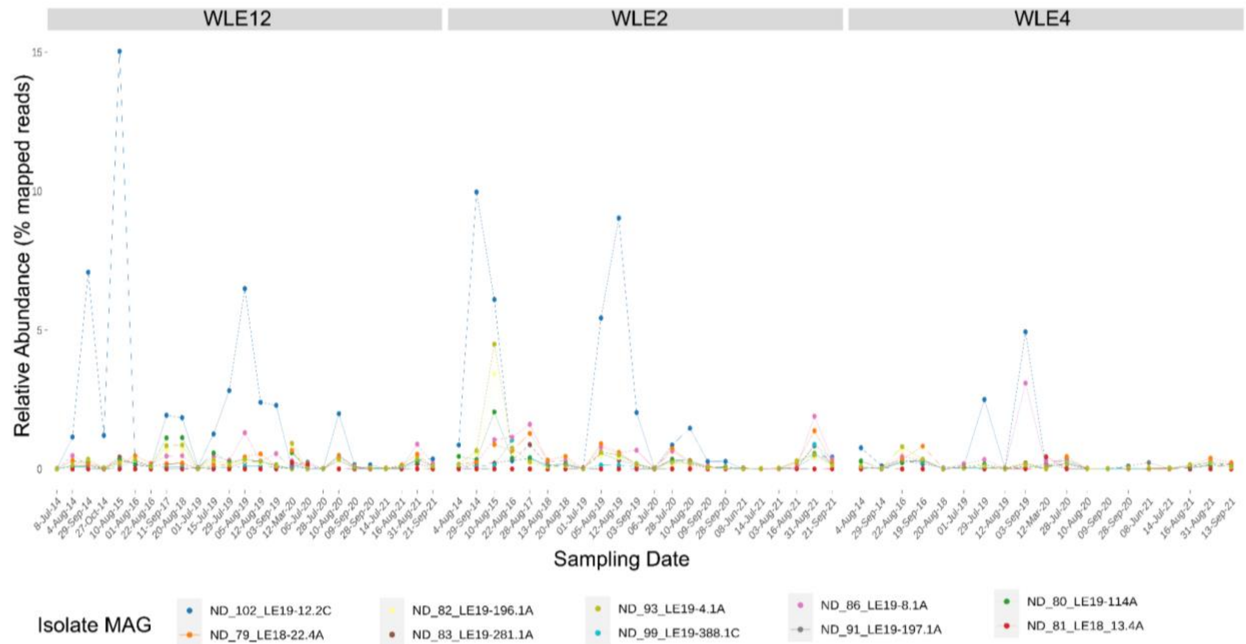


Figure 19. Relative abundance of Lake Erie culture isolate MAGs between 2014-2021. Top row displays all isolate host MAGs. These were dereplicated with wild MAGs and competitively mapped to samples' reads with wild MAGs. Line color pertains to each individual isolate MAG's relative abundance. Bottom row displays only Lake Erie culture isolate *Microcystis* MAGs between 2014-2021. The original 21 *Microcystis*

MAGs from isolates were dereplicated to only the 10 shown here. Note this only shows relative abundance of isolate MAGs at stations WLE12, WLE2 and WLE4 for the >0.22 μm fraction.

4.2.2.5 Isolate vOTU 17663 - LE19-12.2 dynamics track particulate microcystin measurements in cHABs

We characterized eight different bloom seasons in the western basin of Lake Erie using measurements of chlorophyll-a (used as a proxy for primary productivity), particulate phycocyanin (used as a proxy for cyanobacteria) and microcystin (indicative of bloom toxicity) (Fig. 20). Using these bloom proxies, we identified at least 12 toxic bloom peaks across the eight sampling seasons. Microcystin concentrations predominantly mirror phycocyanin and chlorophyll-a measurements across stations WLE12, WLE2 and WLE4, suggesting that as biomass of bacterial community members increases at these locations, so too do microcystin concentrations. While this has been known for some time (Cory et al., 2016; Berry et al., 2017), it reiterates the importance of knowledge regarding phage-host dynamics, as these predator-prey interactions may be key to the release of intracellular host toxins into the system.

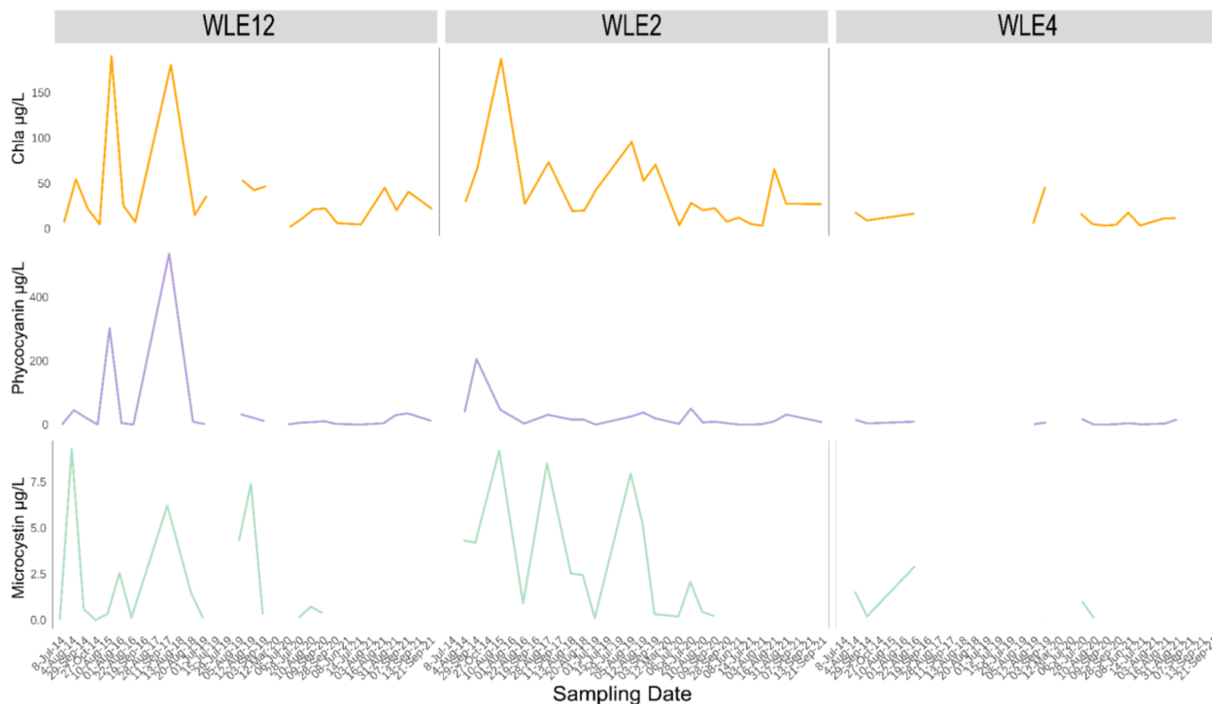


Figure 20. Bloom proxy for Lake Erie cHABs dynamics between 2014-2021. B) Chlorophyll-a, phycocyanin and microcystin measurements across bloom seasons for sampling stations WLE 12, WLE2 and WLE4 (refer to graphical abstract for sampling map).

When comparing the relative abundances of vOTU 17663 and predicted *Microcystis* host LE19-12.2, we see a clear relationship between the abundances of these community members and toxin concentration in the form of microcystin (Fig. 21; SI Table 5). Infection dynamics of vOTU 17663-LE19-12.2 provide a specific example of the importance in monitoring phage-host interactions in cHABs, as this relationship may have substantial impacts in *Microcystis* strain diversification and toxin release in the western basin of Lake Erie. However, LE19-12.2 was found to lack an mcy operon, indicating its lack of potential microcystin production. While the potential infection of LE19-12.2 by vOTU 17663 isn't expected to result in increased microcystin concentrations, this infection may drive down LE19-12.2 abundances, providing an opportunity for toxic *Microcystis* strains to increase in their respective abundances. It is also important to consider this is one of many phage-host interactions in a complex bloom community, leaving the door open to explore the broader implications of such interactions within the intricate web of microbial relationships in cyanobacterial harmful algal blooms (cHABs). While the vOTU 17663-LE19-12.2 interaction is a compelling case study, it represents just a single thread in the rich tapestry of phage-host dynamics within these ecosystems.

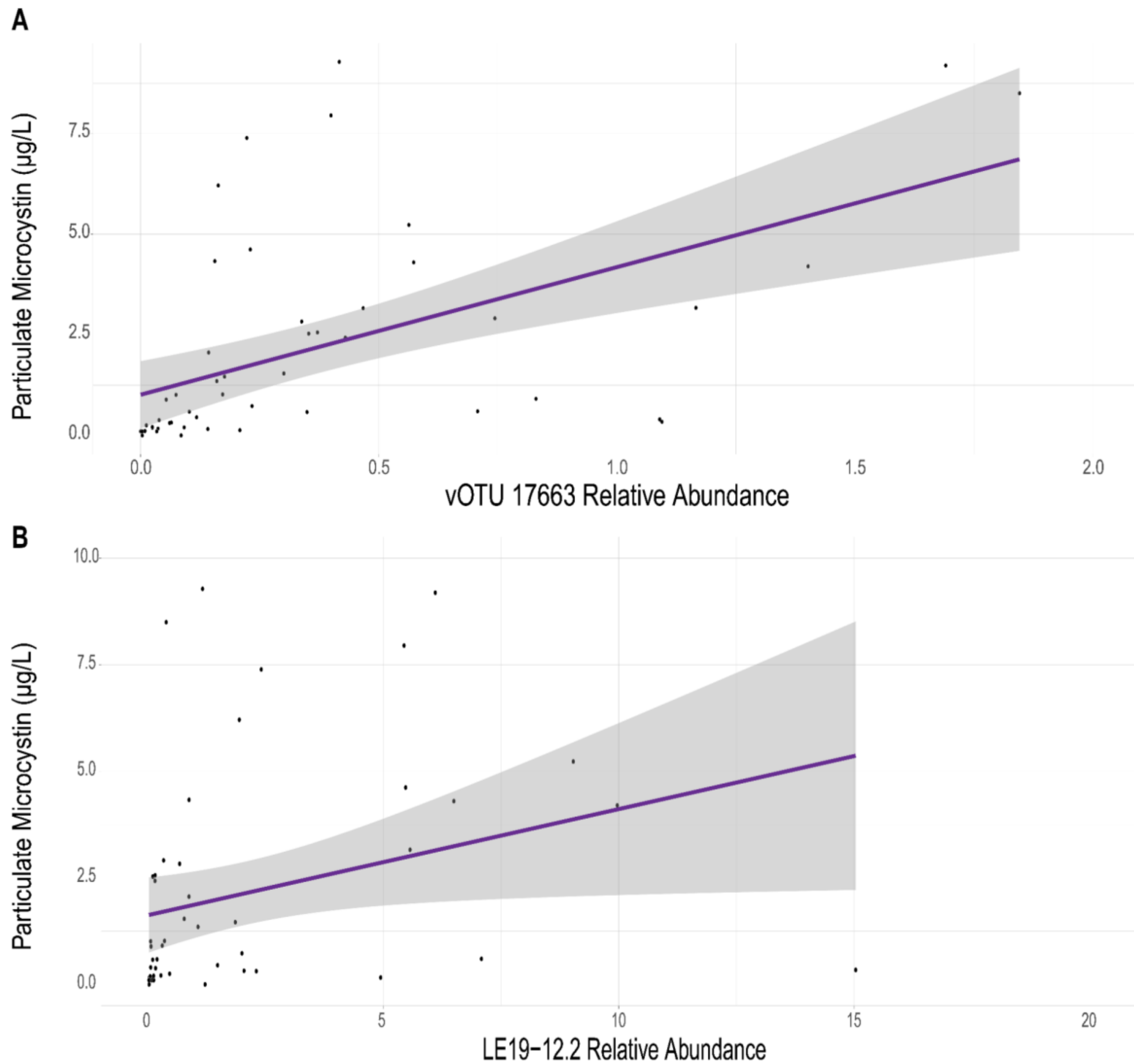


Figure 21. Correlation between microcystin concentration, (A) vOTU 17663 and (B) LE19-12.2 relative abundance. Each point represents the community member at a given date, station and fraction. Purple line depicts linear regression line and gray shadow represents 95% confidence intervals.

Understanding the nuances of virus-host interactions, like the one observed here, can provide valuable insights into the ecology and evolution of bacterial populations and their associated phages, especially as we continue to explore non-cyanobacterial populations like those heterotrophs described in previous studies (Smith et al., 2021; Yancey et al., 2023). These interactions may play a pivotal role not only shaping *Microcystis* strain diversification and toxin release, but ultimately the overall health and stability of cHABs in freshwater systems. Therefore, continued research into the myriad

phage-host interactions in cHABs is essential to grasp the full scope of their impact on these complex microbial communities and the ecosystems they inhabit.

4.3 Conclusion

Our study provides a glimpse into the intricate world of virus-host interactions within cHABs and highlights the importance of considering both genetic and temporal dimensions in these microbial ecosystems. As we continue to uncover the complexities of these interactions, we move one step closer to unraveling the mysteries of cHABs and their ecological implications. We established a correlation between the phylogenetic distances of *Microcystis* isolates and their infection profiles, suggesting that *Microcystis* strains with greater genetic similarity tend to share similar infection profiles. This correlation underscores the importance of shared genetic and physiological characteristics among closely related *Microcystis* strains in influencing their susceptibility or resistance to specific viruses.

Furthermore, our analysis highlighted the significant role of temporal dynamics in *Microcystis*-phage interactions. The collection date emerged as the primary predictor of infection profiles among *Microcystis* strains, emphasizing the importance of considering the temporal dimension when studying these microbial communities. While other environmental parameters were examined, they did not exhibit significant relationships with infection profiles, indicating that additional unexplored factors likely contribute to the variation observed. The study also revealed a specific case of a phage-host interaction between vOTU 17663 and *Microcystis* MAG LE19-12.2, which exhibited remarkable consistency over multiple bloom seasons. This persistence in tracking the host's dynamics raises intriguing questions about the underlying mechanisms and selective pressures governing such associations. It highlights the need for further exploration into the coevolutionary processes and potential adaptations that allow phages like vOTU 17663 to maintain close relationships with specific hosts.

4.4 Future Directions

Our study has paved the way for several promising avenues of research in the field of cyanobacterial harmful algal blooms (cHABs) and virus-host interactions. One critical direction involves expanding our understanding of the intricate phage-host network within these complex microbial communities. While our study focused on a specific phage-host interaction, it is essential to acknowledge that numerous other interactions are occurring concurrently. Thus, future research efforts should aim to uncover and characterize additional phage-host relationships to provide a more comprehensive view of the network's structure and dynamics both in the lab and in the wild.

The results of this work emphasize the importance of continuous monitoring and long-term studies to capture the full extent of temporal variations in these blooms. Understanding how phage-host interactions evolve over time is essential if we are to better understand the importance of viruses in ecosystems. Future investigations should prioritize the incorporation of temporal dimensions into the long-term study of virus-host dynamics.

Exploring the mechanisms behind host range expansion and contraction in phages, particularly in the context of cHABs, represents another crucial area of research. Understanding how phages switch hosts and the ecological consequences of such events is essential for comprehending the adaptability and persistence of these viruses. Investigating the broader microbial community dynamics, including interactions beyond cyanobacteria, such as heterotrophic bacteria, is equally vital. Phages play multifaceted roles in shaping these communities, and exploring these interactions can offer a more holistic view of cHAB stability and function.

Furthermore, efforts should be directed towards isolating more phages and hosts from cHABs. The improvement of inputs for infection networks depends on expanding the catalog of isolated phages and their respective hosts. This initiative will enhance our ability to construct more comprehensive and accurate phage-host interaction networks, allowing for a deeper understanding of the dynamics within these complex microbial ecosystems. In conclusion, our study has shed light on the complexities of virus-host interactions within cHABs, offering valuable insights into genetic and temporal dimensions. As we delve deeper into these intricacies, we move closer to unraveling the

mysteries of cHABs and their ecological implications. Future research endeavors in this field will undoubtedly contribute to a more comprehensive understanding of these vital ecosystems and perhaps guide strategies for managing and mitigating cHABs in freshwater systems.

4.5 Acknowledgements

We would like to thank the field crew at CIGLR/GLERL including Paul Den Uyl, Dack Stuart, Kent Baker, and Holly Kelchner for allowing us to sample alongside them and for collecting environmental parameter measurements.

4.6 Methods

4.6.1 Field Sampling, Culture Collection, Extraction and Sequencing

For the full methods of collection and cultivation of xenic cultures containing *Microcystis* from western Lake Erie during various sampling expeditions in 2017, 2018, and 2019, refer to Yancey et al., 2023 . Briefly, The isolation work involved plating samples, incubating them under specific light and temperature conditions, and serial streaking of *Microcystis* colonies onto agarose plates. Cultures were cryopreserved for long-term storage. Cultures were maintained at room temperature under controlled light conditions and reported successful growth at various temperature and light ranges. Several culture media, including BG-11 2N, LE BG11–2 N, unmodified BG-11, and WC medium, were used during isolation and maintenance.

Cultures were homogenized and centrifuged to obtain DNA. DNA extraction was performed using the DNeasy Blood and Tissue Kit with the QIAshredder adapter. DNA concentrations were quantified using the Quant-iT™ PicoGreen™ DNA Assay Kit. The sequencing was conducted at the University of Michigan's Advanced Genomic Core using an Illumina NovaSeq (S4) platform with 300 cycles for 150bp paired-end reads, maximizing insert size without compromising read quality.

4.6.2 Host Assembly and Binning

For full host assembly and binning methods, refer to Yancey et al., 2023. Briefly, bbtools software was used to remove adapters, quality trim reads, and eliminate contamination with the Univec reference collection. Duplicate reads were removed using clumpify and dedupe tools. Each sample was independently assembled using Megahit with the meta-sensitive parameter. Contigs longer than 1kb were used to create Anvi'o databases for each sample. *Microcystis* and associated bacterial bins were generated and manually refined using Concoct and Anvi'o. The taxonomies of these bins were determined using single-copy genes in Anvi'o and further evaluated with GTDBtk and the GTDB release 202 database. A pangenome analysis involving 159 *Microcystis* reference genomes and 21 obtained genomes identified 26 single-copy genes common to all 180 genomes, and these genes were used to construct a phylogenetic tree. Additionally, genomic pairwise average nucleotide identity (gANI) between each *Microcystis* MAG was calculated using pyani.

4.6.3 Viral Contig Identification, Clustering and Taxonomy

For full viral identification tools and accompanying rule sets to choose a specific set of tools, refer to Hegarty et al., 2023. Briefly, CheckV (v0.9.0) (Nayfach et al., 2020), DeepVirFinder (v1.0) (Ren et al., 2020), Kaiju (v1.9.0) (Menzel et al., 2016), VIBRANT (v1.2.1) (Kieft et al., 2020), VirSorter (v1.0.6) (Roux et al., 2015), and VirSorter2 (v2.2.3) (Guo et al., 2021) were executed on the University of Michigan Great Lakes Supercomputing Cluster to identify viral contigs from wild Lake Erie samples between 2014-2021 in addition to viral contigs in WLECC cultures collected between 2017-2019. For Kaiju taxonomic classification, the Kaiju nr_euk database (updated from NCBI 05-23-2022) was employed. Default parameters were generally used for all tools, with the exception of specifying a 3 kb contig length cutoff. Viruses were then binned using vRhyme default settings (Kieft et al., 2022) to create a collection of viral bins and high-quality unbinned contigs for population clustering. Viral contigs greater than 10kb were clustered (Roux and Bolduc, 2016; stampede-clustergenomes) according to previously established standards defining viral populations (Roux et al. 2019). Contigs sharing an

average nucleotide identity (ANI) of 95% across 85% of the contig length were clustered and the longest sequence of each cluster was considered the representative for a cluster, referred to as a viral OTU (vOTU) moving forward for downstream analyses. Taxonomy of vOTUs was estimated using the Phage Taxonomy Tool approach (PTT; Kieft et al., 2021).

4.6.4 Virus-Host Infection Prediction Network

For full methods of phage-host infection predictions, refer to Bastien et al., 2023. Briefly, a gradient-boosted machine learning model was employed, incorporating various feature classes related to virus-host interactions. These features included calculating the percent G+C content for both viral and host genomes and determining the differences between them (viral%G+C - host%G+C). Additionally, k-mer profiles were generated for viral and host genomes, considering two k-mer lengths: 3 and 6. Two distance metrics, namely euclidean and d2*, were applied to assess the distances between the k-mer frequencies of viruses and hosts. Sequence homology was evaluated between viral and host sequences to identify evidence of prior infections. Viral genome sequences were compared to all bacterial and archaeal sequences using default BLASTn parameters. Rare events involving the presence of hits against either bacterial genomes or spacers were combined into a single feature termed "homology." CRISPRCasFinder was used to identify CRISPR spacers in host species. Spacer sequences were collected, and viruses were subjected to a BLAST search against this spacer database. Hits with either 0 or 1 mismatch were retained as part of the CRISPR feature in the analysis. Model outputs were visualized using Gephi 0.9.0 (https://gephi.org/).

4.6.5 *Microcystis* MAG phylogenetic distance and infection profile correlation

Phylogenetic distances were extracted for 21 *Microcystis* isolate MAGs originally provided from Yancey et al., 2023. Next, the Jaccard method was used to calculate distances between infection profiles (presence/absence data generated from the number of infections/non-infections for all wild and isolate vOTUs against *Microcystis*

MAGs). Spearman Rank Correlation Coefficient on relationship between phylogenetic distances and infection profiles (non-parametric measure used to assess the strength and direction of the monotonic (non-linear) relationship between two variables) was performed using the v2.5-2 package in R v4.0.2 to assess the correlation between these data. PERMANOVA using the adonis function in vegan was used to test the effects of sampling location, sampling date, sampling fraction as well as effects of environmental parameters on the variation between infection profiles.

4.6.6 Isolate MAG and vOTU read mapping and relative abundance

Isolate host MAGs were dereplicated with all wild MAGs between 2014-2021 using dRep default settings (Olm, 2017). Isolate vOTUs were also clustered (see Viral Contig Identification, Clustering and Taxonomy section) to provide a non-redundant set of vOTUs. Filtered and trimmed reads were assembled from the same sample and quantified using Samtools v1.11 (Li et al., 2009). These reads were then competitively mapped to all vOTUs and then all MAGs (both wild and culture-based) using Bowtie2 (Langmead and Salzberg, 2012). Relative abundances of vOTUs and MAGs mapped reads were determined by summarizing reads mapped to vOTUs and MAGs using CoverM v.0.6.1 (Li, 2018).

For a full list of SI Tables, visit: https://docs.google.com/spreadsheets/d/1LAULNraVgo6q-WxefVKCZLuv4fRuWdnaVJiG29KxdZc/edit?usp=sharing

For a full list of SI Figures, visit: https://docs.google.com/document/d/166ohRVRmGyGcQxZe9GIpFY5MIVwvqUe8O91TffgadVA/edit?usp=sharing

4.7 References

Alexova, R., Dang, T. C., Fujii, M., Raftery, M. J., Waite, T. D., Ferrari, B. C., & Neilan, B. A. (2016). Specific global responses to N and Fe nutrition in toxic and non-toxic Microcystis aeruginosa. *Environmental Microbiology*, *18*(2), 401–413. https://doi.org/10.1111/1462-2920.12958

Alexova, R., Haynes, P. A., Ferrari, B. C., & Neilan, B. A. (2011). Comparative Protein Expression in Different Strains of the Bloom-forming Cyanobacterium Microcystis

aeruginosa*. *Molecular & Cellular Proteomics*, *10*(9). https://doi.org/10.1074/mcp.M110.003749
Bastien, E. G., Cable, R. N., Zaman, L., Batterbee, C., Wing, A. J., & Duhaime, M. B. (2023). *Virus-Host Interactions Predictor (VHIP): Machine learning approach to resolve microbial virus-host interaction networks* (p. 2023.11.03.565433). bioRxiv. https://doi.org/10.1101/2023.11.03.565433
Berry, M. A., Davis, T. W., Cory, R. M., Duhaime, M. B., Johengen, T. H., Kling, G. W., Marino, J. A., Den Uyl, P. A., Gossiaux, D., Dick, G. J., & Denef, V. J. (2017). Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial communities. *Environmental Microbiology*, *19*(3), 1149–1162. https://doi.org/10.1111/1462-2920.13640
Bozarth, C. S., Schwartz, A. D., Shepardson, J. W., Colwell, F. S., & Dreher, T. W. (2010). Population Turnover in a Microcystis Bloom Results in Predominantly Nontoxigenic Variants Late in the Season. *Applied and Environmental Microbiology*, *76*(15), 5207–5213. https://doi.org/10.1128/AEM.00001-10
Cazares, D., Cazares, A., Figueroa, W., Guarneros, G., Edwards, R. A., & Vinuesa, P. (2021). A Novel Group of Promiscuous Podophages Infecting Diverse Gammaproteobacteria from River Communities Exhibits Dynamic Intergenus Host Adaptation. *mSystems*, *6*(1), e00773-20. https://doi.org/10.1128/mSystems.00773-20
Cornuault, J. K., Moncaut, E., Loux, V., Mathieu, A., Sokol, H., Petit, M.-A., & De Paepe, M. (2020). The enemy from within: A prophage of Roseburia intestinalis systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *The ISME Journal*, *14*(3), Article 3. https://doi.org/10.1038/s41396-019-0566-x
Cory, R. M., Davis, T. W., Dick, G. J., Johengen, T., Denef, V. J., Berry, M. A., Page, S. E., Watson, S. B., Yuhas, K., & Kling, G. W. (2016). Seasonal Dynamics in Dissolved Organic Matter, Hydrogen Peroxide, and Cyanobacterial Blooms in Lake Erie. *Frontiers in Marine Science*, *3*. https://www.frontiersin.org/articles/10.3389/fmars.2016.00054
Davis, T. W., Harke, M. J., Marcoval, M. A., Goleski, J., Orano-Dawson, C., Berry, D. L., & Gobler, C. J. (2010). Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of Microcystis during cyanobacterial blooms. *Aquatic Microbial Ecology*, *61*(2), 149–162. https://doi.org/10.3354/ame01445
de Jonge, P. A., Nobrega, F. L., Brouns, S. J. J., & Dutilh, B. E. (2019). Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends in Microbiology*, *27*(1), 51–63. https://doi.org/10.1016/j.tim.2018.08.006
De Sordi, L., Khanna, V., & Debarbieux, L. (2017). The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses. *Cell Host & Microbe*, *22*(6), 801-808.e3. https://doi.org/10.1016/j.chom.2017.10.010
Dolan, D. M., & Chapra, S. C. (2012). Great Lakes total phosphorus revisited: 1. Loading analysis and update (1994–2008). *Journal of Great Lakes Research*, *38*(4), 730–740. https://doi.org/10.1016/j.jglr.2012.10.001
Dziallas, C., & Grossart, H.-P. (2011a). Increasing Oxygen Radicals and Water Temperature Select for Toxic Microcystis sp. *PLOS ONE*, *6*(9), e25569. https://doi.org/10.1371/journal.pone.0025569
Dziallas, C., & Grossart, H.-P. (2011b). Temperature and biotic factors influence bacterial communities associated with the cyanobacterium Microcystis sp. *Environmental Microbiology*, *13*(6), 1632–1641. https://doi.org/10.1111/j.1462-2920.2011.02479.x
Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., & Weitz, J. S. (2011). Statistical structure of host–phage interactions. *Proceedings of the National Academy of Sciences*, *108*(28). https://doi.org/10.1073/pnas.1101595108

Flores, C. O., Valverde, S., & Weitz, J. S. (2013). Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal*, *7*(3), Article 3. https://doi.org/10.1038/ismej.2012.135
Fontana, S., Thomas, M. K., Reyes, M., & Pomati, F. (2019). Light limitation increases multidimensional trait evenness in phytoplankton populations. *The ISME Journal*, *13*(5), Article 5. https://doi.org/10.1038/s41396-018-0320-9
Gao, E.-B., Huang, Y., & Ning, D. (2016). Metabolic Genes within Cyanophage Genomes: Implications for Diversity and Evolution. *Genes*, *7*(10), Article 10. https://doi.org/10.3390/genes7100080
Gobler, C. J., Burkholder, J. M., Davis, T. W., Harke, M. J., Johengen, T., Stow, C. A., & Van de Waal, D. B. (2016). The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. *Harmful Algae*, *54*, 87–97. https://doi.org/10.1016/j.hal.2016.01.010
Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, *9*(1), 37. https://doi.org/10.1186/s40168-020-00990-y
Harke, M. J., Steffen, M. M., Gobler, C. J., Otten, T. G., Wilhelm, S. W., Wood, S. A., & Paerl, H. W. (2016). A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae*, *54*, 4–20. https://doi.org/10.1016/j.hal.2015.12.007
Hegarty, B., V, J. R., Bastien, E., Langenfeld, K., Lindback, M., Saini, J. S., Wing, A., Zhang, J., & Duhaime, M. (2023). *Benchmarking informatics approaches for virus discovery: Caution is needed when combining in silico identification methods* (p. 2023.08.07.552334). bioRxiv. https://doi.org/10.1101/2023.08.07.552334
Heineman, R. H., Springman, R., & Bull, J. J. (2008). Optimal foraging by bacteriophages through host avoidance. *The American Naturalist*, *171*(4), E149-157. https://doi.org/10.1086/528962
Holtzman, T., Globus, R., Molshanski-Mor, S., Ben-Shem, A., Yosef, I., & Qimron, U. (2020). A continuous evolution system for contracting the host range of bacteriophage T7. *Scientific Reports*, *10*(1), 307. https://doi.org/10.1038/s41598-019-57221-0
Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H., Andreopoulos, B., Brewer, H. M., Glavina del Rio, T., Adkins, J. N., Paul, S., Sullivan, M. B., & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *The ISME Journal*, *14*(4), Article 4. https://doi.org/10.1038/s41396-019-0580-z
Hwang, Y., Roux, S., Coclet, C., Krause, S. J. E., & Girguis, P. R. (2023). Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. *Nature Microbiology*, *8*(5), Article 5. https://doi.org/10.1038/s41564-023-01347-5
Kardinaal, W. E. A., Tonk, L., Janse, I., Hol, S., Slot, P., Huisman, J., & Visser, P. M. (2007). Competition for Light between Toxic and Nontoxic Strains of the Harmful Cyanobacterium Microcystis. *Applied and Environmental Microbiology*, *73*(9), 2939–2946. https://doi.org/10.1128/AEM.02892-06
Kieft, K., Adams, A., Salamzade, R., Kalan, L., & Anantharaman, K. (2022). vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Research*, *50*(14), e83. https://doi.org/10.1093/nar/gkac341
Kieft, K., Zhou, Z., & Anantharaman, K. (2019). VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of virome function from genomic sequences. *bioRxiv*, 855387. https://doi.org/10.1101/855387
Kieft, K., Zhou, Z., Anderson, R. E., Buchan, A., Campbell, B. J., Hallam, S. J., Hess, M., Sullivan, M. B., Walsh, D. A., Roux, S., & Anantharaman, K. (2021). Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-23698-5

Kuno, S., Yoshida, T., Kaneko, T., & Sako, Y. (2012). Intricate interactions between the bloom-forming cyanobacterium Microcystis aeruginosa and foreign genetic elements, revealed by diversified clustered regularly interspaced short palindromic repeat(CRISPR) signatures. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.00626-12

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Gu, L., Bai, S., Wang, J., Su, L., Wei, B., Zhang, L., & Fang, J. (2021). Characterization of particle-associated and free-living bacterial and archaeal communities along the water columns of the South China Sea. *Biogeosciences*, *18*(1), 113–133. https://doi.org/10.5194/bg-18-113-2021

Malki, K., Kula, A., Bruder, K., Sible, E., Hatzopoulos, T., Steidel, S., Watkins, S. C., & Putonti, C. (2015). Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. *Virology Journal*, *12*(1), 164. https://doi.org/10.1186/s12985-015-0395-0

Malki, K., Sawaya, N. A., Tisza, M. J., Coutinho, F. H., Rosario, K., Székely, A. J., & Breitbart, M. (2021). Spatial and Temporal Dynamics of Prokaryotic and Viral Community Assemblages in a Lotic System (Manatee Springs, Florida). *Applied and Environmental Microbiology*, *87*(18), e0064621. https://doi.org/10.1128/AEM.00646-21

McKindles, K. M., Manes, M. A., DeMarco, J. R., McClure, A., McKay, R. M., Davis, T. W., & Bullerjahn, G. S. (2020). Dissolved Microcystin Release Coincident with Lysis of a Bloom Dominated by Microcystis spp. In Western Lake Erie Attributed to a Novel Cyanophage. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01397-20

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms11257

Meyer, J. R., Dobias, D. T., Medina, S. J., Servilio, L., Gupta, A., & Lenski, R. E. (2016). Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science (New York, N.Y.)*, *354*(6317), 1301–1304. https://doi.org/10.1126/science.aai8446

Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., & Lenski, R. E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science (New York, N.Y.)*, *335*(6067), 428–432. https://doi.org/10.1126/science.1214449

Monchamp, M.-E., Pick, F. R., Beisner, B. E., & Maranger, R. (2014). Nitrogen Forms Influence Microcystin Concentration and Composition via Changes in Cyanobacterial Community Structure. *PLOS ONE*, *9*(1), e85573. https://doi.org/10.1371/journal.pone.0085573

Morimoto, D., Tominaga, K., Nishimura, Y., Yoshida, N., Kimura, S., Sako, Y., & Yoshida, T. (2019). Co-occurrence of broad and narrow host-range viruses infecting the toxic bloom-forming cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, AEM.01170-19. https://doi.org/10.1128/AEM.01170-19

Morimoto, D., Yoshida, N., Sasaki, A., Nakagawa, S., Sako, Y., & Yoshida, T. (2023). Ecological Dynamics of Broad- and Narrow-Host-Range Viruses Infecting the Bloom-Forming Toxic Cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, *89*(2), e02111-22. https://doi.org/10.1128/aem.02111-22

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, *39*(5), 578–585. https://doi.org/10.1038/s41587-020-00774-7

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), Article 12. https://doi.org/10.1038/ismej.2017.126

Paerl, H. W., & Huisman, J. (2009). Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports*, *1*(1), 27–37. https://doi.org/10.1111/j.1758-2229.2008.00004.x

Preece, E. P., Hardy, F. J., Moore, B. C., & Bryan, M. (2017). A review of microcystin detections in Estuarine and Marine waters: Environmental implications and human health risk. *Harmful Algae*, *61*, 31–45. https://doi.org/10.1016/j.hal.2016.11.006

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, *5*(1), 69. https://doi.org/10.1186/s40168-017-0283-5

Rinta-Kanto, J. M., Konopko, E. A., DeBruyn, J. M., Bourbonniere, R. A., Boyer, G. L., & Wilhelm, S. W. (2009). Lake Erie Microcystis: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. *Harmful Algae*, *8*(5), 665–673. https://doi.org/10.1016/J.HAL.2008.12.004

Roux and Bolduc. (2016). *MAVERICLab / stampede-clustergenomes—Bitbucket*. https://bitbucket.org/MAVERICLab/stampede-clustergenomes/src/master/

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Lavigne, R., Brister, J. R., Varsani, A., Amid, C., Aziz, R. K., Bordenstein, S. R., Bork, P., Breitbart, M., Cochrane, G. R., Daly, R. A., Desnues, C., Duhaime, M. B., … Eloe-Fadrosh, E. A. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology*, *37*(1), Article 1. https://doi.org/10.1038/nbt.4306

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., & Sullivan, M. B. (2017). Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*, *5*. https://doi.org/10.7717/peerj.3817

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: Mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. https://doi.org/10.7717/peerj.985

Sant, D. G., Woods, L. C., Barr, J. J., & McDonald, M. J. (2021). Host diversity slows bacteriophage adaptation by selecting generalists over specialists. *Nature Ecology & Evolution*, *5*(3), Article 3. https://doi.org/10.1038/s41559-020-01364-1

Schroeder, D. C., Oke, J., Hall, M., Malin, G., & Wilson, W. H. (2003). Virus Succession Observed during an Emiliania huxleyi Bloom. *Applied and Environmental Microbiology*, *69*(5), 2484–2490. https://doi.org/10.1128/AEM.69.5.2484-2490.2003

Schuurmans, J. M., Brinkmann, B. W., Makower, A. K., Dittmann, E., Huisman, J., & Matthijs, H. C. P. (2018). Microcystin interferes with defense against high oxidative stress in harmful cyanobacteria. *Harmful Algae*, *78*, 47–55. https://doi.org/10.1016/j.hal.2018.07.008

Singh, S., Rai, P. K., Chau, R., Ravi, A. K., Neilan, B. A., & Asthana, R. K. (2015). Temporal variations in microcystin-producing cells and microcystin concentrations in two fresh water ponds. *Water Research*, *69*, 131–142. https://doi.org/10.1016/j.watres.2014.11.015

Smith, D. J., Berry, M. A., Cory, R. M., Johengen, T. H., Kling, G. W., Davis, T. W., & Dick, G. J. (2022). Heterotrophic Bacteria Dominate Catalase Expression during Microcystis Blooms. *Applied and Environmental Microbiology*, *88*(14), e02544-21. https://doi.org/10.1128/aem.02544-21

Smith, D. J., Tan, J. Y., Powers, M. A., Lin, X. N., Davis, T. W., & Dick, G. J. (2021). Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. *Environmental Microbiology*, *23*(6), 3020–3036. https://doi.org/10.1111/1462-2920.15514

Sorensen, G., Baker, A. C., Hall, M. J., Munn, C. B., & Schroeder, D. C. (2009). Novel virus dynamics in an Emiliania huxleyi bloom. *Journal of Plankton Research*, *31*(7), 787–791. https://doi.org/10.1093/plankt/fbp027

Steffen, M. M., Dearth, S. P., Dill, B. D., Li, Z., Larsen, K. M., Campagna, S. R., & Wilhelm, S. W. (2014). Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in Microcystis. *The ISME Journal*, *8*(10), 2080–2092. https://doi.org/10.1038/ismej.2014.78

Thompson, L. R., Zeng, Q., Kelly, L., Huang, K. H., Singer, A. U., Stubbe, J., & Chisholm, S. W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences*, *108*(39), E757–E764. https://doi.org/10.1073/pnas.1102164108

Trainic, M., Koren, I., Sharoni, S., Frada, M., Segev, L., Rudich, Y., & Vardi, A. (2018). Infection Dynamics of a Bloom-Forming Alga and Its Virus Determine Airborne Coccolith Emission from Seawater. *iScience*, *6*, 327–335. https://doi.org/10.1016/j.isci.2018.07.017

Wang, Q., Cai, L., Zhang, R., Wei, S., Li, F., Liu, Y., & Xu, Y. (n.d.). A Unique Set of Auxiliary Metabolic Genes Found in an Isolated Cyanophage Sheds New Light on Marine Phage-Host Interactions. *Microbiology Spectrum*, *10*(5), e02367-22. https://doi.org/10.1128/spectrum.02367-22

Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B., & Hochberg, M. E. (2013). Phage–bacteria infection networks. *Trends in Microbiology*, *21*(2), 82–91. https://doi.org/10.1016/j.tim.2012.11.003

Wendling, C. C., Goehlich, H., & Roth, O. (2018). The structure of temperate phage–bacteria infection networks changes with the phylogenetic distance of the host bacteria. *Biology Letters*, *14*(11), 20180320. https://doi.org/10.1098/rsbl.2018.0320

Wing, A.J, Hegarty, B., Bastien, G.E., Evans, J., Cable, R., Smith, D., NOAA, Denef, V., Dick, G., & Duhaime, M.B. “Tracking Microcystis Viruses and Infection Dynamics in a Multi-Peak Microcystis-Dominated Bloom.” [Submitted to PLoS Biology]

Wing, A.J, Bastien, G.E., Hegarty, B., Michaelson, L., Cable, R., Smith, D., NOAA, Denef, V., Dick, G., & Duhaime, M.B. “Deciphering Viral Community Structure and Metabolic Potential in a Great Lakes Cyanobacterial Harmful Algal Bloom (cHAB)” [in preparation]

Xiao, M., Willis, A., & Burford, M. A. (2017). Differences in cyanobacterial strain responses to light and temperature reflect species plasticity. *Harmful Algae*, *62*, 84–93. https://doi.org/10.1016/j.hal.2016.12.008

Yancey, C. E., Kiledal, E. A., Chaganti, S. R., Denef, V. J., Errera, R. M., Evans, J. T., Hart, L. N., Isailovic, D., James, W. S., Kharbush, J. J., Kimbrel, J. A., Li, W., Mayali, X., Nitschky, H., Polik, C. A., Powers, M. A., Premathilaka, S. H., Rappuhn, N. A., Reitz, L. A., … Dick, G. J. (2023). The Western Lake Erie culture collection: A promising resource for evaluating the physiological and genetic diversity of Microcystis and its associated microbiome. *Harmful Algae*, *126*, 102440. https://doi.org/10.1016/j.hal.2023.102440

Yang, F., Jin, H., Wang, X.-Q., Li, Q., Zhang, J.-T., Cui, N., Jiang, Y.-L., Chen, Y., Wu, Q.-F., Zhou, C.-Z., & Li, W.-F. (2020). Genomic Analysis of Mic1 Reveals a Novel Freshwater Long-Tailed Cyanophage. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.00484

Yoshida, M., Yoshida, T., Kashima, A., Takashima, Y., Hosoda, N., Nagasaki, K., & Hiroishi, S. (2008). Ecological dynamics of the toxic bloom-forming cyanobacterium Microcystis aeruginosa and its cyanophages in freshwater. *Applied and Environmental Microbiology*, *74*(10), 3269–3273. https://doi.org/10.1128/AEM.02240-07

Yoshida, T., Nishimura, Y., Watai, H., Haruki, N., Morimoto, D., Kaneko, H., Honda, T., Yamamoto, K., Hingamp, P., Sako, Y., Goto, S., & Ogata, H. (2018). Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME Journal*, *12*(5). https://doi.org/10.1038/s41396-018-0052-x

Yoshida, T., Takashima, Y., Tomaru, Y., Shirai, Y., Takao, Y., Hiroishi, S., & Nagasaki, K. (2006). Isolation and Characterization of a Cyanophage Infecting the Toxic Cyanobacterium

Microcystis aeruginosa. *Applied and Environmental Microbiology*, *72*(2), 1239–1247. https://doi.org/10.1128/AEM.72.2.1239-1247.2006

Yoshida-Takashima, Y., Yoshida, M., Ogata, H., Nagasaki, K., Hiroishi, S., & Yoshida, T. (2012). Cyanophage infection in the bloom-forming cyanobacteria Microcystis aeruginosa in surface freshwater. *Microbes and Environments*, *27*(4), 350–355. https://doi.org/10.1264/JSME2.ME12037

Zilliges, Y., Kehr, J.-C., Meissner, S., Ishida, K., Mikkat, S., Hagemann, M., Kaplan, A., Börner, T., & Dittmann, E. (2011). The Cyanobacterial Hepatotoxin Microcystin Binds to Proteins and Increases the Fitness of Microcystis under Oxidative Stress Conditions. *PLOS ONE*, *6*(3), e17615. https://doi.org/10.1371/journal.pone.0017615

Chapter 5: Conclusions and the Road Ahead

Summary

This comprehensive investigation into the viral community dynamics within Lake Erie's cyanobacterial harmful algal blooms (cHABs) has yielded profound insights into the intricate interplay between viruses and their hosts. These findings not only provide a deeper understanding of Lake Erie’s cHAB ecology but also contribute to the broader knowledge of viral dynamics in aquatic ecosystems. This dissertation took a bioinformatic approach to expand current knowledge of viral ecology in cHABs by:

- Identifying Lake Erie-specific populations of known globally distributed *Microcystis* viruses.
- Spatiotemporal patterns of *Microcystis* virus Ma-LEF01 in Lake Erie reveal strain-level population dynamics in *Microcystis* viruses.
- Predicting infection interactions between hundreds of viral operational taxonomic units (vOTUs) and *Microcystis*, with most vOTU host ranges within the genus and some spanning phyla.
- Identifying metabolic genes encoded by predicted *Microcystis* vOTUs can be specific to bloom peaks.
- Observing turnover of predicted *Microcystis* vOTUs is related to colony formation and sampling fraction.
- Noting that the diversity of predicted *Microcystis* vOTUs is highest in the viral fraction and lowest in colony-associated fractions during bloom peaks.
- Applying the viral 'Bank' framework, observing that abundant vOTUs are rare and sporadically observed, and suggesting that cHAB viral activity is fraction-specific
- Demonstrating virus-host network turnover through cHAB progression

- Establishing a correlation between the phylogenetic distance of *Microcystis* isolates from Lake Erie and their predicted viral infection profiles.
- Identifying that collection date as a predictor of predicted viral infection profiles for *Microcystis* isolates.
- Analyzing the Western Lake Erie Culture Collection (WLECC) to identify 41 isolate vOTUs, with isolate vOTU_17663 being the most abundant of the isolate vOTUs in Lake Erie.
- Identifying predicted virus-host pair from isolates, *Microcystis* MAG (LE19-12.2) and vOTU_17663, that reach high relative abundances in Lake Erie across eight bloom seasons.
- Observing that the virus-host isolate pair vOTU_17663/LE19-12.2 relative abundances correspond with particulate microcystin measurements in cHABs.

These findings collectively advance our understanding of virus-host interactions, strain-level dynamics, and the ecological complexities of cyanobacterial harmful algal blooms in Lake Erie. They also highlight the significance of temporal factors, fraction-specific viral activity, and the need for long-term monitoring in studying these ecosystems.

5.1 Tracking the viral predators of *Microcystis* through the 2014 cHAB: novel insights from novel viruses

This dissertation adopted a multifaceted approach, utilizing genomic signals of coevolution, to significantly advance our understanding of the complex interactions between viruses and their *Microcystis* hosts within the context of cHABs.

5.1.1 Ma-LEF01: a *Microcystis* cyanophage specific to Lake Erie

Our work identified four Lake Erie viral operational taxonomic units (vOTUs) highly similar to known *Microcystis* viruses from various geographic regions and time periods (Tucker and Pollard, 2005; Yoshida et al., 2008; Ou et al., 2015; Lin et al., 2021; Yang et al., 2020; Cai et al., 2022; Qian et al., 2022; Wang et al., 2022; Zhang et al., 2022). This finding suggests the persistence of predator-prey relationships with broader implications for cHAB dynamics beyond Lake Erie.

A detailed examination of Ma-LEF01, one of the identified viruses, revealed a complex genome structure with characteristics of both lytic and lysogenic replication strategies. Unique gene clusters and loci specific to Ma-LEF01 suggested potential viral strain diversity and adaptation to local conditions. Spatiotemporal dynamics of Ma-LEF01 unveiled distinct patterns in its abundance, emphasizing the importance of considering strain-level population dynamics within viral communities. Furthermore, the strain-level analysis indicated that Ma-LEF01 and MVGF-J19 likely belong to the same vOTU, while MaMV-DC and Ma-LMM01 form a separate vOTU, aligning with the 95% ANI threshold proposed for distinguishing viral species (Roux et al., 2016). This finding underscores the significance of accounting for strain-level diversity when defining viral species. The study contributes significantly to our understanding of viral-host interactions, strain-level diversity, and the potential ecological consequences of *Microcystis* viruses in shaping cHAB dynamics.

The addition of future omics-centric approaches will offer a more comprehensive view of viral communities within Lake Erie, enabling the identification of novel viral genes, metabolic pathways, and potential interactions with other microbial communities. Transcriptomics and metabolomics, can provide an additional layer of insight into viral ecology in cHABs. Transcriptomics can help reveal gene expression patterns in response to viral infections, shedding light on host-virus interactions and the molecular mechanisms involved. Metabolomics, on the other hand, can elucidate the metabolic changes within *Microcystis* populations during viral infections, aiding in our understanding of the broader biochemical consequences of viral predation. Together, these multi-omics approaches can provide a more comprehensive and integrated

understanding of the intricate dynamics between *Microcystis*, Ma-LEF01 and cHAB progression.

5.1.2 Moving beyond marker gene analyses to track viral dynamics

This dissertation challenges the limitations of previous knowledge, which primarily relied on marker gene analyses focused on the gp91 tail sheath gene of the Ma-LMM01/MaMV-DC viral population, by unveiling the extensive and dynamic nature of *Microcystis* virus-host interactions (Takashima et al., 2007; Yoshida et al., 2008; Kimura et al., 2012; Mankiewicz-Boczek et al., 2016; McKindles et al., 2020; Pound et al., 2020). By leveraging the Virus-Host Interaction Predictor (VHIP) (Bastien et al., 2023), a machine learning-based tool that harnesses genomic signals of coevolution, a vast network of predicted virus-host interactions emerged, encompassing numerous viral operational taxonomic units (vOTUs) and bacterial population genomes (MAGs).

The analysis of viral abundance during the August 4 toxic and September 29 non-toxic bloom peaks revealed a substantial number of vOTUs predicted to infect *Microcystis*, challenging previous metagenomic studies and emphasizing the importance of accounting for strain-level diversity and fragmented viral genomes (McKindles et al., 2020; Morimoto et al., 2023, Pound et al., 2020). These predicted associations, while reflecting past infection networks, do not necessarily imply current infections but demonstrate the existence of both narrow and broad host range *Microcystis* viruses. Some vOTUs span different phyla, suggesting a potential for gene flow between host and virus populations. Furthermore, our work identified virus-encoded auxiliary metabolic genes (AMGs) linked to various metabolic processes, such as photosynthesis, nitrogen metabolism, and response to environmental stress, underscoring the role of viruses in potentially rewiring host metabolisms and influencing matter and energy flow during infections (Roux, 2016; Anantharaman et al., 2014; Kieft and Zhou, 2020; Zimmerman et al., 2020; Howard-Varona et al., 2020).

Our work unveiled high turnover among the most abundant *Microcystis* vOTUs during the cHAB, with dominant viral populations peaking in abundance at specific bloom stages and displaying preferences for either colony-associated or free-living *Microcystis* populations. This observation aligns with the "Kill the Winner" hypothesis and highlights the influence of shifts in host availability, environmental conditions, and specific host-virus interactions on viral dynamics (Breitbart, 2005; Hevroni et al., 2020). Importantly, these findings provide evidence that viruses play a role in shaping the succession of *Microcystis* strains during cHABs, impacting the diversity and ecological dynamics of *Microcystis* populations (Berry et al., 2016; Yancey et al., 2023).

Future studies will leverage these findings to ask questions such as:

i) *What are the specific mechanisms that drive strain-level diversity within Microcystis viruses*, *and how do these strains adapt to local conditions?*

Investigating the specific mechanisms that drive strain-level diversity within *Microcystis* viruses can provide insights into the genetic and evolutionary processes that underpin their adaptation to local conditions. This knowledge can aid in predicting how viral populations might respond to changing environmental factors, including temperature, nutrient levels, and host diversity.

ii) *How does strain-level diversity influence virus-host interactions and the overall dynamics of cHABs in Lake Erie?*

The role of strain-level diversity in influencing virus-host interactions and the overall dynamics of cHABs is crucial for managing and mitigating harmful algal blooms. By uncovering how different viral strains interact with their hosts and how this affects *Microcystis* populations, researchers can develop more accurate models to predict the timing, severity, and persistence of cHABs.

iii) *What are the cascading effects of virus-host interactions on nutrient cycling, other planktonic organisms, and overall ecosystem health?*

Understanding the cascading effects of virus-host interactions on nutrient cycling, other planktonic organisms, and overall ecosystem health is essential for comprehending the broader ecological impacts of *Microcystis* viruses. This knowledge can help assess the resilience of Lake Erie's ecosystem to cHABs and inform strategies for preserving water quality and the well-being of aquatic life.

Overall, answering these questions can enhance our ability to predict and manage the ecological and environmental impacts of harmful algal blooms in Lake Erie and similar freshwater systems. It can also contribute to the development of more effective strategies for mitigating the negative consequences of cHABs, ultimately benefiting both the ecosystem and human communities that rely on these waters to survive.

5.2 Broadening horizons beyond *Microcystis*: why all members of the cHABs microbial community matter

The investigation into the viral community ecology during the 2014 Lake Erie cyanobacterial harmful algal bloom (cHAB) has revealed a wealth of novel information about the intricate dynamics of viruses in this ecosystem, highlighting the importance of future considerations aimed at assessing virus-host interactions among all available community members.

5.2.1 Characterizing Lake Erie viral community results in thousands of novel viruses, and more importantly, an emphasis on the importance of size in aquatic matrices

A total of 36 metagenomes were collected from these stations and fractions, resulting in the identification of 27,086 viral contigs, which were further classified into 3,527 novel viral genus clusters and 15,461 viral operational taxonomic units (vOTUs). The observed viral diversity exhibited intriguing patterns across time and filter fractions. The viral fraction consistently displayed the highest diversity, which is in line with previous findings in oceanic viromes (Dart et al., 2023). This trend was attributed to the seed bank theory, where a diverse reservoir of viruses persists in submicron size fractions, contributing to infections when a suitable host is encountered. Interestingly, the viral diversity in colony-associated fractions (53 and 100 µm) was notably lower than that in non-colony-associated fractions, particularly during the bloom peaks. This lower diversity in the colony-associated fractions was attributed to active infections of host cells within colonies, resulting in reduced viral diversity due to the dominance of specific host-virus interactions during these critical periods (Gregory et al., 2019).

Conducting a more extended and continuous temporal analysis could provide insights into the stability and resilience of viral communities over multiple years. Addressing these questions will add the next layer to our understanding of viral ecology in cHABs:

How do viral communities change in response to interannual variations in environmental conditions and Microcystis dynamics? Are there recurrent patterns in viral diversity and community structure?

Our analysis of viral community turnover revealed that both time and fraction significantly correlated with viral community structure, collectively explaining a significant portion of the variation. Temporal shifts in viral communities have been observed in various aquatic environments, reflecting changes in host communities as microbes respond to seasonal environmental fluctuations (Chow & Fuhrman, 2012; Brum et al., 2016). In contrast, the sampling location did not significantly influence viral community structure, aligning with previous studies in Lake Erie that emphasized seasonal rather than spatial variation in microbial communities (Berry et al., 2017a; Smith et al., 2021). This phenomenon could be attributed to the frequent mixing in Lake

Erie, which tends to dissipate station-driven influences on viral communities (Lin et al., 2021).

Furthermore, the distribution of viral-encoded auxiliary metabolic genes (AMGs) was tracked over time and across fractions. Similar to viral community structure, viral AMGs clustered primarily by sampling date and were not significantly affected by sampling location. These findings indicate a potential functional redundancy across different fractions, suggesting that the potential viral community metabolic function remains relatively consistent, even when taxonomic differences are observed. This may be due to the undersampling of functions, as many viral genes identified in this study had limited homology with reference genes in public databases (Deboutte et al., 2020; Gregory et al., 2019). Investigating the functional redundancy observed in viral-encoded auxiliary metabolic genes (AMGs) across fractions and time can lead to the discovery of novel functions, addressing questions such as:

What are the specific metabolic pathways and processes facilitated by these AMGs, and how do they impact microbial communities and ecosystem functioning in Lake Erie?

Comparing the viral community dynamics and AMG distributions in Lake Erie with other aquatic ecosystems can provide insights into the uniqueness or commonality of these patterns, where a future question may be posed such as: *What distinguishes Lake Erie's viral ecology from other freshwater bodies, and what can we learn from cross-system comparisons?*

Examining the interactions between viral communities and other microbial communities, such as bacteria and algae, can provide a more holistic understanding of ecosystem dynamics, leading to the research question: *How do viral infections influence the composition and function of these microbial communities, and how do these interactions shape the overall health of Lake Erie's ecosystem?*

5.2.2 Lake Erie evidence substantiates the Viral Bank model: how to apply this thinking moving forward

One of the central findings of this study is the compelling evidence that substantiates the "Viral Bank" model, as initially proposed by Breitbart and Rohwer (2005). This model provides a framework to comprehend the observed temporal dynamics and population fluctuations of Lake Erie's viral communities. According to this model, the majority of viruses in an ecosystem exist in a non-active state, comprising a viral bank, while only a small subset are actively infecting hosts at any given time, similar to the seed bank theory (Breitbart and Rowher, 2005). This concept is validated in the Lake Erie context, where a mere 5.7% of viral operational taxonomic units (vOTUs) were ever considered highly abundant (>0.5% of viral reads). These highly abundant vOTUs exhibit transient dominance, rising and falling in abundance over time, highlighting the dynamic nature of viral populations. Furthermore, this study reveals that the majority of viruses in Lake Erie's cHAB ecosystem exist at low abundance, contributing to the overall richness and evenness of the viral community, consistent with the observations in studies by Breitbart et al. (2005) and Dart et al. (2023).

The application of the Viral Bank model in this context offers valuable insights into the seasonal succession of vOTUs. Few vOTUs ever reach the "abundant" level, representing the active fraction of the community. As active viruses infect their hosts and are released as free virions, they contribute to the local viral diversity, enriching the viral bank. This phenomenon is further expounded through the "Kill the Winner" framework (Winter et al., 2010), whereby dominant viruses infect the prevailing "winners" within the microbial community, especially during bloom peaks. This pattern is consistent with the findings of Dart et al. (2023). The colonies captured on larger filter fractions, with their dense and diverse microbial populations, serve as hotspots of viral activity, facilitating infection and potentially influencing microbial community structure within the bloom. This observation underscores the importance of considering different ecological niches within a larger ecosystem when studying host-virus interactions.

Further research can focus on refining and expanding the Viral Bank model's applicability in different aquatic ecosystems, answering questions such as: *How does the Viral Bank model manifest in other freshwater bodies, coastal regions, or marine environments? Are there variations in the proportions of active and inactive viral populations, and what drives these differences?*

Investigating the triggers and mechanisms that lead to the activation of specific vOTUs from the viral bank is essential and future questions asked include: *What environmental cues or host-related factors influence the transition from a non-active to an active state for viruses in Lake Erie and other ecosystems? How do these activations impact microbial community dynamics?*

Expanding the temporal scope of studies can provide insights into the stability and persistence of viral banks over multiple years, asking multi-year questions including: *How do viral banks evolve seasonally and interannually, and what are the consequences for ecosystem functioning and resilience? Are there recurrent patterns in the activation of specific vOTUs?*

Investigating the influence of viral banks on microbial community structure, diversity, and succession during cHABs is crucial. *How do active viruses impact the composition and function of microbial communities, and how does this influence ecosystem health? Can viral bank dynamics be integrated into predictive models for cHAB occurrence and impact?*

Exploring the role of ecological niches, such as the microbial hotspots associated with larger filter fractions and bloom-forming colonies, in shaping viral activity and microbial community interactions is vital. *How do these hotspots facilitate viral infections and potentially influence microbial community structure within cHABs? What are the consequences for nutrient cycling and overall ecosystem dynamics?*

Developing innovative methods for monitoring and characterizing viral banks in aquatic ecosystems can improve our ability to study their dynamics. *How can cutting-edge technologies and high-throughput sequencing be leveraged to capture a more comprehensive view of viral populations, including those in the non-active state?*

By addressing these future research questions, researchers can advance our understanding of the Viral Bank model's applicability in various ecosystems, unravel the mechanisms governing viral activation, and gain deeper insights into the dynamic interplay between viruses, hosts, and microbial communities.

5.2.3 *Microcystis*-virus interactions are not alone in cHABs: moving past the *Microcystis* minority

Predictive modeling identified thousands of potential host-virus interactions, with host specificity playing a significant role. Importantly, not all active vOTUs were observed in the viral bank across sampling points, suggesting that metagenomics may not capture all rare members of the viral community, a limitation also acknowledged in studies by Deboutte et al. (2020) and Gregory et al. (2019). The analysis highlighted that while hundreds of vOTUs were predicted to infect *Microcystis*, the dominant cyanobacterial genus in the bloom, only a small fraction of these phages ever exceeded 0.1% in relative abundance. This observation emphasizes the complexity of virus-host relationships and suggests that most phages capable of infecting *Microcystis* remain at low abundance throughout the bloom. The concept of context-dependent virus-host interactions was further supported by the observation that different ecological niches, such as colony-associated and non-colony-associated fractions, were targeted by different sets of active viruses. This context-dependent variation in host populations has implications for our understanding of viral-mediated top-down control in ecosystems.

Furthermore, this study contributes to our understanding of viral community structure and dynamics by considering temporal factors and size fractionation. The observed turnover in virus-host interactions over time underscores the dynamic nature of these

relationships, a pattern also identified in studies by Chow & Fuhrman (2012), Brum (2016), and Malki et al. (2021). While sampling station did not significantly explain the variation in viral community structure, sampling date played a crucial role, indicating the importance of incorporating temporal dynamics into studies of viral communities. Future research in this field can now delve deeper into the specific host-virus interactions, their implications for harmful algal blooms, and their broader significance in aquatic ecosystems.

Future research can explore the detection and ecological roles of rare viral community members that may not be well-captured by metagenomics. *How do these rare viruses contribute to ecosystem dynamics, and what factors govern their emergence and persistence?*
Investigating the context-dependent nature of virus-host interactions in greater detail can provide insights into the factors driving variations in host populations and viral communities. *How do different ecological niches, such as colony-associated and non-colony-associated fractions, shape the diversity and dynamics of active viruses? What environmental cues influence these context-dependent interactions?*

Extending temporal analyses of viral communities can elucidate recurring patterns and long-term trends in virus-host interactions. *How do virus-host interactions change over multiple years, and are there consistent temporal patterns in the dominance and turnover of active viruses?* Advancing methods for viral community analysis, including improved techniques for capturing rare viruses and characterizing their functions, can enhance our ability to study virus-host interactions comprehensively. *What innovative approaches can be developed to overcome limitations in detecting and characterizing rare virus-host interactions and what implications might these interactions have?*

5.3 A worker is only as good as their tools: encouraging phage and host isolation methods in a bioinformatics-driven field

5.3.1 The correlation between evolutionary distance, suites of infection and time

"*There is a unity that makes us all one, and there is a diversity that makes us each our own unique selves.*" - Fred Rogers

The investigation into the relationship between *Microcystis* phylogenetic distances and viral infection profiles within Lake Erie's cyanobacterial harmful algal bloom (cHAB) ecosystem has provided valuable insights into the intricate dynamics of host-virus interactions. This study reveals a statistically significant positive correlation between *Microcystis* isolate phylogenetic distances and infection profile distances. This correlation suggests that shared genetic and physiological characteristics among closely related *Microcystis* isolates may influence their susceptibility or resistance to specific viruses, resulting in similar infection profiles. Importantly, this finding extends the understanding of phage-host interactions beyond broad taxonomic affiliations to finer levels of specificity within bacterial populations.

The role of temporal dynamics in shaping *Microcystis*-phage interactions within the western basin of Lake Erie emerges as a crucial theme in this study. The collection date was identified as the sole statistically significant predictor of *Microcystis* infection profiles, explaining approximately 28% of the variability in infection profiles between *Microcystis* strains. These findings highlight the coevolutionary arms race between *Microcystis* and its predicted phages, where temporal changes in infection profiles may reflect ongoing adaptations and responses to environmental conditions and viral pressures.

While the study explored various environmental variables, such as sampling site, mcy cassette genotype, and several physicochemical parameters, their limited influence on *Microcystis* infection profiles in this specific dataset suggests that other unexamined

factors may be at play. These unexplored variables could potentially contribute to the unexplained variation in infection profiles. Therefore, it is essential to adopt a holistic approach in future investigations, considering a broader range of factors, including biotic and abiotic parameters, microbial interactions, and ecological contexts, to gain a more comprehensive understanding of the complex dynamics governing *Microcystis* strain diversity in Lake Erie.

Future research can delve deeper into the mechanisms underlying the observed correlation between *Microcystis* phylogenetic distances and viral infection profiles. *What specific genetic and physiological characteristics drive susceptibility or resistance to viruses within closely related Microcystis isolates?* Expanding the investigation into the temporal dynamics of *Microcystis*-phage interactions can reveal the ongoing coevolutionary arms race between these organisms. *How do Microcystis strains and their predicted phages adapt and counter-adapt over time, and what role do environmental conditions play in shaping these adaptations?*

Investigating the broader ecological context of *Microcystis*-phage interactions within Lake Erie's cHAB ecosystem can provide a more holistic understanding of community interactions. *How do interactions with other microbes, such as bacteria and protists, influence the infection dynamics of Microcystis strains and their predicted phages?* Integrating functional genomic analyses can uncover the molecular mechanisms underlying virus-host interactions at a finer scale. *How do specific genes or gene clusters in Microcystis strains and their predicted phages contribute to the observed infection profiles? What are the functional consequences of these interactions for both host and virus?*

5.3.2 Expanding our understanding of phage-host interactions and their contributions to overall cHAB dynamics

This dissertation presented a specific case of a phage-host interaction involving vOTU 17663 and *Microcystis* MAG LE19-12.2. This interaction exhibited remarkable consistency over multiple bloom seasons and tracked particulate microcystin during blooms, despite LE19-12.2's lack of an mcy operon associated with microcystin production. The research suggests that the phage-host networks within cHABs are complex and interconnected, extending beyond cyanobacteria to encompass interactions with heterotrophic bacteria. These findings illuminate several promising avenues for future research, including further exploration of phage-host networks, investigation into the mechanisms governing host range expansion and contraction, examination of interactions with other bacterial taxa, and the expansion of the catalog of isolated phages and hosts from cHABs.

The isolation of host organisms and their interacting phages through culturing methods remains a critical avenue for future research. Culturing approaches allow for in-depth characterization of specific phage-host pairs, including their physiology, genetics, and ecological roles. The continued isolation of novel phages and hosts from cHABs can expand our understanding of microbial interactions in these ecosystems. *What insights can be gained from studying the coevolutionary dynamics of isolated phage-host pairs? How do these interactions evolve over time, and how does coevolution shape the genetic and functional traits of both hosts and phages within cHAB environments?*

5.4 References

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M., & Dick, G. J. (2014). Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. *Science*, *344*(6185), 757–760. https://doi.org/10.1126/science.1252229

Bastien, E. G., Cable, R. N., Zaman, L., Batterbee, C., Wing, A. J., & Duhaime, M. B. (2023). *Virus-Host Interactions Predictor (VHIP): Machine learning approach to resolve microbial virus-host interaction networks* (p. 2023.11.03.565433). bioRxiv. https://doi.org/10.1101/2023.11.03.565433

Berry, M. A., Davis, T. W., Cory, R. M., Duhaime, M. B., Johengen, T. H., Kling, G. W., Marino, J. A., Den Uyl, P. A., Gossiaux, D., Dick, G. J., & Denef, V. J. (2017). Cyanobacterial harmful algal blooms are a biological disturbance to Western Lake Erie bacterial

communities. *Environmental Microbiology*, *19*(3), 1149–1162. https://doi.org/10.1111/1462-2920.13640
Berry, M. A., White, J. D., Davis, T. W., Jain, S., Johengen, T. H., Dick, G. J., Sarnelle, O., & Denef, V. J. (2017). Are Oligotypes Meaningful Ecological and Phylogenetic Units? A Case Study of Microcystis in Freshwater Lakes. *Frontiers in Microbiology*, *8*. https://www.frontiersin.org/articles/10.3389/fmicb.2017.00365
Breitbart, M. (2012). Marine viruses: Truth or dare. *Annual Review of Marine Science*, *4*, 425–448. https://doi.org/10.1146/annurev-marine-120709-142805
Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. https://doi.org/10.1016/j.tim.2005.04.003
Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W., & Sullivan, M. B. (2016). Seasonal time bombs: Dominant temperate viruses affect Southern Ocean microbial dynamics. *The ISME Journal*, *10*(2), 437–449. https://doi.org/10.1038/ismej.2015.125
Cai, L., Feng, C., Xie, L., Xu, B., Wei, W., Jiao, N., & Zhang, R. (2022). Ecological dynamics and impacts of viruses in Chinese and global estuaries. *Water Research*, *226*, 119237. https://doi.org/10.1016/j.watres.2022.119237
Chow, C.-E. T., Sachdeva, R., Cram, J. A., Steele, J. A., Needham, D. M., Patel, A., Parada, A. E., & Fuhrman, J. A. (2013). Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *The ISME Journal*, *7*(12), Article 12. https://doi.org/10.1038/ismej.2013.122
Dart, E., Fuhrman, J. A., & Ahlgren, N. A. (2023). Diverse Marine T4-like Cyanophage Communities Are Primarily Comprised of Low-Abundance Species Including Species with Distinct Seasonal, Persistent, Occasional, or Sporadic Dynamics. *Viruses*, *15*(2), Article 2. https://doi.org/10.3390/v15020581
Deboutte, W., Beller, L., Yinda, C. K., Maes, P., de Graaf, D. C., & Matthijnssens, J. (2020). Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(19), 10511–10519. https://doi.org/10.1073/pnas.1921859117
Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., … Sullivan, M. B. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, *177*(5), 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040
Hevroni, G., Flores-Uribe, J., Béjà, O., & Philosof, A. (2020). Rising through the Ranks: Seasonal and Diel Patterns of Marine Viruses. *bioRxiv*, 2020.05.07.082883. https://doi.org/10.1101/2020.05.07.082883
Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H., Andreopoulos, B., Brewer, H. M., Glavina del Rio, T., Adkins, J. N., Paul, S., Sullivan, M. B., & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *The ISME Journal*, *14*(4), Article 4. https://doi.org/10.1038/s41396-019-0580-z
Kieft, K., Zhou, Z., Anderson, R. E., Buchan, A., Campbell, B. J., Hallam, S. J., Hess, M., Sullivan, M. B., Walsh, D. A., Roux, S., & Anantharaman, K. (2021). Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-23698-5
Kimura, S., Sako, Y., & Yoshida, T. (2013). Rapid microcystis cyanophage gene diversification revealed by long- and short-term genetic analyses of the tail sheath gene in a natural pond. *Applied and Environmental Microbiology*, *79*(8), 2789–2795. https://doi.org/10.1128/AEM.03751-12

Lin, S., Boegman, L., & Rao, Y. R. (2021). Characterizing spatial and temporal distributions of turbulent mixing and dissipation in Lake Erie. *Journal of Great Lakes Research*, *47*(1), 168–179. https://doi.org/10.1016/j.jglr.2020.11.014

Malki, K., Sawaya, N. A., Tisza, M. J., Coutinho, F. H., Rosario, K., Székely, A. J., & Breitbart, M. (2021). Spatial and Temporal Dynamics of Prokaryotic and Viral Community Assemblages in a Lotic System (Manatee Springs, Florida). *Applied and Environmental Microbiology*, *87*(18), e0064621. https://doi.org/10.1128/AEM.00646-21

Mankiewicz-Boczek, J., Jaskulska, A., Pawełczyk, J., Gągała, I., Serwecińska, L., & Dziadek, J. (2016). Cyanophages Infection of Microcystis Bloom in Lowland Dam Reservoir of Sulejów, Poland. *Microbial Ecology*, *71*(2), 315–325. https://doi.org/10.1007/s00248-015-0677-5

McKindles, K. M., Manes, M. A., DeMarco, J. R., McClure, A., McKay, R. M., Davis, T. W., & Bullerjahn, G. S. (2020). Dissolved Microcystin Release Coincident with Lysis of a Bloom Dominated by Microcystis spp. In Western Lake Erie Attributed to a Novel Cyanophage. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01397-20

Morimoto, D., Yoshida, N., Sasaki, A., Nakagawa, S., Sako, Y., & Yoshida, T. (2023). Ecological Dynamics of Broad- and Narrow-Host-Range Viruses Infecting the Bloom-Forming Toxic Cyanobacterium Microcystis aeruginosa. *Applied and Environmental Microbiology*, *89*(2), e02111-22. https://doi.org/10.1128/aem.02111-22

Naknaen, A., Suttinun, O., Surachat, K., Khan, E., & Pomwised, R. (2021). A Novel Jumbo Phage PhiMa05 Inhibits Harmful Microcystis sp. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.660351

Ou, T., Gao, X.-C., Li, S.-H., & Zhang, Q.-Y. (2015). Genome analysis and gene nblA identification of Microcystis aeruginosa myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *Journal of General Virology*, *96*(12), 3681–3697. https://doi.org/10.1099/jgv.0.000290

Pound, H. L., & Wilhelm, S. W. (2020). Tracing the active genetic diversity of Microcystis and Microcystis phage through a temporal survey of Taihu. *PLoS ONE*, *15*(12), e0244482. https://doi.org/10.1371/journal.pone.0244482

Qian et al. (2022). *Viruses | Free Full-Text | A Novel Freshwater Cyanophage, Mae-Yong924-1, Reveals a New Family*. https://www-mdpi-com.proxy.lib.umich.edu/1999-4915/14/2/283

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: Mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. https://doi.org/10.7717/peerj.985

Roux, S., Hallam, S. J., Woyke, T., & Sullivan, M. B. (2015). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife*, *4*, e08490–e08490. https://doi.org/10.7554/eLife.08490

Smith, D. J., Tan, J. Y., Powers, M. A., Lin, X. N., Davis, T. W., & Dick, G. J. (2021). Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. *Environmental Microbiology*, *23*(6), 3020–3036. https://doi.org/10.1111/1462-2920.15514

Takashima, Y., Yoshida, T., Yoshida, M., Shirai, Y., Tomaru, Y., Takao, Y., Hiroishi, S., & Nagasaki, K. (2007). Development and Application of Quantitative Detection of Cyanophages Phylogenetically Related to Cyanophage Ma-LMM01 Infecting *Microcystis aeruginosa* in Fresh Water. *Microbes and Environments*, *22*(3), 207–213. https://doi.org/10.1264/jsme2.22.207

Tucker, S., & Pollard, P. (2005). Identification of Cyanophage Ma-LBP and Infection of the Cyanobacterium Microcystis aeruginosa from an Australian Subtropical Lake by the Virus. *Applied and Environmental Microbiology*, *71*(2), 629–635. https://doi.org/10.1128/AEM.71.2.629-635.2005

Wang, F., Li, D., Cai, R., Pan, L., Zhou, Q., Liu, W., Qian, M., & Tong, Y. (2022). A Novel Freshwater Cyanophage Mae-Yong1326-1 Infecting Bloom-Forming Cyanobacterium Microcystis aeruginosa. *Viruses*, *14*(9), 2051. https://doi.org/10.3390/v14092051

Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., Russell, J. M., Bevins, C. L., Adams, L. G., Tsolis, R. M., Roth, J. R., & Bäumler, A. J. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, *467*(7314), Article 7314. https://doi.org/10.1038/nature09415

Yancey, C. E., Kiledal, E. A., Chaganti, S. R., Denef, V. J., Errera, R. M., Evans, J. T., Hart, L. N., Isailovic, D., James, W. S., Kharbush, J. J., Kimbrel, J. A., Li, W., Mayali, X., Nitschky, H., Polik, C. A., Powers, M. A., Premathilaka, S. H., Rappuhn, N. A., Reitz, L. A., … Dick, G. J. (2023). The Western Lake Erie culture collection: A promising resource for evaluating the physiological and genetic diversity of Microcystis and its associated microbiome. *Harmful Algae*, *126*, 102440. https://doi.org/10.1016/j.hal.2023.102440

Yang, F., Jin, H., Wang, X.-Q., Li, Q., Zhang, J.-T., Cui, N., Jiang, Y.-L., Chen, Y., Wu, Q.-F., Zhou, C.-Z., & Li, W.-F. (2020). Genomic Analysis of Mic1 Reveals a Novel Freshwater Long-Tailed Cyanophage. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.00484

Yoshida, M., Yoshida, T., Kashima, A., Takashima, Y., Hosoda, N., Nagasaki, K., & Hiroishi, S. (2008). Ecological dynamics of the toxic bloom-forming cyanobacterium Microcystis aeruginosa and its cyanophages in freshwater. *Applied and Environmental Microbiology*, *74*(10), 3269–3273. https://doi.org/10.1128/AEM.02240-07

Zhang, D., He, Y., & Gin, K. Y.-H. (2022). Genomic Characterization of a Novel Freshwater Cyanophage Reveals a New Lineage of Cyanopodovirus. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.768868

Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., Waldbauer, J. R., & Coleman, M. L. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology*, *18*(1), Article 1. https://doi.org/10.1038/s41579-019-0270-x