ECOLOGICAL CHARACTERIZATION OF VIRAL COMMUNITIES IN CLIMATE-SENSITIVE ENVIRONMENTS USING 'OMICS APPROACHES

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Elizabeth R. Denison
December 2024

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Wilhelm, for giving me the opportunity to work in his lab and for his patience, support, and guidance over the years. I would also like to acknowledge my committee, Drs. Buchan, Talmy, and Weston, for their feedback on my dissertation. I am grateful to our colleagues who shared their sequencing data with us. Without their collaboration, much of the work in this dissertation would not have been possible. I am also grateful to my cohort, especially Gwen and Jake, for their friendship throughout our time here. Finally, I am thankful for the encouragement of my family while pursuing this degree.

ABSTRACT

All of Earth's ecosystems are affected by climate change, and many Northern ecosystems are particularly vulnerable to ecosystem disturbance. Viruses play key roles in ecosystem dynamics; viral infection can regulate microbial abundances and shape community composition, influence microbial evolution and metabolism, dictate the flux of carbon and nutrients, and more. Despite their recognized roles, our understanding of viral responses to global change is severely limited. Metagenomic approaches have expanded our knowledge of the diversity, distribution, and metabolic potential of uncultured viruses and are useful tools for interrogating viral assemblages across diverse environments. This dissertation uses metagenomic and metatranscriptomic approaches to characterize viral communities within the context of two Northern ecosystems sensitive to climate change. Chapter Two concerns the under-ice microbiome in Lake Erie, a seasonally frozen freshwater lake experiencing climate-driven declines in ice cover. Chapters Three and Four examine the microbiome of *Sphagnum* mosses, a keystone plant species in one of Earth's largest terrestrial carbon sinks. Chapter Two presents a metatranscriptomic study comparing viral activity between an ice-covered and ice-free winter in Lake Erie, where we identified significant differences in active viral community composition and diversity. Chapter Three examined paired metagenomes and metatranscriptomes to define viral members of the core *Sphagnum* microbiome, providing evidence for persistent viral entities in association with *Sphagnum* communities. Chapter Four leveraged metagenomes and metatranscriptomes from warmed and ambient conditions to explore the effect of warming on viral community composition and activity in the moss microbiome. Here we found that viral communities were significantly different between warming conditions, but that differences may be dependent upon the type of warming condition and severity (artificial or natural geothermal warming). The chapters presented below span two different environmental systems, but they are linked by the interest in understanding microbial communities in Northern ecosystems disturbed by rapidly changing climate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ATTACHMENTS

Attachment 1

Sheet 1. Supplemental Table 3.1: Sample metadata and library sizes

Sheet 2. Supplemental Table 3.2: Assembly statistics

Sheet 3. Supplemental Table 3.3: Count and relative abundance of microbial taxa

Sheet 4. Supplemental Table 3.4: Coverage values for individual microbial contigs

Sheet 5. Supplemental Table 3.5: Viral contigs identified from the metagenomes

Sheet 6. Supplemental Table 3.6: Bray-Curtis dissimilarity matrix based on viral contig coverage in the metagenomes

Sheet 7. Supplemental Table 3.7: Nucleotide alignment table comparing the core metagenomic viral contigs

Sheet 8. Supplemental Table 3.8: Protein alignment table comparing the core metagenomic viral contigs to Ralstonia phage Raharianne.

Sheet 9. Supplemental Table 3.9: Nucleotide alignment table of the core contigs to the top hit RefSeq genome

Sheet 10. Supplemental Table 3.10: Quality assessment of the NCLDV MAGs

Sheet 11. Supplemental Table 3.11: Coverage tables for the NCLDV MAGs

Sheet 12. Supplemental Table 3.12: Protein alignment table of the NCLDV MAG proteins to RefSeq Virus

Sheet 13. Supplemental Table 3.13: Quality of metatranscriptomic virus contigs with coverage tables

Sheet 14. Supplemental Table 3.14: Bray-Curtis dissimilarity matrix based on metatranscriptomic viral contig coverage

Attachment 2

Sheet 1. Supplemental Table 4.1: Sample metadata

Sheet 2. Supplemental Table 4.2: Assembly statistics

Sheet 3. Supplemental Table 4.3: Viral contigs identified from the metagenome coassembly

Sheet 4. Supplemental Table 4.4: NCLDV major capsid protein sequences taxonomic estimate

Sheet 5. Supplemental Table 4.5: Quality summary of the NCLDV MAGs

CHAPTER ONE: INTRODUCTION

Viral ecology in the context of changing climate

Many of Earth's ecosystems are experiencing unprecedented and rapid changes in climate, where Northern systems are especially affected (Parmesan, 2006; Hoegh-Guldberg and Bruno, 2010). Climate change is apparent through losses in biodiversity and greater species extinction rates, habitat destruction, and the loss of ecosystem services (Mooney *et al*., 2009). The effect of climate on microorganisms may not be as visible, but microbial communities are impacted nonetheless, including (but not limited to) shifts in microbial composition (Smol *et al*., 2005), distribution (Southward *et al*., 1995), and activity (Sturm *et al*., 2005). These shifts can have cascading effects on ecosystem structure and biogeochemical cycles (Sturm *et al*., 2005; Graham *et al*., 2011). Recent efforts have been made to integrate microorganisms into climate change assessments and predictions of future change (Treseder *et al*., 2012; Yang *et al*., 2023). However, viruses have thus far largely been excluded from such efforts.

Viral infection can regulate microbial abundances through lysis and indirectly foster microbial diversity and co-existence by lysing populations of fast-growing, abundant organisms (the 'kill the winner' hypothesis) (Thingstad and Lignell, 1997). The lysis of infected hosts and subsequent release of cellular material greatly impacts carbon and nutrient cycling. The released organic matter and nutrients are utilized by bacteria for growth and recycled at lower trophic levels as opposed to moving up the food web (the 'viral shunt') (Wilhelm and Suttle, 1999). Additionally, in marine systems infected cells can form aggregates that sink the sea floor and contribute to ocean carbon sequestration, a process termed the 'viral shuttle' (Weinbauer, 2004).

Viruses impact more than microbial mortality rates; virus-host interactions represent an important driver of microbial evolution. Viral infection can introduce new genetic information to cells that can lead to diversification and new metabolisms (Marston *et al*., 2012). Viral infection can also (temporarily) alter host metabolism through the expression of virus-encoded auxiliary metabolic genes during infection (Rohwer and Thurber, 2009). A commonly referenced example is the photosystem genes encoded by cyanophage that are expressed during infection (Lindell *et al*., 2005). Similarly, the metabolism of infected cells differs from non-infected cells (Forterre,

2013), and the physiological remodeling of infected cells by viruses likely produces different ecosystem footprints and alters microbial community dynamics (Howard-Varona *et al*., 2020).

Syntheses regarding viral ecology in the context of global change are scarce (Danovaro *et al*., 2011; Jansson and Wu, 2022; Wieczynski *et al*., 2023), highlighting both the limited amount of current knowledge but also the interest in addressing the existing knowledge gap. While not extensively summarized in this introduction, select trends from the literature compiled by Danovaro *et al*., Jansson *et al*., and Wieczynski *et al*. are as follows:

1. In general, higher temperatures increase microbial growth rates as well as virus-host contact rates, both of which can increase viral production. Climate warming may bolster viral production this way. However, viral particles are generally less stable at higher temperatures, and warming may enhance decay rates at the system-level. In soils, warming may alternatively decrease host activity and virus-host contact rates due to decreased soil moisture.
2. In both aquatic and soil systems, temperature change may determine whether certain viral lifestyles are advantageous. In other words, warming may influence the prevalence of lytic phage lifestyles, where viruses replicate and immediately lyse the host, or lysogenic lifestyles, where viruses temporarily integrate into host genomes without causing mortality.
3. Changes in pH are thought to indirectly affects viruses by affecting host growth, as well as more directly through particle stability and infectivity.

Many of the observed trends are accompanied by caveats where studies have reported the complete opposite effect. Overall, this highlights that 1) climate responses may not be universal but instead system- and organism-specific, and 2) that system-specific studies are still informative for understanding climate change and may also be useful in elucidating more general ecosystem responses to climate when considered with the existing literature.

Case Study 1: The relationship between Lake Erie's under-ice microbiome, declining ice cover, and viral activity

Climate warming can impact freshwater lakes in a variety of ways, including the availability of fresh water, quality of drinking water, and habitat availability for aquatic life (Schindler, 2001; Woolway *et al*., 2022). This section will focus on the influence of warming on lake ice phenology and the 'under-ice microbiome' (Bertilsson *et al*., 2013) as it is most relevant to the work presented in Chapter Two. Air temperature has significant influence over ice phenology (Williams and Stefan, 2006). There is an associated decrease in period of ice cover, ice thickness, and greater interannual variability in ice-on and ice-off dates with increased air temperature (Magnuson *et al*., 2000; Imrit and Sharma, 2021; Sharma *et al*., 2021). Furthermore, recent projections warn that some seasonally-frozen lakes may transition to ice-free winters over the course of this century (Sharma *et al*., 2021). Historical lake warming data has shown that while there is a trend in lake warming globally, lake warming rates were not uniform even within geographic regions (O'Reilly *et al*., 2015). Instead, variable climate factors, like air temperature, interact with individual lake properties such as elevation, surface area, and depth. This is one example of the heterogenous responses of lakes to changing climate and, while global syntheses are equally important, highlights the rationale behind studying individual lakes.

The seasonal ice-covered period was once thought to be biologically dormant (Sommer *et al*., 2012). However, efforts to illuminate the 'black box' that is wintertime ecology have revealed active microbial communities underneath the ice, including phytoplankton and grazer communities (Bertilsson *et al*., 2013; Hampton *et al*., 2017). Across many lakes, microbial biomass is often lower in the winter than the summer (Hampton *et al*., 2017). For phototrophic microorganisms, this is in part attributed to a decrease in light availability beneath ice and snow cover (e.g., (Tulonen *et al*., 1994)). In Lake Erie, however, the under-ice microbiome is characterized by dense blooms of diatoms that form directly under the ice (Twiss *et al*., 2012), a position in the water column where light is most available especially since thick snow-cover is uncommon over Lake Erie (Bolsenga and Vanderploeg, 1992). While comparisons between high-

and low-ice years are scarce, declines in abundance and differences in gene expression of bloom-forming diatoms have been reported from low ice cover samples (Beall *et al*., 2016; Zepernick *et al*., 2024).

In various water bodies, viral abundances are often lowest in the winter, when both temperature and productivity are also at their lowest (Jiang and Paul, 1994; Weinbauer *et al*., 1995; Liu *et al*., 2006). Lower viral abundances in the winter, irrespective of ice cover, is sometimes attributed to lower light availability and lower temperatures, both variables known to influence viral production (Mojica and Brussaard, 2014). In addition to light availability, changes in ice cover can influence exposure to ultraviolet radiation, which can inactivate viral particles (Wilhelm *et al*., 1998; Madan *et al*., 2005). Temperate phage may also take refuge as lysogens during the winter, as there are some reports of lysogeny being common in winter months and under ice (McDaniel *et al*., 2002; Laybourn-Parry *et al*., 2007; Maurice *et al*., 2010). During the summer in Lake Erie, phage can act as a significant source of bacterial mortality (Wilhelm and Smith, 2000), but less is known regarding viral activity during the winter. Gene fragments of putative phage (cyanophage) and RNA viruses (of eukaryotes) have been detected in the winter (Matteson *et al*., 2011; Edgar *et al*., 2016). No studies have assessed the relationship between ice cover extent and viral activity in Lake Erie, or in freshwater lakes in generally. Given this scarcity of knowledge, there are some basic questions that need addressing, including: 1) are viral communities active during the winter in Lake Erie and 2) are there differences in viral activity based on ice cover-extent?

Case Study 2: Northern peatlands and viruses in the *Sphagnum* moss microbiome

Peatlands cover only about 3% of Earth's surface, but account for an estimated 30% of the total amount of terrestrial carbon (C) (Gorham, 1991; Yu *et al*., 2010). Over the course of thousands of years, low rates of decomposition relative to net primary production have led to carbon accumulation in the form of peat, or partially decayed organic matter (Bridgham *et al*., 2006; Gorham *et al*., 2012). As global temperatures rise, peatland C stores are at risk of being released back into the atmosphere in the form of greenhouse gases, with subsequent predicted ecological and economic impacts (Dorrepaal *et al*., 2009; Frolking *et al*., 2011). In Northern peatlands, *Sphagnum* mosses

are the dominant plant genus, where they promote an environment favorable to carbon sequestration (Turetsky, 2003; Turetsky *et al*., 2012). Very little is known regarding viral diversity and activity in this system.

This section provides an overview of the mechanisms of C sequestration by peatlands, the link between C storage, *Sphagnum* mosses, and their microbiomes, and how *Sphagnum* microbiomes are impacted by changing climate. Finally, this section will summarize what has been reported regarding viruses in *Sphagnum*-dominated peatlands.

Peatland carbon sequestration

The majority of peatland C stores are within Northern peatlands (boreal and subarctic), with an estimated 550 gigatons stored (Yu *et al*., 2010). Northern peatlands primarily formed during the early Holocene, when increased sunlight and warmer summers coupled with colder winters resulted in increased C capture (plant production) during the summers and reduced peat decomposition during the winter (Yu *et al*., 2010). Peatlands continued to develop over the Holocene, where globally the mean C accumulation rate was estimated to be 142 ± 7 TgC/yr (Gallego-Sala *et al*., 2018). Peat accumulates because the rate of carbon fixation (net primary production) is greater than the rate of carbon degradation by microbial heterotrophic metabolism (Clymo, 1987). Peatland habitats are classified as ombrotrophic or minerotrophic, referred to as bogs and fens, respectively (Clymo, 1987). Bogs are rainwater-fed only and are in general more acidic and nutrient-limited than fens; bogs are dominated by *Sphagnum* mosses and generally sequester more carbon (Granath *et al*., 2010; Rochefort *et al*., 2012). Here, we focus on *Sphagnum*-dominated ombrotrophic bogs and will refer to them as 'peatlands' for brevity.

The harsh conditions in Northern peatlands suppress organic matter degradation, namely the acidity (pH 3-4), anoxia, nutrient limitation, and low mean temperatures (Moore and Basiliko, 2006). The primary mechanism that drives peatland acidification is debated, but contributing factors include the release of protons via cation exchange by *Sphagnum* spp. (Clymo, 1963), the gradual release of humic acids from decaying peat (Siegel *et al*., 2006), and aspects of bog hydrology that prevents cation-rich groundwater from flowing to the surface (*i.e*., acid neutralization capacity is low) (Soudzilovskaia *et*

al., 2010). Anoxia is brought on by high water table levels (*i.e*., water-logged conditions) (Zhong *et al*., 2020).

Sphagnum mosses are peatland ecosystem engineers

Carbon sequestration is directly linked to the ecological dominance and productivity of *Sphagnum* mosses, or peat mosses. *Sphagnum* mosses are 'ecosystem engineers' in Northern peatlands because their unique physiology and biochemical properties shape the ecosystem as a whole (Norby *et al*., 2019). Despite the lack of a root system, *Sphagnum* mosses can outcompete vascular plants in nutrient-limited bogs because of their efficient uptake of nutrients when deposited from the atmosphere (van Breemen, 1995; Fritz *et al*., 2014). *Sphagnum* microbial symbionts may also promote survival in low-nutrient conditions (presented in further detail below). Additionally, the chemical properties of *Sphagnum* engineer peatland biochemistry and facilitate C storage (Turetsky *et al*., 2008; Straková *et al*., 2010). *Sphagnum* mosses have been proposed as climate change indicator species due to their sensitivity to climate change and their importance in peatland ecosystem health (Whinam and Copson, 2006).

The Sphagnum microbiome

Sphagnum harbor a diverse consortium of microorganisms (Kostka *et al*., 2016). Some taxonomic groups have better defined roles and interactions than others. For example, *Sphagnum* mosses, like many other plants (Kraiser *et al*., 2011), have a symbiotic relationship with diazotrophic bacteria and receive plant-available nitrogen (ammonium) from them (Berg *et al*., 2013). The *Sphagnum*-diazotroph interaction is especially important because peatlands can be nitrogen-limited (Aerts *et al*., 1992). Cyanobacterial diazotrophs are commonly found in *Sphagnum*-associated communities (Kolton *et al*., 2022), but alphaproteobacterial diazotrophs have also been described as the dominant nitrogen-fixing taxa, at least under non-warming conditions (Bragina *et al*., 2012b; Bragina *et al*., 2013a; Carrell *et al*., 2019). Furthermore, using *nifH* amplicon sequencing, Bragina *et al*. found that specific alphaproteobacterial diazotrophs were present in the *Sphagnum* sporophyte (Bragina *et al*., 2013a). This important functional guild may be vertically transmitted to progeny and 'maintained' in the moss microbiome. Alphaproteobacterial diazotrophic methanotrophs can provide both nitrogen and carbon

in the form of CO_2 to the plant. In addition to generating CO_2 that can be used for *Sphagnum* growth (Raghoebarsing *et al*., 2005; Larmola *et al*., 2014), methanotrophic symbionts also act as 'methane filters' and can contribute to lowering greenhouse gas emissions from peatlands (Kip *et al*., 2010; Larmola *et al*., 2010).

Sphagnum microbiome recruitment and assembly is not well understood compared to vascular plants (Santoyo, 2022). However, as referenced above, some diazotrophic taxa are thought to have stable associations with *Sphagnum* (meaning stable over generations) because they are vertically transmitted (Bragina *et al*., 2012a; Bragina *et al*., 2013a; Bragina *et al*., 2013b). Other methanotrophic taxa are thought to be ubiquitous in bog water and can at times be acquired horizontally (here meaning from the surrounding environment) (Larmola *et al*., 2010; Putkinen *et al*., 2012). A study by Kolton *et al*. of over 200 *Sphagnum* samples across the United States revealed a 'core' microbiome that consisted of only 12 genera but a large portion of total relative microbial abundance (Kolton *et al*., 2022). While nitrogen-fixing taxa as a functional guild were common and abundant (ca. 15% of the community), no individual diazotrophic or methanotrophic genera were in the core. Furthermore, some *Sphagnum* bacterial symbionts are shared with other bog vegetation, indicative of broader peatland core microbiome (Bragina *et al*., 2015; Tveit *et al*., 2020).

Interactions between *Sphagnum* and protistan communities are not as well defined, but protists are thought to be key elements of food webs within the microbiome nonetheless. For example, phototrophic protists (*e.g*., microalgae, mixotrophic amoeba) can contribute to carbon capture directly via photosynthesis (Falkowski, 2012), and phototrophs make up a large portion of *Sphagnum*-associated protists communities (Hamard *et al*., 2021). Additionally, microalgal exudates can promote bacterial respiration in peatlands (Wyatt and Turetsky, 2015) and grazing protists influence microbial community composition through predation (Unrein *et al*., 2014).

Viral diversity and ecology in peatlands

This section includes a summary of the few studies that have characterized viral communities in peatlands, almost all of which are on viruses in the peat and/or porewater as opposed to those associated with vegetation. Peat samples have been shown to contain

approximately 10^6 to 10^8 viral particles per mL (Ballaud *et al*., 2016; Trubl *et al*., 2016), which is comparable to the 10^7 typical in surface seawater (Fuhrman, 1999). Metagenomic and metaviromic studies have primarily uncovered novel dsDNA virus sequences of putative bacteriophage (class *Caudoviricetes*) (Emerson *et al*., 2018; ter Horst *et al*., 2021; Sun *et al*., 2024). Sequences similar to single-stranded DNA viruses of bacteria (*Microviridae*) and eukaryotes (*Circoviridae*) have also been detected (Quaiser *et al*., 2015; Ballaud *et al*., 2016). The use of metatranscriptomes showed that a large portion of the putative viruses identified were transcriptionally active (56-58%), suggesting on-going infection in the peat active layer the time of sampling (Emerson *et al*., 2018; Sun *et al*., 2024). Although, Ballaud *et al*. (2016) observed that the virus to prokaryote ratio (VPR) was relatively low in their peat samples and attributed it to the generally lower metabolic rates and dormancy in peat. VPR has not been examined in communities associated with living *Sphagnum*.

Some peat viruses may have a wide distribution as phylogenetically related (putative) viruses could be detected in geographically separate peatlands and other soils (Emerson *et al*., 2018; ter Horst *et al*., 2021) as well as in different peatland habitats (i.e., palsas, bogs, and fens) (Sun *et al*., 2024). While the majority of peat viruses may be endemic to a site and/or ephemeral, some viral sequences are persistent over time (Emerson *et al*., 2018; Sun *et al*., 2024). Additionally, surface peat can experience a seasonal shift in viral particle abundance and a succession in viral community composition (Ballaud *et al*., 2016), but over annual time scales viral communities at various peat depths appear to be stable (Sun *et al*., 2024).

Metagenomic approaches have revealed that viral communities in peat are structured by depth (Emerson *et al*., 2018; ter Horst *et al*., 2021; Sun *et al*., 2024). The influence of depth on community composition may in part be driven by water content, as more “aquatic-like” viral sequences were abundant in the more water-logged surface peat (ter Horst *et al*., 2021). Emerson *et al*. similarly reported overall viral community composition to be correlated with soil moisture content (2021).

Comparing extracellular (virome) and cell-associated sequences in peat over seasons (spring to summer) combined with particle counts suggested a seasonal lysogenic to lytic

switch (Ballaud *et al*., 2016). The authors hypothesized that the switch was due to an increase in microbial metabolism from spring to summer. Similarly, active (*i.e.*, replicating) viral populations had a high proportion of temperate phage and prophage in the samples in northern (arctic) peat soils in simulated winter conditions (anoxic and sub-freezing) (Trubl *et al*., 2021). In the peat, lysogeny could be a mechanism for viruses to persist when microbial metabolisms generally low, similar to mechanisms proposed in polar marine systems (Brum *et al*., 2016) and low productivity environments (Williamson *et al*., 2008).

Multiple virome studies have identified putative viral genomes with auxiliary metabolic genes that may be involved in carbon degradation pathways (Emerson *et al*., 2018; ter Horst *et al*., 2021; Trubl *et al*., 2021; Sun *et al*., 2024). Additionally, putative viruses have been bioinformatically linked to methanogenic and methanotrophic bacterial populations in peat (Emerson *et al*., 2018; Sun *et al*., 2024). Emerson *et al*. determined that putative viruses of methanogens could predict soil porewater C chemistry and may influence the dominate types of methane metabolism. Overall, viruses may impact carbon cycling in peatlands via viral-encoded metabolic genes and by potentially lysing bacteria involved in peatland methane cycling.

Only one published study has examined viral communities within the *Sphagnum* microbiome (Stough *et al*., 2018). Stough *et al*. reported diverse populations of actively replicating phage, NCLDV, and single-stranded RNA viruses inferred from viral marker genes detected in peat moss metatranscriptomes. Their findings suggested that viruses may have a role in structuring bacterial and protist communities and influence the recycling of carbon and nutrients in the peat moss microbiome. To conclude this section, although studies of viral populations derived from peatland soil and water samples inform our understanding of how viruses may modulate microbial metabolism in the belowground environment, studies of viruses associated with the living moss tissue may be more helpful in elucidating direct interactions between *Sphagnum* and its microbiome.

Peatland ecosystem vulnerability to climate warming

Peatland ecosystems are impacted by various anthropogenic activities, such as peat extraction, peatland drainage, and agriculture, as well as by increased acid and

nutrient deposition from acid rain (Turetsky and St. Louis, 2006). Whole-ecosystem experimental warming has shown an increasing net loss of C through with warming through the release of CO_2 and CH_4 (Hanson *et al*., 2020; Hopple *et al*., 2020). Climate warming may also increase the frequency of episodic disturbances like wildfires, which release large amounts of C from combustion and post-fire peatlands exhibit increased mineralization and decomposition (Turetsky *et al*., 2015). Warming also impacts *Sphagnum* ecophysiology. *Sphagnum* productivity and dominance have been shown to decline with warming (Jassey *et al*., 2013). *Sphagnum* decline is in part attributed to the decrease in water content and water table height (Bragazza, 2008; Norby *et al*., 2019), and to a decrease in light availability as a result of the warming-induced expansion of vascular plants (Norby *et al*., 2023).

Warming-induced declines in moss productivity have been associated with an increase in microbial enzymatic activity and respiration in peat soil, likely from increased oxygenation of the soil (Bragazza *et al*., 2016). As mentioned above, anoxia is typical feature of peatlands and limits microbial decomposition. Elevated temperatures create drier conditions, which promote the oxygenation of peat and rise in pH that increase microbial activity (Freeman *et al*., 2004; Brouns *et al*., 2014). Conversely, experimental warming showed that while there were large changes in microbial diversity, microbial community ecosystem function (here meaning decomposition of peat) was not different at a short time scale based on their methods (Roth *et al*., 2023).

Effects of warming on the Sphagnum microbiome

Experimental warming studies have shown that warming alters the *Sphagnum* microbiome and moss-microbiome interactions. A common response may be a decline in microbial diversity (bacterial and micro-eukaryotic), as has been observed in *Sphagnum*-associated communities (Carrell *et al*., 2019; Reczuga *et al*., 2020) and in peat soils (Kolton *et al*., 2019; Yang *et al*., 2021) under warming treatment. Although notably an increase in bacterial diversity has been observed in peat with warming (Roth *et al*., 2023). Microbial responses to warming are likely organism-specific, where specific taxonomic groups have varying levels of resilience to warming (Carrell *et al*., 2019) (Roth *et al*., 2023).

Diazotrophic communities in particular are disturbed by warming. For example, Carrell *et al*. observed that *Nostocaceae* (Cyanobacteria) became dominant with warming in the microbiome, while other cyanobacterial and alphaproteobacterial clades declined in relative abundance or were no longer detected with +9°C warming. The decline in prokaryotic and diazotrophic diversity was coupled with a decline in nitrogen fixation rates (~50% decrease). Therefore, warming-induced changes in microbial community diversity and structure may influence *Sphagnum* fitness in terms of nutrient acquisition. In addition to a decrease in nitrogen fixation rates in the microbiome (Carrell *et al*., 2019; Petro *et al*., 2023), warming impacts microbiome methane metabolism in peat soil and water (Kolton *et al*., 2019; Petro *et al*., 2023). Additionally, peat moss metabolites shape microbial communities (Sytiuk *et al*., 2022) and metabolite exchange is important in plant-symbiont interactions (Carrell *et al*., 2021), but the metabolome is sensitive to temperature (Jassey *et al*., 2013).

Micro-eukaryotic community structure and function are affected by warming, e.g. (Jassey *et al*., 2011; Jassey *et al*., 2013; Jassey *et al*., 2015; Reczuga *et al*., 2020). For example, experimental warming resulted in an increase in bacterial biomass that was partially explained by a decrease in biomass of protists predators. The authors hypothesized warming destabilized the food web, and that the loss of top predators ultimately stimulated microbial metabolism (organic matter breakdown) and increased biomass (Jassey *et al*., 2013). A separate study by Jassey *et al*. determined that a reduction (50% loss) in mixotrophic protists (which can be important source of CO_2 capture) led to an overall reduction in carbon net uptake by the microbiome, thus showing that alterations to moss microbiome communities does not only influence microbial processes but has the ability to influence overall peatland C capturing abilities (Jassey *et al*., 2015).

Very few efforts have been made to investigate viruses in experimentally warmed systems. Notably, ter Horst *et al*. found no difference in peat viral communities from the SPRUCE site after *ca*. two years of experimental warming (sampled peat depths ranged from 10 cm to 175 cm) (ter Horst *et al*., 2021). The authors suggested below ground viral communities may have a more delayed response to temperature. This is similar to other

work that showed that microbial communities (bacterial and protistan) at the surface layers were more strongly impact by warming (Jassey *et al*., 2013; Kolton *et al*., 2019). To our knowledge, no studies have examined the effect of warming on viral populations within the *Sphagnum* microbiome.

Applications of viral metagenomics in environmental microbiology

This section presents a brief overview over viral metagenomic methods as they are used throughout this dissertation. The earliest environmental viral metagenomic studies were published nearly twenty years ago, when Breitbart *et al*. described the viral fraction of seawater and marine sediment communities (Breitbart *et al*., 2002; Breitbart *et al*., 2004). Their work not only revealed the diversity and novelty of viral genes in environmental samples, but that shotgun metagenomics could feasibly recover the genomes of uncultured viruses. A common feature of viral metagenomes is that a large portion of the recovered sequences have no similarity to known sequences (Rosario and Breitbart, 2011). These unknowns constitute the "viral dark matter", which both highlights the incredible genetic diversity within viral communities but also obscures our understanding of the activity and ecology of viruses in nature. Nonetheless, insights into the diversity, evolution, biogeography, and functional potential of viruses have been inferred from metagenomic analyses (reviewed in: (Rosario and Breitbart, 2011; Brum and Sullivan, 2015; Zhang *et al*., 2019)).

Although 'viral metagenomics' typically refers to the sequencing of viral particles that have been enriched from environmental samples, viral communities can also be identified and studied from bulk metagenomes (*i.e*., untargeted sequencing of the entire community). In both cases, viral genes and genomes are often identified based on their similarity to known sequences. While viruses do not encode a single conserved gene, certain viral groups share "marker genes" that can be identified in sequencing data through homology searches. Viral marker genes can then be used to reconstruct phylogenies and document viral diversity, assess the distribution and abundance of viral clades, and infer virus-host interactions (*e.g*., (Hingamp *et al*., 2013; Sakowski *et al*., 2014; Wolf *et al*., 2020)). Certain viral markers can be used to investigate viral lifestyles, such as the prevalence and dynamics of temperate phage population (McDaniel *et al*.,

2008; Knowles *et al*., 2016). Additionally, DNA virus marker genes detected in metatranscriptomes can indicate actively replicating viruses within a sample and predict natural hosts based on abundance correlations (Moniruzzaman *et al*., 2017). In addition to conserved marker genes, viral genomes and genome fragments assembled from metagenomic libraries can be informative. Functional annotation can reveal novel viral genome organizations (Krishnamurthy *et al*., 2016) and auxiliary metabolism (Moniruzzaman *et al*., 2020). Genomic features such as CRISPR spacers, evidence of HGT or phage integration (*i.e*., genetic homology), oligonucleotide usage profiles, and GC content have all been used to predict hosts (Edwards *et al*., 2015; Coclet and Roux, 2021).

Despite the many discoveries made over the past two decades by leveraging metagenomics, there are still many challenges and considerations. As mentioned earlier, the majority of viral sequences are unknowns. The lack of database representation ultimately hinders both identification and annotation of viral sequences. Extensive horizontal gene transfer between microorganisms further complicates the interpretation of viral sequences, especially in bulk metagenomes where there is greater uncertainty if sequences represent "true" viruses. Advancements in *in silico* viral detection (Wu *et al*., 2024), sequencing approaches (Warwick-Dugdale *et al*., 2109), and community standards (Roux *et al*., 2017) are improving our ability to investigate viral populations in nature using metagenomics.

References

Aerts, R., Wallen, B., and Malmer, N. (1992). Growth-limiting nutrients in *Sphagnum*-dominated bogs subject to low and high atmospheric nitrogen supply. *Journal of ecology***,** 131-140.

Ballaud, F., Dufresne, A., Francez, A.J., Colombet, J., Sime-Ngando, T., and Quaiser, A. (2016). Dynamics of Viral Abundance and Diversity in a *Sphagnum*-Dominated Peatland: Temporal Fluctuations Prevail Over Habitat. *Frontiers in Microbiology* 6**,** 14.

Beall, B.F.N., Twiss, M.R., Smith, D.E., Oyserman, B.O., Rozmarynowycz, M.J., Binding, C.E., *et al*. (2016). Ice cover extent drives phytoplankton and bacterial community structure in a large north-temperate lake: implications for a warming climate. *Environmental Microbiology* 18(6)**,** 1704-1719.

Berg, A., Danielsson, Å., and Svensson, B.H. (2013). Transfer of fixed-N from N2-fixing cyanobacteria associated with the moss *Sphagnum* riparium results in enhanced growth of the moss. *Plant and Soil* 362(1)**,** 271-278.

Bertilsson, S., Burgin, A., Carey, C.C., Fey, S.B., Grossart, H.P., Grubisic, L.M., *et al*. (2013). The under-ice microbiome of seasonally frozen lakes. *Limnology and Oceanography* 58(6)**,** 1998-2012.

Bolsenga, S.J., and Vanderploeg, H.A. (Year). "Estimating photosynthetically available radiation into open and ice-covered freshwater lakes from surface characteristics; a high transmittance case study": Springer Netherlands), 95-104.

Bragazza, L. (2008). A climatic threshold triggers the die-off of peat mosses during an extreme heat wave. *Global Change Biology* 14(11)**,** 2688-2695.

Bragazza, L., Buttler, A., Robroek, B.J.M., Albrecht, R., Zaccone, C., Jassey, V.E.J., *et al*. (2016). Persistent high temperature and low precipitation reduce peat carbon accumulation. *Global Change Biology* 22(12)**,** 4114-4123.

Bragina, A., Berg, C., and Berg, G. (2015). The core microbiome bonds the Alpine bog vegetation to a transkingdom metacommunity. *Molecular Ecology* 24(18)**,** 4795-4807.

Bragina, A., Berg, C., Cardinale, M., Shcherbakov, A., Chebotar, V., and Berg, G. (2012a). *Sphagnum* mosses harbour highly specific bacterial diversity during their whole lifecycle. *The ISME Journal* 6(4)**,** 802-813.

Bragina, A., Berg, C., Müller, H., Moser, D., and Berg, G. (2013a). Insights into functional bacterial diversity and its effects on Alpine bog ecosystem functioning. *Scientific Reports* 3(1)**,** 1955.

Bragina, A., Cardinale, M., Berg, C., and Berg, G. (2013b). Vertical transmission explains the specific Burkholderia pattern in *Sphagnum* mosses at multi-geographic scale. *Frontiers in Microbiology* 4.

Bragina, A., Maier, S., Berg, C., Müller, H., Chobot, V., Hadacek, F., *et al*. (2012b). Similar Diversity of Alphaproteobacteria and Nitrogenase Gene Amplicons on Two Related *Sphagnum* Mosses. *Frontiers in Microbiology* 2.

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P., *et al*. (2004). Diversity and population structure of a near–shore marine–sediment viral community. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271(1539)**,** 565-574.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., *et al*. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences* 99(22)**,** 14250-14255.

Bridgham, S.D., Megonigal, J.P., Keller, J.K., Bliss, N.B., and Trettin, C. (2006). The carbon balance of North American wetlands. *Wetlands* 26(4)**,** 889-916.

Brouns, K., Verhoeven, J.T.A., and Hefting, M.M. (2014). Short period of oxygenation releases latch on peat decomposition. *Science of The Total Environment* 481**,** 61-68.

Brum, J.R., Hurwitz, B.L., Schofield, O., Ducklow, H.W., and Sullivan, M.B. (2016). Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *Isme Journal* 10(2)**,** 437-449.

Brum, J.R., and Sullivan, M.B. (2015). Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* 13(3)**,** 147-159.

Carrell, A.A., Kolton, M., Glass, J.B., Pelletier, D.A., Warren, M.J., Kostka, J.E., *et al*. (2019). Experimental warming alters the community composition, diversity, and N-2 fixation activity of peat moss (*Sphagnum* fallax) microbiomes. *Global Change Biology* 25(9)**,** 2993-3004.

Carrell, A.A., Veličković, D., Lawrence, T.J., Bowen, B.P., Louie, K.B., Carper, D.L., *et al*. (2021). Novel metabolic interactions and environmental conditions mediate the boreal peatmoss-cyanobacteria mutualism. *The ISME Journal* 16(4)**,** 1074-1085.

Clymo, R. (1987). The ecology of peatlands. *Science Progress (1933-)***,** 593-614.

Clymo, R.S. (1963). Ion Exchange in *Sphagnum* and its Relation to Bog Ecology. *Annals of Botany* 27(2)**,** 309-324.

Coclet, C., and Roux, S. (2021). Global overview and major challenges of host prediction methods for uncultivated phages. *Current Opinion in Virology* 49**,** 117-126.

Danovaro, R., Corinaldesi, C., Dell'Anno, A., Fuhrman, J.A., Middelburg, J.J., Noble, R.T., *et al*. (2011). Marine viruses and global climate change. *Fems Microbiology Reviews* 35(6)**,** 993-1034.

Dorrepaal, E., Toet, S., van Logtestijn, R.S.P., Swart, E., van de Weg, M.J., Callaghan, T.V., *et al*. (2009). Carbon respiration from subsurface peat accelerated by climate warming in the subarctic. *Nature* 460(7255)**,** 616-U679.

Edgar, R.E., Morris, P.F., Rozmarynowycz, M.J., D'Souza, N.A., Moniruzzaman, M., Bourbonniere, R.A., *et al*. (2016). Adaptations to photoautotrophy associated with seasonal ice cover in a large lake revealed by metatranscriptome analysis of a winter diatom bloom. *Journal of Great Lakes Research* 42(5)**,** 1007-1015.

Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2015). Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews* 40(2)**,** 258-272.

Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H.B., *et al*. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* 3(8)**,** 870-880.

Falkowski, P. (2012). Ocean Science: The power of plankton. *Nature* 483(7387)**,** S17-S20.

Forterre, P. (2013). The virocell concept and environmental microbiology. *Isme Journal* 7(2)**,** 233-236.

Freeman, C., Ostle, N., Fenner, N., and Kang, H. (2004). A regulatory role for phenol oxidase during decomposition in peatlands. *Soil Biology and Biochemistry* 36(10)**,** 1663-1667.

Fritz, C., Lamers, L.P.M., Riaz, M., van den Berg, L.J.L., and Elzenga, T.J.T.M. (2014). *Sphagnum* Mosses - Masters of Efficient N-Uptake while Avoiding Intoxication. *PLOS ONE* 9(1)**,** e79991.

Frolking, S., Talbot, J., Jones, M.C., Treat, C.C., Kauffman, J.B., Tuittila, E.-S., *et al*. (2011). Peatlands in the Earth's 21st century climate system. *Environmental Reviews* 19(NA)**,** 371-396.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399(6736)**,** 541-548.

Gallego-Sala, A.V., Charman, D.J., Brewer, S., Page, S.E., Prentice, I.C., Friedlingstein, P., *et al*. (2018). Latitudinal limits to the predicted increase of the peatland carbon sink with warming. *Nature climate change* 8(10)**,** 907-913.

Gorham, E. (1991). Northern peatlands: role in the carbon cycle and probable responses to climatic warming. *Ecological Applications* 1(2)**,** 182-195.

Gorham, E., Lehman, C., Dyke, A., Clymo, D., and Janssens, J. (2012). Long-term carbon sequestration in North American peatlands. *Quaternary Science Reviews* 58**,** 77-82.

Graham, D.E., Wallenstein, M.D., Vishnivetskaya, T.A., Waldrop, M.P., Phelps, T.J., Pfiffner, S.M., *et al*. (2011). Microbes in thawing permafrost: the unknown variable in the climate change equation. *The ISME Journal* 6(4)**,** 709-712.

Granath, G., Strengbom, J., and Rydin, H. (2010). Rapid ecosystem shifts in peatlands: linking plant physiology and succession. *Ecology* 91(10)**,** 3047-3056.

Hamard, S., Céréghino, R., Barret, M., Sytiuk, A., Lara, E., Dorrepaal, E., *et al*. (2021). Contribution of microbial photosynthesis to peatland carbon uptake along a latitudinal gradient. *Journal of Ecology* 109(9)**,** 3424-3441.

Hampton, S.E., Galloway, A.W.E., Powers, S.M., Ozersky, T., Woo, K.H., Batt, R.D., *et al*. (2017). Ecology under lake ice. *Ecology Letters* 20(1)**,** 98-111.

Hanson, P.J., Griffiths, N.A., Iversen, C.M., Norby, R.J., Sebestyen, S.D., Phillips, J.R., *et al*. (2020). Rapid Net Carbon Loss From a Whole-Ecosystem Warmed Peatland. *AGU Advances* 1(3)**,** e2020AV000163.

Hingamp, P., Grimsley, N., Acinas, S.G., Clerissi, C., Subirana, L., Poulain, J., *et al*. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal* 7(9)**,** 1678-1695.

Hoegh-Guldberg, O., and Bruno, J.F. (2010). The impact of climate change on the world's marine ecosystems. *Science* 328(5985)**,** 1523-1528.

Hopple, A.M., Wilson, R.M., Kolton, M., Zalman, C.A., Chanton, J.P., Kostka, J., *et al*. (2020). Massive peatland carbon banks vulnerable to rising temperatures. *Nature Communications* 11(1)**,** 2373.

Howard-Varona, C., Lindback, M.M., Bastien, G.E., Solonenko, N., Zayed, A.A., Jang, H., *et al*. (2020). Phage-specific metabolic reprogramming of virocells. *Isme Journal* 14(4)**,** 881-895.

Imrit, M.A., and Sharma, S. (2021). Climate change is contributing to faster rates of lake ice loss in lakes around the Northern Hemisphere. *Journal of Geophysical Research: Biogeosciences* 126(7)**,** e2020JG006134.

Jansson, J.K., and Wu, R.N. (2022). Soil viral diversity, ecology and climate change. *Nature Reviews Microbiology***,** 16.

Jassey, V.E., Chiapusio, G., Binet, P., Buttler, A., Laggoun-Défarge, F., Delarue, F., *et al*. (2013). Above- and belowground linkages in *Sphagnum* peatland: climate warming affects plant-microbial interactions. *Global Change Biology* 19(3)**,** 811-823.

Jassey, V.E.J., Gilbert, D., Binet, P., Toussaint, M.L., and Chiapusio, G. (2011). Effect of a temperature gradient on *Sphagnum* fallax and its associated living microbial

communities: a study under controlled conditions. *Canadian Journal of Microbiology* 57(3)**,** 226-235.

Jassey, V.E.J., Signarbieux, C., Hattenschwiler, S., Bragazza, L., Buttler, A., Delarue, F., *et al*. (2015). An unexpected role for mixotrophs in the response of peatland carbon cycling to climate warming. *Scientific Reports* 5**,** 10.

Jiang, S.C., and Paul, J.H. (1994). Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. *Marine Ecology Progress Series***,** 163-172.

Kip, N., van Winden, J.F., Pan, Y., Bodrossy, L., Reichart, G.-J., Smolders, A.J.P., *et al*. (2010). Global prevalence of methane oxidation by symbiotic bacteria in peat-moss ecosystems. *Nature Geoscience* 3(9)**,** 617-621.

Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobian-Guemes, A.G., *et al*. (2016). Lytic to temperate switching of viral communities. *Nature* 531(7595)**,** 466.

Kolton, M., Marks, A., Wilson, R.M., Chanton, J.P., and Kostka, J.E. (2019). Impact of Warming on Greenhouse Gas Production and Microbial Diversity in Anoxic Peat From a *Sphagnum*-Dominated Bog (Grand Rapids, Minnesota, United States). *Frontiers in Microbiology* 10.

Kolton, M., Weston, D.J., Mayali, X., Weber, P.K., McFarlane, K.J., Pett-Ridge, J., *et al*. (2022). Defining the *Sphagnum* Core Microbiome across the North American Continent Reveals a Central Role for Diazotrophic Methanotrophs in the Nitrogen and Carbon Cycles of Boreal Peatland Ecosystems. *Mbio* 13(1)**,** 17.

Kostka, J.E., Weston, D.J., Glass, J.B., Lilleskov, E.A., Shaw, A.J., and Turetsky, M.R. (2016). The *Sphagnum* microbiome: new insights from an ancient plant lineage. *New Phytologist* 211(1)**,** 57-64.

Kraiser, T., Gras, D.E., Gutiérrez, A.G., González, B., and Gutiérrez, R.A. (2011). A holistic view of nitrogen acquisition in plants. *Journal of Experimental Botany* 62(4)**,** 1455-1466.

Krishnamurthy, S.R., Janowski, A.B., Zhao, G.Y., Barouch, D., and Wang, D. (2016). Hyperexpansion of RNA Bacteriophage Diversity. *Plos Biology* 14(3)**,** 17.

Larmola, T., Leppänen, S.M., Tuittila, E.-S., Aarva, M., Merilä, P., Fritze, H., *et al*. (2014). Methanotrophy induces nitrogen fixation during peatland development. *Proceedings of the National Academy of Sciences* 111(2)**,** 734-739.

Larmola, T., Tuittila, E.-S., Tiirola, M., Nykänen, H., Martikainen, P.J., Yrjälä, K., *et al*. (2010). The role of *Sphagnum* mosses in the methane cycling of a boreal mire. *Ecology* 91(8)**,** 2356-2365.

Laybourn-Parry, J., Marshall, W.A., and Madan, N.J. (2007). Viral dynamics and patterns of lysogeny in saline Antarctic lakes. *Polar Biology* 30(3)**,** 351-358.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438(7064)**,** 86-89.

Liu, Y.M., Zhang, Q.Y., Yuan, X.P., Li, Z.Q., and Gui, J.F. (2006). Seasonal variation of virioplankton in a eutrophic shallow lake. *Hydrobiologia* 560**,** 323-334.

Madan, N.J., Marshall, W.A., and Laybourn-Parry, J. (2005). Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. *Freshwater Biology* 50(8)**,** 1291-1300.

Magnuson, J.J., Robertson, D.M., Benson, B.J., Wynne, R.H., Livingstone, D.M., Arai, T., *et al*. (2000). Historical Trends in Lake and River Ice Cover in the Northern Hemisphere. *Science* 289(5485)**,** 1743-1746.

Marston, M.F., Pierciey, F.J., Shepard, A., Gearin, G., Qi, J., Yandava, C., *et al*. (2012). Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences* 109(12)**,** 4544-4549.

Matteson, A.R., Loar, S.N., Bourbonniere, R.A., and Wilhelm, S.W. (2011). Molecular enumeration of an ecologically important cyanophage in a Laurentian Great Lake. *Applied and Environmental Microbiology* 77(19)**,** 6772-6779.

Maurice, C.F., Bouvier, T., Comte, J., Guillemette, F., and del Giorgio, P.A. (2010). Seasonal variations of phage life strategies and bacterial physiological states in three northern temperate lakes. *Environmental Microbiology* 12(3)**,** 628-641.

McDaniel, L., Breitbart, M., Mobberley, J., Long, A., Haynes, M., Rohwer, F., *et al*. (2008). Metagenomic Analysis of Lysogeny in Tampa Bay: Implications for Prophage Gene Expression. *PLOS ONE* 3(9)**,** e3263.

McDaniel, L., Houchin, L.A., Williamson, S.J., and Paul, J.H. (2002). Lysogeny in marine Synechococcus. *Nature* 415(6871)**,** 496-496.

Mojica, K.D.A., and Brussaard, C.P.D. (2014). Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *Fems Microbiology Ecology* 89(3)**,** 495-515.

Moniruzzaman, M., Martinez-Gutierrez, C.A., Weinheimer, A.R., and Aylward, F.O. (2020). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nature Communications* 11(1)**,** 11.

Moniruzzaman, M., Wurch, L.L., Alexander, H., Dyhrman, S.T., Gobler, C.J., and Wilhelm, S.W. (2017). Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nature Communications* 8**,** 10.

Mooney, H., Larigauderie, A., Cesario, M., Elmquist, T., Hoegh-Guldberg, O., Lavorel, S., *et al*. (2009). Biodiversity, climate change, and ecosystem services. *Current Opinion in Environmental Sustainability* 1(1)**,** 46-54.

Moore, T., and Basiliko, N. (2006). "Decomposition in boreal peatlands," in *Boreal peatland ecosystems*. Springer), 125-143.

Norby, R.J., Baxter, T., Živković, T., and Weston, D.J. (2023). Shading contributes to *Sphagnum* decline in response to warming. *Ecol Evol* 13(9)**,** e10542.

Norby, R.J., Childs, J., Hanson, P.J., and Warren, J.M. (2019). Rapid loss of an ecosystem engineer: *Sphagnum* decline in an experimentally warmed bog. *Ecology and Evolution* 9(22)**,** 12571-12585.

O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J., *et al*. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters* 42(24)**,** 10,773-710,781.

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* 37(Volume 37, 2006)**,** 637-669.

Petro, C., Carrell, A.A., Wilson, R.M., Duchesneau, K., Noble-Kuchera, S., Song, T., *et al*. (2023). Climate drivers alter nitrogen availability in surface peat and decouple N fixation from CH oxidation in the *Sphagnum* moss microbiome. *Global Change Biology* 29(11)**,** 3159-3176.

Putkinen, A., Larmola, T., Tuomivirta, T., Siljanen, H.M., Bodrossy, L., Tuittila, E.-S., *et al*. (2012). Water dispersal of methanotrophic bacteria maintains functional methane oxidation in *Sphagnum* mosses. *Frontiers in microbiology* 3**,** 15.

Quaiser, A., Dufresne, A., Ballaud, F., Roux, S., Zivanovic, Y., Colombet, J., *et al*. (2015). Diversity and comparative genomics of Microviridae in *Sphagnum*-dominated peatlands. *Frontiers in Microbiology* 6.

Raghoebarsing, A.A., Smolders, A.J.P., Schmid, M.C., Rijpstra, W.I.C., Wolters-Arts, M., Derksen, J., *et al*. (2005). Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature* 436(7054)**,** 1153-1156. d

Reczuga, M.K., Seppey, C.V.W., Mulot, M., Jassey, V.E.J., Buttler, A., Slowinska, S., *et al*. (2020). Assessing the responses of *Sphagnum* micro-eukaryotes to climate changes using high throughput sequencing. *Peerj* 8**,** 26.

Rochefort, L., Strack, M., Poulin, M., Price, J.S., Graf, M., Desrochers, A., *et al*. (2012). *Northern peatlands.* University of California Press: Berkeley, CA, USA.

Rohwer, F., and Thurber, R.V. (2009). Viruses manipulate the marine environment. *Nature* 459(7244)**,** 207-212.

Rosario, K., and Breitbart, M. (2011). Exploring the viral world through metagenomics. *Current Opinion in Virology* 1(4)**,** 289-297.

Roth, S.W., Griffiths, N.A., Kolka, R.K., Oleheiser, K.C., Carrell, A.A., Klingeman, D.M., *et al*. (2023). Elevated temperature alters microbial communities, but not decomposition rates, during 3 years of in situ peat decomposition. *mSystems* 8(5)**,** e00337-00323.

Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., and Sullivan, M.B. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *Peerj* 5**,** 26.

Sakowski, E.G., Munsell, E.V., Hyatt, M., Kress, W., Williamson, S.J., Nasko, D.J., *et al*. (2014). Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. *Proceedings of the National Academy of Sciences* 111(44)**,** 15786-15791.

Santoyo, G. (2022). How plants recruit their microbiome? New insights into beneficial interactions. *Journal of advanced research* 40**,** 45-58.

Schindler, D.W. (2001). The cumulative effects of climate warming and other human stresses on Canadian freshwaters in the new millennium. *Canadian Journal of Fisheries and Aquatic Sciences* 58(1)**,** 18-29.

Sharma, S., Richardson, D.C., Woolway, R.I., Imrit, M.A., Bouffard, D., Blagrave, K., *et al*. (2021). Loss of ice cover, shifting phenology, and more extreme events in Northern Hemisphere lakes. *Journal of Geophysical Research: Biogeosciences* 126(10)**,** e2021JG006348.

Siegel, D.I., Glaser, P.H., So, J., and Janecky, D.R. (2006). The dynamic balance between organic acids and circumneutral groundwater in a large boreal peat basin. *Journal of Hydrology* 320(3)**,** 421-431.

Smol, J.P., Wolfe, A.P., Birks, H.J.B., Douglas, M.S., Jones, V.J., Korhola, A., *et al*. (2005). Climate-driven regime shifts in the biological communities of arctic lakes. *Proceedings of the National Academy of Sciences* 102(12)**,** 4397-4402.

Sommer, U., Adrian, R., Domis, L., Elser, J., Gaedke, U., Ibelings, B., *et al*. (2012). Beyond the Plankton Ecology Group (PEG) model: Mechanisms driving plankton succession. *Annual Review of Ecology, Evolution, and Systematics, Vol 43* 43**,** 429-448.

Soudzilovskaia, N.A., Cornelissen, J.H.C., During, H.J., van Logtestijn, R.S.P., Lang, S.I., and Aerts, R. (2010). Similar cation exchange capacities among bryophyte species refute a presumed mechanism of peatland acidification. *Ecology* 91(9)**,** 2716-2726.

Southward, A., Hawkins, S., and Burrows, M. (1995). Seventy years' observations of changes in distribution and abundance of zooplankton and intertidal organisms in

the western English Channel in relation to rising sea temperature. *Journal of thermal Biology* 20(1-2)**,** 127-155.

Stough, J.M.A., Kolton, M., Kostka, J.E., Weston, D.J., Pelletier, D.A., and Wilhelm, S.W. (2018). Diversity of Active Viral Infections within the *Sphagnum* Microbiome. *Applied and Environmental Microbiology* 84(23)**,** 16.

Straková, P., Anttila, J., Spetz, P., Kitunen, V., Tapanila, T., and Laiho, R. (2010). Litter quality and its response to water level drawdown in boreal peatlands at plant species and community level. *Plant and Soil* 335(1)**,** 501-520.

Sturm, M., Schimel, J., Michaelson, G., Welker, J.M., Oberbauer, S.F., Liston, G.E., *et al*. (2005). Winter biological processes could help convert arctic tundra to shrubland. *Bioscience* 55(1)**,** 17-26.

Sun, C.L., Pratama, A.A., Gazitúa, M.C., Cronin, D., McGivern, B.B., Wainaina, J.M., *et al*. (2024). Virus ecology and 7-year temporal dynamics across a permafrost thaw gradient. *Environmental Microbiology* 26(8)**,** e16665.

Sytiuk, A., Céréghino, R., Hamard, S., Delarue, F., Guittet, A., Barel, J.M., *et al*. (2022). Predicting the structure and functions of peatland microbial communities from *Sphagnum* phylogeny, anatomical and morphological traits and metabolites. *Journal of Ecology* 110(1)**,** 80-96..

ter Horst, A.M., Santos-Medellin, C., Sorensen, J.W., Zinke, L.A., Wilson, R.M., Johnston, E.R., *et al*. (2021). Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations (vol 9, 233, 2021). *Microbiome* 9(1).

Thingstad, T.F., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology* 13(1)**,** 19-27.

Treseder, K.K., Balser, T.C., Bradford, M.A., Brodie, E.L., Dubinsky, E.A., Eviner, V.T., *et al*. (2012). Integrating microbial ecology into ecosystem models: challenges and priorities. *Biogeochemistry* 109(1)**,** 7-18.

Trubl, G., Kimbrel, J.A., Liquet-Gonzalez, J., Nuccio, E.E., Weber, P.K., Pett-Ridge, J., *et al*. (2021). Active virus-host interactions at sub-freezing temperatures in Arctic peat soil. *Microbiome* 9(1)**,** 208.

Trubl, G., Solonenko, N., Chittick, L., Solonenko, S., Rich, V., and Sullivan, M. (2016). Optimization of viral resuspension methods for carbon-rich soils along a permafrost thaw gradient. *Peerj* 4(e1999).

Tulonen, T., Kankaala, P., Ojala, A., and Arvola, L. (1994). Factors controlling production of phytoplankton and bacteria under ice in a humic, boreal lake. *Journal of Plankton Research* 16(10)**,** 1411-1432.

Turetsky, M.R. (2003). The role of bryophytes in carbon and nitrogen cycling. *The Bryologist* 106(3)**,** 395-409.

Turetsky, M.R., Benscoter, B., Page, S., Rein, G., van der Werf, G.R., and Watts, A. (2015). Global vulnerability of peatlands to fire and carbon loss. *Nature Geoscience* 8(1)**,** 11-14.

Turetsky, M.R., Bond-Lamberty, B., Euskirchen, E., Talbot, J., Frolking, S., McGuire, A.D., *et al*. (2012). The resilience and functional role of moss in boreal and arctic ecosystems. *New Phytologist* 196(1)**,** 49-67.

Turetsky, M.R., Crow, S.E., Evans, R.J., Vitt, D.H., and Wieder, R.K. (2008). Trade-Offs in Resource Allocation among Moss Species Control Decomposition in Boreal Peatlands. *Journal of Ecology* 96(6)**,** 1297-1305.

Turetsky, M.R., and St. Louis, V.L. (2006). "Disturbance in boreal peatlands," in *Boreal peatland ecosystems*. Springer), 359-379.

Tveit, A.T., Kiss, A., Winkel, M., Horn, F., Hájek, T., Svenning, M.M., *et al*. (2020). Environmental patterns of brown moss- and *Sphagnum*-associated microbial communities. *Scientific Reports* 10(1)**,** 22412.

Twiss, M.R., McKay, R.M.L., Bourbonniere, R.A., Bullerjahn, G.S., Carrick, H.J., Smith, R.E.H., *et al*. (2012). Diatoms abound in ice-covered Lake Erie: An investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. *Journal of Great Lakes Research* 38(1)**,** 18-30.

Unrein, F., Gasol, J.M., Not, F., Forn, I., and Massana, R. (2014). Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *The ISME Journal* 8(1)**,** 164-176.

van Breemen, N. (1995). How *Sphagnum* bogs down other plants. *Trends in Ecology & Evolution* 10(7)**,** 270-275.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A., Allen, M., *et al*. (2109). Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *Peerj* 7(e6800). dWeinbauer, M.G. (2004). Ecology of prokaryotic viruses. *Fems Microbiology Reviews* 28(2)**,** 127-181.

Weinbauer, M.G., Fuks, D., Puskaric, S., and Peduzzi, P. (1995). Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the Northern Adriatic Sea. *Microbial Ecology* 30**,** 25-41.

Whinam, J., and Copson, G. (2006). *Sphagnum* moss: an indicator of climate change in the sub-Antarctic. *Polar Record* 42(1)**,** 43-49.

Wieczynski, D.J., Yoshimura, K.M., Denison, E.R., Geisen, S., DeBruyn, J.M., Shaw, A.J., *et al*. (2023). Viral infections likely mediate microbial controls on ecosystem responses to global warming. *FEMS Microbiology Ecology* 99(3)**,** fiad016.

Wilhelm, S.W., and Smith, R.E.H. (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. dWilhelm, S.W., and 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.

Wilhelm, S.W., Weinbauer, M.G., Suttle, C.A., and Jeffrey, W.H. (1998). The role of sunlight in the removal and repair of viruses in the sea. *Limnology and Oceanography* 43(4)**,** 586-592.

Williams, S.G., and Stefan, H.G. (2006). Modeling of Lake Ice Characteristics in North America Using Climate, Geography, and Lake Bathymetry. *Journal of Cold Regions Engineering* 20(4)**,** 140-167.

Williamson, S.J., Cary, S.C., Williamson, K.E., Helton, R.R., Bench, S.R., Winget, D., *et al*. (2008). Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *The ISME Journal* 2(11)**,** 1112-1121.

Wolf, Y.I., Silas, S., Wang, Y.J., Wu, S., Bocek, M., Kazlauskas, D., *et al*. (2020). Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nature Microbiology* 5(10)**,** 1262-+. doi: 10.1038/s41564-020-0755-4.

Woolway, R.I., Sharma, S., and Smol, J.P. (2022). Lakes in hot water: The impacts of a changing climate on aquatic ecosystems. *Bioscience* 72(11)**,** 1050-1061.

Wu, L.-Y., Wijesekara, Y., Piedade, G.J., Pappas, N., Brussaard, C.P.D., and Dutilh, B.E. (2024). Benchmarking bioinformatic virus identification tools using real-world metagenomic data across biomes. *Genome Biology* 25(1)**,** 97.

Wyatt, K.H., and Turetsky, M.R. (2015). Algae alleviate carbon limitation of heterotrophic bacteria in a boreal peatland. *Journal of Ecology* 103(5)**,** 1165-1171.

Yang, P.F., Spanier, N., Aldredge, P., Shahid, N., Coleman, A., Lyons, J., *et al*. (2023). Will free-living microbial community composition drive biogeochemical responses to global change? *Biogeochemistry* 162(3)**,** 285-307.

Yang, S., Liebner, S., Svenning, M.M., and Tveit, A.T. (2021). Decoupling of microbial community dynamics and functions in Arctic peat soil exposed to short term warming. *Molecular Ecology* 30(20)**,** 5094-5104.

Yu, Z.C., Loisel, J., Brosseau, D.P., Beilman, D.W., and Hunt, S.J. (2010). Global peatland dynamics since the Last Glacial Maximum. *Geophysical Research Letters* 37**,** 5.

Zepernick, B.N., Chase, E.E., Denison, E.R., Gilbert, N.E., Truchon, A.R., Frenken, T., *et al*. (2024). Declines in ice cover are accompanied by light limitation responses and community change in freshwater diatoms. *The ISME Journal* 18(1).

Zhang, Y.-Z., Chen, Y.-M., Wang, W., Qin, X.-C., and Holmes, E.C. (2019). Expanding the RNA Virosphere by Unbiased Metagenomics. *Annual Review of Virology* 6(Volume 6, 2019)**,** 119-139.

Zhong, Y., Jiang, M., and Middleton, B.A. (2020). Effects of water level alteration on carbon cycling in peatlands. *Ecosystem Health and Sustainability* 6(1)**,** 1806113.

CHAPTER TWO: METATRANSCRIPTOMIC ANALYSIS REVEALS DISSIMILARITY IN VIRAL COMMUNITY ACTIVITY BETWEEN AN ICE-FREE AND ICE-COVERED WINTER IN LAKE ERIE

Publication Note

A version of Chapter Two was published in a peer-reviewed journal:

Denison ER, Zepernick BN, McKay RML, Wilhelm SW. 2024. Metatranscriptomic analysis reveals dissimilarity in viral community activity between an ice-free and ice-covered winter in Lake Erie. mSystems 9:e00753-24. https://doi.org/10.1128/msystems.00753-24

All authors contributed to the project design. RNA extractions were performed by BNZ. BNZ provided the metatranscriptome assembly and gene predictions. Data analysis and interpretation of results were carried out by ERD. The manuscript was drafted by ERD. All authors reviewed and edited the final version of the manuscript.

Abstract

Winter is a relatively under-studied season in freshwater ecology. The paucity of wintertime surveys has led to a lack of knowledge regarding microbial community activity during the winter in Lake Erie, a North American Great Lake. Viruses shape microbial communities and regulate biogeochemical cycles by acting as top-down controls, yet very few efforts have been made to examine active virus populations during the winter in Lake Erie. Furthermore, climate change-driven declines in seasonal ice cover have been shown to influence microbial community structure, but no studies have compared viral community activity between different ice cover conditions. We surveyed surface water metatranscriptomes for viral hallmark genes as a proxy for active virus populations and compared activity metrics between ice-covered and ice-free conditions from two sampled winters. Transcriptionally active viral communities were detected in both winters, spanning diverse phylogenetic clades of putative bacteriophage (*Caudoviricetes*), giant viruses (*Nucleocytoviricota*, or NCLDV), and RNA viruses (*Orthornavirae*). However, viral community activity metrics revealed pronounced differences between the ice-covered and ice-free winter. Viral community composition was distinct between winters and viral hallmark gene richness was reduced in the ice-covered relative to the ice-free conditions. Additionally, the observed differences in viral communities correlated with microbial community activity metrics. Overall, these findings contribute to our understanding of the viral populations that are active during the winter in Lake Erie and suggest that viral community activity may be associated with ice cover extent.

Importance

As seasonal ice cover is projected to become increasingly rare on large temperate lakes, there is a need to understand how microbial communities might respond to changing ice cover conditions. Although it is widely recognized that viruses impact microbial community structure and function, there is little known regarding wintertime viral activity or the relationship between viral activity and ice cover extent. Our metatranscriptomic analysis indicated that viruses were transcriptionally active in the winter surface waters of Lake Erie. These findings also expanded the known diversity of

viral lineages in the Great Lakes. Notably, viral community activity metrics were significantly different between the two sampled winters. The pronounced differences we observed in active viral communities between the ice-covered and ice-free samples merits further research regarding how viral communities will function in future, potentially ice-free, freshwater systems.

Introduction

Decades of research have demonstrated that viruses are key biological components of aquatic ecosystems (1). Through top-down control, viruses influence microbial community structure and the movement of organic matter *via* the "viral shunt" and "viral shuttle" (2-4). Viruses also modulate the metabolism of infected cells by hijacking host cell physiology and through virus-encoded metabolic genes (5), which can have far-reaching implications for biogeochemical cycles (6, 7). As many of Earth's aquatic ecosystems are faced with unprecedent changes in climate (8, 9), there is growing interest in how environmentally relevant viruses are influenced by climate change, such as how climate modulates viral diversity and production rates (10). The winter period in temperate freshwater lakes is particularly vulnerable to climate change, where one apparent response is the decline in ice cover duration and thickness (9, 11). In fact, the frequency of ice-free winters is projected to increase globally in current climate trajectories (12). Seasonally ice-covered lakes are not biologically dormant during the winter (13), and yet there is relatively little documented about viral activity during this period (14).

The Great Lakes contain about 20% of the global fresh water supply (15) and have substantial contributions to regional socioeconomic services (16). Winter in the Laurentian Great Lakes is undergoing pronounced changes due to a warming climate (17). Since the 1970's, Lake Erie has exhibited significant declines in wintertime ice cover extent and duration (18, 19). Furthermore, ice cover extent drastically alters microbial community structure by shifting the winter ecosystem from one characterized by dense accumulations of filamentous diatoms (20-22) to one dominated by picoplankton (23). Putative phage gene fragments (24) and RNA virus sequences (25) have been detected during the mid-winter in Lake Erie, but there has been no

comprehensive characterization of winter viral activity in the Great Lakes to-date. Additionally, the relationship between viral community activity and ice cover extent has not previously been explored. Viral community composition and activity are shaped by the surrounding microbial community (26, 27). Given that ice cover can influence microbial population structure, it is reasonable to expect a relationship between ice cover extent and viral activity. Viral activity could also be directly impacted by other ice cover-dependent factors as well (*e.g.*, the effect of UV exposure on viral decay rates (28, 29)). Characterizing active viral populations during the winter is an important step towards understanding how virus-mediated processes may be impacted by changing ice conditions.

Targeted gene amplification and shotgun metagenomic approaches have been informative in the study of viruses in the Great Lakes (24, 30, 31), but a limitation of DNA-based approaches is that they do not reveal virus activity (*i.e.*, replicating viruses) or capture RNA viruses. Metatranscriptomics can facilitate both the discovery of RNA viruses as well as estimate *in situ* activity levels of viruses using transcript abundance as a proxy (32). Here, we surveyed metatranscriptomes from filtered (> 0.2 μm) samples collected from Lake Erie surface waters during two contrasting winters: one winter with high-ice cover (2019, 95% mean maximum ice cover) and a subsequent relatively ice-free winter (2020, 15% mean maximum ice cover) (33). Samples from the spring months following the ice-free winter were also collected and serve as an outgroup in this study. Using hallmark genes as proxies for putative viral populations, we identified phylogenetically diverse viral communities that were transcriptionally active in the winter surface waters. Viral community activity metrics (alpha and beta diversity indices) revealed significant differences in community composition and viral richness (*i.e.*, the number of different viral hallmark gene variants) between the ice-covered and ice-free samples. Additionally, viral activity metrics were correlated with microbial community activity metrics, namely with microbial richness and the proportion of diatom reads within the winter libraries. This suggested that the differences in viral activity between the winters may in part be linked to differences in the surrounding microbial community. Although the cause(s) of the observed differences in viral community activity remain

uncertain, they coincided with the substantial variation in ice cover extent between the winters.

Materials and Methods

Sample collection

Opportunistic surface water samples were collected by USCGC *Neah Bay* between February and March of two consecutive years (2019 and 2020). Additional spring samples in 2020 were collected by MV *Orange Apex*. Samples were collected from 0.5 meters below the surface as either whole-water (n = 20) or net-concentrated (n = 57, 64-μm net opening). Samples intended for RNA extraction were passed through 0.22-μm nominal pore-size filters, flash-frozen, and stored at -80° C until further processing. Details regarding RNA extraction and sequence processing can be found in an associated *Microbiology Resource Announcement* (34). Meteorological and surface water conditions, ice thickness, nutrient concentration, chlorophyll *a* (> 20 μm and > 0.22 μm size classes), phytoplankton taxonomy, and phytoplankton cell abundance were recorded. Chlorophyll *a* measurements were performed using the method described by Twiss *et al*. (21). Nutrient measurements were performed by the National Center for Water Quality Research (Heidelberg University, Tiffin, OH, USA). Phytoplankton identification and enumeration were performed as described in Zepernick *et al*. (35). Sample metadata can be accessed through the Biological and Chemical Oceanography Data Management Office (BCO-DMO) under dataset number 809945.

Metatranscriptome assembly and gene prediction

This study used the metatranscriptome co-assembly and gene predictions generated by Zepernick *et al*. (35). Briefly, in Zepernick *et al*., quality-filtered reads were co-assembled using MEGAHIT v1.2.9 (https://github.com/voutcn/megahit) and the resulting co-assembly was assessed using QUAST v5.0.2 (Supplemental **Table 2.1**). All supplemental tables are located within the Chapter Two Appendix. Gene sequences were predicted from the co-assembly contigs using MetaGeneMark v3.38 (36). Here, to detect and quantify the relative transcript abundance for genes within a sample, we mapped trimmed reads to gene sequences with ≥ 90% identity and ≥ 90% read length fraction thresholds using CoverM v0.6.1 (https://github.com/wwood/CoverM). To remove poorly

covered genes, at least 50% of the gene must be covered to be considered detected in a metatranscriptome library (37). Read counts for genes with less than 50% coverage were reset to zero on a per-library basis. Read counts were normalized using the transcripts per million (TPM) method (38).

Viral hallmark gene discovery

Conserved viral hallmark genes were identified as proxies for viral populations to estimate viral diversity and activity, similar to as performed previously (32, 39). All hallmark genes were identified through protein sequence similarity searches. Predicted protein sequences from MetaGeneMark were aligned to databases of viral hallmark genes (detailed below) using DIAMOND BLASTp (https://github.com/bbuchfink/diamond) (E-value threshold of $1e^{-5}$). Protein sequences that aligned to translated hallmark genes were then aligned to the RefSeq v213 database and sequences with a top hit to a viral protein were retained as putatively viral (32). As a second quality control measure, only viral sequences with an appropriate protein domain identified by a Pfam domain search (database v32) were retained for analysis (*i.e.*, a putative Gp23 sequence retrieved from the BLASTp that had no capsid domain was not included). RdRp-like protein sequences were further filtered using Palmscan (-rdrp option) and only high confidence RdRp sequences retained (https://github.com/rcedgar/palmscan). As an additional measure to remove non-viral sequences, the metatranscriptome contig from which the hallmark gene originated was piped through VirSorter2 v2.2.3 (https://github.com/jiarong/VirSorter2) and CheckV v1.0.1 (https://bitbucket.org/berkeleylab/CheckV). While the majority of contigs were too short to determine if they were non-viral, no hallmark gene contigs had any cellular genes called by VirSorter2 or CheckV.

Hallmark gene databases

The terminase large subunit (TerL) was used as a general phage marker (*i.e.*, a marker that is present in all *Caudoviricetes*) (Supplemental **Table 2.2**) (40). Relatively few TerL were detected, most of which resembled T4-like myophage based on sequence homology. Therefore, the T4-like major capsid protein (Gp23), DNA polymerase B (Gp43), and tail sheath domains were used to detect myophage. The Gp23 had the greatest richness compared to the other phage markers and was therefore chosen as the

representative hallmark gene for myophage for further analysis (Supplemental **Table 2.2**). Viral integrase, excisionase, CI repressor, and Cro repressor were also screened for as these markers have been used to detect lysogenic phage in metagenomic data (41, 42). All *Caudoviricetes* marker sequences were retrieved from Interpro, where only viral sequences were downloaded.

The major capsid protein (MCP) was used as a proxy for active *Nucleocytoviricota*, or NCLDV, populations. We chose the MCP marker because 1) it is conserved in many NCLDV genomes (43) (all known groups except for *Pandoraviridae* and *Pithoviridae*), and 2) it is most likely to be detectable since the MCP can be highly expressed relative to other NCLDV core genes (44). We also searched for DNA polymerase B (PolB) as an alternative NCLDV marker, but only nine PolB sequences were detected in the metatranscriptomes *via* read mapping (Supplemental **Table 2.2**). Therefore, we only discuss the MCP as a proxy for NCLDV with the caveat that the MCP may not be the most robust phylogenetic marker for NCLDV (45). NCLDV marker sequences were retrieved from the NCVOG database (https://ftp.ncbi.nih.gov/pub/wolf/COGs/NCVOG/).

RNA-dependent RNA polymerase (RdRp) was used as the hallmark gene for RNA viruses (*Orthornavirae*) (46), and the RdRp-Scan database was used for RdRp searches (https://github.com/JustineCharon/RdRp-scan). Replicase sequences compiled by Kazlauskas *et al*. (47) and Moniruzzaman *et al*. (32) were used to screen for circular replicase-encoding single-stranded (CRESS) DNA viruses (*Cressdnaviricota*).

Phylogenetic analysis

Hallmark protein sequences were aligned to reference alignments using Clustal Omega v1.2.4 (76). Positions with >90% gaps were trimmed from the alignments by trimAl v1.4 (https://github.com/inab/trimal). Trees were constructed using FastTree v2.1.11 (-lg option) (http://www.microbesonline.org/fasttree) or IQ-TREE v2.2.0.3 (-m TEST option) (http://www.iqtree.org) and were visualized in iTOL (80). NCLDV reference genomes were retrieved from Gilbert *et al*. (81) and the MCP sequences were predicted using ViralRecall v2 (https://github.com/faylward/viralrecall). RdRp-Scan

alignments were used for the RdRp tree references. References for all other trees were manually curated.

Microbial community characterization

Taxonomic and functional annotation of predicted cellular genes were retrieved from Zepernick *et al*. (35). Briefly, taxonomy was estimated using EUKulele v1.0.6 (https://github.com/AlexanderLabWHOI/EUKulele), where protein sequences were aligned to the PhyloDB database v1.076 by DIAMOND BLASTp. Cellular genes were functionally annotated by eggNOG-mapper v2.1.6 using the eggNOG 5.0 orthology database with a DIAMOND BLASTp cutoff of $1e^{-10}$ (https://github.com/eggnogdb/eggnog-mapper).

Statistical analyses

Community diversity metrics were calculated using the R package vegan (https://github.com/vegandevs/vegan) and in PRIMER v7 (48, 49). Normalized transcript abundance (TPM) tables were used as the input for all diversity analyses. For beta diversity statistics, Bray-Curtis dissimilarity matrices were generated from the standardized TPM table (Wisconsin double standardization). Similarity percentage (SIMPER) analyses were used to quantify the contribution of individual marker genes to average dissimilarity between sample groups.

Spearman's correlations were performed using R (cor.test) (48). The virus-host hallmark gene interaction network was generated using FlashWeave ($\alpha < 0.01$) (https://github.com/meringlab/FlashWeave.jl), where genes annotated as DNA-directed RNA polymerase subunits *rpb1* (KEGG orthology K03006) and *rpoB* (K03043) were used as cellular hallmark genes (32, 50). A TPM table of all viral and cellular hallmark genes was used as input for the correlation analysis.

Data availability

Raw and quality-filtered reads are publicly available through the JGI Genomes Online Database (GOLD) under GOLD Study ID Gs0142002.

Results

Preliminary comparison of viral community activity between whole-water and net-concentrated samples

Beta diversity analysis based on the relative transcript abundance of viral hallmark genes showed that the composition of active viral communities was significantly different between the whole-water (no filtration) and net-concentrated (64-μm opening) samples (ANOSIM R = 0.91, $P < 0.001$) (Supplemental **Figure 2.5A**). All supplemental figures are located within the Chapter Two Appendix. The total number of putative viral hallmark gene variants detected within a library was significantly lower in the net-concentrated relative to the whole-water samples (ANOVA, $P < 0.001$) (Supplemental **Figure 2.5B**). Examination of the relative abundance of viral hallmark genes within the libraries revealed that phage (Gp23), NCLDV (MCP), and RNA virus (RdRp) transcripts were detected in each whole-water sample, but that phage and NCLDV transcripts were relatively absent in the net-concentrated metatranscriptomes (Supplemental **Figure 2.5C**). Since we observed more hallmark gene variants in the unfiltered samples, both in terms of the number of hallmark genes detected and the different virus types detected (phage, NCLDV, and RNA viruses), we focused on the whole-water samples and the net-concentrated metatranscriptomes are not discussed here further.

Physiochemical parameters and phytoplankton abundance estimates

The whole-water samples spanned temporal, climatic, and spatial gradients. Whole-water samples were collected during February – March of 2019 ($n = 4$), February – March of 2020 ($n = 10$), and May – June of 2020 ($n = 6$) across twelve sites (Supplemental **Figure 2.6**, Supplemental **Table 2.3**). The winter of 2019 was considered a high-ice year (95% mean maximum ice cover) while the winter of 2020 was largely ice-free (15% mean maximum ice cover) (33). Ice cover ranged from 90 – 100% at the 2019 winter (*i.e.*, February/March) sample sites and was 0% at all 2020 winter sample sites (Supplemental **Table 2.3**). Ice thickness ranged from approximately 2.5 to 15 cm at the ice-covered sites. The concentration of dissolved ammonia, chlorine, silicate, soluble reactive phosphorus, and the silica: nitrate ratio were not significantly different between

the 2019 (ice-covered) and 2020 (ice-free) winters (two-tailed unpaired *t*-test, $P > 0.1$) (Supplemental **Figure 2.7**, Supplemental **Table 2.4**). Dissolved nitrate was on average significantly higher in the ice-covered relative to the ice-free winter ($P = 0.03$) (Supplemental **Figure 2.7**). Chlorophyll *a* (both > 20 µm and > 0.22 µm size classes) and diatom cell abundance were on average lower in the ice-free samples, but not significantly so ($P > 0.1$) (Supplemental **Figure 2.8**, Supplemental **Table 2.5**). An in-depth analysis on phytoplankton dynamics is outside of the scope of this study, but details regarding phytoplankton taxonomy and distribution in these samples can be found in the associated study by Zepernick *et al*. (35).

Signatures of active viral communities within the winter surface water metatranscriptomes

Viral hallmark genes detected in the metatranscriptomes were interpreted to represent transcriptionally active or cell-associated viral populations. This was presumed because 1) DNA viruses must be replicating intracellularly to synthesize mRNA and 2) samples were filtered (> 0.2 µm) prior to nucleic acid extraction at a size cutoff that would enrich for cell-associated RNA viruses. Hereafter we refer to the detected virus populations as "active" with the consideration that this approach may detect cell-associated RNA virus genomes that are not replicating.

Putative phage (Gp23), NCLDV (MCP), and RNA virus (RdRp) hallmark genes were detected in the winter surface water samples **(Figure 2.1A**). Overall, active winter viral communities displayed high hallmark gene richness (*i.e.*, the number of hallmark gene variants detected) and were phylogenetically diverse (**Figure 2.1A-D**). The majority of Gp23 detected in the winter metatranscriptomes *via* read mapping (439 predicted gene sequences) were most closely related to uncultured myophage environmental sequences with uncertain hosts as opposed to the established cultured clades, although some Gp23 were related to isolated cyanophage (two sequences) and *Pelagibacter* phage (six sequences) (**Figure 2.1B**, Supplemental **Table 2.6**). The NCLDV MCP detected in the winter libraries (427 sequences) represented five of the six established NCLDV orders

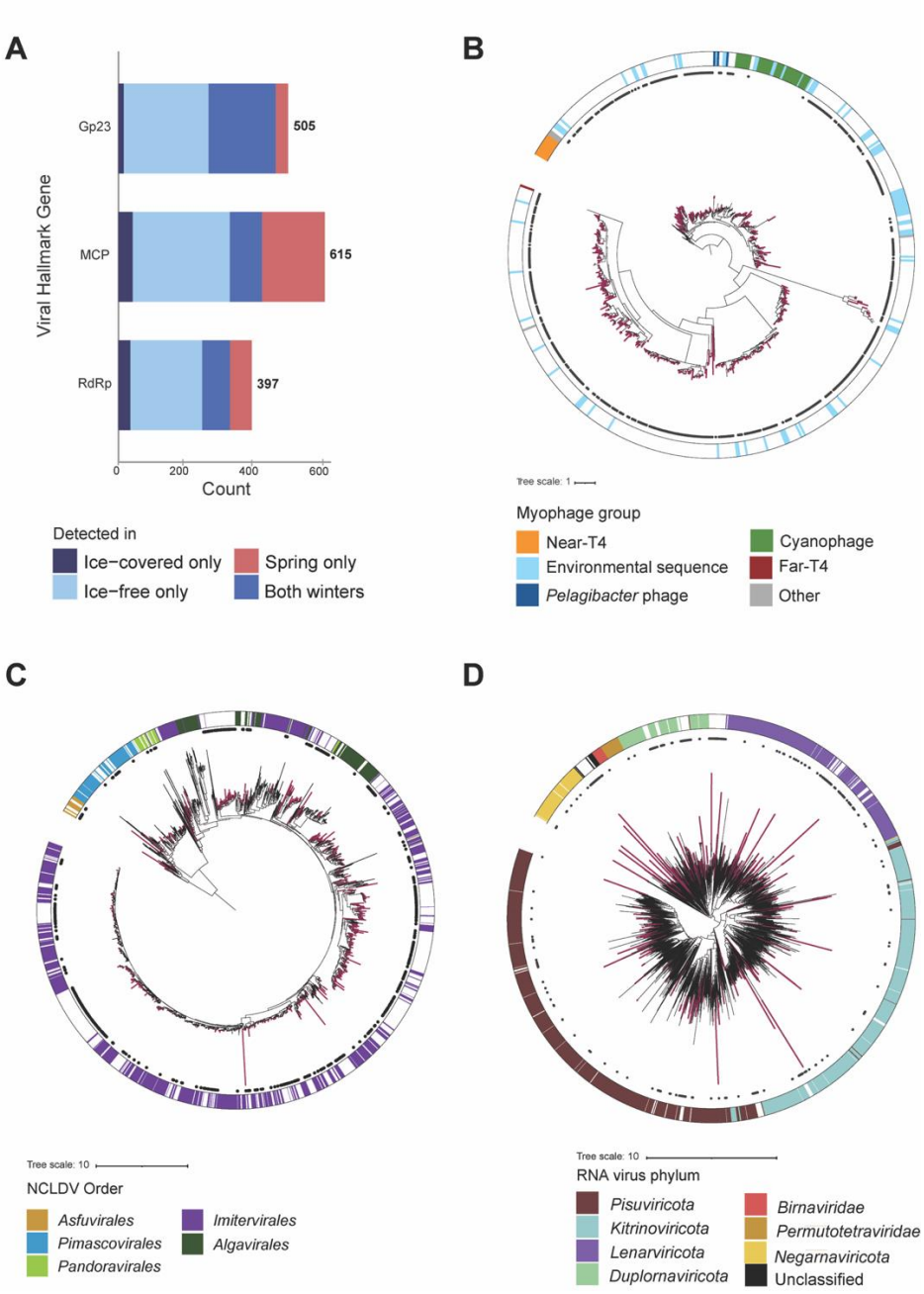


Figure 2.1. Viral hallmark genes detected in the surface water metatranscriptomes with phylogenetic placement.
A) Total count of genes discovered for Gp23, MCP, and RdRp. Maximum-likelihood trees for **B)** Gp23, **C)** MCP, and **D)** RdRp. Branches for Lake Erie sequences are in color, reference sequence branches are in black. Inner ring circles represent Lake Erie sequences detected in at least one winter library. Outer ring shows taxonomy of reference sequence.

(**Figure 2.1C**) (71), most of which were assigned to the *Imitervirales* (395 sequences) (Supplemental **Table 2.6**). The relatively higher *Imitervirales* MCP richness may reflect the broad host range of this group (92, 93). The phylogenetic diversity of RdRp present in the winter libraries (332 sequences) was similarly high, spanning the known phylum-level diversity of the *Orthornavirae* (**Figure 2.1D**, Supplemental **Table 2.6**). Furthermore, only a single lysogenic marker (integrase) was considered viral based on its top BLASTp hit to uncultured freshwater *Caudoviricetes* (GenBank accession: CAB5220699.1) and was only present in one library in the ice-free conditions (MT6, 02/14/2020).

The viral hallmark gene transcript pool was consistently dominated by Gp23 and RdRp transcripts across both winters, although MCP representation increased in the spring, specifically the May 1-22 samples (Supplemental **Figure 2.9**). Only eight single-stranded DNA virus replicase markers were detected with low representation in the water column libraries (Supplemental **Table 2.2**, Supplemental **Figure 2.9**) and they were not analyzed further.

Viral community activity metrics differed between the ice-covered and ice-free conditions

Alpha and beta diversity metrics were calculated based on the detected viral hallmark genes in order to compare viral community activity between the two winters. Community activity was highly dissimilar between the ice-covered and ice-free conditions when comparing the relative abundance (TPM) of viral hallmark genes (73% dissimilar on average, SIMPER analysis). Furthermore, non-metric multidimensional scaling (nMDS) plots revealed that viral community composition clustered by season and that composition was significantly different between the winters based on both hallmark gene relative abundance (ANOSIM R = 0.79, $P = 0.003$) and presence-absence matrices (ANOSIM R = 0.86, $P = 0.002$) (**Figure 2.2A**). Mean viral community evenness (Pielou's) and diversity (Shannon's H) were not significantly different between seasons (**Figure 2.2B**, Supplemental **Table 2.7**). While community evenness was high in all winter samples (Pielou's $J > 0.6$), evenness was on average lower in the ice-free samples. Mean hallmark richness, however, was significantly lower (Tukey HSD $P < 0.001$) in the

ice-covered condition compared to the ice-free (**Figure 2.2B**). In other words, there were significantly fewer viral hallmark genes detected via read mapping in the ice-covered winter samples. Viral hallmark richness was still significantly lower when differences in sequencing depth were accounted for (*i.e.*, when hallmark gene counts were normalized to library size) (Tukey HSD $P < 0.001$).

Viral community activity was significantly different between winters when examining virus hallmark types individually. The separate Gp23, MCP, and RdRp communities each clustered by season and were significantly different between the ice-covered and ice-free winters based on relative abundance (Supplemental **Figure 2.10**). Similarly, the mean hallmark gene richness of all three virus types was significantly lower in the ice-covered winter relative to the ice free (Supplemental **Figure 2.11**, Supplemental **Table 2.8**). Overall, active viral communities displayed pronounced differences in diversity when compared between the ice-covered and ice-free winter samples and these differences were observed in a broad range of virus types.

Differences in viral community activity were correlated with microbial community activity metrics

Collectively, viral hallmark genes comprised a larger portion of the transcript pool (*i.e.*, total TPM) in the ice-free samples compared to the ice-covered (**Figure 2.3A**). Higher relative viral transcript abundance in the ice-free conditions coincided with lower abundance of the bloom-forming diatoms (*Bacillariophyta*) in terms of both diatom relative transcript abundance (**Figure 2.3B**) and cell counts (Supplemental **Figure 2.8**). Conversely, prokaryotic transcripts had higher representation in the ice-free conditions (13-36% versus 30-67%, respectively) (**Figure 2.3B**). In fact, collective prokaryotic relative abundance was significantly negatively correlated with diatom relative abundance across the winter samples (Spearman's $\rho = -0.83$, $P < 0.001$) (Supplemental **Table 2.9**). Moreover, both prokaryotic (*rpoB*) (Tukey HSD $P < 0.001$) and eukaryotic (*rpb1*) (Tukey HSD $P = 0.02$) hallmark gene richness were lower in the ice-covered winter (Supplemental **Figure 2.12**).

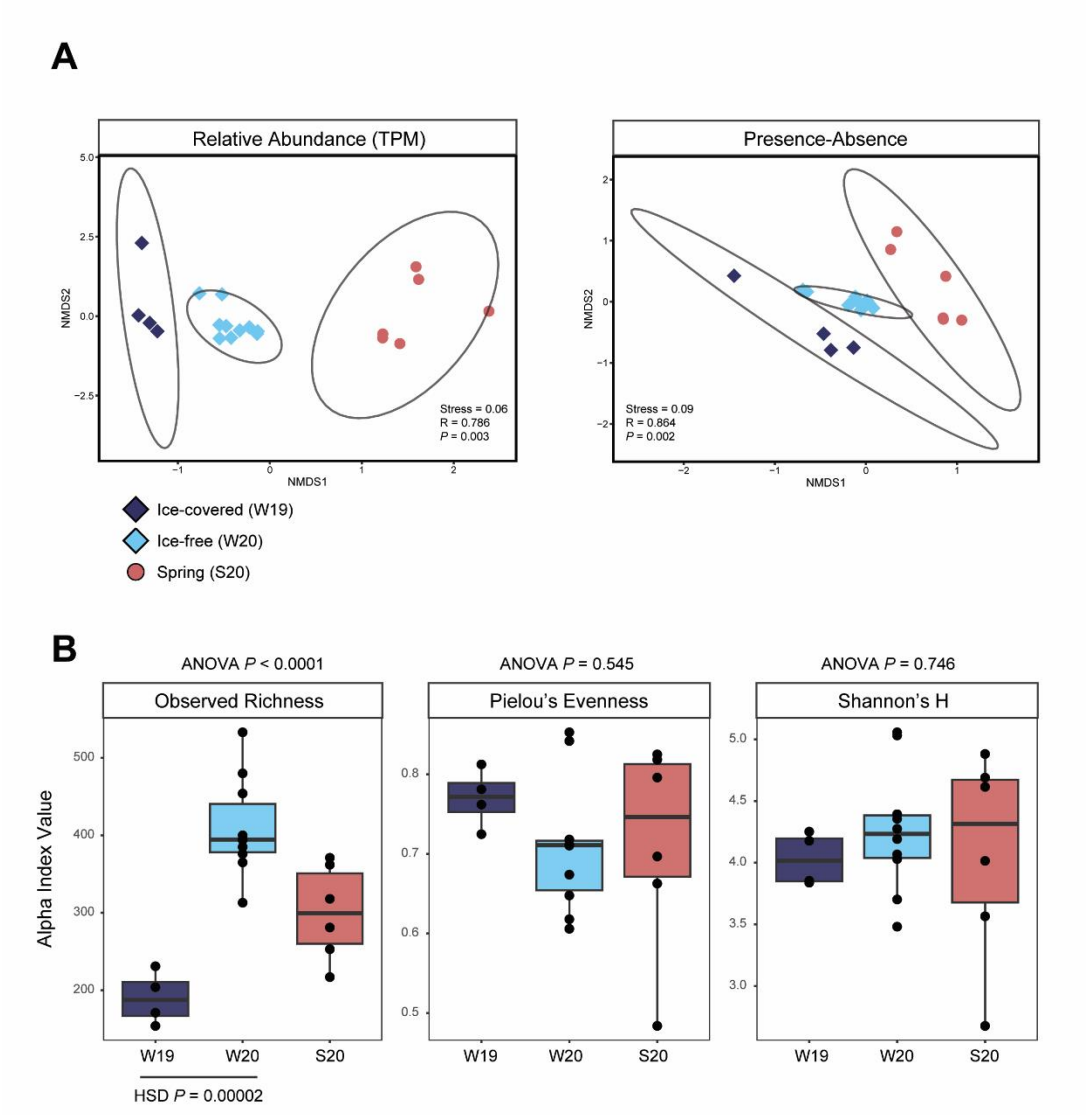


Figure 2.2. Comparison of alpha and beta diversity metrics between winters.

Non-metric multidimensional scaling (nMDS) plot illustrating similarity between samples is based on **A)** relative transcript abundance and **B)** presence-absence transformed abundance of viral markers. ANOSIM results are shown for the pairwise winter comparison. Ellipses represent 95% confidence intervals for the three seasons sampled. **C)** Alpha diversity metrics grouped by season. Tukey HSD is shown for the ice-covered winter 2019 (W19) and ice-free winter 2020 (W20) sample comparison when applicable. One-way ANOVA results are shown for the comparison between the three seasons.

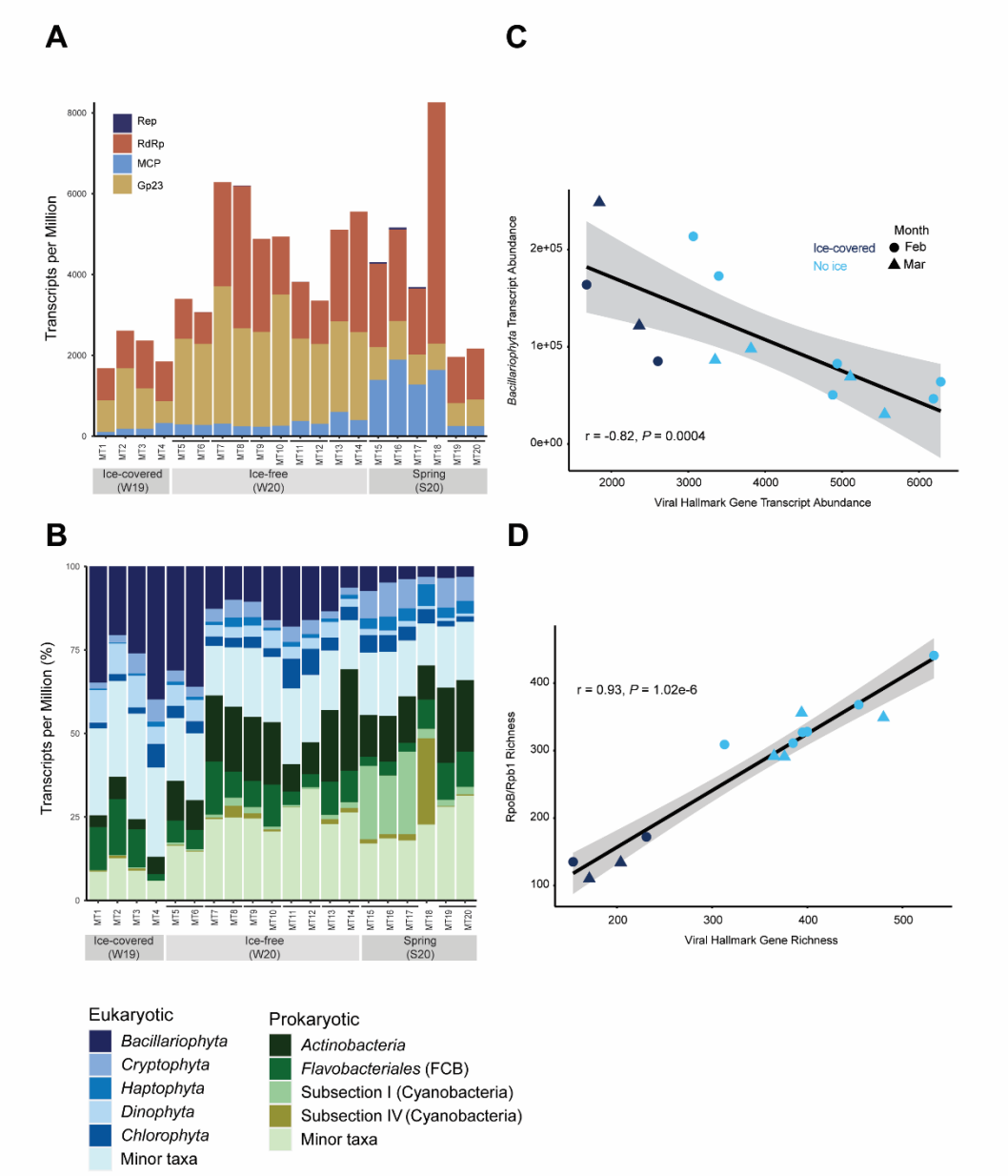


Figure 2.3. Relative transcript abundance of viral and microbial groups with Spearman's correlations.
A) Relative transcript abundance (TPM) of the viral hallmark genes myophage (Gp23), NCLDV (MCP), RNA viruses (RdRp) and CRESS DNA viruses (Rep). **B)** Relative transcript abundance of the dominant eukaryotic and prokaryotic taxa. Spearman's correlations between **C)** viral hallmark and diatom (*Bacillariophyta*) transcript abundance and **D)** viral and cellular hallmark gene richness. Biological replicates are connected by horizontal bars on the bar charts on panels A and B.

The shifts in microbial community metrics correlated with the observed differences in viral hallmark gene abundance and richness. Viral hallmark gene relative transcript abundance was significantly negatively correlated with diatom relative transcript abundance ($\rho = -0.82$, $P < 0.001$) (**Figure 2.3C**) and positively so with prokaryotic abundance (Supplemental **Table 2.9**). Furthermore, viral richness was positively correlated with microbial community richness (RpoB/Rpb1) ($\rho = 0.93$, $P < 0.001$) (Figure 2.3D). Altogether, lower diatom representation in the ice-free metatranscriptomes was associated with higher microbial community richness and prokaryotic representation, and in turn with higher viral richness and representation.

Hallmark genes with the largest contributions to dissimilarity between winter conditions

Although viral richness and representation (collective TPM) were higher in the ice-free conditions, only a few individual viral markers increased in relative abundance and contributed to the dissimilarity between the ice cover conditions. Out of the total 1,525 viral hallmark genes, only 11 had greater that 0.5% contribution to average dissimilarity and amounted to the cumulative 10% average dissimilarity between the ice cover conditions based on the SIMPER analysis (Supplemental **Table 2.10**). While this cut-off is somewhat arbitrary, it captures the viral markers with the largest shifts in relative abundance between the winters, and the remaining markers have low individual contributions. Eight of the eleven markers of interest (four Gp23 and four RdRp) had higher average representation in the ice-free relative to the ice-covered winter where they "spiked" in relative transcript abundance in the ice-free conditions, while the remaining three (one Gp23 and two RdRp) had the opposite trend (Figure 2.4). Notably, the three markers that had higher average representation in the ice-covered winter had an overall low relative abundance (Figure 2.4), which is in line with the overall low viral TPM in the ice-covered winter samples noted previously.

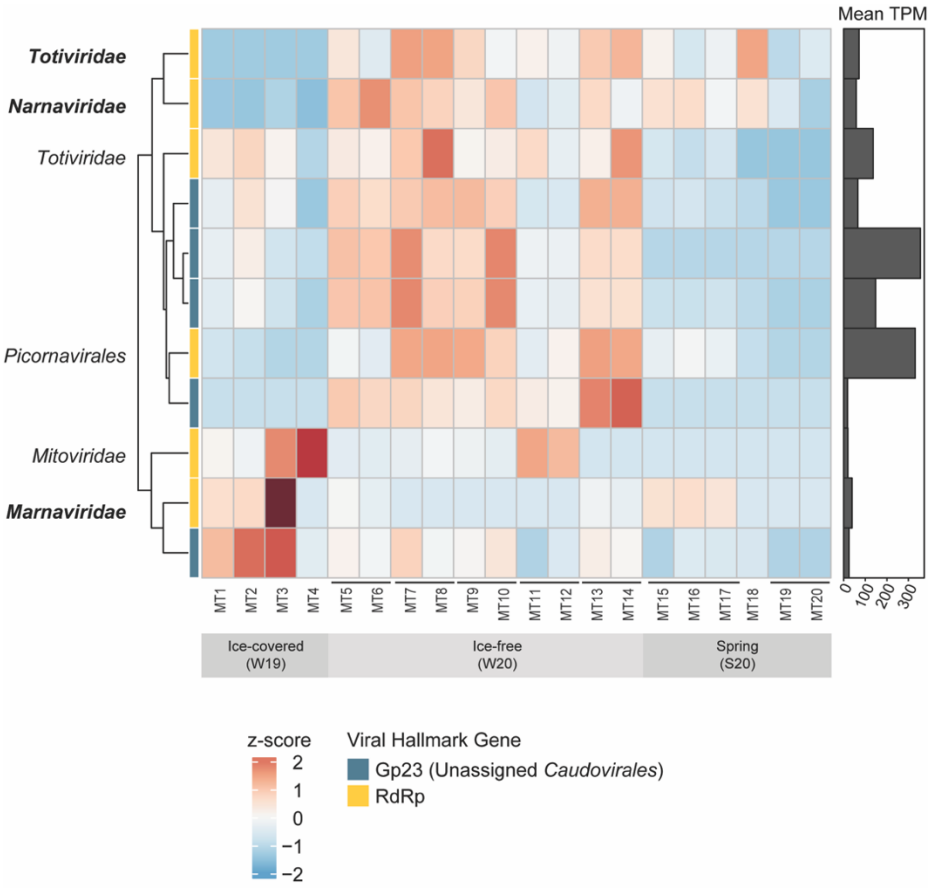


Figure 2.4. Heatmap illustrating the shifts in relative abundance of viral hallmark genes between the ice-covered and ice-free samples.
Abundance values are shown as the z-scored TPM for the top eleven contributors to dissimilarity in viral communities between winters. Rows are clustered by abundance pattern. Bar plot shows mean TPM across all twenty samples. Biological replicates are connected by horizontal bars. Phylogenetic placement for RdRp (see Supplemental Figures 2.13 through 2.15) are indicated by text labels and diatom-associated RdRp clades are in bold.

All five of the top Gp23 contributors were placed within a clade containing an uncultured freshwater Gp23 sequence with uncertain host range (GenBank accession CAB5226458.1). The RdRp of interest spanned three RNA virus phyla (**Figure 2.4**). Interestingly, three RdRp were related to diatom viruses within the *Bacillarnavirus* (order *Picornavirales*) (Supplemental **Figure 2.13**) and diatom colony associated viruses within the *Narnaviridae* (phylum *Lenarviricota*) (Supplemental **Figure 2.14**) and *Totiviridae* (phylum *Duplornaviricota*) (Supplemental **Figure 2.15**). To further infer virus-host pairs, we assessed correlations in relative transcript abundance between the cellular marker genes (*rpoB*/*rpb1*) and the virus hallmark genes, but no significant interactions ($\alpha < 0.01$) were found between the diatom Rpb1 and putative diatom virus RdRp markers.

Major capsid protein transcript abundance was consistently low throughout both winters, and subsequently no single MCP marker had a large contribution to the dissimilarity between winters. However, MCP had higher representation in the spring samples (May 1-22) relative to both winters (Supplemental **Figure 2.9**) largely due to the relative increase in two MCP. Their dominance was also likely responsible for the lower evenness within the MCP communities in the spring samples (Supplemental **Figure 2.11**). Although host range is uncertain, these MCP resembled isolated Chloroviruses (order *Algavirales*) that infect green algae and *Mimiviridae* (order *Imitervirales*) isolates that infect amoeba (**Figure 2.1C**). It was noteworthy that the Chlorovirus-like MCP had a significant positive correlation with a Rpb1 annotated as Chlorophyta (*Trebouxiophyceae*) ($\alpha < 0.01$, correlation coefficient > 0.8).

Discussion

Viral activity during the winter months in the Great Lakes is largely undocumented. Here, viral hallmark genes were detected in metatranscriptomes from two contrasting winters in Lake Erie, indicating that there are viral communities “active” in the surface waters. Our findings expanded the known diversity of viruses that have been detected during the winter in Lake Erie. In one of the few studies to quantify viruses in winter samples collected from the Great Lakes, Matteson *et al*. enumerated putative cyanophage in mid-winter samples and found them to be a substantial component of the standing stock of viruses (24). However, the qPCR-based detection approach that was used only

characterized a narrow group of viruses (cyanomyophage of *Synechococcus* spp.) and could not reveal if those viruses were transcriptionally active. While our metatranscriptomic approach also detected cyanomyophage-like hallmark genes, they represented a small component of the total myophage phylogenetic diversity. Instead, most myophage markers detected here belonged to clades of putative aquatic viruses with unknown hosts. It was previously also hypothesized that phage may "overwinter" as prophage in the Lake Erie water column (24). While we do not dismiss lysogeny as a component of Lake Erie winter microbial communities, we detected little evidence of lysogenic lifestyles based on the transcripts assembled here. Metagenomic sequencing may enable the assembly of phage genomes that could help to clarify the prevalence of temperate phage in Lake Erie.

Diverse populations of putative eukaryotic viruses within the NCLDV and RNA virus groups were also discovered in the winter samples. Previously, only eleven dsRNA viruses within the *Partitiviridae* (*Pisuviricota*), *Totiviridae* (*Duplornaviricota*), and *Birnaviridae* (*incertae sedis*) families were identified in a single Lake Erie metatranscriptome by Edgar *et al*. (25), whereas here we detected hundreds of putative RNA viruses belonging to all five currently established phyla. While perhaps a limitation of sequencing depth and sample size, Edgar *et al*. detected no NCLDV transcripts using a similar hallmark gene approach (25). Our survey, however, revealed putative NCLDV spanning most of the established diversity (five of the six NCLDV orders). Regarding viral discovery, one consideration when using a hallmark gene approach is that viral sequence detection is limited by the diversity of the database (*i.e.*, highly divergent viruses may not be detected by sequence homology searches). Additionally, metatranscriptomes represent "snapshots" of the environment at the time and conditions sampled, therefore the true diversity of active viruses in the surface waters was likely not captured.

Based on the putative viral sequences we identified, active virus communities were compositionally distinct between winters and viral richness was significantly lower in the ice-covered winter. Our findings showed that the observed differences in active viral communities strongly correlated with the shifts in the microbial community,

suggesting a direct relationship between active viral and microbial community composition. Similar to previous observations (23), we saw evidence of a decrease in magnitude of the winter diatom bloom in the ice-free conditions compared to the ice-covered state determined by both the representation in the metatranscriptomes (diatom relative transcript abundance) and cell count data. Other studies in aquatic environments have attributed lower viral abundance or richness in the winter to lower potential host pool abundance (51, 52), and in principle viral diversity and abundance is shaped by the surrounding susceptible host diversity and density (53). The drivers behind the observed differences in viral activity between winters is uncertain, but perhaps the decrease in diatom representation and increase in microbial richness contributed to the greater viral richness in the ice-free winter metatranscriptomes.

The higher levels of viral activity (viral TPM) in the ice-free winter relative to the ice-covered winter was primarily due to increased representation of a few hallmark gene variants. In other words, most viral hallmark genes did not display higher relative transcript abundance in the ice-free conditions and were instead consistently low. Interestingly, the subset of viral hallmark genes that did display higher relative transcript abundance contained hallmark genes similar to those of diatom-associated RNA viruses. This included one RdRp related to lytic *Bacillarnavirus* (order *Marnaviridae*) isolates that infect marine diatoms (54), as well as RNA viruses associated with an environmental diatom colony (55). Culture-dependent studies have shown titers of some diatom viruses increased with the proliferation of diatom blooms and have been speculated to control the bloom crash (56), but also coexist during the bloom course (57). None of the putative diatom RNA viruses interacted with diatom (or any) Rpb1 markers based on our abundance-based network analysis, thus the hosts of these putative RNA viruses remain uncertain.

We note that NCLDV did not have large contributions to the differences in active viral community composition between ice-cover states, and that overall NCLDV displayed relatively low transcript abundance during both winters. Low counts (gene copy number) of various algal viruses (*Algavirales*) have been detected during the wintertime in previous studies surveying the Great Lakes (58, 59). Our findings support

the idea that NCLDV activity is generally low during winter months in the Great Lakes, but continued winter surveys are needed to resolve this.

In all, we show that diverse viral populations were an active component of the winter surface waters in Lake Erie, emphasizing the need for any future winter surveys to account for viruses when investigating microbial dynamics during the winter. While we documented distinctions in viral activity between the ice-covered and ice-free winter samples, we cannot definitively attribute those difference to ice cover extent. However, we posit that the stark difference in ice cover extent (0% versus 90-100% cover) is likely a contributing factor to the variation in microbial and viral activity since ice cover is considered a "master variable" due to its influence on biological, biogeochemical, and physical processes in freshwater lakes (17). Aspects such as the reduced active virus richness observed in the ice-covered samples could be characteristic of ice-covered waters in Lake Erie, but ultimately further winter sampling efforts spanning temporal and climatic gradients are needed to resolve trends in viral activity associated with ice cover extent.

Acknowledgements

This work was supported by National Institutes of Health, NIEHS (1P01ES02328939-01), National Science Foundation (OCE-1840715), NSERC (RGPIN-2019-03943), and the U.S. Department of Energy Joint Genome Institute (JGI) (project 503851). We acknowledge the Kenneth & Blaire Mossman Endowment to UTK. We are thankful for the members of the U.S. Coast Guard Cutter Neah Bay and the MV Orange Apex along with Dan Peck and Thijs Frenken for assistance with sample collection and processing. We thank Derek Niles of Orange Force Marine Ltd. for coordinating sample collection during spring 2020 in the face of COVID disruptions that prevented conventional surveillance of the Great Lakes by federal agencies. Finally, we thank George Bullerjahn, Christa Pennacchio (JGI), and Justin Chaffin for their collaboration.

References

1. Suttle CA. Marine viruses - Major players in the global ecosystem. Nat Rev Microbiol. 2007;5(10):801-12.
2. Brussaard CPD. Viral control of phytoplankton populations - a review. J Eukaryot Microbiol. 2004;51(2):125-38.
3. Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea - Viruses play critical roles in the structure and function of aquatic food webs. Bioscience. 1999;49(10):781-8.
4. Sullivan MB, Weitz JS, Wilhelm S. Viral ecology comes of age. Environ Microbiol Rep. 2017;9(1):33-5.
5. Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, *et al*. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. Nat Rev Microbiol. 2020;18(1):21-34.
6. Sieradzki ET, Ignacio-Espinoza JC, Needham DM, Fichot EB, Fuhrman JA. Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. Nat Commun. 2019;10:9.
7. Coutinho FH, Cabello-Yeves PJ, Gonzalez-Serrano R, Rosselli R, López-Pérez M, Zemskaya TI, *et al*. New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. Microbiome. 2020;8(1):163.
8. Hoegh-Guldberg O, Bruno JF. The impact of climate change on the world's marine ecosystems. Science. 2010;328(5985):1523-8.
9. Woolway RI, Sharma S, Smol JP. Lakes in hot water: The impacts of a changing climate on aquatic ecosystems. Bioscience. 2022;72(11):1050-61.
10. Danovaro R, Corinaldesi C, Dell'Anno A, Fuhrman JA, Middelburg JJ, Noble RT, *et al*. Marine viruses and global climate change. Fems Microbiol Rev. 2011;35(6):993.
11. Li X, Long D, Huang Q, Zhao F. The state and fate of lake ice thickness in the Northern Hemisphere. Science Bulletin. 2022;67(5):537-46.
12. Filazzola A, Blagrave K, Imrit MA, Sharma S. Climate change drives increases in extreme events for lake ice in the Northern Hemisphere. Geophys Res Lett. 2020;47(18):10.

13. Hampton SE, Galloway AWE, Powers SM, Ozersky T, Woo KH, Batt RD, *et al*. Ecology under lake ice. Ecol Lett. 2017;20(1):98-111.
14. Bertilsson S, Burgin A, Carey CC, Fey SB, Grossart HP, Grubisic LM, *et al*. The under-ice microbiome of seasonally frozen lakes. Limnol Oceanogr. 2013;58(6):1998.
15. U.S. EPA. Great Lakes Facts and Figures 2023. Available from: https://www.epa.gov/greatlakes/great-lakes-facts-and-figures.
16. Rau E, Riseng C, Vaccaro L, Read JG. The dynamic Great Lakes economy: employment trends from 2009 to 2018. Available from: https://www.michiganseagrant.org/topics/resilient-coastal-communities/great-lakes-and-the-economy/2020.
17. Ozersky T, Bramburger AJ, Elgin AK, Vanderploeg HA, Wang J, Austin JA, *et al*. The changing face of winter: Lessons and questions from the Laurentian Great Lakes. J Geophys Res-Biogeosci. 2021;126(6):25.
18. Wang J, Kessler J, Bai XZ, Clites A, Lofgren B, Assuncao A, *et al*. Decadal variability of Great Lakes ice cover in response to AMO and PDO, 1963-2017. J Clim. 2018;31(18):7249-68.
19. Wang J, Kessler J, Hang F, Hu H, Clites AH, Chu P. Great Lakes ice climatology update of winters 2012-2017: Seasonal cycle, interannual variability, decadal variability, and trend for the period 1973-2017. 2017.
20. Saxton MA, D'Souza NA, Bourbonniere RA, McKay RML, Wilhelm SW. Seasonal Si:C ratios in Lake Erie diatoms - Evidence of an active winter diatom community. J Gt Lakes Res. 2012;38(2):206-11.
21. Twiss MR, McKay RML, Bourbonniere RA, Bullerjahn GS, Carrick HJ, Smith REH, *et al*. Diatoms abound in ice-covered Lake Erie: An investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. J Gt Lakes Res. 2012;38(1):18-30.
22. Twiss MR, Smith DE, Cafferty EM, Carrick HJ. Phytoplankton growth dynamics in offshore Lake Erie during mid-winter. J Gt Lakes Res. 2014;40(2):449-54.
23. Beall BFN, Twiss MR, Smith DE, Oyserman BO, Rozmarynowycz MJ, Binding CE, *et al*. Ice cover extent drives phytoplankton and bacterial community structure in a

large north-temperate lake: implications for a warming climate. Environ Microbiol. 2016;18(6):1704-19.
24. Matteson AR, Loar SN, Bourbonniere RA, Wilhelm SW. Molecular enumeration of an ecologically important cyanophage in a Laurentian Great Lake. Appl Environ Microbiol. 2011;77(19):6772-9.
25. Edgar RE, Morris PF, Rozmarynowycz MJ, D'Souza NA, Moniruzzaman M, Bourbonniere RA, *et al*. Adaptations to photoautotrophy associated with seasonal ice cover in a large lake revealed by metatranscriptome analysis of a winter diatom bloom. J Gt Lakes Res. 2016;42(5):1007-15.
26. Bekliz M, Pramateftaki P, Battin TJ, Peter H. Viral diversity is linked to bacterial community composition in alpine stream biofilms. ISME Communications. 2022;2(1):27.
27. Coclet C, Sorensen PO, Karaoz U, Wang S, Brodie EL, Eloe-Fadrosh EA, *et al*. Virus diversity and activity is driven by snowmelt and host dynamics in a high-altitude watershed soil ecosystem. Microbiome. 2023;11(1):237.
28. Long AM, Short SM. Seasonal determinations of algal virus decay rates reveal overwintering in a temperate freshwater pond. Isme J. 2016;10(7):1602-12.
29. Madan NJ, Marshall WA, Laybourn-Parry J. Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. Freshw Biol. 2005;50(8):1291.
30. Wilhelm SW, Carberry MJ, Eldridge ML, Poorvin L, Saxton MA, Doblin MA. Marine and freshwater cyanophages in a Laurentian Great Lake: Evidence from infectivity assays and molecular analyses of g20 genes. Appl Environ Microbiol. 2006;72(7):4957-63.
31. Mohiuddin M, Schellhorn HE. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. Front Microbiol. 2015;6:13.
32. Moniruzzaman M, Wurch LL, Alexander H, Dyhrman ST, Gobler CJ, Wilhelm SW. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. Nat Commun. 2017;8:10.
33. NOAA-GLER. Historical ice cover. Available from: https://www.glerl.noaa.gov/data/ice/#historical.

34. Zepernick BN, Denison ER, Chaffin JD, Bullerjahn GS, Pennacchio CP, Frenken T, *et al*. Metatranscriptomic sequencing of winter and spring planktonic communities from Lake Erie, a Laurentian Great Lake. Microbiol Resour Announc. 2022;11(7):3.
35. Zepernick BN, Chase EE, Denison ER, Gilbert NE, Truchon AR, Frenken T, *et al*. Declines in ice cover are accompanied by light limitation responses and community change in freshwater diatoms. ISME J. 2024;18(1).
36. Zhu WH, Lomsadze A, Borodovsky M. *Ab initio* gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38(12):15.
37. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, *et al*. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018;15(11):962-+.
38. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131(4):281-5.
39. Kaneko H, Blanc-Mathieu R, Endo H, Chaffron S, Delmont TO, Gaia M, *et al*. Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. iScience. 2021;24(1):45.
40. Adriaenssens EM, Cowan DA. Using signature genes as tools to assess environmental viral ecology and diversity. Appl Environ Microbiol. 2014;80(15):4470.
41. Luo EL, Eppley JM, Romano AE, Mende DR, DeLong EF. Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. Isme J. 2020;14(5):1304-15.
42. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, *et al*. Lytic to temperate switching of viral communities. Nature. 2016;531(7595):466-+.
43. Koonin EV, Yutin N. Multiple evolutionary origins of giant viruses. F1000Research. 2018;7.
44. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, *et al*. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res. 2010;20(5):664-74.

45. Aylward F, Moniruzzaman M, Ha A, Koonin E. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLoS Biol. 2021;19(10):18.
46. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology. 2015;479:2-25.
47. Kazlauskas D, Varsani A, Krupovic M. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. Viruses. 2018;10(4):11.
48. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
49. Clarke KR, Gorley RN. PRIMER V7: User manual/tutorial. PRIMER-E2015.
50. Ogier JC, Pages S, Galan M, Barret M, Gaudriault S. rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. BMC Microbiol. 2019;19:16.
51. Liu YM, Zhang QY, Yuan XP, Li ZQ, Gui JF. Seasonal variation of virioplankton in a eutrophic shallow lake. Hydrobiologia. 2006;560:323-34.
52. Sandaa RA, Larsen A. Seasonal variations in virus-host populations in Norwegian coastal waters: Focusing on the cyanophage community infecting marine *Synechococcus* spp. Appl Environ Microbiol. 2006;72(7):4610-8.
53. Jacquet S, Miki T, Noble R, Peduzzi P, Wilhelm S. Viruses in aquatic ecosystems: Important advancements of the last 20 years and prospects for the future in the field of microbial oceanography and limnology. Adv in Oceanogr and Limnol. 2010;1(1):97-141.
54. Tomaru Y, Shirai Y, Toyoda K, Nagasaki K. Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus*. Aquat Microb Ecol. 2011;64(2):175-84.
55. Urayama S-i, Takaki Y, Nunoura T. FLDS: A Comprehensive dsRNA sequencing method for intracellular RNA virus surveillance. Microbes and Environments. 2016;31(1):33-40.
56. Bettarel Y, Kan J, Wang K, Williamson KE, Cooney S, Ribblett S, *et al*. Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom *Chaetoceros* cf. *gracilis*. Aquat Microb Ecol. 2005;40(2):103-14.

57. Tomaru Y, Fujii N, Oda S, Toyoda K, Nagasaki K. Dynamics of diatom viruses on the western coast of Japan. Aquat Microb Ecol. 2011;63(3):223-30.
58. Short SM, Short CM. Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. Environ Microbiol. 2009;11(10):2639-48.
59. Short CM, Rusanova O, Short SM. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. ISME J. 2011;5(5):810-21.

Appendix

Supplemental Tables

Supplemental Table 2.1. Quality assessment for the metatranscriptome co-assembly. Co-assembly statistics were generated by QUAST v5.0.2. All QUAST statistics are based on contigs of size >= 500 bp unless otherwise noted.

# contigs (>= 0 bp)	1677102
# contigs (>= 1000 bp)	148925
# contigs (>= 5000 bp)	2139
# contigs (>= 10000 bp)	288
# contigs (>= 25000 bp)	8
# contigs (>= 50000 bp)	0
Total length (>= 0 bp)	966763097
Total length (>= 1000 bp)	247639904
Total length (>= 5000 bp)	15925927
Total length (>= 10000 bp)	4001501
Total length (>= 25000 bp)	258591
Total length (>= 50000 bp)	0
# contigs	639361
Largest contig	40227
Total length	574417227
GC (%)	44.46
N50	891
N75	641
L50	190932
L75	383548
# N's per 100 kbp	0

Supplemental Table 2.2. Number of hallmark genes identified in the metatranscriptome co-assembly *via* BLASTP.

Viral Group	**Marker**	**BLASTP Database Source**	**No. Identified**
Caudoviricetes	Gp23	PF07068	505
	Tail Sheath	PF17481, PF17482, and PF04984	239
	PolB (gp34)	PF03175 and PF00136	33
	Gp20	PF06810	1
	Integrase	PF00589	1
	Excisionase	PF06806	0
	CI Repressor	PF07022	0
	Cro Repressor	IP000655	0
	TerL	PF05876 and PF20454	1
Nucleocytoviricota	MCP	NCVOG0022	615
	PolB	NCVOG0038	9
Orthornavirae	RdRp	RdRp-Scan	397
Cressdnaviricota	Rep	Kazlauskas *et al*. and Moniruzzaman *et al*.	8

Supplemental Table 2.3. Sampling site locations and water conditions for the 20 whole-water samples.
Site ID corresponds to the map in Figure A1.2. ND indicates data was not collected.

Sample ID	Collection date	Latitude	Longitude	Site ID	Ice cover (%)	Ice thickness (cm)	Snow cover (%)	Water temp (°C)	Air temp (°C)	Baro press (Hg)	Wind (kt)	Wind (°)	DO (mg/L)	DO (%)	Cond uS/cm	pH
MT1	2-26-2019	41.839333	-82.5215	S1	100	15	0	ND	-7.8	29.85	3	90	ND	ND	ND	ND
MT2	2-6-2019	41.806	-82.383833	S2	100	2.5	0	ND	-7.8	29.85	6	70	ND	ND	ND	ND
MT3	3-11-2019	41.7497	-81.8352	S3	100	8	0	ND	-1.1	29.57	6	260	ND	ND	ND	ND
MT4	3-11-2019	41.8658	-82.6407	S4	90	10	0	ND	-1.1	29.59	16	260	ND	ND	ND	ND
MT5	2-14-2020	41.8285666	-82.5011166	S5	0	0	0	0.0	-8.3	ND	ND	ND	12.3	89.1	248.0	9.4
MT6	2-14-2020	41.8285666	-82.5011166	S5	0	0	0	0.0	-8.3	ND	ND	ND	12.3	89.1	248.0	9.4
MT7	2-14-2020	41.7385166	-82.2743166	S6	0	0	0	1.89	-8.9	ND	ND	ND	12.4	89.4	240.0	9.4
MT8	2-14-2020	41.7385166	-82.2743166	S6	0	0	0	1.89	-8.9	ND	ND	ND	12.4	89.4	240.0	9.4
MT9	2-14-2020	41.6662833	-82.0816333	S7	0	0	0	2	-9.4	ND	ND	ND	12.3	88.9	254.0	9.4
MT10	2-14-2020	41.6662833	-82.0816333	S7	0	0	0	2	-9.4	ND	ND	ND	12.3	88.9	254.0	9.4
MT11	3-2-2020	41.6659166	-82.0791166	S8	0	0	0	2.22	2.8	29.07	ND	ND	ND	ND	237.0	9.96
MT12	3-2-2020	41.6659166	-82.0791166	S8	0	0	0	2.22	2.8	29.07	ND	ND	ND	ND	237.0	9.96
MT13	3-2-2020	41.7703333	-82.3516666	S9	0	0	0	1.67	2.2	29.05	ND	ND	ND	ND	255.0	9.89
MT14	3-2-2020	41.7703333	-82.3516666	S9	0	0	0	1.67	2.2	29.05	ND	ND	ND	ND	255.0	9.89
MT15	5-1-2020	42.43045	-81.2061666	S10	0	0	0	9.67	11.3	ND	14	315	ND	ND	273.0	9.13
MT16	5-1-2020	42.43045	-81.2061666	S10	0	0	0	9.67	11.3	ND	14	315	ND	ND	273.0	9.13
MT17	5-1-2020	42.43045	-81.2061666	S10	0	0	0	9.67	11.3	ND	14	315	ND	ND	273.0	9.13
MT18	5-22-2020	42.429533	-81.20435	S11	0	0	0	9.67	11.3	ND	5.3	127	12.43	108.5	266	9.03
MT19	6-8-2020	42.2817	-81.6717166	S12	0	0	0	15.66	14.8	ND	9	140	ND	ND	266	9.03
MT20	6-8-2020	42.2817	-81.6717166	S12	0	0	0	15.66	14.8	ND	9	140	ND	ND	266	9.03

Supplemental Table 2.4. Dissolved (< 0.2 µm) and particulate (unfiltered) nutrient measurements for the 20 whole-water samples.

	Dissolved							Particulate	
Sample ID	**NH3 (mg/l)**	**CL (mg/l)**	**S04 (mg/l)**	**NO2 (mg/l)**	**NO3 (mg/l)**	**SIO2 (mg/l)**	**SRP (mg/l)**	**TP (mg/l)**	**TN (mg/l)**
MT1	0.014	15.4	19.2	0	0.36	0.01	0.003	0.0349	0.442
MT2	0.01	15.4	18.9	0	0.29	0	0.003	0.0201	0.292
MT3	0.006	19.3	18.9	0	0.28	0	0.003	0.0149	0.276
MT4	0.024	13.8	15.4	0	0.6	1.5	0.002	0.0127	0.253
MT5	0.0411	12.9	17.6	0	0.08	0.1583	0.0059	ND	0.27
MT6	0.0411	12.9	17.6	0	0.08	0.1583	0.0059	ND	0.27
MT7	0.0181	11.4	16.8	0	0.08	0.0653	0.0035	ND	0.369
MT8	0.0181	11.4	16.8	0	0.08	0.0653	0.0035	ND	0.369
MT9	0.0262	19.2	24.1	0	0.11	0.1737	0	ND	0.274
MT10	0.0262	19.2	24.1	0	0.11	0.1737	0	ND	0.274
MT11	0.0682	12.1	17	0	0.34	1.7211	0	ND	0.245
MT12	0.0682	12.1	17	0	0.34	1.7211	0	ND	0.245
MT13	0.0137	11.6	16.5	0	0.09	0.2136	0	ND	0.403
MT14	0.0137	11.6	16.5	0	0.09	0.2136	0	ND	0.403
MT15	0.1921	16.6	17.2	0.063	0	0	0.0015	0.0075	0.208
MT16	0.1921	16.6	17.2	0.063	0	0	0.0015	0.0075	0.208
MT17	0.1921	16.6	17.2	0.063	0	0	0.0015	0.0075	0.208
MT18	0.0125	15.6	16.6	0.065	0.0777	0.0635	0	0.0044	0.196
MT19	0.0316	15.6	16.8	0.062	0.083	0.115	0.001	0.0027	0.173
MT20	0.0316	15.6	16.8	0.062	0.083	0.115	0.001	0.0027	0.173

Supplemental Table 2.5. Phytoplankton biomass estimates and diatom enumeration.
Diatom abundance shown as cell/mL.

Sample ID	Chl *a* (> 0.2 µm) (µg/L)				Chl *a* (>20 µm) (µg/L)				*Aulacoseira islandica*	*Stephanodiscus* spp.	*Fragilaria* spp.	*Nitzschia*	*Asterionella*	centric diatoms (5-20 µm)
MT1	16.47	26.58	ND	ND	18.09	23.00	ND	ND	1.44E+03	2.35E+03	0.00E+00	ND	ND	ND
MT2	12.95	13.89	13.04	13.2	8.15	10.55	12.04	11.93	9.62E+02	2.35E+03	3.60E+01	ND	ND	ND
MT3	14.23	12.45	11.41	ND	9.8	8.01	9.94	ND	7.60E+02	2.48E+03	5.40E+01	ND	ND	ND
MT4	2.87	2.24	3.23	ND	1.1	0.82	0.76	ND	1.20E+01	1.02E+03	ND	ND	ND	ND
MT5	8.18	7.76	8.23	ND	4.6	4.32	4.66	ND	3.89E+02	1.13E+03	4.00E+01	1.10E+01	4.00E+00	2.64E+02
MT6	8.18	7.76	8.23	ND	4.6	4.32	4.66	ND	3.89E+02	1.13E+03	4.00E+01	1.10E+01	4.00E+00	2.64E+02
MT7	10.96	9.15	6.94	ND	4.86	5.29	5.73	ND	3.87E+02	8.26E+02	1.49E+02	1.40E+01	1.60E+01	1.49E+02
MT8	10.96	9.15	6.94	ND	4.86	5.29	5.73	ND	3.87E+02	8.26E+02	1.49E+02	1.40E+01	1.60E+01	1.49E+02
MT9	17.43	8.95	14.27	ND	5.38	4.36	ND	ND	4.52E+02	6.28E+02	7.90E+01	1.50E+01	1.40E+01	1.70E+02
MT10	17.43	8.95	14.27	ND	5.38	4.36	ND	ND	4.52E+02	6.28E+02	7.90E+01	1.50E+01	1.40E+01	1.70E+02
MT11	6.43	7.4	5.78	ND	2.15	2.50	2.69	ND	1.50E+02	2.32E+02	5.80E+01	1.00E+01	1.70E+01	3.56E+03
MT12	6.43	7.4	5.78	ND	2.15	2.50	2.69	ND	1.50E+02	2.32E+02	5.80E+01	1.00E+01	1.70E+01	3.56E+03
MT13	5.68	5.56	5.44	ND	3.41	3.64	3.99	ND	3.85E+02	2.32E+02	2.60E+01	1.00E+00	3.10E+01	1.20E+02
MT14	5.68	5.56	5.44	ND	3.41	3.64	3.99	ND	3.85E+02	2.32E+02	2.60E+01	1.00E+00	3.10E+01	1.20E+02
MT15	2.49	2.65	2.05	ND	0.82	0.78	0.77	ND	9.60E+01	3.12E+02	4.40E+02	0	0	ND
MT16	2.49	2.65	2.05	ND	0.82	0.78	0.77	ND	9.60E+01	3.12E+02	4.40E+02	0	0	ND
MT17	2.49	2.65	2.05	ND	0.82	0.78	0.77	ND	9.60E+01	3.12E+02	4.40E+02	0	0	ND
MT18	1.24	0.82	1.23	ND	0.19	0.21	0.32	ND	0	1.90E+01	2.40E+01	1.00E+00	5.50E+01	ND
MT19	0.78	0.82	0.74	ND	0.06	0.08	0.06	ND	0	0	1.18E+02	6.00E+00	1.53E+02	ND
MT20	0.78	0.82	0.74	ND	0.06	0.08	0.06	ND	0	0	1.18E+02	6.00E+00	1.53E+02	ND

Supplemental Table 2.6. Taxonomic estimate and detection of the viral hallmark genes. Taxonomy was estimated by ML phylogenies (see Figure 2.1). Viral hallmark gene detection was determined by read mapping.

			Detected in	
	Taxonomy estimate	**No. hallmark genes identified**	**Winter**	**Spring only**
Gp23	Uncultured myophage	468	439	29
	Cyanomyophage	2	1	1
	Pelagibacter phage	6	6	0
	Uncertain	29	22	7
	Total	**505**	**468**	**37**
MCP	*Pimascovirales*	24	10	14
	Asfuvirales	9	8	1
	Pandoravirales	4	3	1
	Algavirales	18	9	9
	Imitervirales	557	395	162
	Uncertain	3	2	1
	Total	**615**	**427**	**188**
RdRp	*Pisuviricota*	92	76	16
	Duplornaviricota	99	84	15
	Lenarviricota	84	72	12
	Kitrinoviricota	53	43	10
	Negarnaviricota	17	13	4
	Birnaviridae	1	1	0
	Permutotetraviridae	1	1	0
	Uncertain	50	42	8
	Total	**397**	**332**	**65**

Supplemental Table 2.7. ANOVA results for alpha diversity metrics for the total active viral community compared between the three seasons sampled.
Diversity metrics were generated from the relative transcript abundance (TPM) table. Significant test results ($P < 0.05$) are in bold.

	Comparison	Observed richness	Pielou's evenness	Shannon's H
Main test	**Season**	**2.20E-05**	0.545	0.746
Tukey's HSD	**Ice-covered/Ice-free**	**0.0000205**	NA	NA
	Ice-free/spring	**0.0242167**	NA	NA
	Ice-covered/spring	**0.0056772**	NA	NA

Supplemental Table 2.8. ANOVA results for alpha diversity metrics compared between the three seasons sampled for the separate viral hallmark gene types**.**
Diversity metrics were generated from the relative transcript abundance (TPM) table. Significant test results ($P < 0.05$) are in bold.

		Gp23		
	Comparison	**Observed richness**	**Pielou's evenness**	**Shannon's H**
Main test	**Season**	**5.26E-06**	**7.36E-04**	**3.58E-02**
Tukey's HSD	**Ice-covered/Ice-free**	**1.93E-05**	7.42E-02	9.78E-01
	Ice-free/spring	3.50E-01	2.78E-01	1.37E-01
	Ice-covered/spring	**1.03E-04**	**5.91E-04**	**3.42E-02**

		MCP		
	Comparison	**Observed richness**	**Pielou's evenness**	**Shannon's H**
Main test	**Season**	**1.38E-02**	**2.50E-02**	**1.25E-03**
Tukey's HSD	**Ice-covered/Ice-free**	**3.12E-02**	7.27E-01	**9.04E-04**
	Ice-free/spring	**1.37E-02**	2.27E-01	**1.17E-02**
	Ice-covered/spring	7.12E-01	**1.96E-02**	5.63E-01

		RdRp		
	Comparison	**Observed richness**	**Pielou's evenness**	**Shannon's H**
Main test	**Season**	**2.73E-05**	1.09E-01	**3.53E-02**
Tukey's HSD	**Ice-covered/Ice-free**	**1.16E-04**	NA	4.34E-01
	Ice-free/spring	5.80E-01	NA	5.18E-01
	Ice-covered/spring	**3.03E-04**	NA	**2.89E-02**

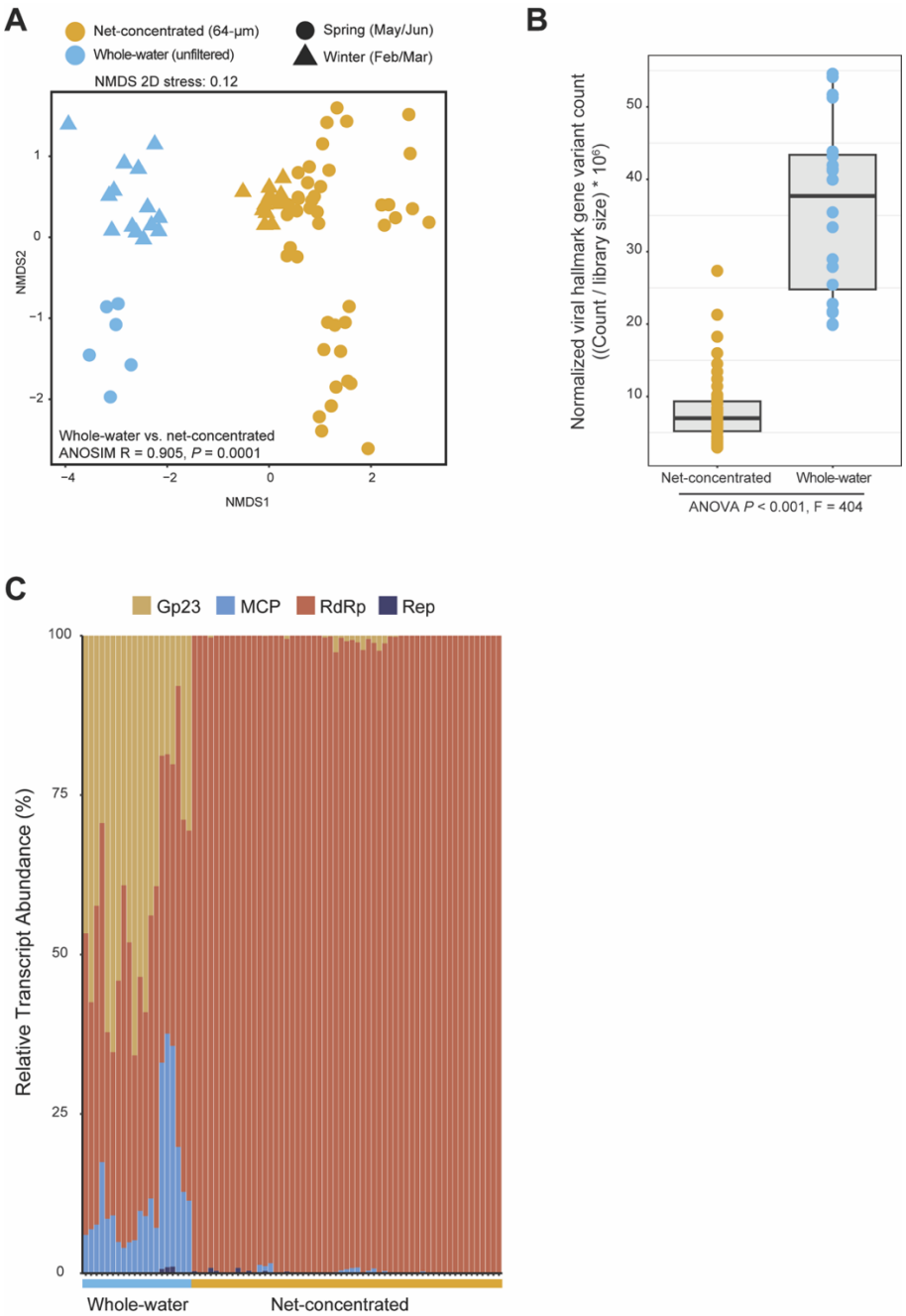
Supplemental Table 2.9. Spearman correlations between viral and microbial diversity and abundance metrics.
Significant correlations ($P < 0.05$) are in bold.

		Spearman	
Comparison		**rho**	**p**
***Bacillariophyta* TPM**	**Prokaryotic TPM**	-0.8329670	**0.0002166**
***Bacillariophyta* TPM**	**RpoB TPM**	-0.7098901	**0.004451**
Bacillariophyta TPM	RpoB richness	-0.2615385	0.3664
Bacillariophyta TPM	Rpb1 richness	-0.0044249	0.988
Bacillariophyta TPM	RpoB TPM	0.1868132	0.5225
Bacillariophyta TPM	Rpb1 TPM	-0.2615385	0.3664
Prokaryotic TPM	**Viral hallmark gene richness**	0.9384615	**6.86E-07**
***Bacillariophyta* TPM**	**Viral hallmark gene richness**	-0.8153846	**0.0003791**
***Bacillariophyta* TPM**	**Gp23 TPM**	-0.7318681	**0.002925**
Bacillariophyta TPM	MCP TPM	-0.1340659	0.6477
***Bacillariophyta* TPM**	**RdRp TPM**	-0.8769231	**3.83E-05**
Bacillariophyta TPM	Viral hallmark gene richness	-0.1956044	0.5028
Bacillariophyta TPM	Gp23 richness	-0.1870188	0.522
Bacillariophyta TPM	MCP richness	-0.1584159	0.5886
***Bacillariophyta* TPM**	**RdRp richness**	-0.6072611	**0.02127**
Prokaryotic TPM	Gp23 richness	0.4576460	0.09988
Prokaryotic TPM	**MCP richness**	0.5808584	**0.02939**
Prokaryotic TPM	**RdRp richness**	0.6226626	**0.01739**
RpoB/Rpb1 richness	**Viral hallmark gene richness**	0.9340659	**1.028E-06**
RpoB richness	**Gp23 richness**	0.9636970	**3.06E-08**
RpoB richness	**MCP richness**	0.7876792	**8.23E-04**
RpoB richness	**RdRp richness**	0.6600664	**1.02E-02**
Rpb1 richness	**MCP richness**	0.7497346	**0.002018**
Rpb1 richness	RdRp richness	0.4507267	0.1058
Rpb1 richness	**Gp23 richness**	0.8815196	**3.08E-05**

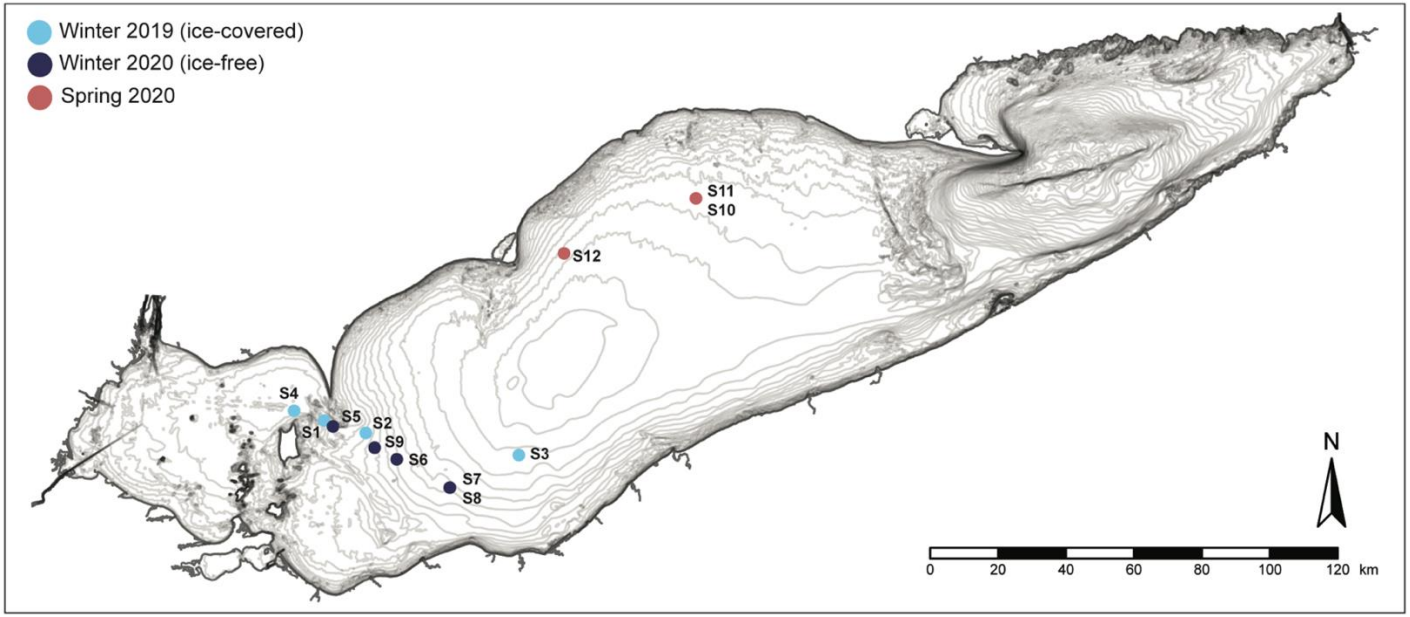
Supplemental Table 2.10. Select output from similarity percentage (SIMPER) analysis. Viral hallmark genes contributing to the top ~10% cumulative dissimilarity between the ice-covered (2019) and ice-free (2020) winter are shown (see Figure 2.4). The abundance values shown are based on the square root transformed TPM table.

Gene ID	Viral marker	Av. Abund. (2019)	Av. Abund. (2020)	Av. Diss.	Diss/SD	Contrib. (%)	Cumul. (%)
gene_458515	RdRp	1.94	22.95	1.46	2.25	2.06	2.06
gene_293376	Gp23	8.46	24.46	1.2	1.65	1.7	3.76
gene_678216	Gp23	4.87	15.75	0.81	1.71	1.14	4.9
gene_1385399	RdRp	0	9.83	0.68	3.12	0.97	5.86
gene_26037	RdRp	9.23	0.87	0.62	1.07	0.87	6.73
gene_1263456	RdRp	2.68	8.71	0.43	2.3	0.61	7.35
gene_430917	RdRp	10.14	13.89	0.41	1.19	0.58	7.93
gene_1656192	Gp23	5.26	10.08	0.41	1.64	0.58	8.51
gene_626655	Gp23	0	5.56	0.39	2.7	0.56	9.06
gene_1534634	RdRp	6.68	2.38	0.38	1.25	0.53	9.6
gene_118358	Gp23	8.14	3.73	0.37	1.81	0.52	10.12

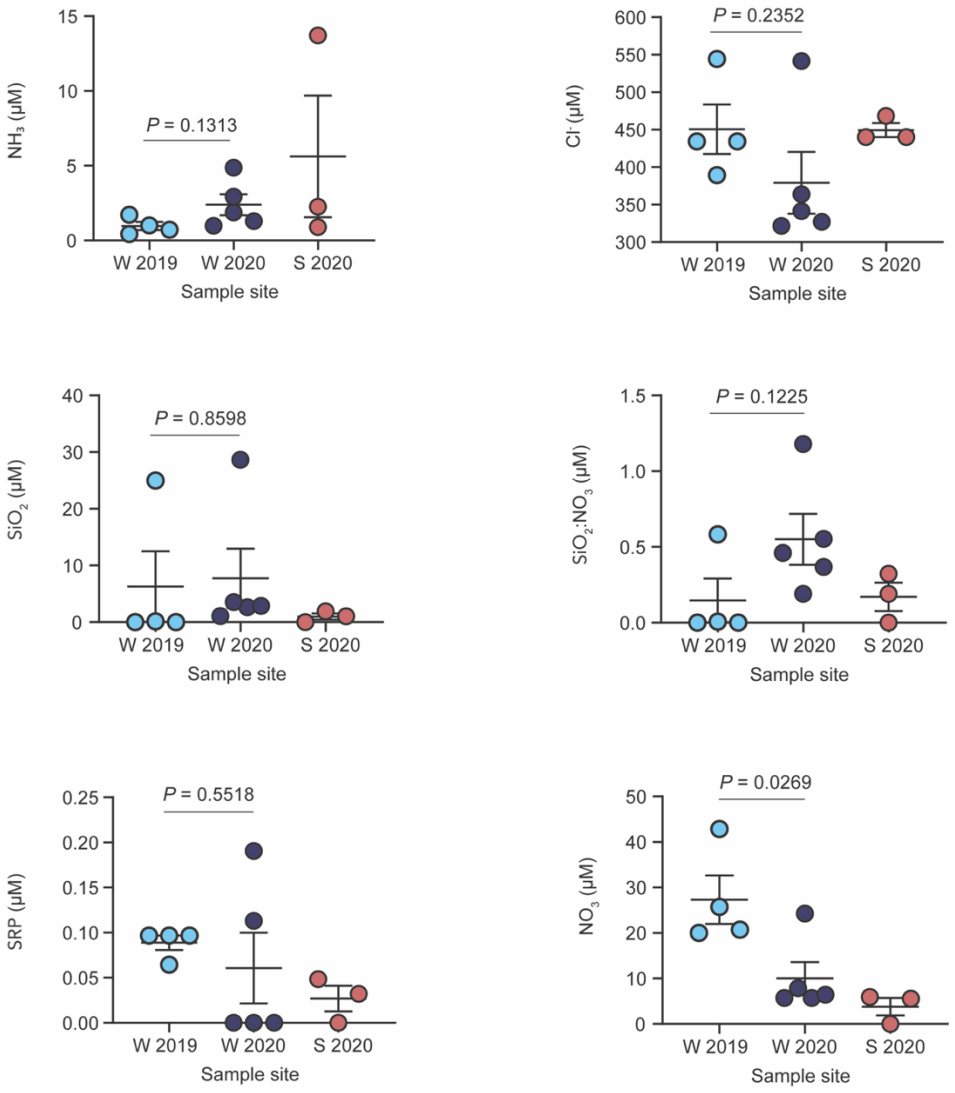
Supplemental Figures



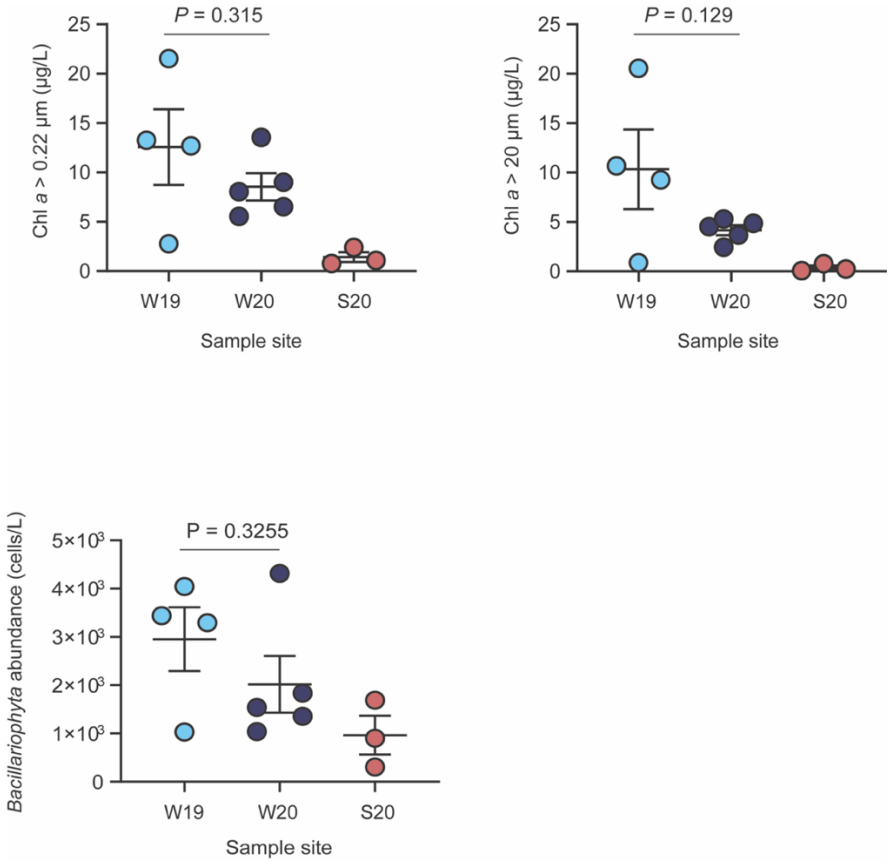
Supplemental Figure 2.5. Differences in active viral communities based on collection method. **A)** nMDS plot showing clustering of active viral communities by collection method based on the TPM of viral hallmark genes (Bray-Curtis dissimilarity matrix). **B)** Count of viral hallmark gene variants grouped by collection method and ANOVA test results. Total counts per sample were normalized by sequencing depth and scaled. **C)** Proportion of the viral hallmark genes (Gp23, MCP, RdRp, and Rep) within each library based on summed abundance (TPM).



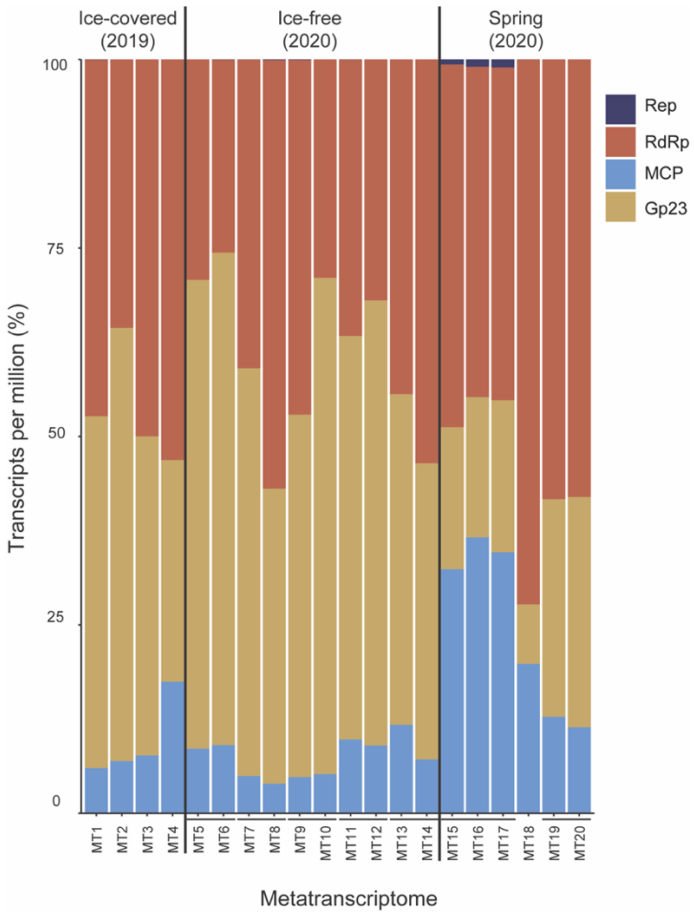
Supplemental Figure 2.6. Map of Lake Erie depicting locations of the whole-water sampling sites.
Site color represents the sampling period (winter 2019, winter 2020, or spring 2020). Please refer to Supplemental Table 3 for exact site coordinates.



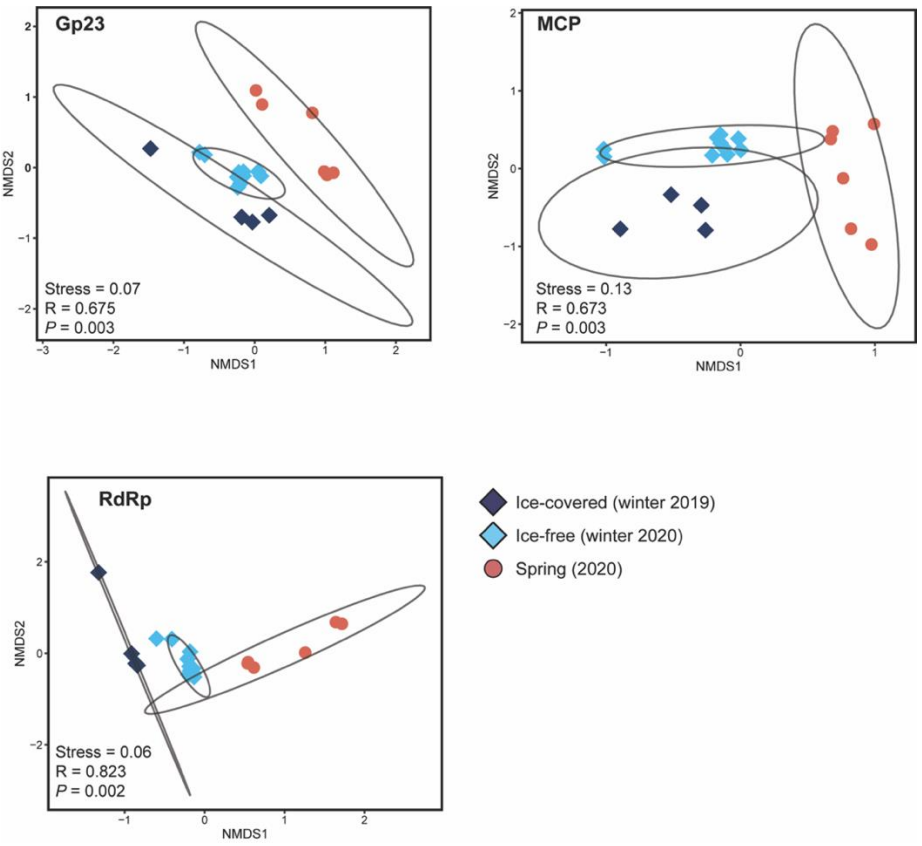
Supplemental Figure 2.7. Nutrient measurements corresponding to each water column sample site.
Two-tailed unpaired *t*-test results are shown for the winter 2019 (ice-covered) versus winter 2020 (ice-free) comparisons.



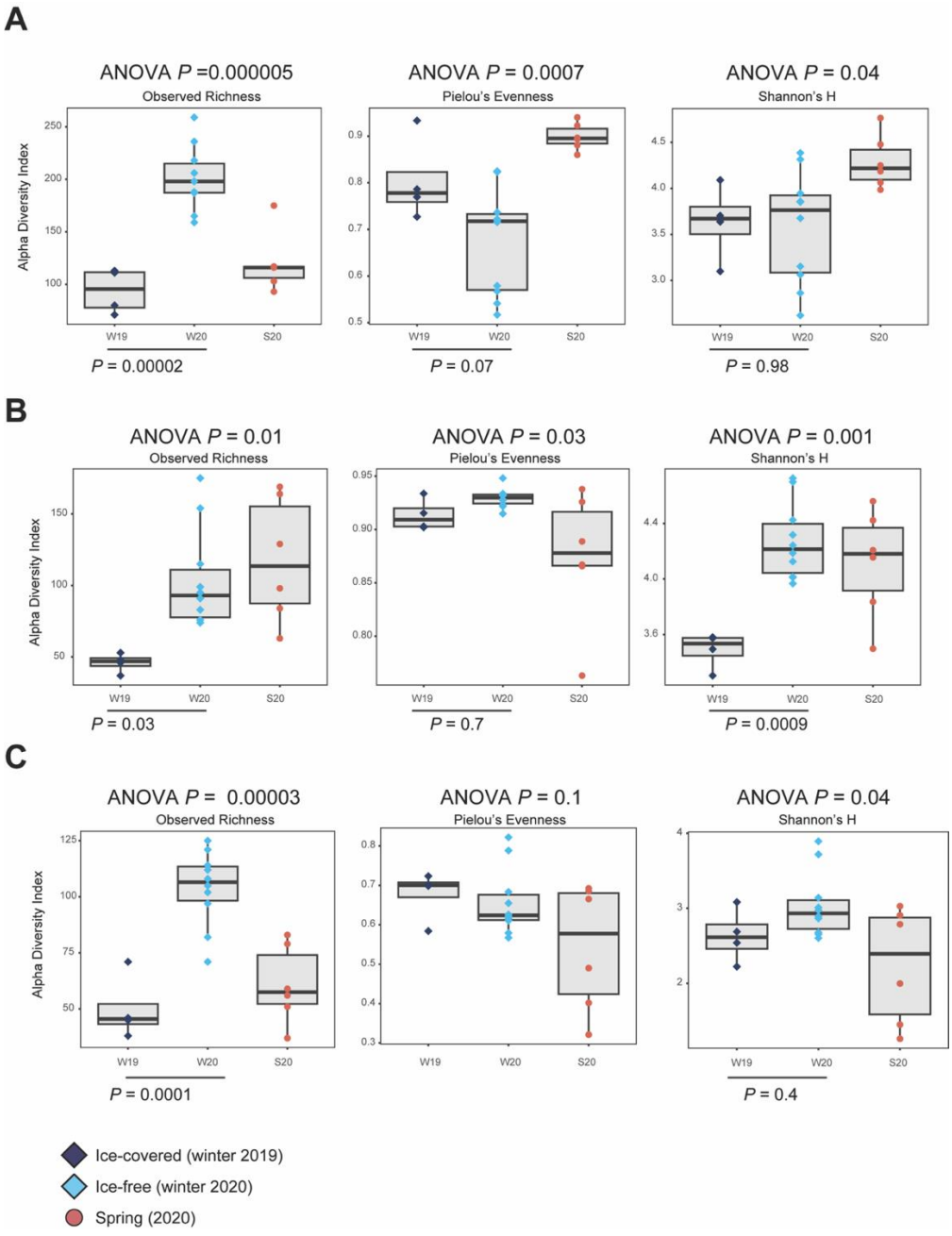
Supplemental Figure 2.8. Phytoplankton biomass measurements corresponding to each water column sample site.
Phytoplankton biomass shown as Chl *a* (µg/L, > 0.22 and > 0.20 µm) and diatoms (*Bacillariophyta*) as cell counts (cells/L) at each site. Chl *a* values are shown as the mean of the biological replicates (refer to Supplemental Table 1.5). Diatom cell counts represent the sum of *Aulacoseira islandica*, *Stephanodiscus* spp., centric diatoms 5-20 µm, *Fragilaria* spp., *Asterionella formosa*., and *Nitzschia* spp. Two-tailed unpaired *t*-test results are shown for the winter 2019 (ice-covered) versus winter 2020 (ice-free) comparisons.



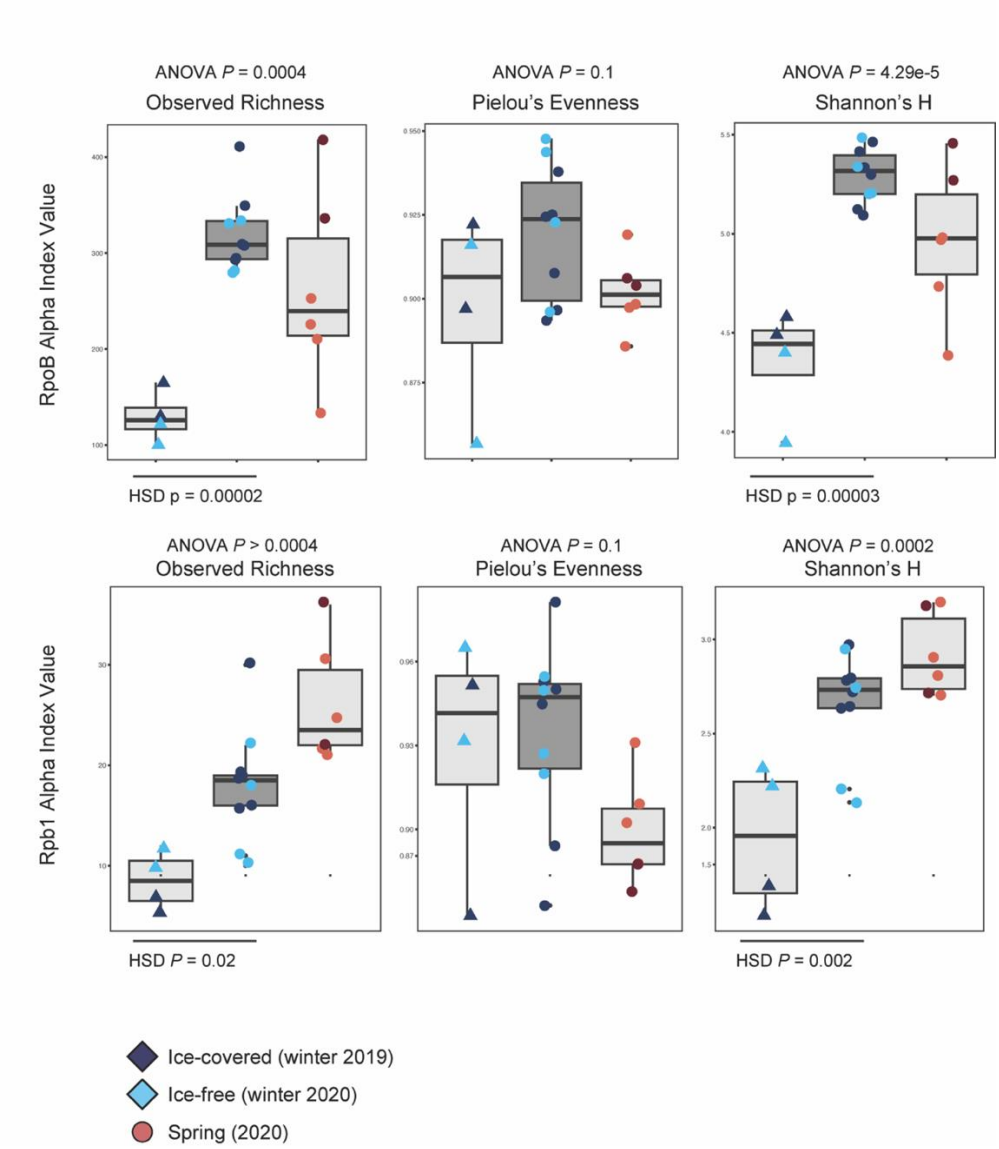
Supplemental Figure 2.9. Relative transcript abundance of viral hallmark genes in the water column samples.
TPM values were summed by hallmark gene type and are shown as a percentage.
Biological replicates are connected by horizontal bars.



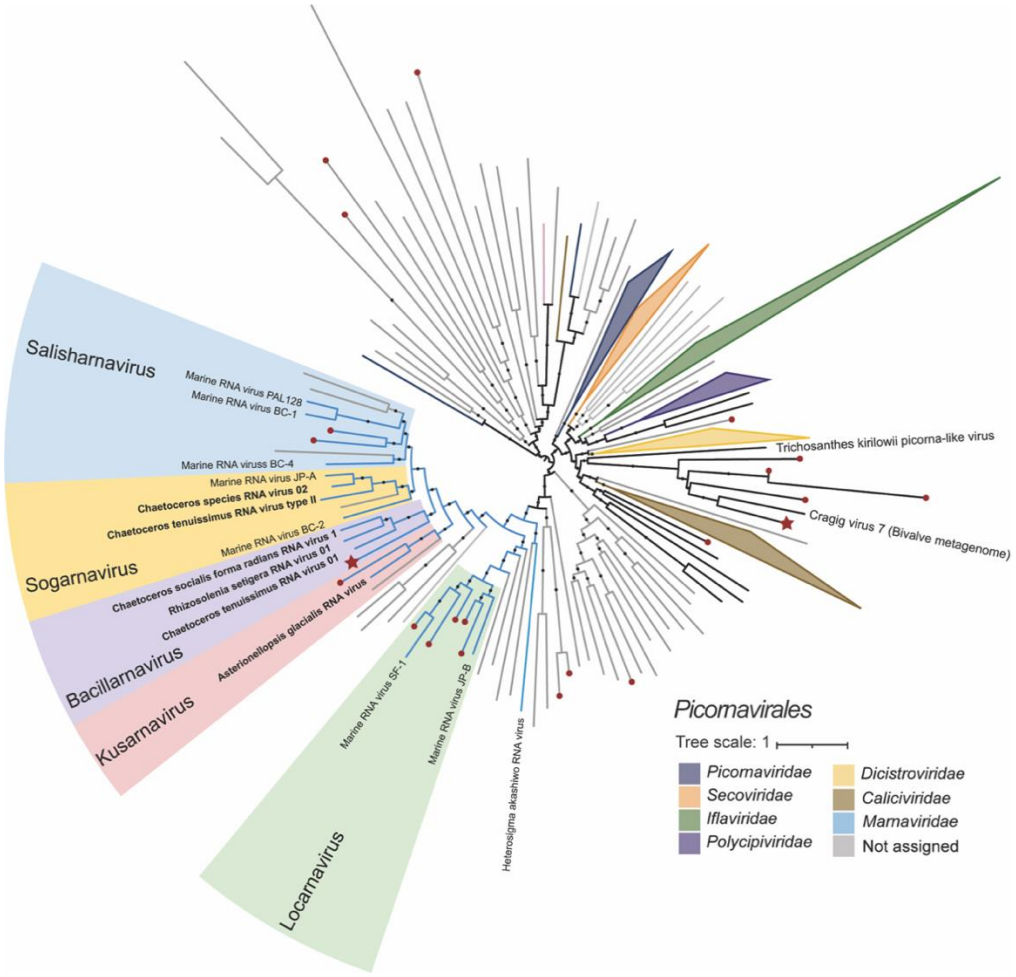
Supplemental Figure 2.10. Ordinate plot of active viral communities for each viral hallmark gene type.
nMDS plot based on Bray-Curtis similarity matrix of hallmark gene relative transcript abundance (TPM) of individual viral hallmark gene types. Ellipses represent 95% confidence intervals for the three seasons sampled (winter 2019, winter 2020, and spring 2020).



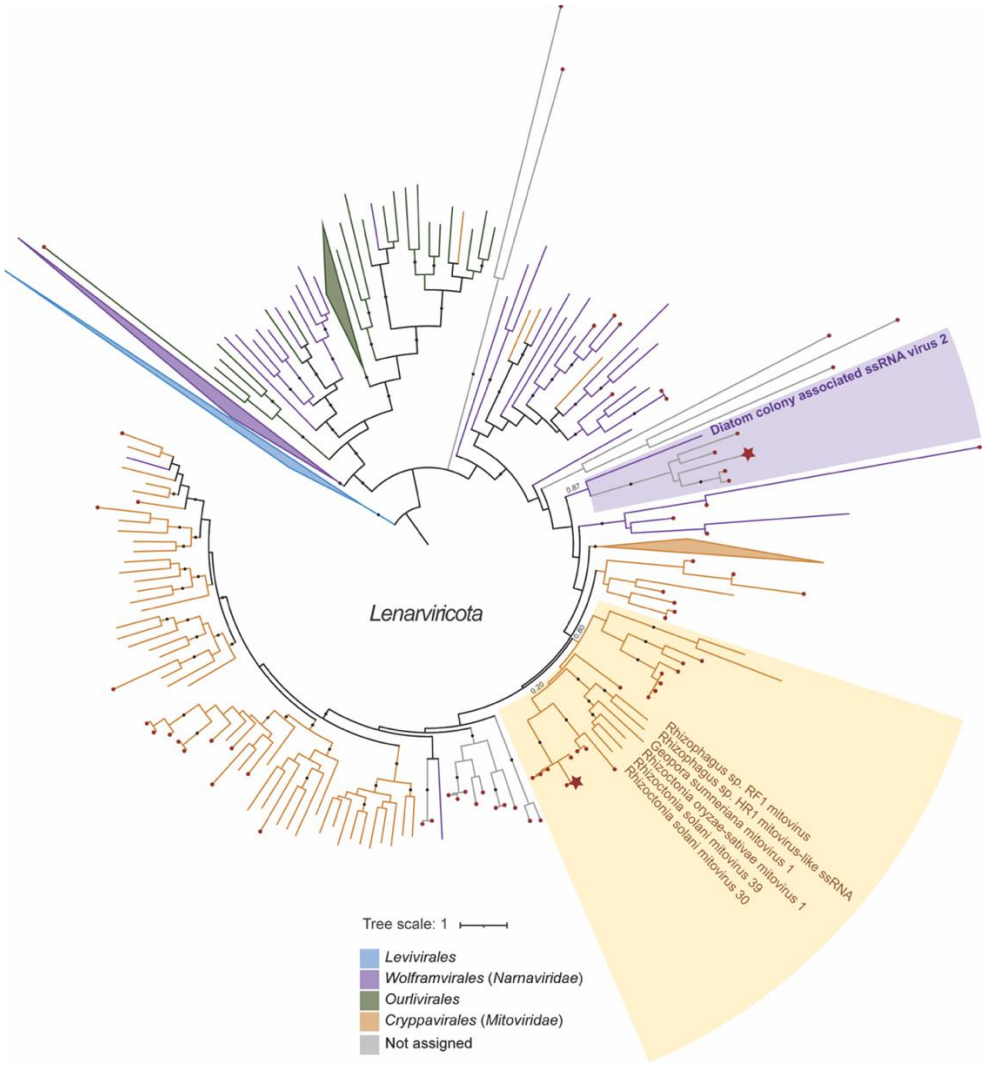
Supplemental Figure 2.11. Alpha diversity metric comparisons between seasons for each viral hallmark gene type.
A) Gp23, **B)** MCP, and **C)** RdRp. Tukey's HSD is shown for the ice-covered (winter 2019, W19) and ice-free (winter 2020, W20) comparison when applicable.



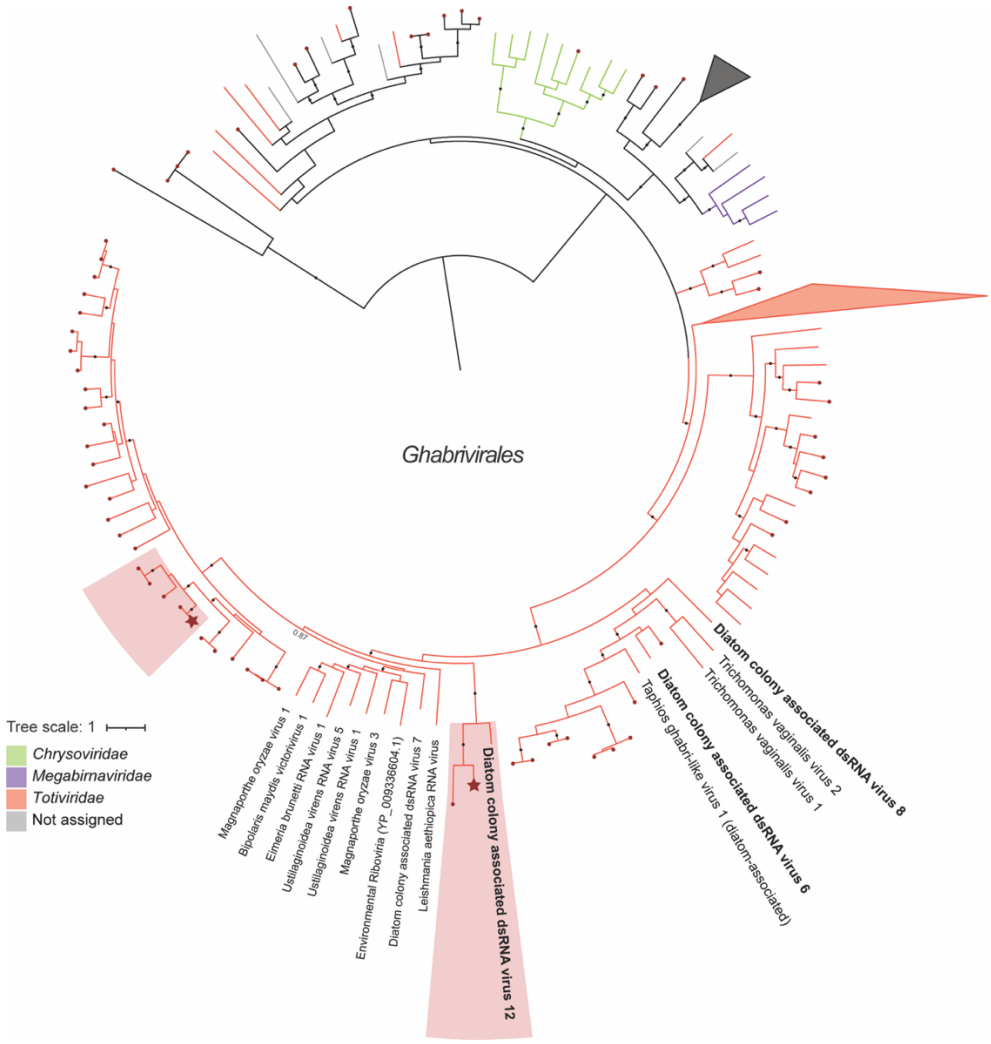
Supplemental Figure 2.12. Alpha diversity metric comparisons between seasons for microbial marker gene profiles.
Tukey's HSD is shown for the ice-covered (winter 2019, W19) and ice-free (winter 2020, W20) comparison when applicable. Top and bottom rows show plots for RpoB and RPB1 alpha diversity metrics, respectively.



Supplemental Figure 2.13. Maximum-likelihood (ML) phylogenetic placement of *Picornavirales* RdRp hallmark genes.
Nodes ending in red dots represent Lake Erie RdRp. Nodes ending in red stars represent Lake Erie RdRp with greater that 0.5% individual contribution to average dissimilarity between the ice-covered and ice-free sample groups.



Supplemental Figure 2.14. ML phylogenetic placement of *Lenarviricota* RdRp hallmark genes.
Nodes ending in red dots represent Lake Erie RdRp. Nodes ending in red stars represent Lake Erie RdRp with greater that 0.5% individual contribution to average dissimilarity between the ice-covered and ice-free sample groups.



Supplemental Figure 2.15. ML phylogenetic placement of *Ghabrivirales* RdRp hallmark genes.
Nodes ending in red dots represent Lake Erie RdRp. Nodes ending in red stars represent Lake Erie RdRp with greater that 0.5% individual contribution to average dissimilarity between the ice-covered and ice-free sample groups.

CHAPTER THREE: SHARED VIRAL SEQUENCES IN PEAT MOSS METAGENOMES REVEAL ELEMENTS OF THE *SPHAGNUM* CORE VIROME

Publication Note

Content from Chapter Three is included in a submission to a peer-reviewed journal. The contributing authors are Elizabeth R. Denison, Helena Pound, Eric Gann, Naomi Gilbert, Dale Pelletier, David Weston, and Steven W. Wilhelm.

DP and DW contributed the sequencing data. ERD conducted the analysis and drafted the manuscript. HP, EG, and NG provided feedback on the analysis. All authors contributed to revising the manuscript.

Abstract

Viruses are an understudied component of plant microbiomes despite the potential for viral infection to shape microbial composition, metabolism, and genome evolution. Viruses also have important roles in carbon and nutrient cycling at both microbial and global scales. Defining shared viral populations across plant microbiomes, *i.e.*, 'core' viromes, may reveal persistent top-down controls over plant-associated microbial communities as well as illuminate potential sources of genetic exchange for plant symbionts. We sought to identify components of the core virome associated with *Sphagnum* moss, a keystone species in Northern peatlands with direct ties to the fate of global soil carbon stocks. Complementary metagenomic and metatranscriptomic analysis uncovered shared viral contigs from *Sphagnum* field samples collected over a ten-month period, including sequences of likely phage (*Caudoviricetes*), giant virus (*Nucleocytoviricota*), and RNA virus (*Riboviria*) origin. Gene annotation suggested that certain viral elements may be maintained in the core microbiome as 'molecular hitchhikers', for example as cryptic phage or strictly intracellular RNA viruses. Additionally, sequence similarity searches linked core *Caudoviricetes* contigs to viral regions within the genomes of bacterial taxa known to dominate the peat moss core microbiome. These findings expand our understanding of the microbiome of an ecologically and climatically relevant plant species as well as provide justification for the consideration of viruses when defining plant core microbiomes.

Introduction

Peatlands serve as one of Earth's largest carbon sinks by sequestering about one-third of all terrestrial carbon in the form of peat, or partially decayed organic matter (1). The majority of the global peatland carbon stock is stored in northern peatlands, where low mean temperatures, nutrient limitation, and anoxic conditions slow the rate of decomposition and promote the accumulation of organic matter (2). However, climate warming threatens to reduce carbon capture, mobilize existing carbon stores, and, ultimately, transition peatlands from carbon sinks to sources of atmospheric carbon (3-5). Carbon sequestration in northern peatlands is largely engineered by *Sphagnum* mosses, or peat mosses, which are a dominant feature of peatland vegetation (6). *Sphagnum* spp.

both thrive in and promote the nutrient-poor and acidic conditions that contribute to slow decay rates in peatlands (7, 8). Recalcitrant *Sphagnum* litter and *Sphagnum*-derived antimicrobials are also thought to promote peat accumulation (9, 10), further supporting its role in ecosystem stability and global carbon sequestration.

The productivity and ecological dominance of *Sphagnum* is tied to its associated microbial community (11, 12). *Sphagnum* spp. harbor diverse communities of bacteria and microbial eukaryotes (protists and fungi) (11, 13), where certain functional guilds are known to perform important roles for moss health. For example, clades of diazotrophic bacteria supply fixed nitrogen to the moss in return for carbohydrates and sulfur-rich metabolites (14, 15), whereas methanotrophic symbionts provide plant-available carbon while promoting methane recycling and carbon storage in peatlands (16). Far less is known about the eukaryotic communities associated with *Sphagnum*, but they are proposed to influence the moss microbiome via grazing and nutrient transformation (17, 18). *Sphagnum* symbionts are involved in a range of other roles including pathogen suppression (19), stress tolerance (12, 20), and disease (21). Given that the microbiome directly influences moss productivity and the peatland carbon balance, factors that influence the composition and function of the *Sphagnum* microbiome are of interest.

Viral infection shapes microbial community structure and function through lysis (22), lateral gene transfer (23), and by altering the physiology of infected cells (24). Viruses are recognized mediators of carbon and nutrient cycles in aquatic and soil systems (25, 26), yet are an often-overlooked component of embryophyte microbiomes. Research on plant 'viromes' has historically focused on viruses that infect the host plant (27), whereas viruses of the plant-associated microorganisms are far less studied. However, recent work has begun to uncover the roles of viruses within plant microbiomes (28), where viruses can influence bacterial diversity (29), suppress plant pathogens (30), and may modulate diazotroph-plant interactions (31). Identifying 'core' guilds within plant microbiomes has the potential to reveal important microbial functions in plant health as well as plant-microbe evolutionary relationships (32). In *Sphagnum*, defining core bacterial communities has further highlighted the relationship between peat moss and beneficial diazotrophic clades (33, 34). No viruses that infect *Sphagnum* have

been reported to date, but metatranscriptomic analysis has suggested that viruses actively infect bacteria and microbial eukaryotes in the moss microbiome (35). Furthermore, no studies have examined shared viral communities across different *Sphagnum* samples and thus the core virome is currently unknown. Expanding our definition of the core *Sphagnum* microbiome to include viruses may help elucidate regulators of and interactions within the peat moss microbiome (28).

Here, we leveraged complementary bulk metagenomes and metatranscriptomes from living *Sphagnum* tissue collected over a ten-month period (August 2016 to June 2017) to estimate viral community composition and identify shared viral signatures between samples. Metagenomic assembly allowed for the identification of shared DNA virus contigs and *Nucleocytoviricota*, or NCLDV, metagenome-assembled genomes (MAGs). Additionally, metatranscriptomes were used to infer transcriptional activity of core DNA viruses and to identify assembled RNA virus genomes. Contig presence was generally patchy across the samples, but a small portion of the total viral contigs were shared between them and may constitute part of a *Sphagnum* core virome. While the natural hosts remain unknown, the *Caudoviricetes* core contigs were linked to phage-like regions within the genomes of *Acetobacteraceae* and *Acidobacteraceae*, two bacterial clades that dominate *Sphagnum*-associated communities. Additionally, phylogenetic analysis suggested that core NCLDV and RNA viruses could infect a wide range of micro-eukaryotes. Lastly, gene annotation of core sequences suggested possible modes of transmission and persistence of viruses within the microbiome.

Materials and Methods

Site description and sample collection

Samples were collected from the Spruce and Peatland Responses Under Changing Environments (SPRUCE) field site located at the Marcell Experimental Forest S1 bog in northern Minnesota, USA (https://mnspruce.ornl.gov). The S1 bog is a nutrient-deficient and acidic (pH 3.5 – 4) *Sphagnum*-dominated bog. Eight individual plants were collected at four different dates over a ten-month period: 1 August 2016 (n = 1), 10 November 2016 (n = 3), 18 April 2017 (n = 2), and 26 June 2017 (n = 2) and from hollow (n = 6) and lawn (n = 2) micro-topographies (Supplemental **Table 3.1**). All supplemental tables

referenced in Chapter Three can be found in Attachment 1. Only living *Sphagnum* tissue was collected (*i.e.*, capitula and approximately 3 cm of living stem). Tissue samples were flash frozen in dry ice-alcohol baths and shipped to Oak Ridge National Laboratory on dry ice where they were stored at -80°C until further processing.

Nucleic acid extraction, sequence processing and contig assembly

Genomic DNA extractions were performed using DNeasy Plant Pro Kit (QIAGEN) according to the manufacturer's instructions. Total RNA was extracted using a method combining CTAB lysis buffer and the Spectrum Total Plant RNA Kit (Sigma) as described previously (36). Library construction and sequencing were performed by Hudson Alpha (Huntsville, AL, USA). Eight PCR-free metagenome libraries (one per tissue sample) were constructed using the TruSeq DNA PCR-Free LT Library Prep Kit (Illumina). Sixteen metatranscriptome libraries (two replicates per tissue sample) were prepared using the ScriptSeq RNA-Seq Library Preparation Kit and were depleted with Ribo-Zero yeast, Ribo-Zero bacteria and Ribo-Zero plant (Illumina). Libraries were quantified by Qubit (Invitrogen) for concentration, by Agilent Bioanalyzer (Agilent Technologies) for fragment size, and by qPCR for determining the optimal loading concentration. Libraries were sequenced on the Illumina HiSeq 2500 platform (2 x 250 bp reads). DNA libraries were pooled two per channel and RNA libraries were pooled four per channel.

Raw reads were trimmed to remove adapters and quality scores < 0.3 using CLC Genomics Workbench v12.0 (QIAGEN). To reduce the number of *Sphagnum* reads prior to assembly, metagenome and metatranscriptome reads that mapped to the *S. fallax* v1.1 (GCA_021442195.1) or *S. magellanicum* v1.1 (GCA_021904315.1) genome were removed from the libraries. Read mapping was performed using BBMap v38.90 with the default settings (37). Unmapped reads comprised 64.0% ± 4.8% (mean ± standard error (SE)) of reads in the metagenome libraries and 28.5% ± 2.8% of reads in the metatranscriptome libraries (Supplemental **Table 3.1**). The unmapped reads were pooled and de novo assembled using MEGAHIT v1.2.9 (--k-min 23 --k-max 123) (metagenome and metatranscriptome reads were kept separate and assembled separately) (38). Assembly quality was assessed using QUAST v5.0.2 (39) (Supplemental **Table 3.2**).

Virus contig identification

Viral contigs were identified and annotated using geNomad, which employs a hybrid neural network- and marker-based approach to detect viral sequences in metagenomic data (40). The metagenomic and metatranscriptomic contigs were piped through geNomad to identify putative DNA virus and RNA virus genomic sequences, respectively. A length cutoff of 10 kbp for metagenomic contigs (41) and 2 kbp for metatranscriptomic contigs (42) was used. Contigs that did not meet the geNomad conservative thresholds (confidence score > 0.8, at least one viral marker gene, and a marker enrichment score > 1.5) were not included. Contigs with no viral genes called by CheckV (43) were considered false positives and removed in accordance with the developer's recommendations. Viral contigs were dereplicated at ≥ 99% average nucleotide identity using CD-HIT-EST to account for any redundant contigs (44). Viral contigs were categorized into Minimum Information about an Uncultivated Virus Genome (MIUViG) quality tiers using CheckV v0.8.1 (43). Here, we refer to the 'virome' as the set of viral sequences identified bioinformatically as opposed to in the technical sense where viral particles are concentrated prior to sequencing.

Generating NCLDV MAGs

NCLDV MAGs were binned using a workflow similar to that established by Moniruzzaman *et al*. for recovering high-confidence NCLDV genomes from metagenomes (45). Bins were generated with MetaBAT2 v2.15 (-s 100000, -m 10000, -minS 75, -maxEdges 75) using the metagenome contigs and coverage profiles from the metagenome libraries (46). Bins were screened using ViralRecall v2 (--contiglevel flag) and were retained as putative NCLDV MAGs if the mean ViralRecall contig score was positive (47). The putative NCLDV MAGs were decontaminated by removing any cellular contigs as described in Moniruzzaman *et al*. (ViralRecall score < 0 or < 3 hits to the Giant Virus Orthologous Group (GVOG) database) (45). The putative NCLDV MAGs were further decontaminated by querying translated open reading frames (ORFs) against the complete RefSeq protein database (release 213) using DIAMOND v2.0.14 BLASTP (48). Contigs were removed if they had zero ORFs with hits of NCLDV origin within the top five hits (45). NCLDV MAG completeness was estimated using both

CheckV (43) and the method described in Schulz *et al*. (49), which is based on the expected number of conserved nucleocytoplasmic virus orthologous genes (NCVOGs) for a given NCLDV clade.

Read mapping to viral contigs and NCLDV MAGs

Read mapping was used to determine the presence or absence of viral contigs within a sample. Reads were mapped to viral contigs using CoverM v0.6.1 with ≥ 90% identity and ≥ 90% read aligned thresholds (50). Contig coverage tables were generated by mapping metagenomic reads to DNA virus contigs and metatranscriptomic reads to RNA virus contigs independently. Contig coverage was calculated as the trimmed mean, where values represent the average number of mapped reads to each base-pair after trimming the most deeply and shallowly covered positions (--trim-min 5 and --trim-max 95). Coverage values were then normalized to library size using the calculation described in Emerson *et al*. (41). To account for poorly covered contigs and spurious mappings, contigs were only considered detected in a library if ≥ 75% of the contig length was covered by mapped reads following benchmarked thresholds (51). Coverage values were reset to zero for contigs with < 75% of the length covered on a per-library basis. Metatranscriptomic reads were mapped to DNA virus contigs to estimate transcriptional activity using ≥ 90% identity and ≥ 90% read aligned thresholds. The number of transcripts mapping to predicted ORFs was assessed using featureCounts from the Subread package (https://subread.sourceforge.net/). For the NCLDV MAGs, read mapping and normalization were performed using the same method as above. A MAG was considered present in a metagenome library if reads mapped to ≥ 10% of the total MAG length, as in Tithi *et al*. (52). A MAG was considered transcribed in a sample if ≥ 10% of predicted ORFs had metatranscriptome reads mapped (53).

Virus contig taxonomic annotation and phylogenies

Taxonomic assignments were estimated by geNomad v1.7.0, where contigs are assigned taxonomy based on marker gene content (40). The average nucleotide identity (ANI) between the MAGs generated here and a database of available NCLDV genomes was calculated using FastANI v1.33 (54). A database of 1,177 high-quality NCLDV

genomes (isolates and environmental MAGs) compiled by Gilbert *et al*. (55) was used for the FastANI comparisons. Taxonomic assignments were also estimated based on phylogenetic placement for the core NCLDV and RNA viruses. For the NCLDV MAGs, a multi-locus tree was constructed in accordance with the phylogenetic framework presented by Aylward *et al*. (56). The concatenated alignment of seven NCLDV marker protein sequences (DNA polymerase B, A32 packaging enzyme, superfamily II helicase, VLTF3 transcription factor, RNA polymerase large subunit, topoisomerase family II, and transcription factor IIB) was generated with ncldv_markersearch v1.1 using the default settings (56). Gaps in the alignment were trimmed with trimAl v1.2 (-gt 0.1) (57) and the tree was constructed with IQ-TREE v2.2.0.3 (LG+F+I+G4 substitution model with 1,000 ultrafast bootstraps) (58). For the RNA virus contigs, phylogenies were constructed using the RNA-dependent RNA polymerase (RdRp) gene. RdRp sequences predicted by geNomad were aligned to a custom set of reference sequences using Clustal Omega v1.2.4 (59). Trees were constructed using FastTree2 v2.1.10 (-lg) (60). All trees were visualized using iTOL.

Characterization of microbial community composition

DNA-dependent RNA polymerase was used as a marker for prokaryotic (*rpoB*) and eukaryotic (*rpb1*) genomes (61, 62). Contigs containing *rpoB*/*rpb1*-like sequences were retrieved from the metagenome co-assembly by DIAMOND v2.0.14 (BLASTX, e-value < 1e-5) using a database of *rpoB* (TIGR02013, TIGR03670) and RPB1 (CD02733, CD02584) protein sequences. To confirm the presence of *rpoB* /*rpb1*, proteins were predicted on metagenomic contigs using Prodigal v2.6.3 (63) and annotated with eggNOG-mapper (64) and the eggNOG v5.0 database (65). The top (*i.e.*, lowest e-value) BLASTP hits for the *rpoB* /*rpb1* gene were used for taxonomic estimates. Reads were mapped to *rpoB* /*rpb1* using CoverM v0.6.1 with ≥ 90% identity and ≥ 90% read aligned thresholds (50).

Results

Microbial community structure and core taxa

Prokaryotic community composition in the *Sphagnum* metagenomes was similar as to what has previously been described using 16S rRNA gene sequencing approaches

(12, 34, 66). Prokaryotic communities were consistently dominated by Proteobacteria based on both observed richness (326 *rpoB*-containing contigs) and relative abundance in the metagenomes (63.5% ± 1.4%, mean relative read abundance ± SE), followed by Acidobacteria (79 contigs, 15.5% ± 0.8%) (Supplemental **Figure 3.6**, Supplemental **Table 3.3**). All supplemental figures are located in the Chapter Three appendix. The prokaryotic community activity estimates reflected the metagenome structure, where Proteobacteria generally dominated the pool of mapped *rpoB* transcripts (42.7% ± 2.5%) (Supplemental **Figure 3.6**). Cyanobacteria displayed larger representation in the metatranscriptomes (24.9% ± 19.3%) relative to the corresponding metagenomes, a trend that has been observed previously (34). Bray-Curtis dissimilarity indices ranged from 0.29 – 0.72 between the metagenomes based on *rpoB* structure (*i.e.*, abundance of *rpoB* contigs) and from 0.11 – 0.41 based on composition (*i.e.*, presence-absence of contigs). *rpoB* community structure and composition did not cluster by micro-topography (hollow or lawn) (Supplemental **Figure 3.7**).

A 'core' prokaryotic community comprised largely of Proteobacteria (125 *rpoB* -containing contigs) and Acidobacteria (27 contigs) were detected in all eight metagenomes via read mapping (Supplemental **Figure 3.8**, Supplemental **Table 3.4**). The majority of the core *rpoB* were also detected in the metatranscriptomes and considered transcriptionally active (160 of 171 contigs) (Supplemental **Table 3.4**). Genera within the families *Rhodospiralles* (*Acetobacteraceae,* Alphaproteobacteria) and *Acidobacteriales* (Acidobacteria) have been found to dominate the North American *Sphagnum* core microbiome (34). We detected related genera in the core based on sequence similarity searches and *rpoB* phylogenetic placement, including *Acidocella*, *Acidsphaerea*, and *Acidosoma* (*Rhodospiralles*), as well as *Granulicella* (*Acidobacteriales*). Core taxa also included diazotrophic methanotrophs within the family *Beijerinckiaceae* (Alphaproteobacteria).

Micro-eukaryotic communities were dominated by phototrophic taxa, including *Chlorophyta* (green algae, 35.2% ± 5.9% of *rpb1* relative abundance) and *Bacillariophyta* (diatoms, 27.7% ± 7.0%), followed by the generally mixotrophic *Chrysophyceae* (golden algae, 9.3% ± 3.0%) and *Cryptophyta* (cryptomonads, 3.2% ± 1.5%) (Supplemental

Figure 3.6). Ascomycota fungi (8.8% ± 2.4%) and Evosea (4.8% ± 2.5%) sequences were also detected. For the micro-eukaryotic community, Bray-Curtis dissimilarity indices ranged from 0.49 – 0.95 and 0.25 – 0.89 based on *rpb1* structure and composition, respectively. Similar to the *rpoB* community, *rpb1* structure and composition did not cluster by micro-topography (Supplemental **Figure 3.7**). *Chlorophyta* and Ascomycota fungi were collectively detected in every metagenome, while *Bacillariophyta*, *Chrysophyceae*, Euglenozoa, and Evosea were detected in seven of the eight (Supplemental **Figure 3.6**). One contig (*Chlorophyta*) was detected in all eight metagenomes (Supplemental **Table 3.4**). The composition of core micro-eukaryotic assemblages in *Sphagnum* are not well defined, but *Chlorophyta*, *Chrysophyceae*, and Ascomycota fungi were similarly found to be common constituents of *Sphagnum* microbiomes sampled across various climates (13).

Identification of core DNA virus contigs

In total, 506 contigs from the metagenome assembly met the filtering thresholds (see Methods) (Supplemental **Figure 3.9,** Supplemental **Table 3.5**). Contig lengths ranged from 10.0 kb to 163.1 kb, with a mean length of 21.9 kb. The majority of the DNA virus contigs were classified by geNomad as *Caudoviricetes* (tailed dsDNA phage, 487 contigs) in addition to *Tectiliviricetes* (tail-less dsDNA phage, 5 contigs) and *Nucleocytoviricota* (NCLDV, 5 contigs) (Supplemental **Figure 3.9**). Contigs classified as *Maveriviricetes* (virophage, 1 contig), which obligately co-infect with NCLDV, and *Polintoviricetes* (polinton-like viruses, 4 contigs) were also identified. Eighteen *Caudoviricetes* contigs were predicted to be prophage by CheckV (Supplemental **Table 3.5**). The vast majority of contigs were categorized as genome fragments (~95% of contigs) (Supplemental **Figure 3.9**), as is common with short-read metagenomic sequencing (67). We note that taxonomic assignment of fragments should be interpreted with caution and we therefore regard it only as an estimate.

At the class-level, *Caudoviricetes* were a core component of the *Sphagnum* metagenomes, where collectively they dominated the set of viral contigs in every sample both in terms of count and relative abundance (**Figure 3.1**). However, viral community composition at the contig-level varied between the samples. Bray-Curtis dissimilarity

indices ranged from 0.42 – 0.90 between samples when calculated from viral contig presence-absence in the metagenomes (Supplemental **Table 3.6**). Additionally, the distribution of viral contigs in the metagenomes was generally patchy (**Figure 3.1**), where nearly 40% of all viral contigs were detected in only a single sample (Supplemental **Figure 3.10**). Five *Caudoviricetes* contigs were detected in every *Sphagnum* metagenome. We then examined the gene content, taxonomic classification, and possible lifestyle of the core viral contigs.

All five of the core *Caudoviricetes* contigs were categorized as genome fragments (Supplemental **Table 3.5**). Indeed, no individual core contig encoded all of the genomic modules that are generally present in sequenced phage genomes and required for replication (*i.e.*, the DNA replication, structural, packaging, and lysis modules) (**Figure 3.2**) (68, 69). Furthermore, no core contigs contained any predicted gene products with sequence similarity to those implicated in phage DNA replication mechanisms (*e.g.*, initiators, helicases/primases, DNA polymerases) (70). Collectively the contigs were predicted to encode genes that may belong to structural and packaging modules, as well as genes potentially involved in cell lysis and gene expression (**Figure 3.2**). Although gene content predictions were similar among the core contigs, most contigs shared little sequence similarity at the nucleotide level (0 – 11% contig coverage). Two contigs (C3 and C4) did display a higher degree of overlap (58% and 93% coverage, respectively), but with < 70% nucleotide identity and were therefore treated as distinct sequences (Supplemental **Figure 3.11**, Supplemental **Table 3.7**).

We examined shared protein content between the core contigs and RefSeq viral genomes using vConTACT2 to provide additional taxonomic resolution and to assess the relatedness of the contigs to other viral genomes (71). Three of the five contigs (C1, C3, C4) were placed in a viral cluster with Ralstonia phage Raharianne, a strain of lytic *Rahariannevirus* phages that were isolated from the agricultural plant pathogen *Ralstonia solanacearum* (order *Burkholderiales*) (72). The two remaining contigs did not cluster with any RefSeq viral genomes, but an additional contig (C2) displayed homology to *Rahariannevirus* protein sequences, namely putative structural genes (Supplemental **Figure 3.12**, Supplemental **Table 3.8**).

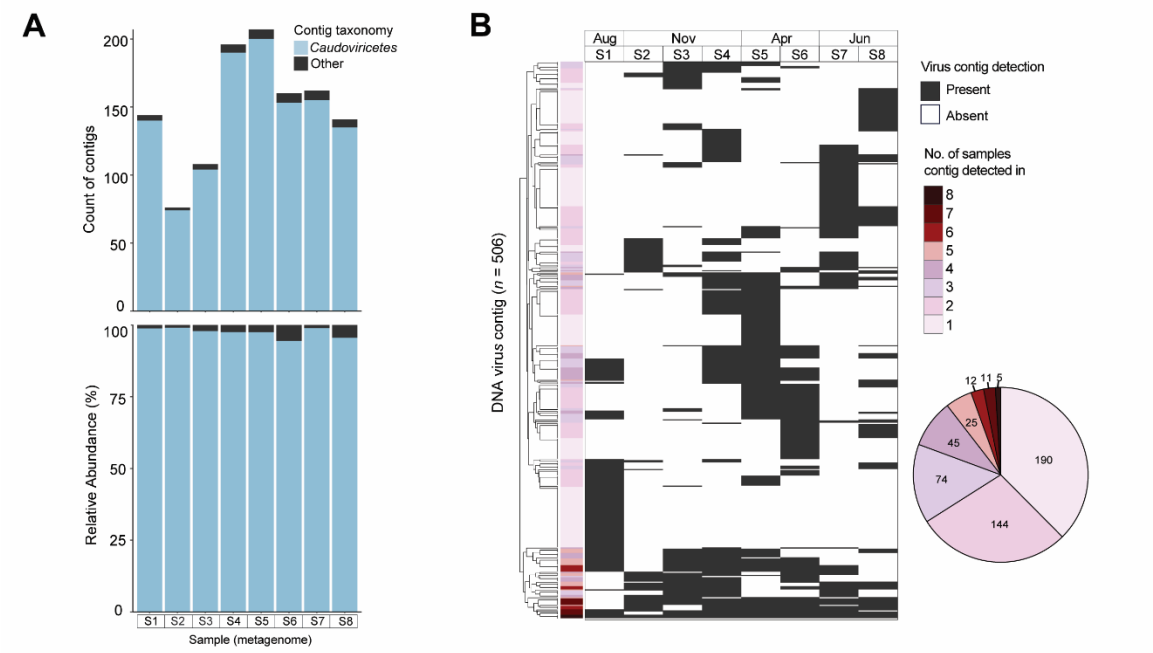


Figure 3.1. Dominate viral taxa and distribution of viral contigs in the metagenomes.

A) Contigs were grouped by geNomad-estimated taxonomy and summed by number detected in each metagenome and contig coverage values. B) Plot shows the presence or absence of individual contigs. Average linkage hierarchical clustering was applied based on contigs presence-absence. Color-coded rows indicate the number of metagenomes each contig was detected in. Pie chart shows the total number of contigs within each detection category (i.e., the number of metatranscriptomes a contig was detected in).

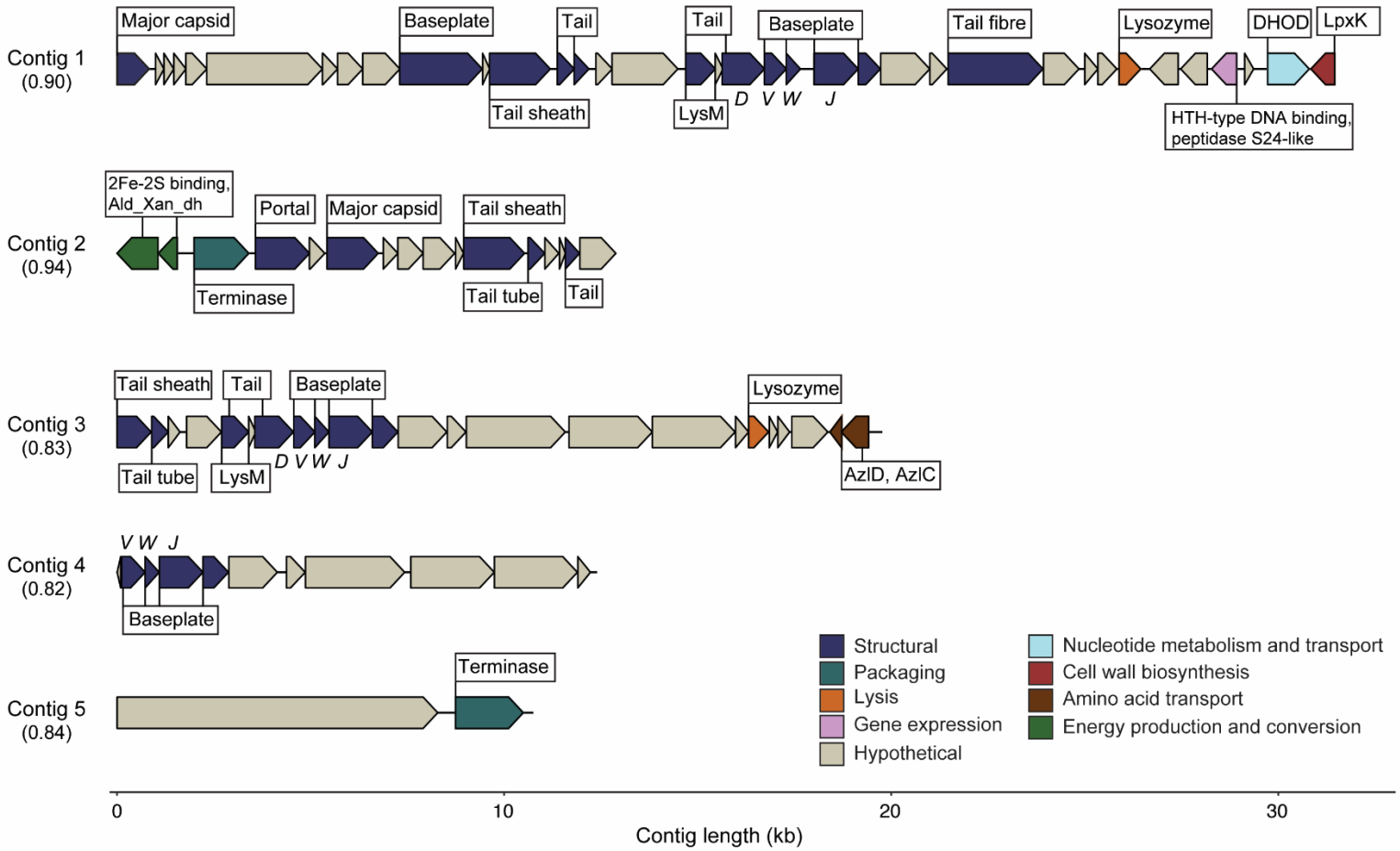


Figure 3.2. Predicted gene content of the core metagenomic viral contigs.
The value in parentheses represents the geNomad virus score. The figure was made using the gggenomes R package.

Core viral contigs share features with putative prophage or cryptic phage regions

Certain features of the core contigs suggested they may represent fragmented sequences of phage genomes integrated into host genomes (prophage) or non-inducible remnants of integrated phage (cryptic phage). Prophage can be predicted from metagenomic contigs by identifying putative microbial genes flanking viral regions (43). Three core contigs (C1, C2, C3) were categorized as prophage by CheckV and contained predicted metabolic genes that flanked virus-specific structural genes on one side of the contig (though we cannot rule out that these are phage-encoded accessory metabolic genes) (**Figure 3.2**). The putative host-encoded genes included those involved in pyrimidine biosynthesis (DHOD), amino acid transport (AzlCD), energy production and conversion (subunits of Aldehyde, CO, or xanthine dehydrogenase), and lipid A biosynthesis (LpxK) (**Figure 3.2**).

None of the core contigs contained a putative complete lysogeny control module (based on the lack of hits to integrases/recombinases, excisionases, transposases, or repressor/antirepressor gene products), but one contig (C1) did encode a putative S24-type peptidase that may function in the microbial SOS response or as a phage lytic cycle repressor. In addition, four contigs (C1 through C4) contained putative baseplate domains with homology to those conserved in P2-like and Mu-like temperate phages (gpVWJ domains) (**Figure 3.2**) (73, 74).

Although cryptic phage genes can be expressed (75), evidence of transcriptional activity of phage structural genes may indicate lytic replication of a functional virus as opposed to a defunct cryptic element (76). Metatranscriptome reads mapped to the core contigs, but no coding sequences displayed significant read recruitment or coverage to indicate they were transcribed in our samples (data not shown). We note that these contigs may indeed represent active phage populations, but that activity was not captured or was below the limit of detection at the time of sampling.

Finally, the majority of the predicted gene products on all five contigs had significant amino acid sequence similarity to bacterial sequences. We subsequently investigated if the BLASTP hits were to prophage or cryptic phage regions within RefSeq bacterial genomes. Four contigs (C1 through C4) aligned to single, continuous regions

within *Acetobacteraceae* genomes (Alphaproteobacteria), where for three of which the majority of the contig aligned at >65% percent nucleotide identity (56% to 93% contig coverage) (**Figure 3.3**, Supplemental **Table 3.9**). C1, the longest core contig, had relatively low coverage (10%) at ~73% nucleic acid identity. The final and shortest contig (C5) aligned to a region of an *Acidobacteriaceae* genome (Acidobacteria) with ~75% nucleic acid identity and 78% contig coverage (**Figure 3.3**, Supplemental **Table 3.9**).

Using the top BLASTN RefSeq genome hit as a case study, all of the RefSeq genomes (either complete genomes or scaffolds) had prophage regions identified by geNomad that corresponded to the core contig alignment region (**Figure 3.3**). Examining protein-protein comparisons (BLASTP) revealed that the predicted viral regions contained putative phage structural, DNA packaging, and lysis genes (Supplemental **Figure 3.13 – 3.16**). Aside from the shortest contig that contained only two ORFs (C5), the core contigs were similar in length to the prophage region (Figure 3.3). None of the prophage regions had the gene modules generally necessary for temperate phage excision and replication and may represent cryptic phage.

Phylogenomics and distribution of NCLDV MAGs

Five metagenomic viral contigs were classified as Nucleocytoviricota by geNomad. The length of each contig was less than 50 Kbp and no contigs were detected in all eight metagenomes (Supplemental **Table 3.5**). Since NCLDVs have characteristically large genomes that may not be contained on a single contig (77, 78), metagenome-assembled genomes (MAGs) were generated to further investigate core NCLDV in the *Sphagnum* microbiome. Three NCLDV MAGs were generated from the *Sphagnum* metagenomes, each larger than 100 Kbp and containing less than 15 contigs. The MAGs were categorized as genome fragments and ranged from 42 – 65% complete based on the presence of conserved marker genes (NCVOGs). Further details regarding the completeness and quality of each NCLDV MAG are provided in Supplemental **Table 3.10.**

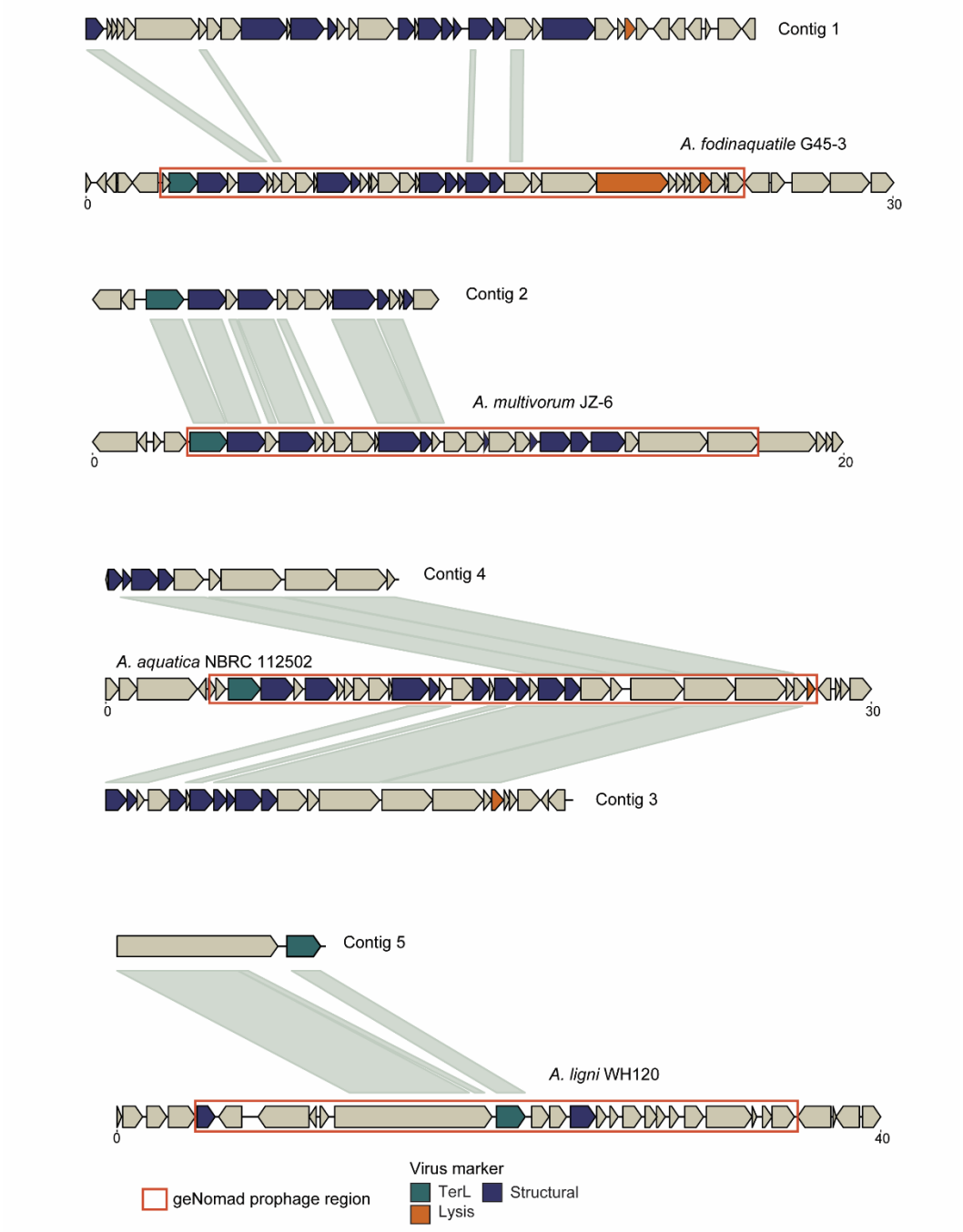


Figure 3.3. Nucleotide alignment of the core phage-like contigs to RefSeq bacterial genomes.
Regions of significant alignment are depicted by green lines. Predicted viral marker genes are in color. Predicted prophage regions in the RefSeq bacterial genomes are outlined in red..

The MAGs were most closely related to viruses of the orders *Pimascovirales* and *Asfuvirales* (**Figure 3.4**). MAG-1 fell within a sister clade to the *Pimascovirales* that contained no cultivated NCLDVs. Although it did not pass the detection thresholds in every sample, MAG-1 was present in seven of the eight metagenomes and was transcriptionally active in three samples (**Figure 3.4**, Supplemental **Table 3.11**). Notably, MAG-1 was closely related to and displayed 98% ANI with a reference NCLDV MAG (GVMAG-M-3300027807-3). The reference MAG was generated from a *Sphagnum* metagenome collected in 2015 from the same site sampled in this study (IMG Taxon ID 3300027807) (49). MAG-2 fell within a clade of cultivated viruses of the families *Iridoviridae* and *Ascoviridae* (order *Pimascovirales*) (**Figure 3.4**). Thirteen of the 18 MAG-2 proteins with top hits to RefSeq viruses were to the *Iridoviridae* (with BLASTP scores ranging from 24 to 46% percent amino acid identity), further suggesting MAG-2 is a relative of iridoviruses (Supplemental **Table 3.12**. The iridovirus-like lineage identified here may be a satellite member of the *Sphagnum* microbiome as it was detected in only a single sample (**Figure 3.4**, Supplemental **Table 3.11**).

Notably, *Pimascovirales* homologues have been discovered within NCLDV-like regions of the *Physcomitrella patens* genome, a moss species related to the *Sphagnales* (79). Remnants of pimasco-like genomes within *P. patens* suggests that the ancestors of this NCLDV lineage interacted with, and possibly infected, moss ancestors (79). Given the presence of NCLDV homologues in a moss genome along with the discovery that large segments of NCLDV genomes exist within unicellular algal genomes (76), we considered that the *Pimascovirales*-like MAGs represented endogenized viruses. Phylogenetic analysis of the DNA polymerase marker revealed that MAG-1 and MAG-2 were placed within a clade containing the P. patens NCLDV remnants, but that they were more closely related to extant *Pimascovirales* (Supplemental **Figure 3.17**). Additionally, Maumus *et al*. found that the NCLDV-like regions were transcriptionally inactive (79). Our findings that transcripts map to MAG-1 and MAG-2 provides further evidence that they are representative of a contemporary NCLDV lineage as opposed to remnants integrated into a cellular genome (MAGs ranged from 32 – 50% of ORFs with reads mapped within a sample).

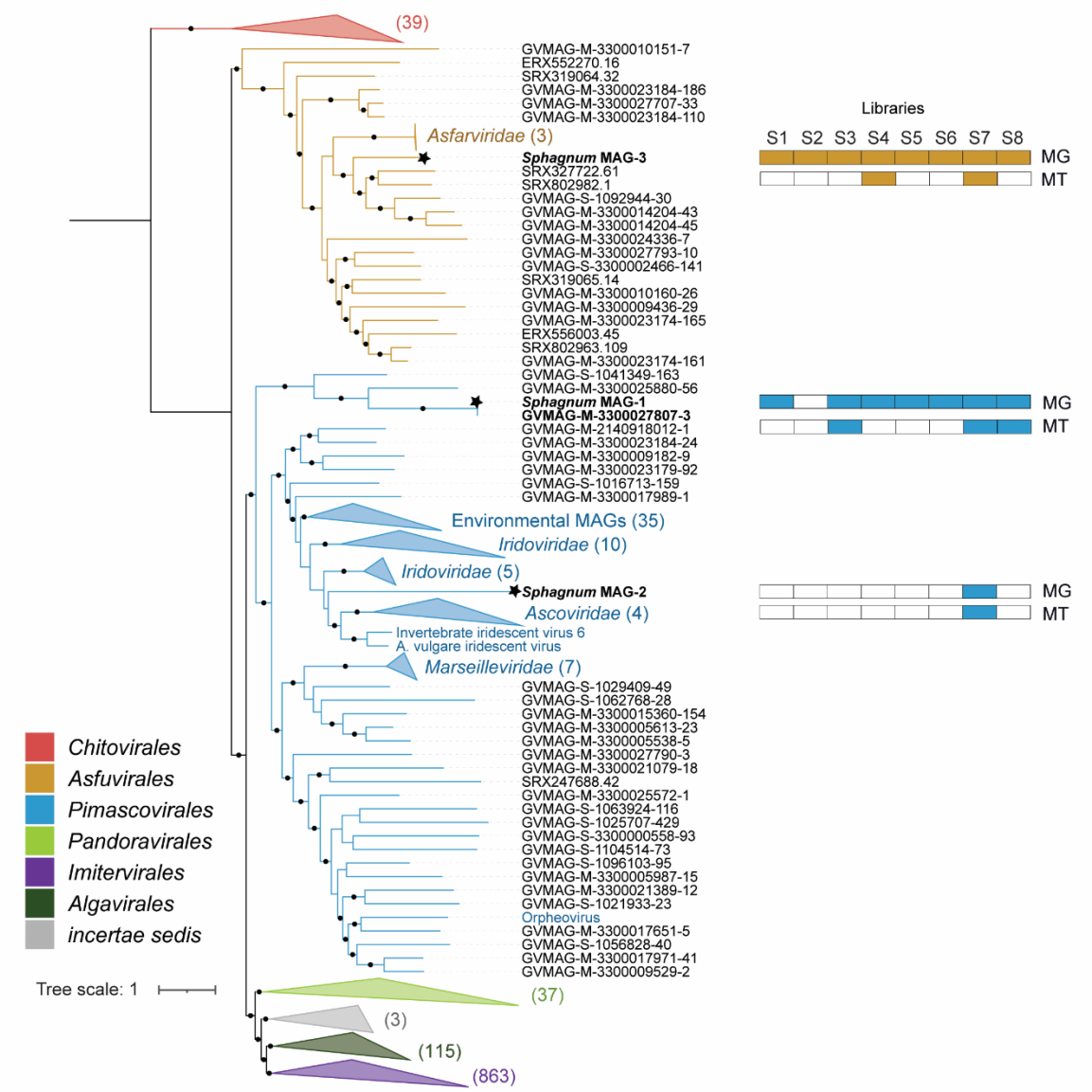


Figure 3.4. Phylogenomic analysis and distribution of the NCLDV MAGs.
The MAGs generated from this study are denoted by stars. Established NCLDV orders are color-coded. Values in parentheses represent the number of collapsed nodes. Nodes with labels beginning with GVMAG and ERX/SRX represent environmental MAGs generated from Schulz *et al*. (49) and Moniruzzaman *et al*. (45), respectively. Bootstraps ≥ 80 are denoted by dots. Filled in squares represent detection in the metagenome (MG) and metatranscriptome (MT) libraries.

Lastly, MAG-3 was placed within a clade of environmental *Asfuvirales* MAGs related to the cultivated *Asfarviridae* (represented here by the African swine fever virus (ASFV) strains) (**Figure 3.4**). Based on protein sequence similarity MAG-3 is likely related to Pacmanvirus A32 (28/36 of proteins with top hit to RefSeq virus, Supplemental **Table 3.12**), an asfarvirus relative isolated from the amoeba *Acanthamoeba castellanii* (80). MAG-3 was detected in all eight metagenomes and deemed a core NCLDV, however it was transcriptionally active in only two samples (**Figure 3.4**).

RNA virus genomes detected in metatranscriptome assembly

Screening the metatranscriptome assembly recovered 1,477 *Riboviria* contigs in total (Supplemental **Figure 3.18**, Supplemental **Table 3.13**). *Riboviria* contig lengths ranged from 2.0 kb to 16.4 kb, with a mean length of 4.4 kb. Overall, all five phyla within the kingdom *Orthornavirae* (RdRp-encoding viruses) were identified as well as a small number of *Pararnavirae* (reverse-transcriptase-encoding viruses). Within samples, most contigs were assigned to the positive-sense, single-stranded RNA virus phyla *Lenarviricota* (25-39% of contigs), *Pisuviricota* (20-33% of contigs, contains some dsRNA viruses), and *Kitrinoviricota* (17-24% of contigs) (Supplemental **Figure 3.20**).

Although the RNA virus communities detected within the metatranscriptomes were generally dominated by these three phyla, they were dissimilar based on contig composition. Samples were 61 – 95% dissimilar based on contig presence-absence (Supplemental **Table 3.14**). While the majority of contigs (~51%) were not shared between samples, nine *Riboviria* contigs were detected in every *Sphagnum* sample (Supplemental **Figure 3.19-3.20**). All contigs but one (k123_1073360) were considered high-quality draft genomes and contained the RdRp conserved catalytic motifs, which can be indicative of 'true' RNA viruses as opposed to endogenized or non-functional sequences (Supplemental **Table 3.13**). Based on the geNomad-assigned taxonomy, the core RNA virus contigs belonged to the *Lenarviricota*, *Kitrinoviricota*, and *Pisuviricota* phyla.

Based of RdRp phylogenies, five contigs were related to the *Mitoviridae* (mitoviruses) and *Narnaviridae* (narnaviruses) (**Figure 3.5A**). We note that while the core mito-like sequences were closely related, they displayed low amino acid sequence

similarity (38% identity). Mito- and narna-viruses are simple RNA viruses that were thought to exclusively infect fungi (81), but deep sequencing has revealed related sequences that may infect unicellular algae (82), invertebrate metatranscriptomes (83), and vascular plants (84). Typical for these groups, none of these core mito- and narna-like contigs encoded putative capsid proteins.

Two core RNA virus contigs were assigned to the *Durnavirales*, an order containing double-stranded (ds) RNA viruses (**Figure 3.5B**). One contig, k123_1277865, was a partiti-like sequence related to a recently proposed clade of *Ulvophyceae* (green algae)-associated dsRNA viruses (85). This clade contains Bryopsis mitochondira-associated dsRNA and related partiti-like sequences discovered from unicellular algal transcriptomes. While related to the *Ulvophyceae*-associated viruses, the RdRp was also related to partiti-like viruses that infect fungi and oomycetes (**Figure 3.5B**). The second core *Durnavirales* contig (k123_5391772) was picobirna-like and most similar to sequences associated with invertebrates and environmental sequences (**Figure 3.5B**). Although the predicted RdRp contained the catalytic motifs, the contig length and number of ORFs was irregular for picobirna-like genomes. A small nucleotide segment displayed high sequence similarity to *Sphagnum* genomes (129 bp segment with 96% nucleotide identity). Since non-retroviral RNA virus sequences have been found integrated into plant genomes (86), we investigated if the picobirna-like core sequence was endogenized in the moss genome. Mapping the un-partitioned metagenome libraries (*i.e.*, *Sphagnum* reads and microbial reads) showed that reads only mapped to the 129 bp region and not to the viral like region (data not shown). Additionally, the hypothetical proteins had no match in the *S. fallax* or *S. magellanicum* protein sequences nor to any proteins in the NCBI NR database (BLASTP, e-value < 1e-4). Neither Durnavirales contig encoded recognizable coat proteins.

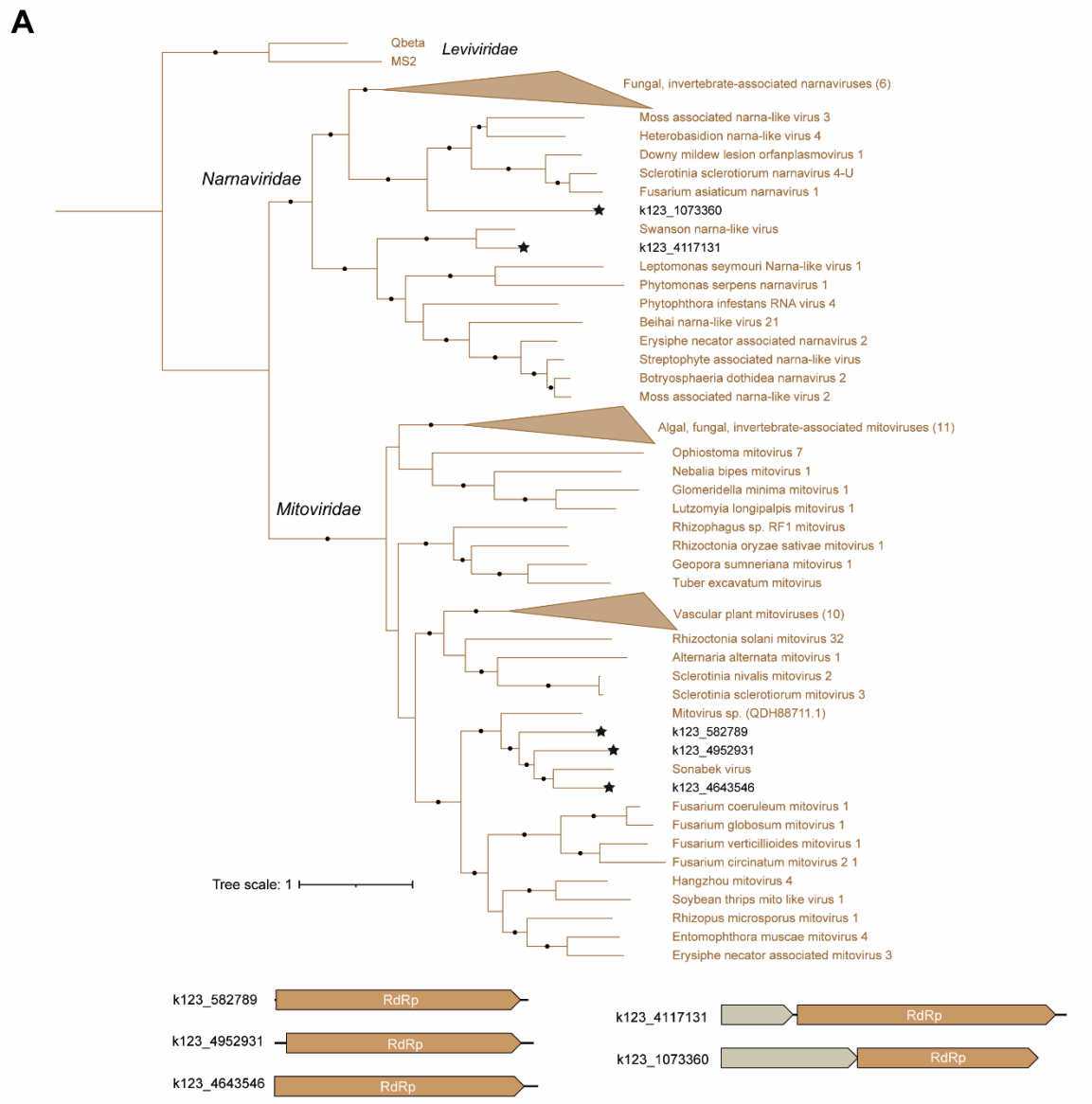


Figure 3.5. Phylogenies and genome organization of the core RNA virus contigs. ML phylogenetic trees are shown for the following phyla: A) *Narnaviridae*/*Mitoviridae*, B) *Durnavirales*, C) *Hepelivirales*, and D) *Picornavirales*. Bootstraps ≥ 80 are depicted as dots. Some monophyletic clades are collapsed for visibility of the core RdRp. The number in parentheses outside of collapsed clade represents the number of collapsed nodes. NCBI protein accession is provided in parenthesis for references with non-descript names. Contigs were annotated using geNomad and BLASTP hits and visualized using the gggenomes R package.

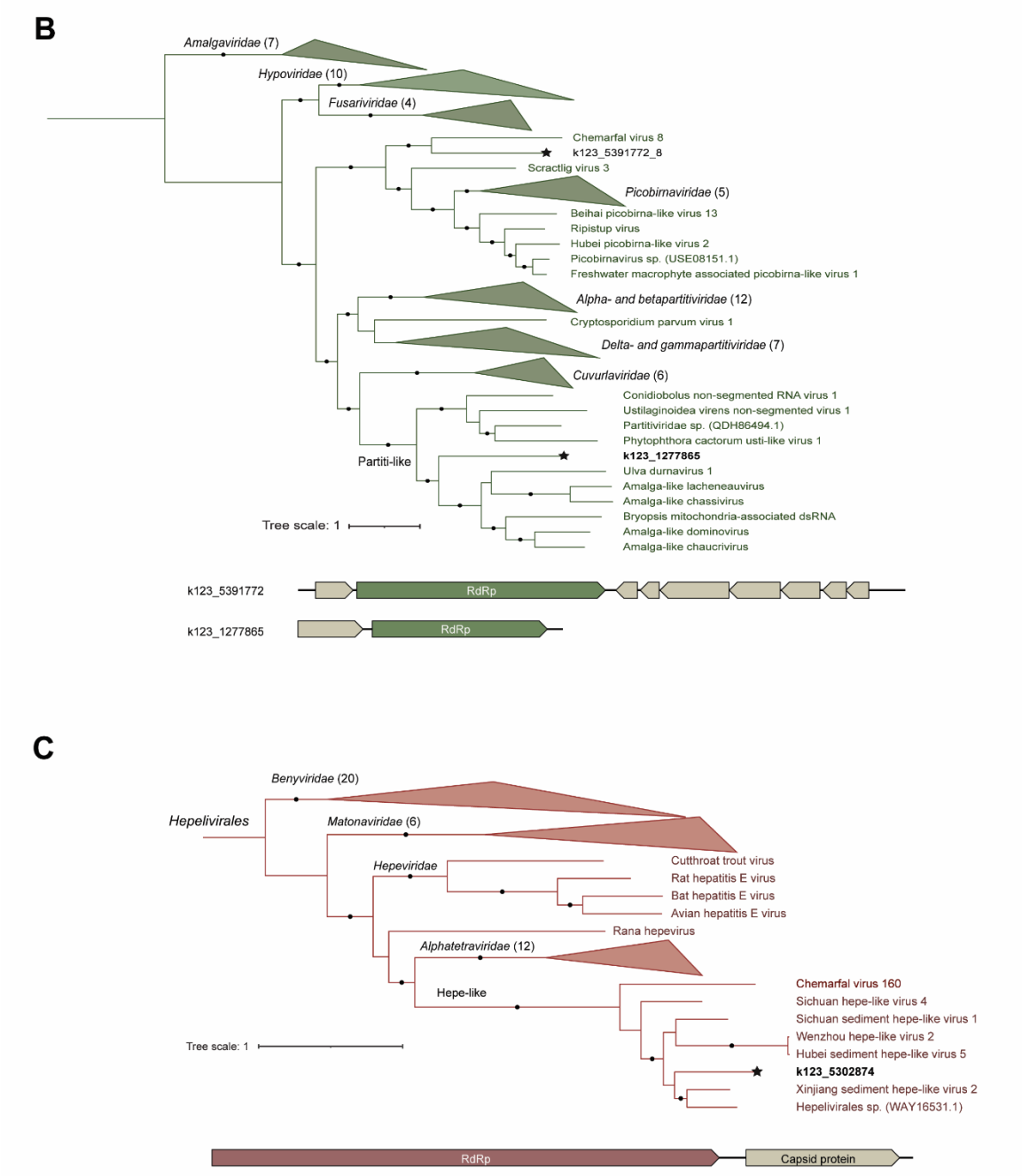


Figure 3.5. Phylogenies and genome organization of the core RNA virus contigs.
Continued from page 93.

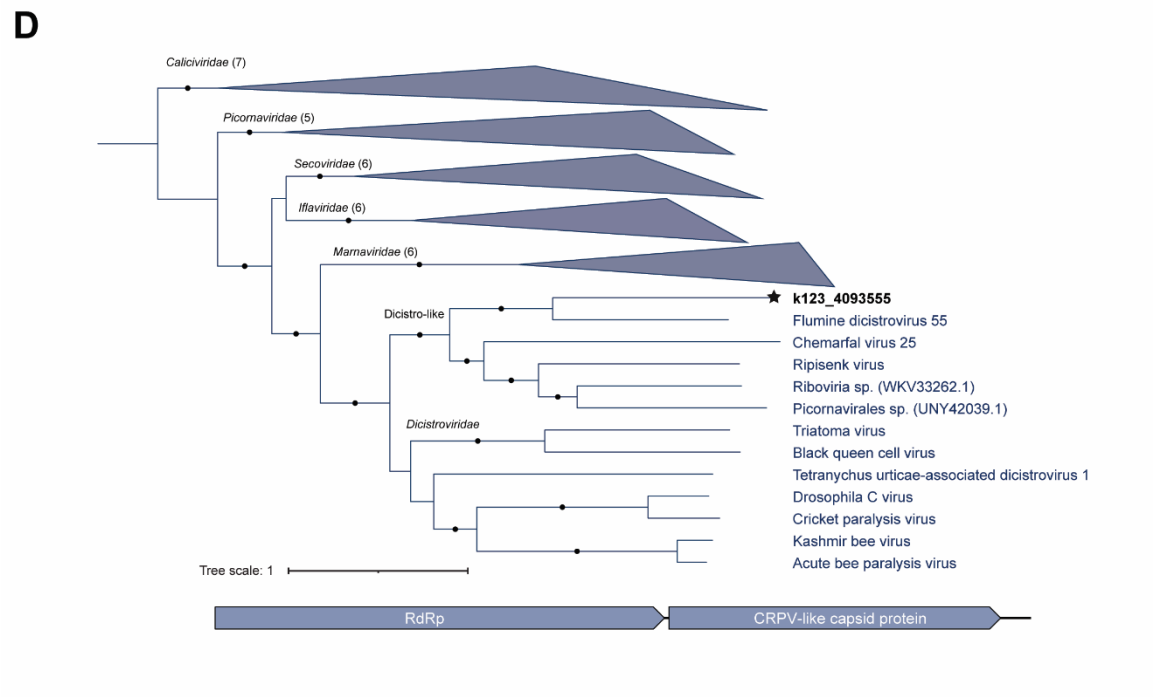


Figure 3.5. Phylogenies and genome organization of the core RNA virus contigs.
Continued from page 93.

One contig (k123_4093555) was dicistro-like (*Pisuviricota*) based on phylogenetic analysis of the RdRp (**Figure 3.5D**). The *Dicistroviridae* encompass ssRNA (+) viruses known to infect invertebrates, although the core sequence here was most closely related to a sequence from a freshwater metagenome where the natural host is unknown. Lastly, the final contig (k123_5302874) was placed within the *Hepelivirales* (phylum *Kitrinoviricota*) (**Figure 3.5C**), ssRNA (+) viruses whose members are known to infect vertebrates, invertebrates, and plants. The RdRp was placed within a clade of divergent environmental and invertebrate-associated hepe-like sequences. Both the dicistro- and hepe-like contigs encoded putative capsid proteins.

Discussion

Tissue samples from individual moss collected over a ten-month period allowed us to investigate the presence of a 'core virome' present in association with *Sphagnum*. Our metagenomic and metatranscriptomic approach identified viral sequences that were detected in all samples and putatively originated from phage, NCLDV, and RNA viruses, indicating a previously neglected viral element of the core *Sphagnum* microbiome.

Features of the core *Caudoviricetes* contigs suggested they may be representative of either prophage or cryptic phage elements. These features included the presence of flanking metabolic genes, conserved temperate phage baseplate motifs, and, perhaps most convincingly, the presence of highly similar regions in sequenced *Acetobacteraceae* and *Acidobacteriaceae* genomes. As mentioned previously, genera within the *Acetobacteraceae* (*Acidocella*, *Acidisphaera*, and *Acidisoma*) and *Acidobacteriaceae* (*Granulicella*) are dominant members of the core *Sphagnum* microbiome (34) and were also detected in the core defined in this study. While phage elements within these taxa have not been studied in peat moss symbionts specifically, prophage regions have been reported in *Acetobacteraceae* genomes (87) and Acidobacteria harbor a recognized phage mobilome that is implicated in host survival in harsh environments (88, 89). Similarly, Bragina *et al*. observed that *Sphagnum* metagenomes were enriched in mobile genetic elements, including temperate phage genes, suggesting that the peat moss microbiome may be highly plastic in part due to phage-mediated gene transfer (12). From our analysis it is unclear if the putative phage regions would provide any specific selective advantage

for microbial symbionts that harbor them, but in general prophage and cryptic phage have the ability to impact host cell physiology as well as provide protection against phage infection (90, 91). While we observed potentially cryptic phage regions in the core, our analysis utilized RefSeq genomes. Future investigations into viral regions within sequenced genomes from peat moss cultivated isolates may provide more insight into the prevalence of lysogeny and phage-mediated gene transfer in the *Sphagnum* microbiome. Nonetheless, homology between the *Caudoviricetes* contigs and RefSeq bacterial genomes suggests a gene sharing network between bacteria and phage in the core microbiome that may have the potential to influence plant-symbiont interactions.

The *Caudoviricetes* contigs shared predicted gene content with lytic phages of the phytopathogen *Ralstonia solanacearum* (order *Burkholderiales*) (72). No phytopathogenic *Burkholderiales* have been described in association with peat mosses, but other members of the order are commonly associated with *Sphagnum* and may suppress pathogens, produce plant growth-promoting hormones, and supply plant-available forms of phosphorus and nitrogen (34, 92, 93). *Burkholderiales* phages in the core moss microbiome may therefore have indirect influence over moss growth by modulated these processes through infection of beneficial symbionts. However, phage genomes have been shown to contain homologues to other viruses that infect distantly related host lineages (69, 94). Therefore, the sequence similarity to *Burkholderiales* phages could indicate that the core contigs also infect related members of this bacterial lineage, or that lateral gene transfer has occurred between distinct phages with relatively unrelated hosts. An additional consideration is that while protein clustering can indicate viral genus-level relationships (71), the core contigs likely represent fragments of phage genomes. Since phage genomes are often mosaic in nature it is unknown if the genomic architecture of the core contigs would further resemble Rahariannevirus-like phages if complete genomes were assembled (94).

The presence of putative integrated phage elements and capsid-less RNA viruses in the core virome prompts hypotheses regarding the modes of acquisition and transmission of viruses into the core moss microbiome. For example, prophage or cryptic phage elements may have be detected in the core because they are present in the genomes

of conserved bacterial taxa, in this case *Acetobacteraceae* and *Acidobacteriaceae* symbionts. Similarly, capsid-less RNA viruses often perform strictly intracellular lifestyles and may be retained in the microbiome as 'molecular hitchhikers' (95) of core eukaryotic taxa, where, for example, most RNA mycoviruses are only transmitted via hyphae or asexual reproduction (96). For viruses with extracellular stages of replication, which may include the viruses related to the capsid-encoding contigs detected in the core, viral particles could plausibly be transmitted vertically from the moss sporophyte to gametophyte or be acquired from the surrounding peat and water, as is thought to occur in the case of certain *Sphagnum* bacterial symbionts (33, 66). Efforts to compare the virome of living *Sphagnum* tissue to the surrounding soils as well as screening for viral sequences within sporophytes could help to infer methods of acquisition and persistence of core viruses.

Our analysis identified putative lineages of *Pimascovirales* and *Asfuvirales* that may be shared members of moss-associated communities. *Asfuvirales*-like transcripts have also been previously reported in *Sphagnum* metatranscriptomes (30), further suggesting that asfarvirus-like NCLDV are persistent and actively replicating in the *Sphagnum* microbiome. While the cultivated *Asfarviridae* were originally isolated from pigs, recent work has expanded the known host range of asfarviruses to include unicellular dinoflagellates and amoeba (97, 98), and sequence similarity searches linked the *Asfuvirales* MAG to distantly related viruses of amoeba. The detection of a divergent *Pimascovirales* lineage in nearly all of the metagenomes here combined the assembly of a highly similar MAG from a separate moss metagenome provides evidence that this divergent *Pimascovirales* lineage may also be a shared viral constituent of the *Sphagnum* microbiome. The *Pimascovirales* are known to infect a broad range of organisms, including animals (invertebrates and vertebrates) and amoeba, and their detection in marine metagenomes suggests the order may also include viruses of other undetermined protists (99-101). Additionally, the *Pimascovirales* lineages may not only be a part of the contemporary *Sphagnum* core virome but may also represent extant relatives of an ancestral NCLDV lineage that could have a long history of interacting with *Sphagnum* in

some capacity. Additional work is required to determine if there is evidence of past interactions and integration into the *Sphagnum* genome as is the case for *P. patens*.

Since the hosts of the NCLDV MAGs and RNA viruses are unknown, the potential roles of eukaryotic viruses in the core microbiome are difficult to discern without this crucial context. In general, micro-eukaryotes in the *Sphagnum* microbiome are understudied relative to bacterial symbionts (11), but some protistan guilds likely impact microbial dynamics and plant-microbiome interactions through grazing (102). Non-grazing protists may also impact the structure and function of the microbiome. For example, algal exudates have been shown to alleviate carbon limitation and stimulate heterotrophic bacterial metabolism in peatland surface soils (103). It remains unclear how exactly eukaryotic viruses may modulate microbial structure, activity, and trophic interactions in the *Sphagnum* microbiome, but further investigations into the core lineages identified here may illuminate their roles.

In conclusion, we identified shared viral elements across different *Sphagnum* samples from a ten-month period, indicative of a core moss virome. One consideration of this study is that our defined core is based on specific cut-offs and what was detectable at the time of sampling. In addition, sequence homology-based searches some very divergent viruses potentially in the core may not be detected. Therefore, this core is likely not a comprehensive list but instead highlights persistent viral elements in the *Sphagnum* microbiome. Future work incorporating more *Sphagnum* samples, larger geographic areas, and longer timeframes may illuminate how widespread certain viral lineages and elements are. The presence of core viruses warrants their inclusions into future studies of core microbiomes, both of *Sphagnum* and other plants, as it could help to understand ecological and evolutionary relationships between plants and their microbiomes.

Acknowledgements

This work was supported by a US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program Grant under Award Number DE-SC0020362. A portion of this research was sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S.

Department of Energy under contract DE-AC05-00OR22725. SPRUCE is a cooperative venture between ORNL and the USDA Forest Service with interested university cooperators. We also acknowledge the University of Tennessee-Oak Ridge Innovation Institute's Science Alliance for their support through the Graduate Advancement, Training and Education (GATE) program.

References

1. Yu ZC, Loisel J, Brosseau DP, Beilman DW, Hunt SJ. Global peatland dynamics since the Last Glacial Maximum. Geophys Res Lett. 2010;37:5.

2. Limpens J, Berendse F, Blodau C, Canadell JG, Freeman C, Holden J, *et al*. Peatlands and the carbon cycle: from local processes to global implications – a synthesis. Biogeosciences. 2008;5(5):1475-91.

3. Hugelius G, Loisel J, Chadburn S, Jackson RB, Jones M, MacDonald G, *et al*. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. Proceedings of the National Academy of Sciences. 2020;117(34):20438-46.

4. Gallego-Sala AV, Charman DJ, Brewer S, Page SE, Prentice IC, Friedlingstein P, *et al*. Latitudinal limits to the predicted increase of the peatland carbon sink with warming. Nature climate change. 2018;8(10):907-13.

5. Hopple AM, Wilson RM, Kolton M, Zalman CA, Chanton JP, Kostka J, *et al*. Massive peatland carbon banks vulnerable to rising temperatures. Nat Commun. 2020;11(1):2373.

6. Rydin H, Sundberg S, Gunnarsson U. The Role of *Sphagnum* in Peatland Development and Persistence. In: Wieder RK, Vitt, D.H., editor. Boreal Peatland Ecosystems. 188: Springer, Berlin, Heidelberg; 2006.

7. Clymo RS, Hayward PM. The Ecology of *Sphagnum*. In: Smith AJE, editor. Bryophyte Ecology. Dordrecht: Springer Netherlands; 1982. p. 229-89.

8. Malmer N, Svensson BM, Wallén B. Interactions between *Sphagnum* Mosses and Field Layer Vascular Plants in the Development of Peat-Forming Systems. Folia Geobotanica & Phytotaxonomica. 1994;29(4):483-96.

9. Hájek T, Ballance S, Limpens J, Zijlstra M, Verhoeven JTA. Cell-wall polysaccharides play an important role in decay resistance of *Sphagnum* and actively depressed decomposition in vitro. Biogeochemistry. 2011;103(1):45-57.

10. Verhoeven J, Liefveld W. The ecological significance of organochemical compounds in *Sphagnum*. Acta Botanica Neerlandica. 1997;46(2):117-30.

11. Kostka JE, Weston DJ, Glass JB, Lilleskov EA, Shaw AJ, Turetsky MR. The *Sphagnum* microbiome: new insights from an ancient plant lineage. New Phytol. 2016;211(1):57-64.
12. Bragina A, Oberauner-Wappis L, Zachow C, Halwachs B, Thallinger GG, Muller H, *et al*. The *Sphagnum* microbiome supports bog ecosystem functioning under extreme conditions. Mol Ecol. 2014;23(18):4498-510.
13. Singer D, Metz S, Unrein F, Shimano S, Mazei Y, Mitchell EAD, *et al*. Contrasted Micro-Eukaryotic Diversity Associated with *Sphagnum* Mosses in Tropical, Subtropical and Temperate Climatic Zones. Microbial Ecology. 2019;78(3):714-24.
14. Berg A, Danielsson Å, Svensson BH. Transfer of fixed-N from N2-fixing cyanobacteria associated with the moss *Sphagnum riparium* results in enhanced growth of the moss. Plant Soil. 2013;362(1):271-8.
15. Carrell AA, Veličković D, Lawrence TJ, Bowen BP, Louie KB, Carper DL, *et al*. Novel metabolic interactions and environmental conditions mediate the boreal peatmoss-cyanobacteria mutualism. The ISME Journal. 2021;16(4):1074-85.
16. Raghoebarsing AA, Smolders AJP, Schmid MC, Rijpstra WIC, Wolters-Arts M, Derksen J, *et al*. Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. Nature. 2005;436(7054):1153-6.
17. Jassey VEJ, Signarbieux C, Hattenschwiler S, Bragazza L, Buttler A, Delarue F, *et al*. An unexpected role for mixotrophs in the response of peatland carbon cycling to climate warming. Sci Rep. 2015;5:10.
18. Wieczynski DJ, Yoshimura KM, Denison ER, Geisen S, DeBruyn JM, Shaw AJ, *et al*. Viral infections likely mediate microbial controls on ecosystem responses to global warming. FEMS Microbiol Ecol. 2023;99(3):fiad016.
19. Shcherbakov AV, Bragina AV, Kuzmina EY, Berg C, Muntyan AN, Makarova NM, *et al*. Endophytic bacteria of *Sphagnum* mosses as promising objects of agricultural microbiology. Microbiology. 2013;82(3):306-15.
20. Carrell AA, Lawrence TJ, Cabugao KGM, Carper DL, Pelletier DA, Lee JH, *et al*. Habitat-adapted microbial communities mediate *Sphagnum* peatmoss resilience to warming. New Phytol. 2022.

21. Tsuneda A, Chen MH, Currah RS. Characteristics of a disease of *Sphagnum fuscum* caused by *Scleroconidioma sphagnicola*. Canadian Journal of Botany. 2001;79(10):1217-24.
22. Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? Environ Microbiol. 2004;6(1):1-11.
23. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405(6784):299-304.
24. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. Nature. 2003;424(6950):741-.
25. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. Nature. 1999;399(6736):541-8.
26. Albright MBN, Gallegos-Graves LV, Feeser KL, Montoya K, Emerson JB, Shakya M, *et al*. Experimental evidence for the impact of soil viruses on carbon cycling during surface plant litter decomposition. ISME Communications. 2022;2(1):24.
27. Dolja VV, Krupovic M, Koonin EV. Deep Roots and Splendid Boughs of the Global Plant Virome. In: Leach JE, Lindow SE, editors. Annual Review of Phytopathology, Vol 58. Annual Review of Phytopathology. 58. Palo Alto: Annual Reviews; 2020. p. 23-53.
28. Koskella B, Taylor TB. Multifaceted Impacts of Bacteriophages in the Plant Microbiome. In: Leach JE, Lindow SE, editors. Annual Review of Phytopathology, Vol 56. Annual Review of Phytopathology. 56. Palo Alto: Annual Reviews; 2018. p. 361-80.
29. Morella NM, Gomez AL, Wang G, Leung MS, Koskella B. The impact of bacteriophages on phyllosphere bacterial abundance and composition. Mol Ecol. 2018;27(8):2025-38.
30. Yang K, Wang X, Hou R, Lu C, Fan Z, Li J, *et al*. Rhizosphere phage communities drive soil suppressiveness to bacterial wilt disease. Microbiome. 2023;11(1):16.
31. Joglekar P, Ferrell Barbra D, Jarvis T, Haramoto K, Place N, Dums Jacob T, *et al*. Spontaneously Produced Lysogenic Phages Are an Important Component of the Soybean *Bradyrhizobium* Mobilome. mBio. 2023;14(2):e00295-23.

32. Lemanceau P, Blouin M, Muller D, Moënne-Loccoz Y. Let the Core Microbiota Be Functional. Trends Plant Sci. 2017;22(7):583-95.
33. Bragina A, Cardinale M, Berg C, Berg G. Vertical transmission explains the specific *Burkholderia* pattern in *Sphagnum* mosses at multi-geographic scale. Front Microbiol. 2013;4.
34. Kolton M, Weston DJ, Mayali X, Weber PK, McFarlane KJ, Pett-Ridge J, *et al*. Defining the *Sphagnum* Core Microbiome across the North American Continent Reveals a Central Role for Diazotrophic Methanotrophs in the Nitrogen and Carbon Cycles of Boreal Peatland Ecosystems. mBio. 2022;13(1):17.
35. Stough JMA, Kolton M, Kostka JE, Weston DJ, Pelletier DA, Wilhelm SW. Diversity of Active Viral Infections within the *Sphagnum* Microbiome. Appl Environ Microbiol. 2018;84(23):16.
36. Timm CM, Pelletier DA, Jawdy SS, Gunter LE, Henning JA, Engle N, *et al*. Two Poplar-Associated Bacterial Isolates Induce Additive Favorable Responses in a Constructed Plant-Microbiome System. Front Plant Sci. 2016;7:10.
37. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. United States2014.
38. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6.
39. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.
40. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, *et al*. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2023.
41. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, *et al*. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol. 2018;3(8):870.
42. Camargo AP, Nayfach S, Chen IMA, Palaniappan K, Ratner A, Chu K, *et al*. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. Nucleic Acids Res. 2023;51(D1):D733-D43.

43. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39(5):578-+.
44. Li WZ, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.
45. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat Commun. 2020;11(1):11.
46. Kang DWD, Li F, Kirton E, Thomas A, Egan R, An H, *et al*. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:13.
47. Aylward FO, Moniruzzaman M. ViralRecall - A flexible command-line tool for the detection of giant virus signatures in ‘omic data. Viruses-Basel. 2021;13(2):13.
48. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366-+.
49. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denef VJ, *et al*. Giant virus diversity and host interactions through global metagenomics. Nature. 2020;578(7795):432-+.
50. Aroney STN, Newell RJP, Nissen J, Camargo AP, Tyson GW, Woodcroft BJ. CoverM: Read coverage calculator for metagenomics. 2024.
51. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5:26.
52. Tithi SS, Aylward FO, Jensen RV, Zhang LQ. FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. PeerJ. 2018;6:18.
53. Ha AD, Moniruzzaman M, Aylward FO. High Transcriptional Activity and Diverse Functional Repertoires of Hundreds of Giant Viruses in a Coastal Marine System. mSystems. 2021;6(4):15.

54. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114.
55. Gilbert NE, LeCleir GR, Pound HL, Strzepek RF, Ellwood MJ, Twining BS, *et al*. Giant virus infection signatures are modulated by euphotic zone depth strata and iron regimes of the subantarctic Southern Ocean. mSystems. 2023.
56. Aylward F, Moniruzzaman M, Ha A, Koonin E. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLoS Biol. 2021;19(10):18.
57. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3.
58. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74.
59. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. In: Russell DJ, editor. Multiple Sequence Alignment Methods. Methods in Molecular Biology. 1079. Totowa: Humana Press Inc; 2014. p. 105-16.
60. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):10.
61. Ogier JC, Pages S, Galan M, Barret M, Gaudriault S. rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. BMC Microbiol. 2019;19:16.
62. Stiller JW, Harrell L. The largest subunit of RNA polymerase II from the *Glaucocystophyta*: Functional constraint and short-branch exclusion in deep eukaryotic phylogeny. BMC Evol Biol. 2005;5:17.
63. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:11.

64. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825-9.
65. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, *et al*. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309-D14.
66. Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V, Berg G. *Sphagnum* mosses harbour highly specific bacterial diversity during their whole lifecycle. The ISME Journal. 2012;6(4):802-13.
67. Roux S, Paez-Espino D, Chen IMA, Palaniappan K, Ratner A, Chu K, *et al*. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. Nucleic Acids Res. 2021;49(D1):D764-D75.
68. Brüssow H, Desiere F. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol. 2001;39(2):213-23.
69. Comeau AM, Bertrand C, Letarov A, Tétart F, Krisch HM. Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery. Virology. 2007;362(2):384-96.
70. Weigel C, Seitz H. Bacteriophage replication modules. Fems Microbiol Rev. 2006;30(3):321-81.
71. Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, *et al*. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37(6):632-+.
72. Trotereau A, Boyer C, Bornard I, Pecheur MJB, Schouler C, Torres-Barcelo C. High genomic diversity of novel phages infecting the plant pathogen *Ralstonia solanacearum*, isolated in Mauritius and Reunion islands. Sci Rep. 2021;11(1):10.
73. Christie GE, Calendar R. Bacteriophage P2. Bacteriophage. 2016;6(1):e1145782.
74. Büttner CR, Wu Y, Maxwell KL, Davidson AR. Baseplate assembly of phage Mu: Defining the conserved core components of contractile-tailed phages and related

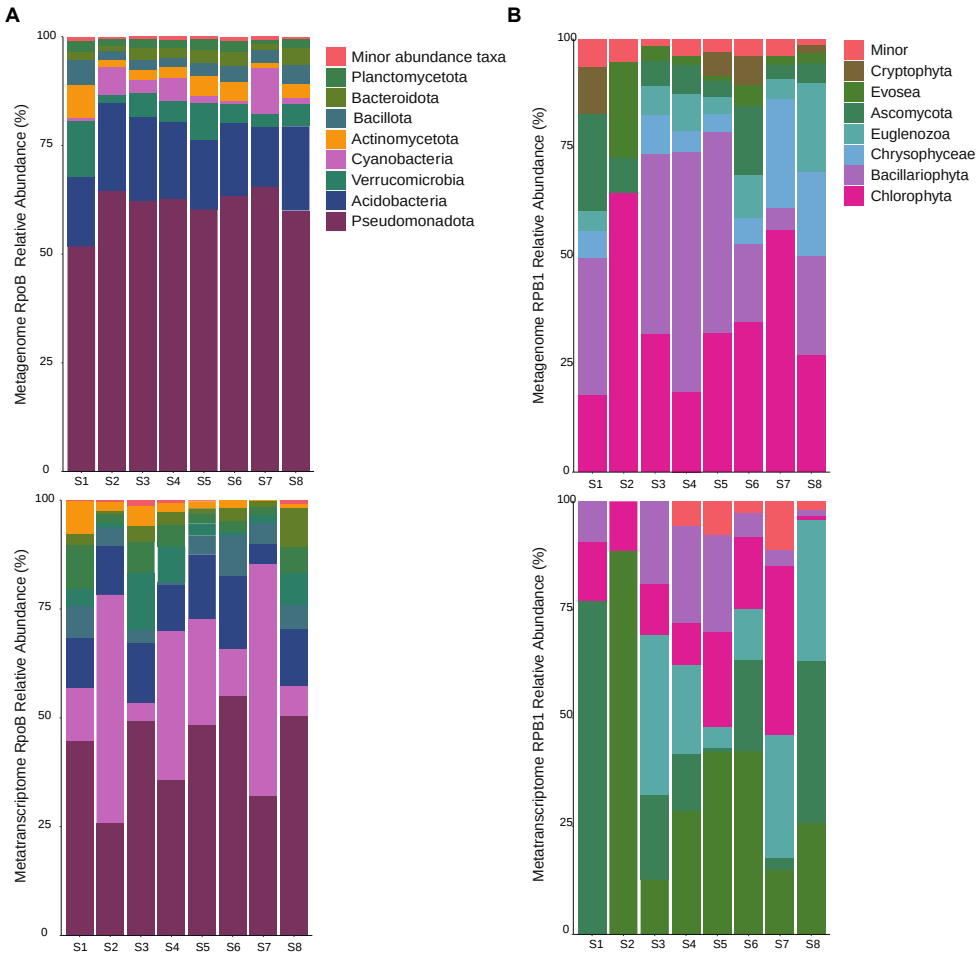
bacterial systems. Proceedings of the National Academy of Sciences. 2016;113(36):10174-9.
75. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, *et al*. Cryptic prophages help bacteria cope with adverse environments. Nat Commun. 2010;1(1):147.
76. Moniruzzaman M, Weinheimer AR, Martinez-Gutierrez CA, Aylward FO. Widespread endogenization of giant viruses shapes genomes of green algae. Nature. 2020;588(7836):141-5.
77. Schulz F, Andreani J, Francis R, Boudjemaa H, Khalil JYB, Lee J, *et al*. Advantages and Limits of Metagenomic Assembly and Binning of a Giant Virus. mSystems. 2020;5(3):10.
78. Fischer MG. Giant viruses come of age. Curr Opin Microbiol. 2016;31:50-7.
79. Farzad R, Ha AD, Aylward FO. Diversity and genomics of giant viruses in the North Pacific Subtropical Gyre. Front Microbiol. 2022;13.
80. Koonin EV, Yutin N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology. 2010;53(5):284-92.
81. Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. New dimensions of the virus world discovered through metagenomics. Trends Microbiol. 2010;18(1):11-9.
82. Maumus F, Epert A, Nogué F, Blanc G. Plant genomes enclose footprints of past infections by giant virus relatives. Nat Commun. 2014;5(1):4268.
83. Ogata H, Toyoda K, Tomaru Y, Nakayama N, Shirai Y, Claverie JM, *et al*. Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. Virol J. 2009;6:8.
84. Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, *et al*. Faustovirus, an Asfarvirus-Related New Lineage of Giant Viruses Infecting Amoebae. J Virol. 2015;89(13):6585-94.
85. Andreani J, Khalil JYB, Sevvana M, Benamar S, Di Pinto F, Bitam I, *et al*. Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between *Asfarviridae* and Faustoviruses. J Virol. 2017;91(14):11.

86. Hillman BI, Cai GH. The Family *Narnaviridae*: Simplest of RNA Viruses. In: Ghabrial SA, editor. Advances in Virus Research, Vol 86: Mycoviruses. Advances in Virus Research. 86. San Diego: Elsevier Academic Press Inc; 2013. p. 149-76.
87. Charon J, Murray S, Holmes EC. Revealing RNA virus diversity and evolution in unicellular algae transcriptomes. Virus Evol. 2021;7(2):18.
88. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, *et al*. Redefining the invertebrate RNA virosphere. Nature. 2016;540(7634):539-+.
89. Nibert ML, Vong M, Fugate KK, Debat HJ. Evidence for contemporary plant mitoviruses. Virology. 2018;518:14-24.
90. Charon J, Marcelino VR, Wetherbee R, Verbruggen H, Holmes EC. Metatranscriptomic Identification of Diverse and Divergent RNA Viruses in Green and *Chlorarachniophyte* Algae Cultures. Viruses. 2020;12(10):1180.
91. Chiba S, Kondo H, Tani A, Saisho D, Sakamoto W, Kanematsu S, *et al*. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. PLoS Pathog. 2011;7(7):e1002146.
92. Li L, Liu Z, Zhang M, Meng D, Liu X, Wang P, *et al*. Insights into the Metabolism and Evolution of the Genus *Acidiphilium*, a Typical Acidophile in Acid Mine Drainage. mSystems. 2020;5(6):10.1128/msystems.00867-20.
93. Challacombe J, Kuske C. Mobile genetic elements in the bacterial phylum Acidobacteria. Mob Genet Elements. 2012;2(4):179-83.
94. Eichorst SA, Trojan D, Roux S, Herbold C, Rattei T, Woebken D. Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ Microbiol. 2018;20(3):1041-63.
95. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. Isme J. 2017;11(7):1511.
96. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, *et al*. Prophages mediate defense against phage infection through diverse mechanisms. The ISME Journal. 2016;10(12):2854-66.

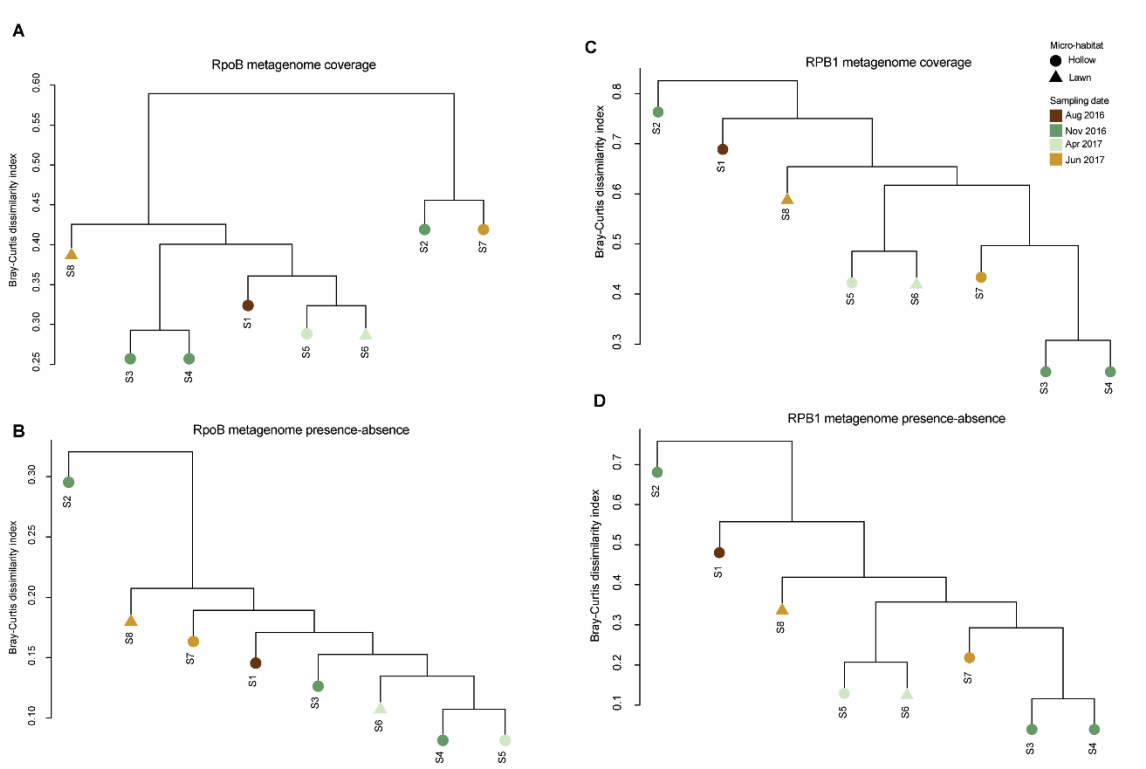
97. Opelt K, Chobot V, Hadacek F, Schönmann S, Eberl L, Berg G. Investigations of the structure and function of bacterial communities associated with *Sphagnum* mosses. Environ Microbiol. 2007;9(11):2795-809.
98. Bragina A, Berg C, Müller H, Moser D, Berg G. Insights into functional bacterial diversity and its effects on Alpine bog ecosystem functioning. Sci Rep. 2013;3(1):1955.
99. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proceedings of the National Academy of Sciences. 1999;96(5):2192-7.
100. Roossinck MJ. Persistent Plant Viruses: Molecular Hitchhikers or Epigenetic Elements? In: Witzany G, editor. Viruses: Essential Agents of Life. Dordrecht: Springer Netherlands; 2012. p. 177-86.
101. Son M, Yu J, Kim K-H. Five Questions about Mycoviruses. PLoS Pathog. 2015;11(11):e1005172.
102. Gao ZL, Karlsson I, Geisen S, Kowalchuk G, Jousset A. Protists: Puppet Masters of the Rhizosphere Microbiome. Trends Plant Sci. 2019;24(2):165-76.
103. Wyatt KH, Turetsky MR. Algae alleviate carbon limitation of heterotrophic bacteria in a boreal peatland. J Ecol. 2015;103(5):1165-71.

Appendix

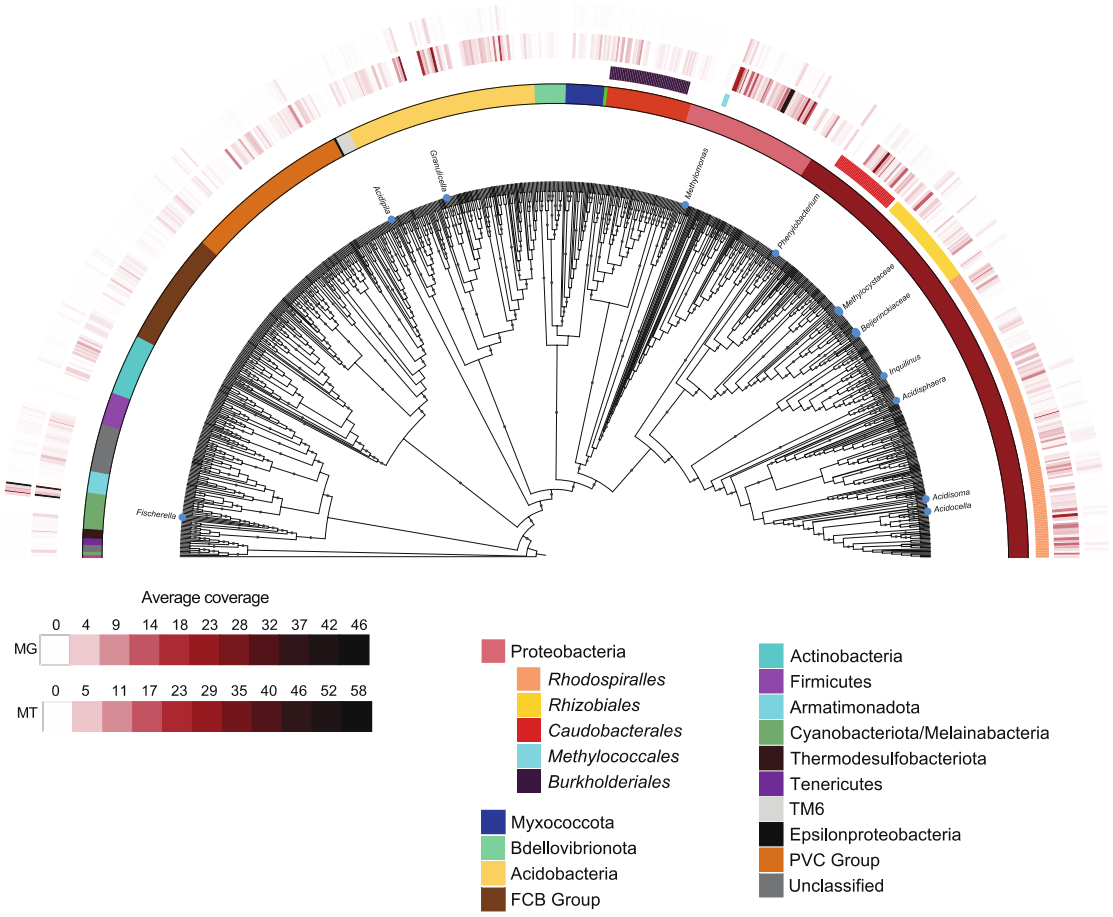
Supplemental Figures



Supplemental Figure 3.6. Relative abundance of microbial taxa in the metagenomes and metatranscriptomes.
A) Estimate of prokaryotic abundance using mapped reads to *rpoB*-containing contigs. B) Estimate of micro-eukaryotic abundance using *rpb1*. Abundance values were an average of the two replicates for metatranscriptomes. Minor prokaryotic taxa include: *Armatimonadetes*, *Chlamydiae*, *Gemmatimonadetes*, *Tenericutes*, *Euryarchaeota*, and unclassified bacteria. Minor abundance eukaryotic taxa include: Basidiomycota, Ciliophora, and Ichthyosporea.

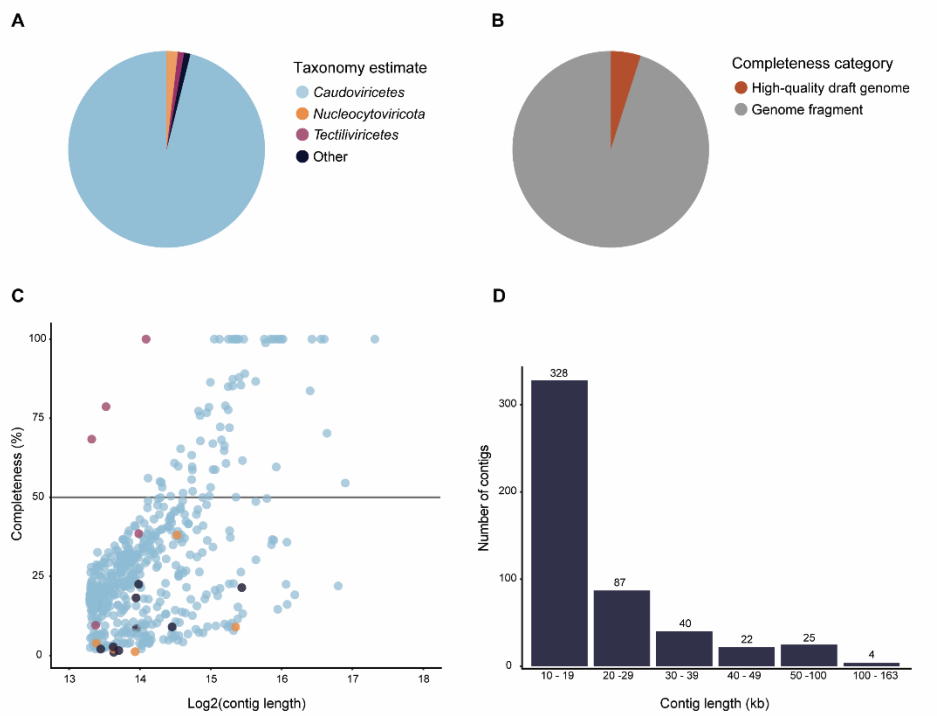


Supplemental Figure 3.7. Hierarchical clustering of microbial communities by structure and composition.
Clustering is based on A) coverage values and B) presence-absence of rpoB-containing contigs in the metagenomes. Micro-eukaryotic community clustering based on C) coverage values and D) presence-absence of rpb1-containing contigs in the metagenomes. All clustering was performed using the average linkage method. Metagenomes are color-coded by collection date and shapes represent sample micro-topography of sample site.

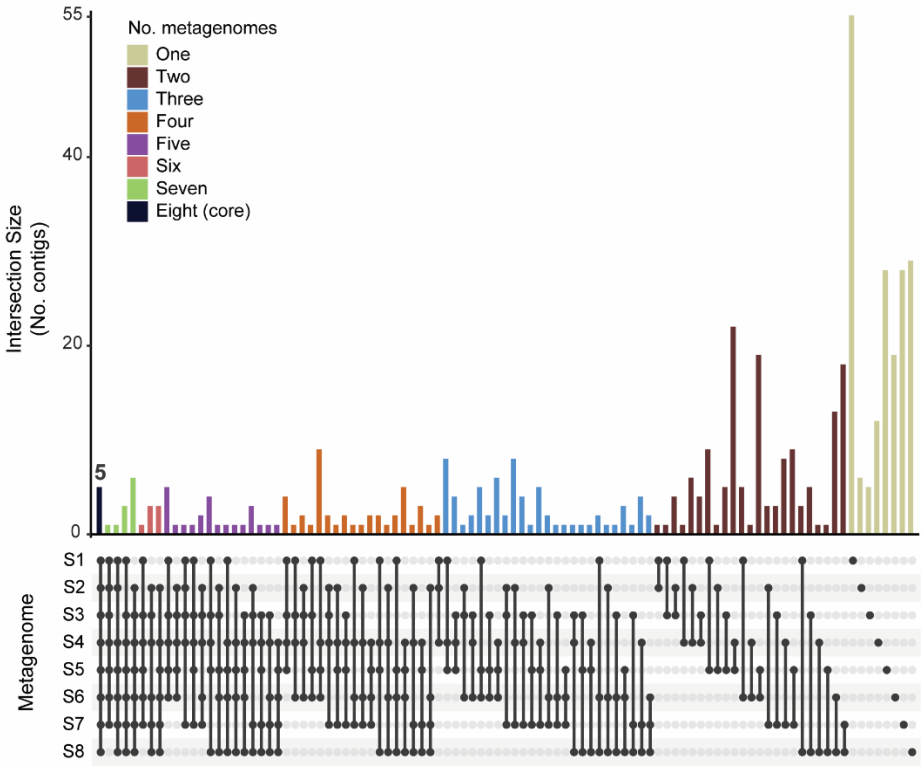


Supplemental Figure 3.8. Cladogram of maximum likelihood phylogeny of RpoB protein sequences.

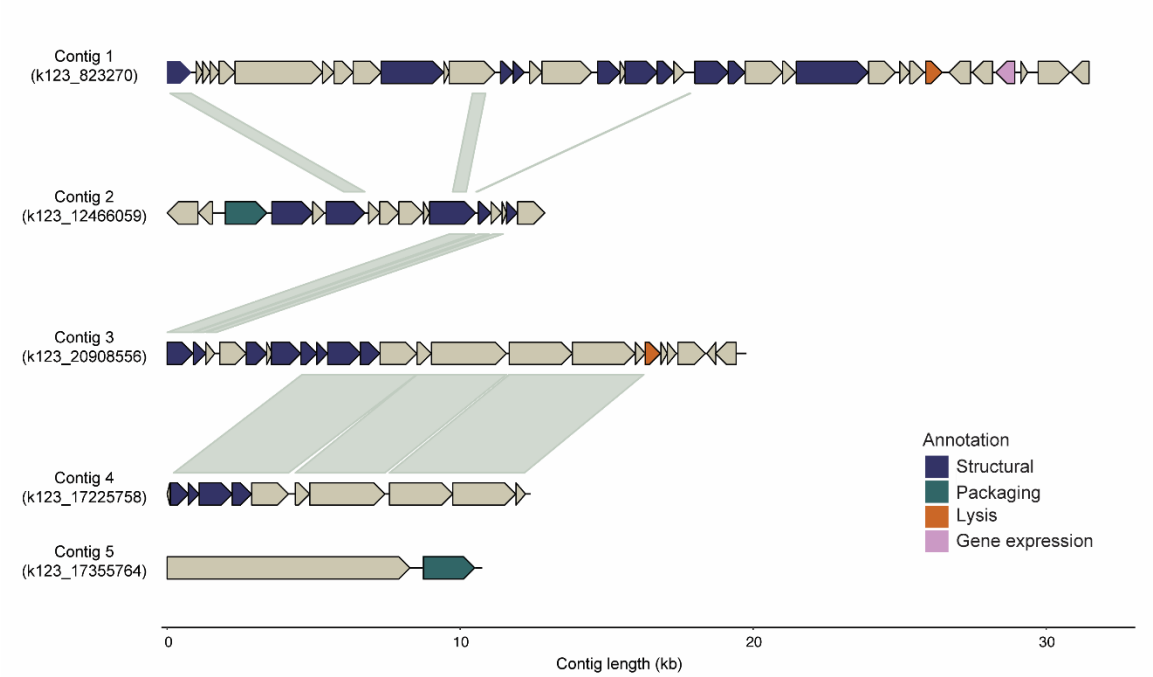
Core taxa identified in Kolton *et al*. (2022) are labeled and indicated by nodes ending in circles. Core RpoB detected in this dataset are bolded branches Inner ring represents prokaryotic phylum. Order is shown for clades of interest containing core prokaryotic members. Abundance for each RpoB is represented as average coverage. FastTree bootstraps > 0.9 are displayed as dots.



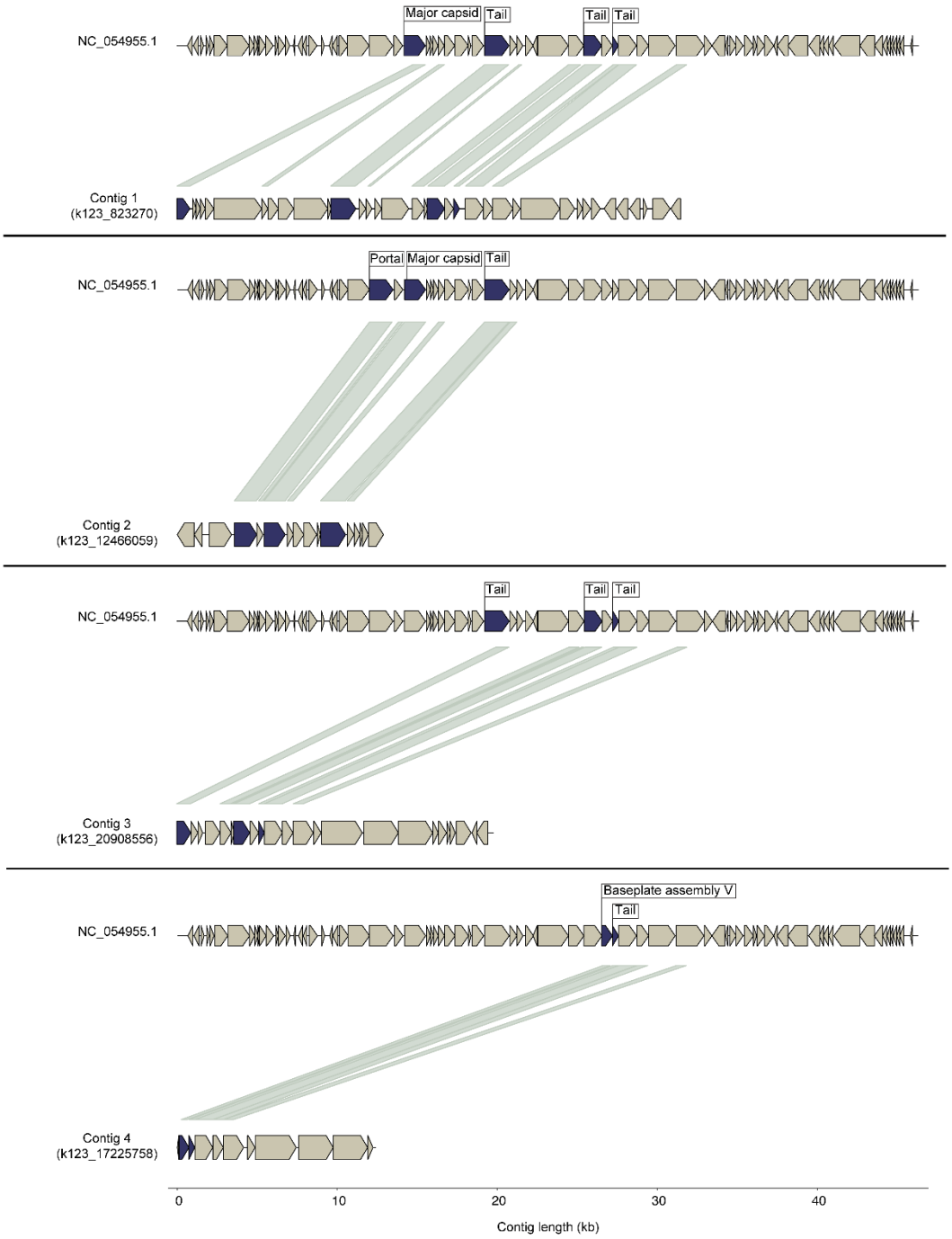
Supplemental Figure 3.9. Overview of the metagenomic viral contigs.
Pie charts show the proportion of A) taxonomic groups and B) completeness categories. C) Dot plot displays the quality estimates for the individual DNA virus contigs plotted as completeness by contig length (length shown as Log2-transformed kb). Below the line at 50% indicates contigs in the CheckV "low quality" category. D) Distribution of contig sizes. 'Other' taxa shown on panel A and C include: *Maveriviricetes*, *Polintoviricetes*, *Naldaviricetes*, unclassified *Bamfordvirae*, and unclassified virus.



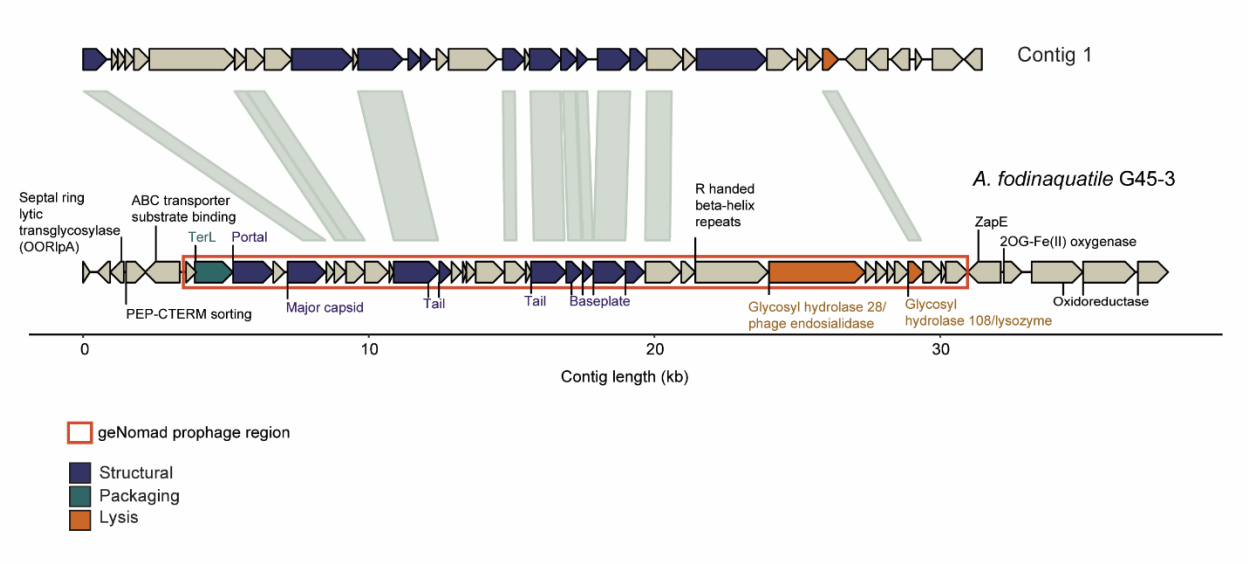
Supplemental Figure 3.10. Plot showing the number of viral contigs shared between metagenomes.
Connected dots indicate contig(s) shared among metagenomes and the intersection size is the number of unique contigs shared. Bars are color-coded by the number of metagenomes each intersection (*i.e.*, group of contigs) was detected in.



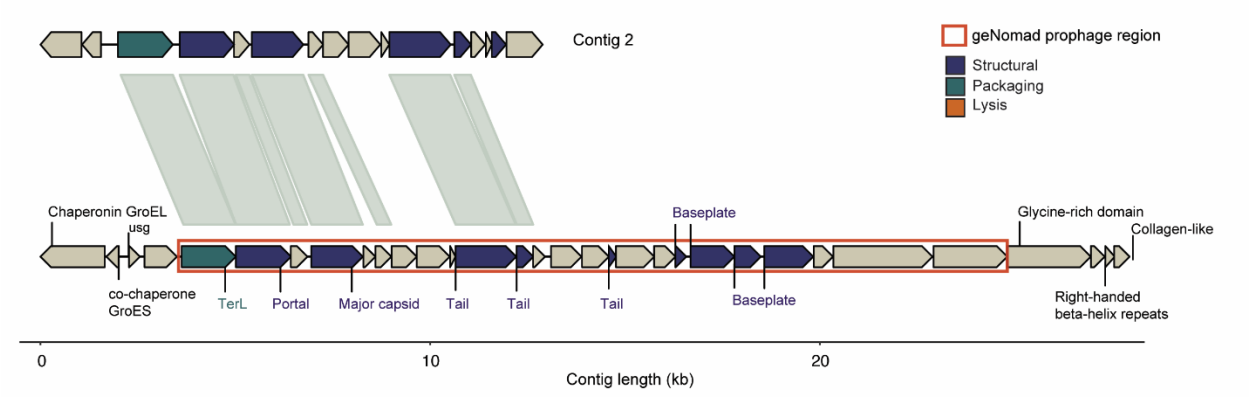
Supplemental Figure 3.11. Visualization of nucleotide similarity between the core phage-like contigs.
The BLASTN table used to construct his figure is available at Supplementary Table 7.



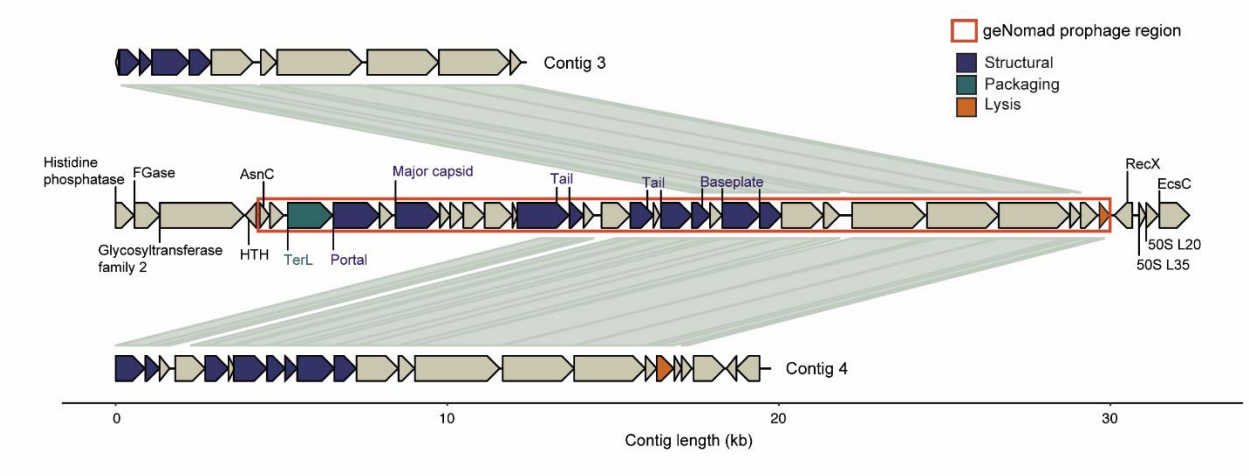
Supplemental Figure 3.12. Visualization of protein sequence similarity between the core phage-like contigs and Ralstonia phage Raharianne (NC_054955.1).
The BLASTP table used to construct this figure is available at Supplementary Table 8.



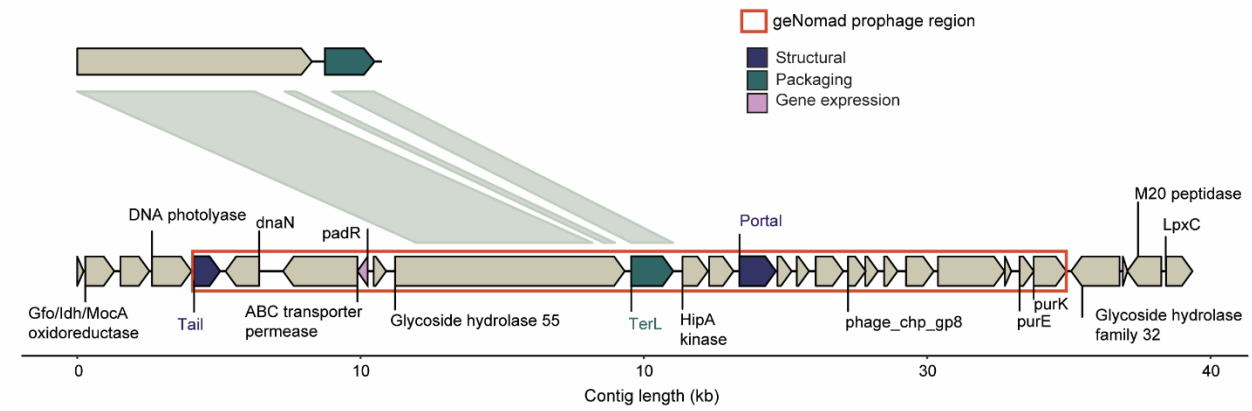
Supplemental Figure 3.13. Visualization of protein sequence similarity between the core phage-like contig 1 (C1) and the top hit RefSeq bacterial genome.



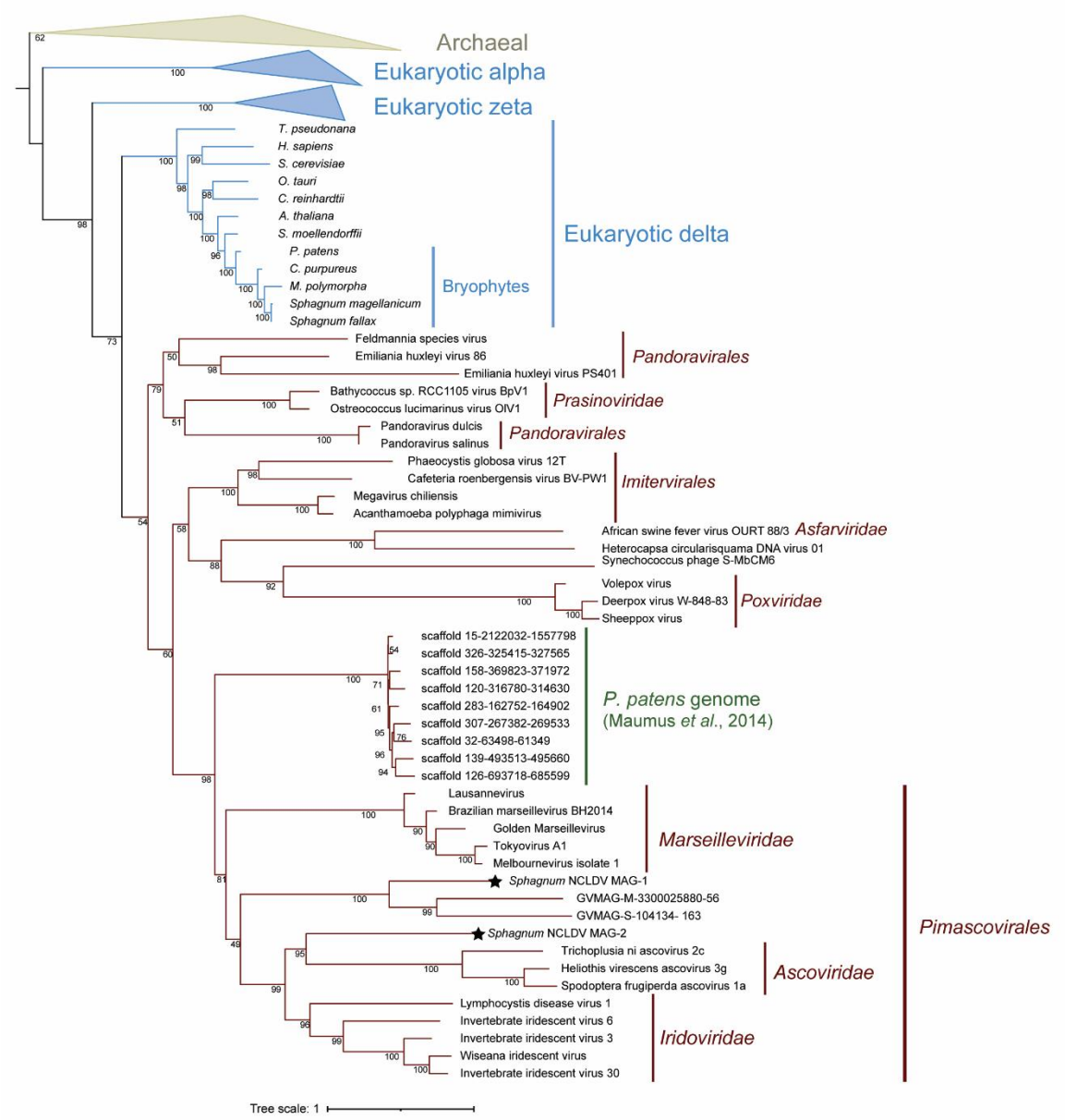
Supplemental Figure 3.14. Visualization of protein sequence similarity between the core phage-like contig 2 (C2) and the top hit RefSeq bacterial genome.



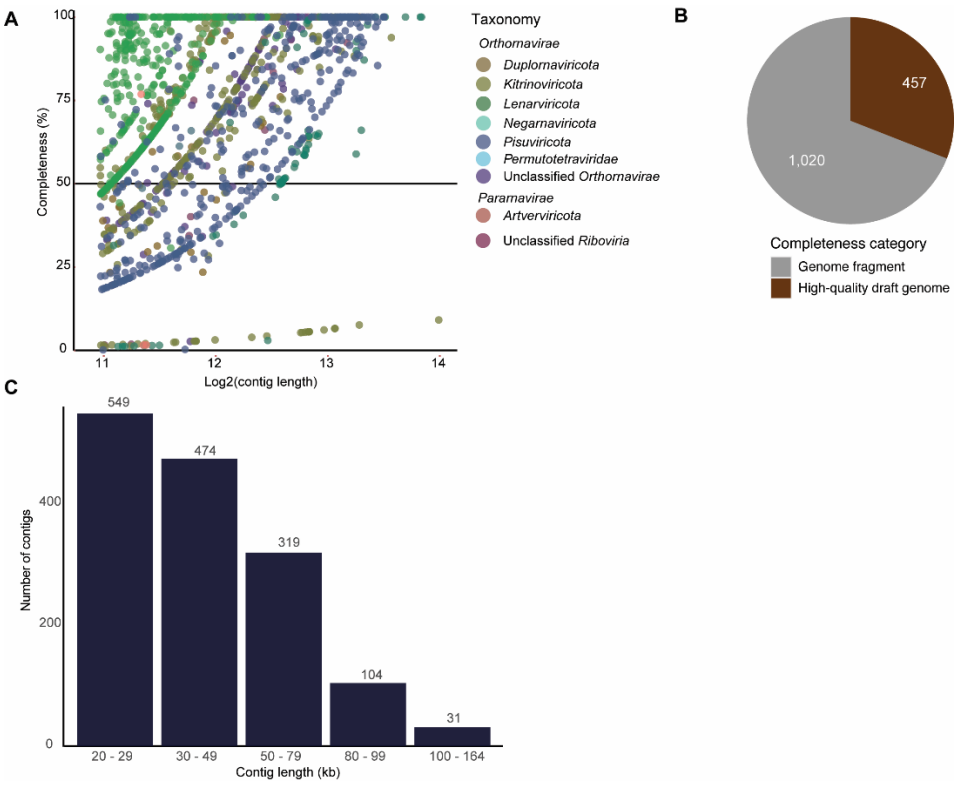
Supplemental Figure 3.15. Visualization of protein sequence similarity between core phage-like contigs 3 (C3) and 4 (C4) and the top hit RefSeq bacterial genome.



Supplemental Figure 3.16. Visualization of protein sequence similarity between the core phage-like contig 5 (C5) and the top hit RefSeq bacterial genome.



Supplemental Figure 3.17. Maximum likelihood phylogeny of DNA polymerase protein sequences.
The tree was inferred using the best-fit model determined by IQTREE (Q.pfam+F+I+G4). DNA polymerase sequences from the *Sphagnum* NCLDV MAGs, *S. fallax* and *S. magellanicum* genomes, and additional *Pimascovirales* genomes were aligned to the alignment retrieved from Maumus *et al*. (2014).



Supplemental Figure 3.18. Viral contigs detected in the metatranscriptomes.
A) Scatter plot shows each contig represented by a point plotted as completeness by contig length (shown as Log2-transformed kb). Below the line at 50% indicates contigs in the CheckV "low quality" category. B) Pie chart shows a summary of the completeness category of all RNA virus contigs. C) Distribution of contig sizes.

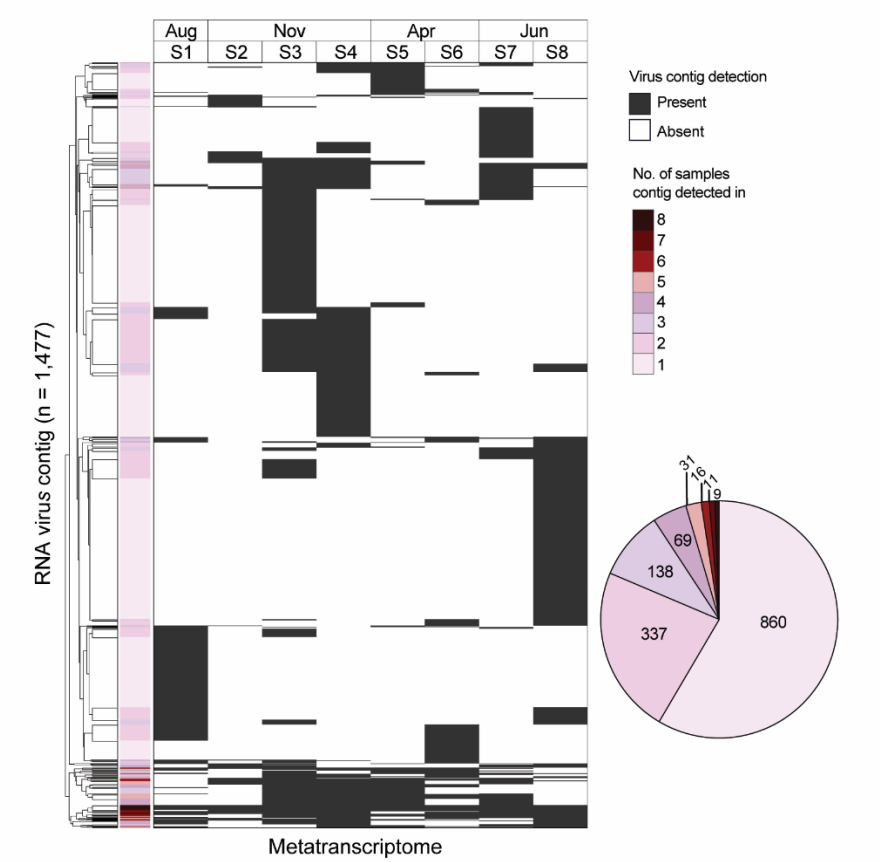


Figure 3.19. Distribution of RNA virus contigs in the metatranscriptomes. Heatmap shows the presence or absence of individual contigs. Average linkage hierarchical clustering was applied based on contigs presence-absence. Color-coded rows indicate the number of metatranscriptomes each contig was detected in. Pie chart shows the total number of contigs within each detection category (i.e., the number of metatranscriptomes a contig was detected in).

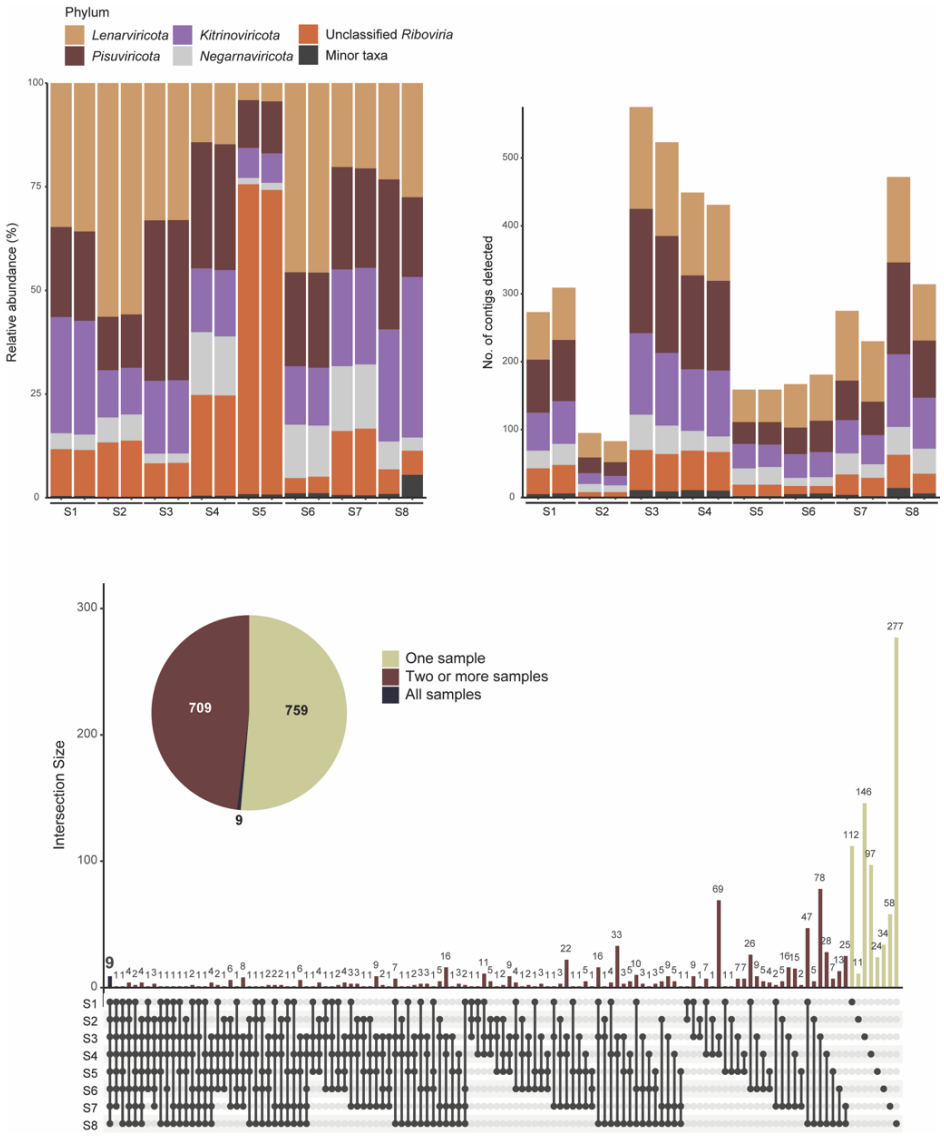


Figure 3.20. Relative abundance and composition RNA virus phyla in the metatranscriptomes.
Replicates are indicated by black bars. UpSet plot depicting the number of RNA virus contigs shared between or unique to the *Sphagnum* metatranscriptome libraries. Intersection size represents the number of contigs shared between libraries in each column (represented as filled in circles).

CHAPTER FOUR: INVESTIGATING THE EFFECTS OF WARMING ON VIROME COMPOSITION AND STRUCTURE IN *SPHAGNUM*

Publication Note

Content from Chapter Four is included in a manuscript that is in preparation for submission to a peer-reviewed journal.

Sequencing data was provided by David Weston and colleagues. Sequencing processing and data analysis were performed by Elizabeth Denison.

Abstract

Sphagnum mosses, or peat mosses, are a keystone species in Northern peatlands where they play integral roles in the sequestration and storage of terrestrial carbon. Warming is known to impact the structure and metabolism of the *Sphagnum* microbiome. Viral activity may shape the peat moss microbiome and modulate microbial responses to warming, but no studies have examined the effect of warming on viral communities associated with the *Sphagnum* microbiome. This chapter used paired metagenomes and metatranscriptomes to investigate the effects of warming on DNA virus populations in the peat moss microbiome. Samples originated from ambient, artificially warmed, and geothermally warmed sites across Europe. We found that viral community structure, composition, and activity were significantly different between warming conditions, but only for the moss species that were sampled from geothermally warmed conditions (*S. teres* and *S. medium*). The difference in warming severity between the artificially warmed and geothermally warmed samples compared to the ambient temperature, with an average difference of +2.1° C and +15.3° C, respectively, may in part explain our observations. These findings suggest that the viral community is not noticeably shaped by warming until a temperature threshold is passed.

Background

Northern peatlands perform an important global 'carbon service' by sequestering large amounts of atmospheric carbon (C) and causing a net cooling effect on Earth's climate (1, 2). Over the Holocene, Northern peatlands have accumulated carbon at an average rate of 22.9 g C m^{-2} $year^{-1}$, resulting in an estimated 500 gigatons of carbon currently sequestered globally (3, 4). However, the peatland C balance, or the amount of C stored versus emitted, is vulnerable to climate warming (5, 6). While predictions have been made regarding the future of peatland carbon stores (7, 8), there are still many unknowns regarding how different components of peatland ecosystems will respond to warming.

Sphagnum mosses are critical to peatland ecosystem function and stability (9). Yet, *Sphagnum* dominance is threatened by warming (10, 11), which has been attributed to changes in hydrology and moss desiccation (12) and a decline in light availability (13,

14). The moss microbiome, consisting of bacteria, microbial eukaryotes, and viruses, is an important but understudied link between *Sphagnum* ecology and climate warming responses. Previous work has shown warming to impact microbial community structure, including key diazotrophic symbionts, with corresponding declines in nitrogen-fixation rates by the microbiome (15). Warming-adapted moss microbiomes may also harbor thermotolerant symbionts that can influence moss resilience to warming (16).

Viral activity may mediate microbiome response to warming (17). Viruses can influence microbial community structure through lysis (18) and can impact key plant-associated function guilds that influence plant growth (19). Viral activity may also shape microbial communities by altering the flow of carbon and nutrients (20, 21). These types of dynamics have been explored in soil and aquatic systems (22-24), and researches have called for integrating viruses in soil food webs under climate change (25), but less so for plant-associated communities. Nonetheless, the top-down and bottom-up pressures exerted by viruses may modulate microbiome responses to warming. Studies into viruses in the moss microbiome are scarce (26), and to our knowledge no studies have assessed moss-associated viral communities in relation to temperature.

Methods

Description of Sphagnum samples

Sphagnum tissue samples were collected by Piatkowski *et al*. from seven peatland sites across Europe (Supplemental **Table 4.1**) (27). All supplemental tables referenced in this chapter are available in Attachment 2. This included two experimental field sites, the Forbonnet (France) and Degerö Stormyr (Sweden) peatlands, where *Sphagnum* was collected from both the ambient and artificially warmed plots. Samples were also collected from five different sites in Iceland that included ambient and geothermally warmed conditions. At each site, the dominant *Sphagnum* species was sampled (11 different species total). The average moss temperature was 12.2° C, 14.7° C, and 27.2° C under ambient, artificially warmed, and geothermally warmed conditions, respectively (Supplemental **Table 4.1**). The average difference in temperature between ambient and artificially warmed condition comparisons was 2.1° C, while the average difference between ambient and geothermally warmed condition comparisons was 15.3° C.

Nucleic acid extraction from plant matter and sequence processing

Sphagnum samples were flash frozen using liquid nitrogen until nucleic acid extraction. DNAA was extracted from the moss samples using the DNeasy Plant Pro DNA extraction kit (Qiagen) by Piatkowski *et al*. (27). Sequencing was carried out at the Department of Energy (DOE) Joint Genome Institute (JGI). Libraries were constructed using KAPA DNA library kits and sequencing was performed using the Illumina NovaSeq S4 platform (2x151-bp sequencing). Read filtering was performed at JGI using the JGI Metagenome Workflow (28). For metatranscriptomes, ribosomal RNA reads were also removed in silico. The filtered reads were downloaded from the JGI Genome Portal. In total, 49 metagenomes and 42 metatranscriptomes were retrieved, 34 of which were paired libraries (*i.e.*, a metagenome and metatranscriptome generated from the same sample).

Here, we removed host (peat moss) reads by mapping the filtered reads to the *S. fallax* v1.1 (GCA_021442195.1) and *S. magellanicum* v1.1 (GCA_021904315.1) genomes using BBMap v38.90 with the default settings (29). Unmapped reads on average made up 80.6% ± 1.82% (SE) of the metagenomes. The remaining reads were pooled (metagenomes and metatranscriptomes were kept separate) and assembled using MEGAHIT v1.2.9 (--k-min 31 --k-max 121) on the University of Tennessee high-performance computing cluster (ISAAC-NG) (30). Assembly quality was assessed with QUAST v5.0.2 (31). The metagenome and metatranscriptome assemblies contained 32,695,709 and 3,759,860 contigs ≥ 500 bp, respectively (Supplemental **Table 4.2**).

Identification of viral contigs from the metagenome assembly

Viral-like contigs were identified using geNomad with the default settings (32). Only metagenomic contigs ≥ 10 kb were used as input. Viral contigs were then filtered to retain those with a geNomad confidence score > 0.8, at least one viral marker gene, and a marker enrichment score >1.5. Finally, contigs were retained if they had at least one viral marker gene called by CheckV (33) were considered viral-like contigs. The filtering thresholds were chosen based on the developer's recommendations and current community standards (34). CheckV was used to assign quality categories to the viral contigs: low-quality (0-50% complete), medium-quality (50-90%), high-quality (> 90%),

or undetermined (33). Gene annotation and contig taxonomic assignments were performed by geNomad. Finally, prophage regions (temperate phage integrated into host genomes) were predicted by geNomad and CheckV.

NCLDV major capsid protein search

NCLDV-like major capsid protein (MCP) sequences were identified using protein sequence homology searches. The metagenomic contigs > 1kb were queried against the Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOGs) database using DIAMOND BLASTp (e-value < 1e-10, subject cover >60%) (35, 36). Contigs with hits to MCP reference (NCVOG0022) were retained. Genes were predicted using Prodigal, then the translated sequences were queried against the RefSeq v213 database using (BLASTp, e-value < 1e-10). Only protein sequences with a top hit to a NCLDV MCP were used for further analysis. Both the top RefSeq BLASTp hit were used to assign taxonomy to the MCPs. For the phylogenetic tree, MCPs a custom set of MCPs were retrieved from GenBank belonging to sequenced NCLDV genomes. If the genome contained multiple predicted MCP copies, they were all used to build the tree. Pandoraviruses were not included since they lack the typical NCLDV MCP (37). The reference and *Sphagnum* metagenome MCPs were aligned using Clustal Omega v1.2.4 (38) and the alignment was trimmed with trimAl v.1.2 (-gt 0.1) (39). The tree was constructed using FastTree v.2.1.10 (LG model) (40).

NCLDV metagenome-assembled genomes

NCLDV MAGs were binned and curated approach similar to workflow developed by Moniruzzaman *et al*. (41) and as in Chapter Two. A phylogenetic tree was constructed to assess the relatedness of the MAGs to isolated NCLDV genomes and other environmental NCLDV MAGs. A set of seven conserved marker genes (42) were predicted from the European MAGs and from > 1,000 reference NCLDV genomes (compiled by Gilbert *et al*. (43)), which included both cultivated virus genomes and uncultivated MAGs, to build the tree. Marker genes from the NCLDV MAGs generated from the North American *Sphagnum* metagenomes in Chapter Three were also included in the alignment. The concatenated alignment of the marker genes was generated using ncldv_markersearch and the alignment was trimmed using trimAl v.1.2 (-gt 0.1) (39). A

phylogenetic tree was constructed using IQTREE v.2.2.0.3 (LG+F+I+G4 substitution model with 1,000 ultrafast bootstraps) (44).

Read mapping analysis and community diversity statistics

Reads were mapped to viral contigs to gauge the distribution and abundance of viral sequences in the *Sphagnum* metagenomes. Reads were mapped using CoverM at 90% identity and 90% read length aligned thresholds. A contig was considered detected in a metagenome if >75% of the contig length was covered.

The coverage tables from read mapping were used for community-level statistical analysis. Square root-transformed or presence-absence transformed coverage values were used as the input to generate Bray-Curtis similarity matrices. PERMANOVA tests were ran from the similarity matrices. Statistical tests were carried out in PRIMER and using the R package vegan. Throughout, community 'structure' refers to things calculated from abundance tables, and 'composition' refers to things calculated from presence-absence transformed abundance table.

Results

Overview of viral contigs identified from the metagenome assembly

geNomad predicted 6,200 virus contigs ≥ 10kb from the metagenome assembly. After quality filtration, 3,756 viral contigs were retained for further analysis (Supplemental **Table 4.3**). Taxonomic assignment showed that the majority of viral contigs were phage-like (3,682 *Caudoviricetes* contigs, or 98%). *Megaviricetes* (giant viruses, 62 contigs), *Tectiliviricetes* (tail-less dsDNA phage, 14 contigs), *Polintoviricetes* (polinton-like viruses, 8 contigs), and *Maveriviricetes* (virophage, 3 contigs) were also identified. Additionally, 26% of the *Caudoviricetes* contigs (985 contigs) were predicted to be prophage by geNomad and/or CheckV (Supplemental **Table 4.3**).

Differences in viral communities between warming condition are dependent upon plant species and warming type

Beta diversity statistics based on abundance in the metagenomes showed that, overall, viral community structure was significantly different between warming conditions (PERMANOVA, $p = 0.001$) and between *Sphagnum* species ($p = 0.001$). However, there was a significant interaction between warming condition and *Sphagnum*

species (p = 0.021), signifying that the difference between warming conditions was dependent upon the plant species (**Figure 4.1**) All figures referenced in this chapter are located in the Chapter Four Appendix). We therefore compared viral community structure between warming conditions within a plant species (*i.e.*, pairwise tests using the warming condition X species interaction term). Since some plants were only sampled from one warming condition (*S. warnstorfii, S. subnitens, S. girgensohnii*, and *S. angustifolium*) and no intra-species comparison would be possible, the interaction term was only relevant for the *S. fallax, S. lindbergii, S. balticum*, and *S. divinum* samples from the ambient and artificial warming sites, and the *S. teres* and *S. medium* samples from the Icelandic geothermal and non-geothermal sites. Viral community structure was only significantly different between warming conditions for *S. teres* and *S. medium* (p = 0.014 and p = 0.023, respectively).

Community composition (*i.e.*, viral contig presence-absence) was also significantly different between warming conditions (p = 0.001) and *Sphagnum* species (p = 0.001), and the difference between warming conditions was dependent upon plant species (p = 0.004) **(Figure 4.1**). Community composition was still only significantly different between warming conditions for *S. teres* and *S. medium* (p = 0.004 and p = 0.027, respectively).

Comparison of active viral communities between warming conditions

We assumed that viral contigs detected in the metatranscriptomes represent replicating, or "active", viral populations. Here, we assessed samples for which there was a paired metagenome and metatranscriptome (34 paired libraries). Community-level trends were similar to what was observed in the metagenomes. Both the structure and composition of the active viral communities were significantly different between warming condition and *Sphagnum* species (PERMANOVA p = 0.001) but were dependent on each other (warming condition-species interaction, p = 0.001 and p = 0.009) (**Figure 4.2**). None of the intra-species pairwise comparisons between warming conditions were significant based on abundance or presence-absence (p > 0.05). However, since only ten permutations were possible we used Monte-Carlo p-values,

which determined that active viral communities warming conditions were significantly different within *S. teres* and *S. medium*.

Overview of NCLDV populations recovered from the metagenomes

We used the MCP marker gene to estimate the diversity and distribution of NCLDV in the *Sphagnum* metagenomes since it is conserved in most NCLDV lineages. In all, 86 putative NCLDV MCPs were identified from the metagenome assembly (Supplemental **Table 4.4**). The amino acid percent similarity of the *Sphagnum* MCP to the RefSeq database hits ranged from only 22% up to 72%. Based on the taxonomic assignment of the top MCP hit in the RefSeq database (v213), most of the MCPs were similar to those belonging to NCLDV within the *Imitervirales* order (45 MCP), including 28 MCP within the *Mimiviridae* family (Supplemental **Table 4.4**). The best hits were to NCLDV known to infect amoeba. The remaining *Imitervirales* (17 MCP) most resembled algae-infecting *Mimiviridae* relatives. The MCP also included *Algavirales* (25 MCP) viruses that infect unicellular algae, *Iridoviridae* (10 MCP), which are known to infect insects and cold-blooded vertebrates, and *Asfurvirales* (six MCP), whose isolated viruses infect animals.

NCLDV beta diversity comparisons between warming conditions

MCP community structure was significantly different between warming conditions (PERMANOVA, $p = 0.001$) and *Sphagnum* species ($p = 0.001$). The difference in warming condition was dependent upon the species ($p = 0.015$). Community structure was only significantly different between warming conditions for *S. teres* and *S. medium* ($p = 0.008$ and $p = 0.025$, respectively). The average dissimilarity in MCP community structure between the geothermal warming and no geothermal warming samples was 76% and 80% for *S. teres* and *S. medium*, respectively (based on SIMPER results). MCP community composition (*i.e.*, presence-absence) was also significantly different between warming condition ($p = 0.001$) and *Sphagnum* species ($p = 0.001$), as well as for the interaction between warming condition and species ($p = 0.049$) (**Figure 4.3**). Community composition was only significantly different between warming conditions for *S. teres* and *S. medium* ($p = 0.001$ and $p = 0.035$, respectively). The average dissimilarity in MCP community composition between the geothermal warming

and no geothermal warming samples was 63.41% and 63.46% for *S. teres* and *S. medium*, respectively.

Hierarchical clustering based on MCP presence-absence showed a pattern in the distribution of individual MCP across all of the metagenomes. Based on the beta diversity statistics, we focused on the subset of samples where an intra-species warming comparison was possible (**Figure 4.4**). Some MCP were detected across many groups (cluster 1, seven contigs), which all had top RefSeq hits to viruses of algae within the *Algavirales* and *Imitervirales*. Other MCP were detected almost exclusively in *S. teres* and/or *S. medium* (cluster 2) or were generally absent from these species (cluster 3). *S. teres* and *S. medium* samples collected from the same site (Siðmúli, Iceland) clustered together.

Within *S. teres*, six MCP were detected in all metagenomes from the ambient condition but none from the warming condition. Conversely, seven MCP were detected in all metagenomes from the warming condition but none from the ambient condition. The MCP detected only in the warming condition were similar to that from algae- and amoeba-infecting NCLDV within the *Imitervirales*. The MCP found only in the ambient condition included *Imitervirales*-like as well as *Pimascovirales*- and *Asfuvirales*-like MCP. Three MCP, one Algavirales and two *Imitervirales* were found in all S. teres samples.

Within *S. medium*, two *Imitervirales*-like MCP were detected only in the warming condition while one *Asfuvirales*-like MCP was only found in the ambient condition. Four MCP, one *Algavirales* and three *Imitervirales*, were found in all *S. medium* metagenomes. All three of the *Imitervirales* MCP were predicted from the same contig, suggested they are multiple copies originating from the same NCLDV population as opposed to three different populations.

Here we used metatranscriptomes for which there was a paired metagenome (34 pairs of metagenomes/metatranscriptomes). At the community level, the structure (based on mapped transcript abundance) and composition (based on presence-absence) of actively infecting NCLDV were significantly different between warming condition and *Sphagnum* species (PERMANOVA $p = 0.001$ for both), but dependent on each other

(warming condition X species interaction, PERMANOVA p = 0.001). None of the intra-species pairwise comparisons between warming condition were significant (p > 0.05).

Overview of the NCLDV MAGs: completeness and phylogenetic placement

In total, sixteen NCLDV-like MAGs were generated from the metagenome co-assembly, including eight medium-quality (50-90% complete) and one high-quality (> 90% complete) MAG (Supplemental **Table 4.5**). We retained the seven low-quality MAGs (< 50% complete) or further analysis since genome fragments can still be used to estimate the diversity and distribution of NCLDV in our samples.

The MAGs were placed within three defined orders: the *Asfuvirales* (two MAGs), *Imitervirales* (five MAGs), and *Pimascovirales* (eight MAGs) (**Figure 4.4**). The two *Asfuvirales* MAGs fell within a sister clade to the *Asfarviridae* (contains the African swine fever viruses). This clade contains NCLDV isolated via amoeboid protists (Pacmanvirus and Kaumoebavirus) and related environmental MAGs. Notably, this branch contained a NCLDV MAG that we previously generated from North American *Sphagnum* microbiome metagenomes (see Chapter Three). Within the *Imitervirales*, four MAGs were placed on the *Mimiviridae* family branch (IM_16) that includes protist-infecting NCLDV, while one MAG was placed within a proposed family containing only uncultured MAGs. The MAGs placed within the *Pimascovirales* were related to the amoeba-infecting *Marseilleviridae* (four MAGs) family and to the *Iridoviridae-Ascoviridae* giant virus family. Generally, these MAGs were not closely related to the isolated viruses within these families but instead fell near environmental, uncultured MAGs.

Similarities between the European and North American NCLDV MAGs

Phylogenetic analysis showed that similar NCLDV lineages were present in both European and North American *Sphagnum* samples. Two MAGs were placed within a sister clade to the *Asfarviridae* (*Asfurvirales*) that contained viruses isolated from amoeba (Pacmanvirus A23 and Kaumoebavirus) as well as environmental MAGs with unknown hosts. This clade contained SPRUCE MAG-3, which was found to be a ‘core’ NCLDV MAG in our previous work (please refer to Chapter Two for details on the North American NCLDV MAGs). Although related, the MAGs were not identical. They were

too divergent for an ANI comparison (below 80% ANI, suggesting they are not the same species).

References

1. Harris LI, Richardson K, Bona KA, Davidson SJ, Finkelstein SA, Garneau M, et al. The essential carbon service provided by northern peatlands. Frontiers in Ecology and the Environment. 2022;20(4):222-30.
2. Frolking S, Roulet NT. Holocene radiative forcing impact of northern peatland carbon accumulation and methane emissions. Glob Change Biol. 2007;13(5):1079-88.
3. Yu ZC, Loisel J, Brosseau DP, Beilman DW, Hunt SJ. Global peatland dynamics since the Last Glacial Maximum. Geophys Res Lett. 2010;37:5.
4. Gorham E. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. Ecol Appl. 1991;1(2):182-95.
5. Hugelius G, Loisel J, Chadburn S, Jackson RB, Jones M, MacDonald G, et al. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. Proceedings of the National Academy of Sciences. 2020;117(34):20438-46.
6. Hopple AM, Wilson RM, Kolton M, Zalman CA, Chanton JP, Kostka J, et al. Massive peatland carbon banks vulnerable to rising temperatures. Nat Commun. 2020;11(1):2373.
7. Frolking S, Talbot J, Jones MC, Treat CC, Kauffman JB, Tuittila E-S, et al. Peatlands in the Earth's 21st century climate system. Environmental Reviews. 2011;19(NA):371-96.
8. Zhuang Q, Melillo JM, Sarofim MC, Kicklighter DW, McGuire AD, Felzer BS, et al. CO2 and CH4 exchanges between land ecosystems and the atmosphere in northern high latitudes over the 21st century. Geophys Res Lett. 2006;33(17).
9. van Breemen N. How *Sphagnum* bogs down other plants. Trends in Ecology & Evolution. 1995;10(7):270-5.
10. Bragazza L, Buttler A, Robroek BJM, Albrecht R, Zaccone C, Jassey VEJ, et al. Persistent high temperature and low precipitation reduce peat carbon accumulation. Glob Change Biol. 2016;22(12):4114-23.
11. Bragazza L. A climatic threshold triggers the die-off of peat mosses during an extreme heat wave. Glob Change Biol. 2008;14(11):2688-95.

12. Norby RJ, Childs J, Hanson PJ, Warren JM. Rapid loss of an ecosystem engineer: *Sphagnum* decline in an experimentally warmed bog. Ecol Evol. 2019;9(22):12571-85.
13. Norby RJ, Baxter T, Živković T, Weston DJ. Shading contributes to *Sphagnum* decline in response to warming. Ecol Evol. 2023;13(9):e10542.
14. van der Wal R, Pearce IS, Brooker RW. Mosses and the struggle for light in a nitrogen-polluted world. Oecologia. 2005;142(2):159-68.
15. Carrell AA, Kolton M, Glass JB, Pelletier DA, Warren MJ, Kostka JE, et al. Experimental warming alters the community composition, diversity, and N-2 fixation activity of peat moss (*Sphagnum* fallax) microbiomes. Glob Change Biol. 2019;25(9):2993-3004.
16. Carrell AA, Lawrence TJ, Cabugao KGM, Carper DL, Pelletier DA, Lee JH, et al. Habitat-adapted microbial communities mediate *Sphagnum* peatmoss resilience to warming. New Phytol. 2022.
17. Wieczynski DJ, Yoshimura KM, Denison ER, Geisen S, DeBruyn JM, Shaw AJ, et al. Viral infections likely mediate microbial controls on ecosystem responses to global warming. FEMS Microbiol Ecol. 2023;99(3):fiad016.
18. Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? Environ Microbiol. 2004;6(1):1-11.
19. Joglekar P, Ferrell Barbra D, Jarvis T, Haramoto K, Place N, Dums Jacob T, et al. Spontaneously Produced Lysogenic Phages Are an Important Component of the Soybean Bradyrhizobium Mobilome. mBio. 2023;14(2):e00295-23.
20. Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea - Viruses play critical roles in the structure and function of aquatic food webs. Bioscience. 1999;49(10):781-8.
21. Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. Nat Rev Microbiol. 2020;18(1):21-34.
22. Wang Y, Liu Y, Wu Y, Wu N, Liu W, Wang X. Heterogeneity of soil bacterial and bacteriophage communities in three rice agroecosystems and potential impacts of bacteriophage on nutrient cycling. Environmental Microbiome. 2022;17(1):17.

23. Albright MBN, Gallegos-Graves LV, Feeser KL, Montoya K, Emerson JB, Shakya M, et al. Experimental evidence for the impact of soil viruses on carbon cycling during surface plant litter decomposition. ISME Communications. 2022;2(1):24.
24. Braga LPP, Orland C, Emilson EJS, Fitch AA, Osterholz H, Dittmar T, et al. Viruses direct carbon cycling in lake sediments under global change. Proceedings of the National Academy of Sciences. 2022;119(41):e2202261119.
25. Carreira C, Lønborg C, Acharya B, Aryal L, Buivydaite Z, Borim Corrêa F, et al. Integrating viruses into soil food web biogeochemistry. Nat Microbiol. 2024;9(8):1918.
26. Stough JMA, Kolton M, Kostka JE, Weston DJ, Pelletier DA, Wilhelm SW. Diversity of Active Viral Infections within the *Sphagnum* Microbiome. Appl Environ Microbiol. 2018;84(23):16.
27. Piatkowski BT, Carper DL, Carrell AA, Chen IMA, Clum A, Daum C, et al. Draft Metagenome Sequences of the *Sphagnum* (Peat Moss) Microbiome from Ambient and Warmed Environments across Europe. Microbiology Resource Announcements. 2022;11(10).
28. Clum A, Huntemann M, Bushnell B, Foster B, Foster B, Roux S, et al. DOE JGI Metagenome Workflow. mSystems. 2021;6(3).
29. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. United States2014.
30. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6.
31. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.
32. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2023.
33. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39(5):578.

34. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5:26.
35. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. Virol J. 2009;6:13.
36. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366.
37. Krupovic M, Yutin N, Koonin E. Evolution of a major virion protein of the giant pandoraviruses from an inactivated bacterial glycoside hydrolase. Virus Evol. 2020;6(2):veaa059.
38. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. In: Russell DJ, editor. Multiple Sequence Alignment Methods. Methods in Molecular Biology. 1079. Totowa: Humana Press Inc; 2014. p. 105-16.
39. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3.
40. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):10.
41. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat Commun. 2020;11(1):11.
42. Aylward F, Moniruzzaman M, Ha A, Koonin E. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLoS Biol. 2021;19(10):18.
43. Gilbert NE, LeCleir GR, Pound HL, Strzepek RF, Ellwood MJ, Twining BS, et al. Giant virus infection signatures are modulated by euphotic zone depth strata and iron regimes of the subantarctic Southern Ocean. mSystems. 2023.
44. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74.

Appendix

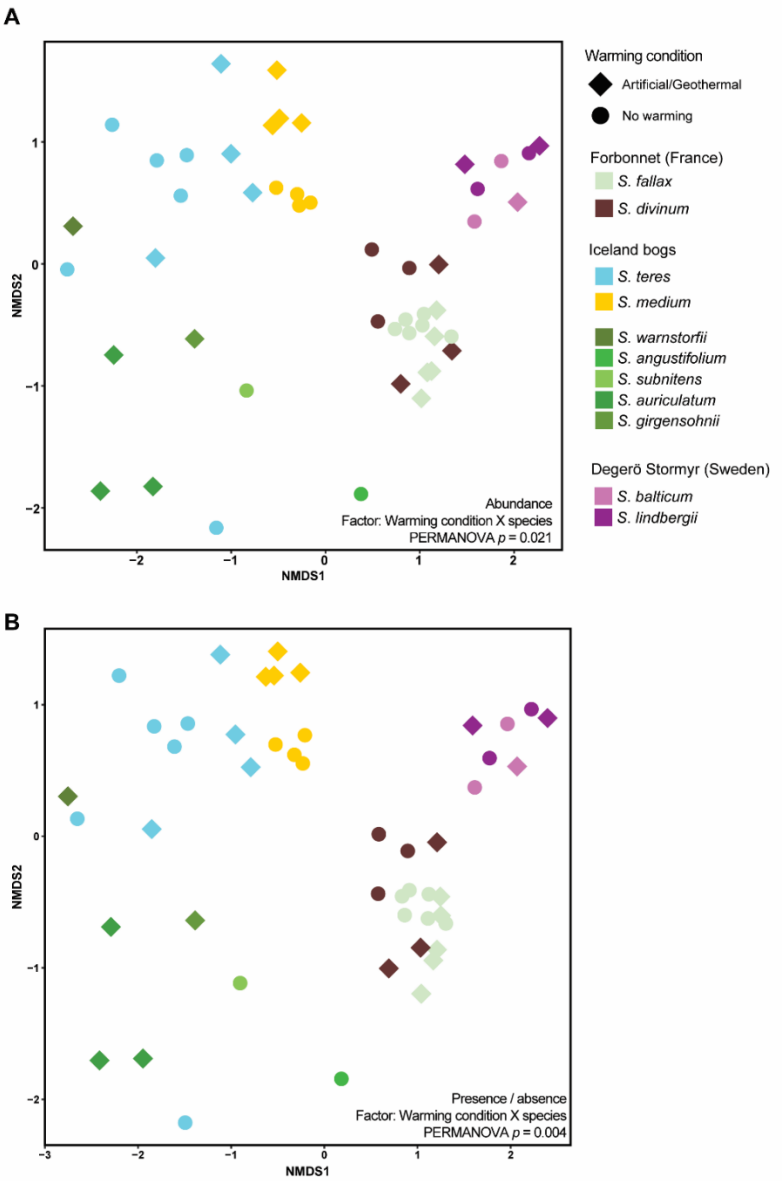


Figure 4.1. Ordinate plots of viral community structure and composition. nMDS plots were generated from Bray-Curtis dissimilarity matrices based on A) abundance values and B) presence-absence transformed values.

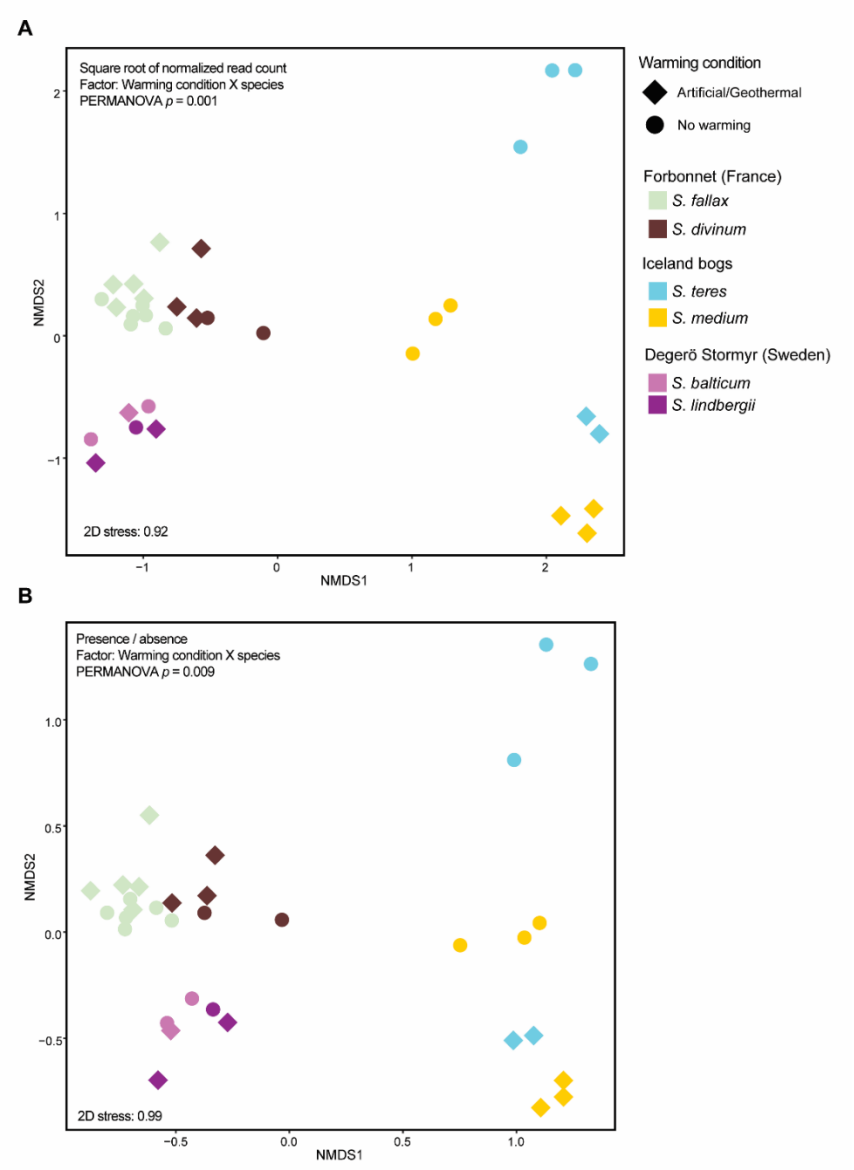


Figure 4.2. Ordinate plots of active viral community structure and composition. nMDS plots were generated from Bray-Curtis dissimilarity matrices based on A) abundance values and B) presence-absence transformed values.

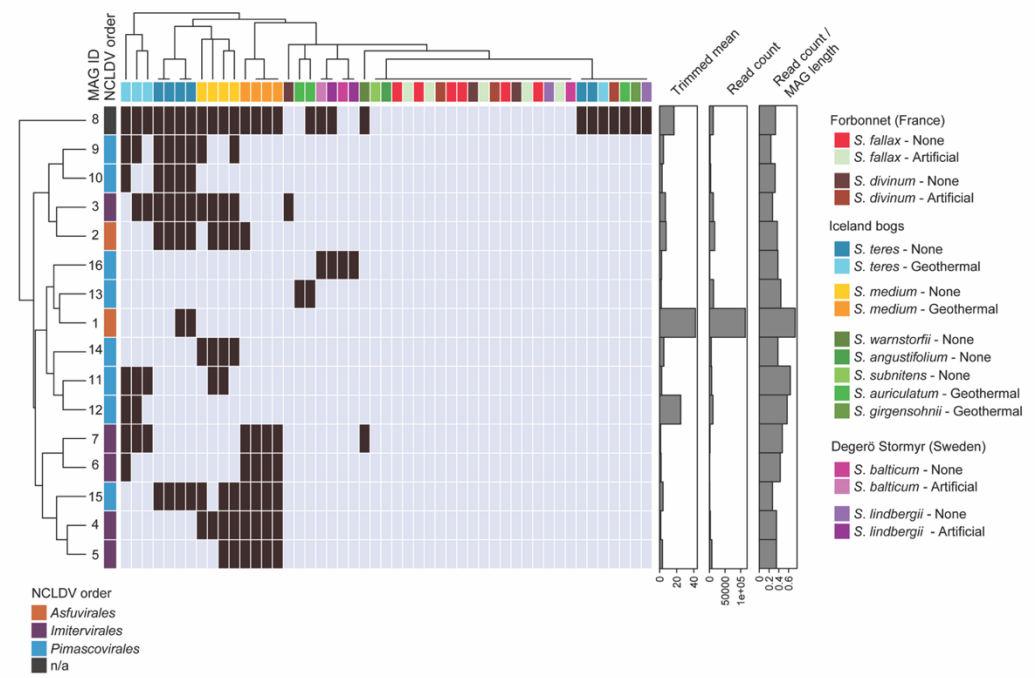


Figure 4.3. Distribution of the NCLDV MAGs in the metagenomes.
Heatmap shows if a NCLDV MAG was detected in a sample. Rows are MAGs (color-coded by taxonomic estimate) and columns are samples. Row branches are labeled with the MAG ID. Samples are color-coded by isolation source/plant species/warming condition. Adjacent bar plots show abundance (trimmed mean, read count, and length-normalized read count) averaged across the samples.

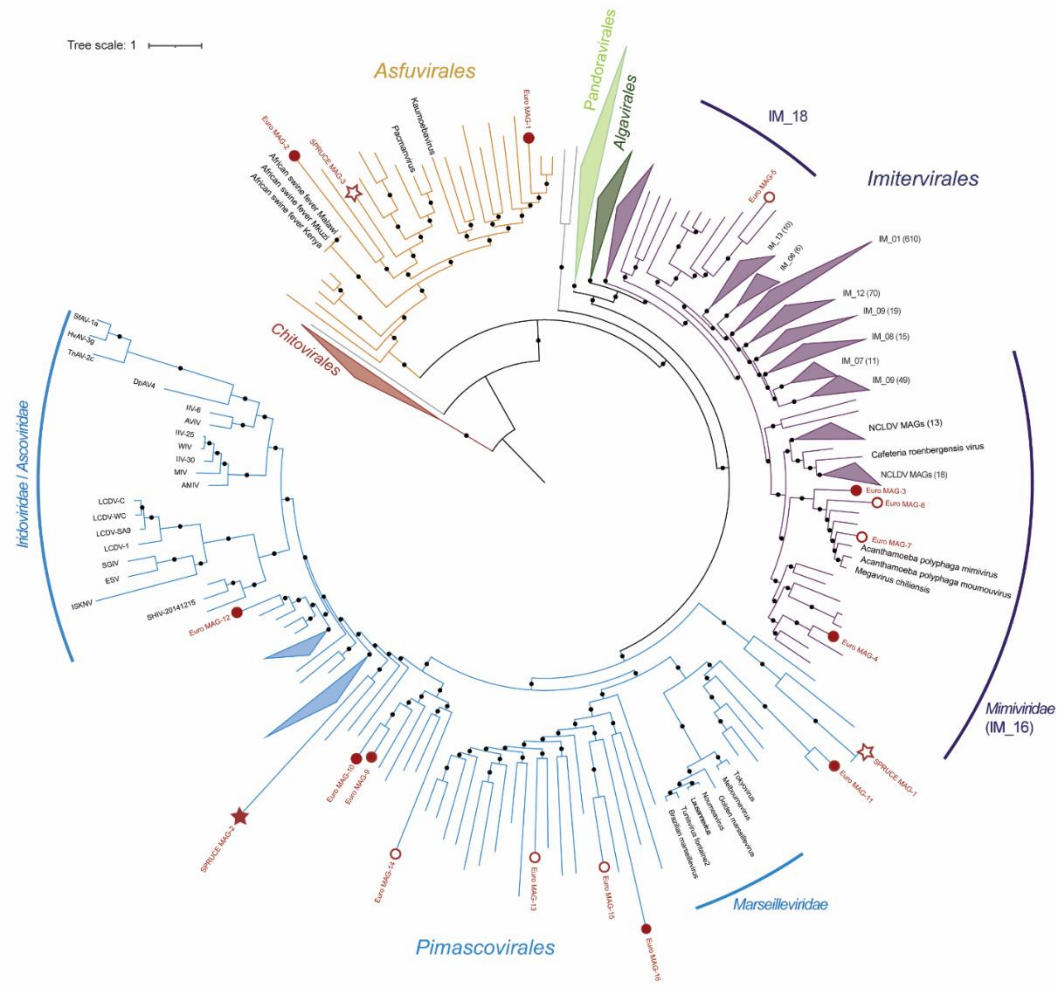


Figure 4.4. Phylogenetic placement of *Sphagnum* microbiome NCLDV MAGs. Nodes with red symbols indicate placement of the *Sphagnum* NCLDV MAGs. Symbols represent either European (circles) or SPRUCE NCLDV MAGs (stars). Filled symbols represent high-quality and medium-quality MAGs, unfilled are low-quality.

CHAPTER FIVE: CONCLUSION

Viral ecology research can illuminate many aspects of ecosystems, from the composition, function, and evolution of microbial communities to global carbon and nutrient cycles. It follows that the current lack of knowledge regarding viral responses to changing climate hinders our understanding of how Earth's ecosystems will function under new climate regimes. This is especially relevant for many ecosystems in northern latitudes that are experiencing changing climate at an accelerated rate. Overall, the work presented in this dissertation helps to characterize viral communities in novel, climate-relevant contexts and may provide a foundation for future research and hypothesis testing. Below is a brief summary of each chapter's conclusions along with additional research questions that have arisen from our findings.

Chapter Two entails a comparative analysis of active viral communities between an ice-free and ice-covered winter in the surface waters of Lake Erie. We found phylogenetically diverse viral communities active during the winter, wherein active viral community composition and diversity were significantly different between ice cover conditions. This work describes viruses in a context not previously investigated before, as well as contributes to our knowledge of how viral communities are impacted by changing climate, in this case declining ice cover over freshwater lakes.

The stark differences we observed in viral activity at the community-level based on ice-cover conditions raises additional questions. For example, while we hypothesize that ice cover extent is a primary driver of the observed difference in active microbial communities, our analysis was restricted to two winters. Continued surveys of microbial communities (including viruses) in Lake Erie from different years with varying ice cover conditions may strengthen or challenge our findings in Chapter Two. The mechanism behind the decrease in viral diversity and relative abundance in the ice-covered samples is still unknown. Experimental manipulation of ice cover extent could also aid in understand the relationship between ice cover and viral activity. For certain aspects this would be difficult (*e.g.*, assessing the direct effects of ice on viral particles or microbial cells), but experiments could address indirect effects of ice-cover on viral activity by manipulating light availability or UV exposure. These open questions also highlight the lack of virus-host systems in culture from the Great Lakes. Overall, viral dynamics are

likely sensitive to ice cover. Therefore, the inclusion of viruses is warranted in any future investigations into the relationship between changing climate, declining ice cover, and freshwater ecology.

Chapter Three describes shared viral sequences in *Sphagnum* moss metagenomes, indicative of viral constituents within the peat moss core microbiome. Our analysis revealed that the core 'virome' consisted of a small portion of the total identified viral diversity. This core was comprised of putative viruses of micro-eukaryotes (NCLDV and RNA viruses) and bacteria (phage), where most phage-like sequences were possibly integrated viral elements in core bacterial symbiont genomes. The integrated elements along with capsid-less putative RNA viruses would likely replicate strictly intracellularly and may reflect modes of persistence in the microbiome. In Chapter Four, we present an investigation into viral community composition and diversity in the *Sphagnum* microbiome across different warming conditions and *Sphagnum* species. We found that there was a significant difference in viral communities by warming condition, but that it was dependent upon the plant species. Notably, the warming difference was only significant for plants sampled from the naturally geothermally warmed sites as opposed to the artificially warmed experimental sites.

While Chapters Three and Four contribute to our knowledge of the *Sphagnum* microbiome, where previously very little was known regarding viruses, these works also raised many more questions regarding the ecology of viruses in peatlands. One question is how viruses, especially core viruses, are acquired and maintained by the moss microbiome. For example, are virus particles or lysogens transmitted vertically as is the case for some bacterial symbionts? Viral metagenomics approaches could be used to compare viral sequences between the sporophyte and gametophyte. Studies of viral communities in the peat versus living plant tissue have thus far been independent. A study with different spatial and depth gradients including both peat soil and the living moss-associated communities may reveal if viruses are acquired from the surrounding peat matrix.

In the case of the putative cryptic phage elements, it is unknown how prevalent they are in *Sphagnum* symbionts and if they confer any selective advantage. Furthermore,

these specific integrated elements may be common outside of *Sphagnum*, as similar regions where found in genomes of organisms for other environments. Comparative genomics and phylogenomics could help to reveal how prevalent they are in nature as well as their evolutionary origin in *Sphagnum* symbionts.

There are still many questions regarding if/how viruses shape *Sphagnum* microbiomes or modulate moss-microbiome interactions. While no core viral sequences were classified as viruses of key moss symbionts, such as diazotrophs or methanotrophs/gens, there may be yet undescribed viruses of these functional guilds. In peatlands soils, bioinformatics has linked viruses to putative hosts and discovered viral-encoded metabolic genes that are important in methane metabolism and carbon degradation. In the moss microbiome, viruses of key symbionts could sever moss-symbiont interactions by lysing hosts or altering host metabolism. Alternatively, they may temporarily enhance certain nutrient metabolisms through the expression of viral-encoded metabolic genes. However, at present these are merely possibilities based on observations from aquatic systems and soils, and future work would be needed to explore these dynamics in the moss microbiome.

In terms of methodology, our knowledge of the moss virome is limited by the fragmented nature of assembled microbial and viral genomes from bulk metagenomes. Long-read sequencing technology can improve the recovery and length of assembled viral genomes. This may be helpful in trying to differentiate between ‘true’ viral genomes and degraded viral elements in microbial genomes. Also, combining bulk metagenomes with technical viromes, where samples are enriched for viral particles and microbial cells are removed, would also aid in disentangling the origin (viral or microbial) of virus-like sequences.

VITA

Elizabeth (Liz) Denison was raised in Knoxville, Tennessee. In 2016, she obtained a Bachelor of Science degree in Biology from the University of Tennessee. Liz then moved to the beach and earned a Master of Science from the University of North Carolina Wilmington. She returned to the mountains in 2019 to pursue a doctorate in Steven Wilhelm's lab at the University of Tennessee.