VIRAL GENOMIC DIVERSITY AND THE EVOLUTION OF HOST RESISTANCE IN *MICROMONAS*-VIRUS SYSTEMS FROM THE TROPICAL NORTH PACIFIC OCEAN

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
OCEANOGRAPHY

AUGUST 2024

By Anamica Bedi de Silva

Dissertation Committee:

Kyle F. Edwards, Chairperson
Marcia F. Marston
Grieg F. Steward
Karen E. Selph
Rebecca A. Chong

Keywords: marine virus, prasinovirus, prasinophyte, cost of resistance

I dedicate this work to the women in my family who were denied higher education.
Thank you for all the sacrifices that you made so that I can be here.

Acknowledgements

I was lucky enough to have Dr. Kyle F. Edwards as my PhD advisor and mentor. My graduate school tenure coincided with some of the most difficult events in my personal life- I lost several family members suddenly, watched as my hometown of Minneapolis burned after the terrifying murder of George Floyd, survived a pandemic, and went through two major surgeries after tearing the ligaments in my ankle. Kyle got me through all of that. He made sure I had financial support for my academics and was uncompromising in his moral support for me. When I felt I could not do it, Kyle made it clear to me that there was only a question of when I would do it. Kyle, I cannot thank you enough.

My dissertation committee was made up of amazing scientists and role models. Drs Rebecca Chong, and Marcie Marston, Karen Selph, and Grieg Steward all provided a wonderful sense of humor alongside invaluable scientific expertise I am extremely lucky that this was the group guiding me through my PhD journey. I am forever grateful. Thank you, all.

The University of Hawai‘i at Mānoa Department of Oceanography has changed significantly for the better during my grad school experience. This was due to the tireless efforts towards diversity, equity, and inclusion by the Women of SOEST. Drs. Barbara Bruno, Jennifer Engels, Margaret McManus, Alison Nugent, Kathleen Ruttenberg and so many others, thank you for being Mana Wahine and making academia a better place for women of color.

Mahalo nui loa to Anne Lawyer, Kristin Momohara, Auntie Catalpa Kong, Mireya Inga, and Pamela Petras for keeping me employed, paid, and on track to graduate. Thanks for lending me your ears when I was stressed about navigating university life.

I am grateful to the members of Black Women in Ecology, Evolution, and Marine Sciences (BWEEMS) for providing a safe space and financial support for me and other Black Women. In particular, thank you Dr. Nikki Traylor Knowles and Keiko Wilkins.

Thank you to my fellow MarVEL lab members, Drs Qian Li, Julie Thomy, Nerissa Fisher, Laetitia Dadaglio, as well as Kelsey Allen, Petra Byl, Andrian “Adi” Gajigan, Amanda Laughlin, and Jess Leaning for providing a supportive and fun environment to conduct research in. A special thank you to Dr. Christopher Schvarcz, as I could not have done any of my research without the culture collection that you worked so tirelessly to establish.

My heart is full thinking about all the amazing friends and colleagues I met during graduate school. A special thank you to my office mates Dr. Steven Scherer and Liz Miller for providing a pleasant workspace and for all your kindness. A big shout out to my sisters-in-science Dr. Emily Young, Dr. Amie Dobracki, and Clarisse Sullivan. You are all so wonderful.

Of course, thank you to my family. Mom, I am so grateful that you are the woman who raised me and for giving me your intelligence. I would not be here without you. Dr. Jessica Bedi thank you for being my first science teacher and an excellent role model. Derek, thank you for your unwavering support and for making a happy home with me. Dr. Elzie McCord, thank you for always cheering me on, you are my grandfather in science. I love you all so much.

Funding for this dissertation was generously provided by the National Sciences Foundation (Awards OCE 1559356, OCE 2129697 and RII Track-2 FEC 1736030), the ARCS Foundation, BWEEMS, and UH Mānoa Department of Oceanography Teaching Assistantships.

Abstract

Viruses in the ocean outnumber microbes by approximately an order of magnitude. These pathogens shape microbial communities via manipulation of host metabolism, through the mortality of host cells, and by causing evolutionary change in the host population. While the dynamics of cyanobacteria and their phages have provided a foundation of knowledge on coevolutionary dynamics through the lens of genetics and phenotypic selection, there is a smaller foundation for marine eukaryotic phytoplankton and their associated viruses. The important role of eukaryotic phytoplankton in marine primary production and biogeochemical cycling merits further investigations into how these phytoplankton interact with their pathogens. The question of how viruses evolve to successfully infect their hosts may be answered in part by examination of viral gene content and the evolutionary origins of such genes. Furthermore, virus-induced mortality selects for host phenotypes that are resistant to lysis. It has been hypothesized that the evolution of resistance results in a fitness cost that could alter phytoplankton productivity, but the magnitude of this cost, and how it varies under different resource regimes, is not clear. Such selection is presumably tied to genomic change in resistant cells, but resistance mutations and their mechanistic effects are poorly understood. We sought to examine questions on the dynamics of marine viruses infecting eukaryotic hosts using isolates of the common and ecologically relevant alga *Micromonas commoda* and strains of its double-stranded DNA viruses in the genus *Prasinovirus*. These isolates represent the first prasinophyte-prasinovirus systems isolated from the North Pacific Subtropical Gyre. Through four new genomic assemblies of *Micromonas commoda* virus strains, we found 61 putative genes not seen in other prasinoviruses. Additionally, 192 putative genes varied in occurrence among the four virus strains, despite the fact that they have overlapping host ranges. Across prasinoviruses, 25% of gene content is strongly correlated with host genus, and the functions of these genes suggests that successful lytic infection is achieved through a diversity of genetic strategies. We subsequently used experimental evolution to create 88 resistant *M. commoda* cell lines and found a large decrease in fitness, particularly when cells were grown at a higher light level. This fitness cost attenuated after 15

months even while cell populations maintained resistance, suggesting compensatory mutations can ameliorate the cost of resisting infection. The genomes of resistant cell lines had a larger number of non-synonymous variants than susceptible control lines. The genes affected by such variants were dependent on the identity of the ancestral cell line, and indicated a diverse suite of mechanisms of resistance among closely related isolates. Both resistance mutations and viral gene content imply that host stress responses, such as programmed cell death, are an important site of coevolutionary antagonism, as are genes potentially related to viral attachment and entry. This work demonstrates a complex network of coevolutionary strategies among marine eukaryotes and their viruses.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction: Expanding the knowledge of viruses of marine protists

As humans, our relationship with viruses is nuanced at best, particularly in the wake of the global COVID-19 pandemic. The danger to our health during this historic time was exacerbated by the lack of public education on the biology of viruses. This paucity of knowledge is understandable, given the relative difficulty of observing and manipulating viruses in a controlled environment. While public health unequivocally takes immediate priority, research on viruses that shape the global scale environment is also valuable when considering a holistic world view. The marine environment, in particular, is home to the young and arguably understudied field of marine viral ecology. This field of study only started to gain widespread attention in the late 1980s when marine scientists developed methods to reliably quantify the number of virus-like particles in the ocean (Bergh et al., 1989). Today, we know that the number of viruses in the ocean outnumber any other portion of the marine microbial community by at least an order of magnitude (Wommack and Colwell, 2000). The majority of these viruses infect single-celled organisms, including bacteria, archaea, and eukaryotic phytoplankton (Proctor and Fuhrman, 1990; Suttle et al., 1990; Philosof et al., 2017). It is through these organisms that viruses have a global impact on the health of our ecosystems (Fuhrman, 1999).

1.1: VIRAL BIOLOGY AND BASIC ECOLOGY

All viruses, marine or otherwise, have basic structural components of a protein “shell”, called a capsid, that encapsulates and protects genetic material, and the viral genome itself, which can take the form of DNA or RNA. Viruses cannot reproduce on their own, and instead they parasitize cells to synthesize and assemble viral progeny. The parasitization cycle begins when a virus encounters a compatible host cell. At this point, the virus attaches to the cell and insert their genetic material into it through one of a variety of entry mechanisms (Dimmock et al., 2016). Once genetic material has entered the host cell the virus may lay dormant, perhaps by integrating its genome into the host genome, or may continue the replication cycle, co-opting cellular materials and machinery used for protein synthesis and genome replication to make the structural and genetic components of new viruses. These new viruses are assembled and then released into the environment, often through a process of cellular disintegration referred

to as “lysis.” Viral infection that ends in lysis is called lytic infection, which is the focus of my work.

Through lysis viruses contribute to microbial mortality as much as zooplankton grazing (Evans et al., 2003). While grazing activities allow for the passage of carbon and nutrients from microbes to larger organisms at higher trophic levels, lytic infection results in the release of both dissolved organic matter (DOM) and particulate organic matter (POM) that was once contained inside cells (Suttle et al., 1990; Våge et al., 2013). This pool of organic matter can be consumed by heterotrophic bacteria in a process called the viral shunt, so named because viral lysis “shunts” organic matter away from higher trophic levels. The shunting of energy and nutrients to heterotrophic microbes reduces the growth of larger organisms and may result in lower productivity of the system (Suttle et al., 1990; Våge et al., 2013). In addition, the process of lysis releases large polymers, some of which are “sticky” substances that can aggregate with other marine detritus, such as fecal pellets. Aggregated material tends to sink more rapidly than individual components and thus important elements may be exported out of the euphotic zone at a faster rate in the presence of viral lysis (Guidi et al., 2016). This process is called the “viral shuttle.”

Viruses can also impact biogeochemical cycling through genetics. Viruses of microbes carry a plethora of genes originating from cells and other viruses that are obtained through horizontal gene transfer (Bachy et al., 2021). These genes are often involved in nutrient uptake and cellular metabolism (Monier et al. 2012). The implication is that viruses carry a suite of genes that can manipulate the host to allocate resources towards processes beneficial for viral replication. Therefore, the metabolism of infected “virocells” may differ in important ways from other cells in the environment, which could affect ecosystem function (Rosenwasser et al., 2016). Viral alteration of cellular metabolism could be particularly significant in resource-poor pelagic environments.

Horizontal transfer of metabolic genes is just one example of the broader range of evolutionary processes involving viruses that could influence biogeochemistry. The direction of viral evolution is towards successful infection of the host, with the counterbalance being a continual evolutionary drive of hosts toward better resistance to infection (Martiny et al., 2014). These opposing selection pressures lead to antagonistic

coevolution, which can occur at a rapid pace because selection in host-virus systems is strong: viruses depend entirely on their hosts for survival, while host cells will die if a lytic infection is successful (Buckling and Rainey, 2002; Weitz et al., 2005). Over the long term, these processes may have important ecological/ecosystem consequences, because the evolution of traits altering viral success will alter rates of mortality and lysis, and because resisting viral infection may diminish host growth rates (Lennon et al., 2007). Therefore, host-virus coevolution may ultimately have large effects on productivity and element cycling (Våge et al., 2013).

1.2: PRASINOPHYTES AND PRASINOVIRUSES

The majority of our knowledge of marine viral ecology is derived from cyanobacteria and their associated viruses, known as cyanophage (Breitbart, 2012). While the cyanobacteria make up the most numerous fraction of photosynthetic organisms in the ocean, eukaryotic phytoplankton dominate total biomass, and make a greater contribution to carbon flux out of the euphotic zone and into the deep ocean (Falkowski et al., 2004). Given the influence of viruses on carbon and nutrient cycling, the expansion of viral ecology necessitates a more comprehensive understanding of infection dynamics of marine protists. One group of eukaryotic hosts and viruses that have allowed for notable contributions to the field are the prasinophytes and their double-stranded DNA (dsDNA) viruses, the prasinoviruses.

The first virus isolated on a unicellular marine eukaryote caused lysis to members of the prasinophyte genus *Micromonas* (Mayer and Taylor, 1979). This bacterium-sized (~2 μm) alga belongs to the order of marine picoeukaryotes called the Mamiellales, which, along with their prasinoviruses, have been the subject of foundational research concerning marine viruses, due in part to their amenability to laboratory manipulation. Included in this order is the genus *Bathycoccus*, as well as *Ostreococcus*, the smallest known free-living eukaryote. These three genera are thought to be at the phylogenetic base of the green algae-terrestrial plant divide (Worden et al., 2009), and often dominate the picoeukaryotic fraction of phytoplankton in both coastal and open ocean environments (Worden et al., 2009; Bachy et al., 2021; Ha et al., 2023).

The known dsDNA viruses infecting the Mamiellales are in the genus *Prasinovirus*, within the family Phycodnaviridae, in the phylum *Nucleocytoviricota*, which are colloquially referred to as 'giant viruses' (ICTV, 2023; Ha and Aylward, 2024). Prasinoviruses possess icosahedral capsids ~104-118 nm in diameter with ~150-200 kbp genomes (Brandes and Linial, 2019; Bachy et al., 2021). Like other members of the virus family Phycodnaviridae, prasinoviruses replicate their genome in the host nucleus and assemble virions in the cytoplasm (Weynberg et al., 2017). Prasinophyte-prasinovirus systems have provided evidence of horizontal gene transfer between host and virus (Moreau et al., 2010; Finke et al., 2017; Weynberg et al., 2017). Researchers have also used such systems to assess fitness costs associated with viral immunity, to analyze comparative genomics of hosts and viruses, to establish phylogenies, and to test how resource availability alters infection dynamics (Brown et al., 2007; Thomas et al., 2011; Demory et al., 2017; Heath et al., 2017; Bachy et al., 2021).

1.3: THE NEED FOR MORE ISOLATES

Isolation and cultivation of viruses and their hosts allows for controlled experiments and for the establishment of reference material, including whole genomes of viruses with known host identity. Therefore, isolates are invaluable to our ability to form well-supported theories about ecological and evolutionary processes in host-virus systems. The majority of prasinophyte-virus systems in global culture collections were isolated in the Mediterranean or the Atlantic, with less representation from the South Pacific (e.g., Moreau et al., 2010; Weynberg et al., 2017; Bachy et al., 2021; see additional strain information for cells in Bigelow National Center for Algae and Microbiota (NCMA) and Roscoff Culture Collection (RCC) catalogs). While there are some cellular isolates from coastal California, there are no prasinovirus strains isolated from the North Pacific Subtropical Gyre (NPSG), which is the largest contiguous ecosystem on earth (Karl 1999). Ecotypic partitioning of hosts and viruses has been seen in cyanobacteria and eukaryotic algae, suggesting that geography and depth can influence phytoplankton and virus populations (Clerissi et al., 2014; Marston and Martiny, 2016). A continued effort to examine isolates from underrepresented environments is then merited.

The University of Hawai‘i at Mānoa (UHM) culture collection, which houses phytoplankton and associated viruses isolated from the coastal and pelagic waters of Hawai‘i, is a valuable resource for answering questions regarding marine viral ecology. The UHM collection contains seven strains of *Micromonas*, with one prasinovirus isolated on each host strain. Six strains were isolated from the surface waters (< 2 m) of Kāne‘ohe Bay in the springs of 2010 and 2011 (Schvarcz, 2018). Kāne‘ohe Bay is an oligotrophic coastal environment on the windward side of the island of O‘ahu influenced by nutrient pulses, in the form of storm events, from a well-populated watershed (see McKenzie, 2018 and references therein). One *Micromonas* and virus pair was isolated from the oligotrophic Station ALOHA, located 100 km north of O‘ahu (Schvarcz, 2018). This pelagic station exemplifies the permanently stratified, low nutrient waters of the North Pacific Subtropical Gyre, with a phytoplankton community dominated by *Prochlorococcus* and small eukaryotes such as prasinophytes (Karl and Church, 2014; Rii et al., 2016). The Station ALOHA *Micromonas* host was isolated in 2012, while the virus was isolated in 2015.

1.4: RESEARCH OBJECTIVES

The multiple strains of *Micromonas* and prasinovirus in the UHM collection facilitate comparative research (comparing novel viruses and their hosts to each other, as well as to previous isolates) and experimental research (utilizing multiple virus-host pairs to conduct controlled tests). In this dissertation I use these host-virus systems to gain insight into the evolutionary processes affecting viruses and their hosts, and the ecological consequences of evolutionary change. Because *Micromonas* and other Mamiellales are ecologically important algae, understanding their interactions with lytic viruses is relevant in the broader context of phytoplankton community dynamics and impacts on marine biogeochemical cycling. This brings us to the overarching questions that will be examined in this body of work.

1.4.1: Genome evolution of *Micromonas* viruses

- How has the gene content of *Micromonas* viruses and other prasinoviruses evolved over time, based on comparison of four novel viral genomes to previously published isolates?

- What are core genes among UHM *Micromonas* virus isolates? Which genes are unique to each strain? Are there genes in our isolates that have not been found in related virus strains?
- What genes are shared with *Micromonas* cell lines?
- When comparing related viruses with our *Micromonas* virus strains, which genes, if any, are associated with virus host genus?

To answer these questions, I begin this body of work by examining the first genomic assemblies of prasinoviruses isolated in Hawaiian waters, in **Chapter 2: Comparative genomics of four newly sequenced viruses infecting the picoeukaryote *Micromonas***. The genomes of four *Micromonas* virus strains in our collection were compared to each other, to their hosts' genomes, and to previously published prasinovirus genomes.

1.4.2: Ecological consequences of resisting viral infection in *Micromonas*

- In response to the presence of lytic viruses, what consequences does selection for resistance have on the fitness of *Micromonas* populations?
- Do resistant populations have observable changes to growth rate?
- Are these changes impacted by availability of resources?
- Are such impacts consistent over time?
- What are the ecological implications of viral selection pressure?

Chapter 3: Transient, context-dependent fitness costs accompanying viral resistance in isolates of the marine microalga *Micromonas* sp. (class Mamiellophyceae) goes on to address these questions about the consequences of resisting lytic viral infection, using experimental evolution to select for resistant cells. Resistant cells are then compared to susceptible counterparts to determine if there is a fitness cost to viral immunity shortly after isolation, and whether this fitness cost persists 15 months later.

1.4.3: The genomic basis of resisting viral infection in *Micromonas*

- Are there genetic signatures of resistance in the genomes of *Micromonas* that are immune to lytic viral infection?
- If so, are these signatures composed of small-scale changes to individual genes or large-scale structural variations in chromosomes?
- What cellular mechanisms are implicated in causing resistance to viral infection?

In **Chapter 4: Genetic signatures of resistance to viral infection in the marine picoeukaryote *Micromonas*,** I address the genomic basis of resistance by examining the genomes of cell lines experimentally selected for resistance, using short-read Illumina and long-read PacBio technology, to examine putative genetic signatures of resistance in the form of small polymorphisms and large structural variants.

1.4.4: Synthesis

Finally, in **Chapter 5: Conclusions**, I synthesize what novel insights were gained from the three data chapters, about the dynamics of eukaryotic phytoplankton and their viruses, adding to both the genetic and phenotypic understanding of these bipartite systems.

1.5 REFERENCES

Bachy, C., Yung, C.C.M., Needham, D.M., Gazitúa, M.C., Roux, S., Limardo, A.J., et al. (2021) Viruses infecting a warm water picoeukaryote shed light on spatial co-occurrence dynamics of marine viruses and their hosts. ISME J 15: 3129–3147.

Bergh, Ø., Børsheim, K.Y., Bratbak, G., and Heldal, M. (1989) High abundance of viruses found in aquatic environments. Nature 340: 467–468.

Brandes, N. and Linial, M. (2019) Giant Viruses - Big Surprises, LIFE SCIENCES.

Breitbart, M. (2012) Marine Viruses: Truth or Dare. Annu Rev Mar Sci 4: 425–448.

Brown, C.M., Campbell, D.A., and Lawrence, J.E. (2007) Resource dynamics during infection of *Micromonas* pusilla by virus MpV-Sp1. Environ Microbiol 9: 2720–2727.

Buckling, A. and Rainey, P.B. (2002) Antagonistic coevolution between a bacterium and a bacteriophage. Proc R Soc Lond B 269: 931–936.

Clerissi, C., Grimsley, N., Subirana, L., Maria, E., Oriol, L., Ogata, H., et al. (2014) Prasinovirus distribution in the Northwest Mediterranean Sea is affected by the environment and particularly by phosphate availability. Virology 466–467: 146–157.

Demory, D., Arsenieff, L., Simon, N., Six, C., Rigaut-Jalabert, F., Marie, D., et al. (2017) Temperature is a key factor in *Micromonas*–virus interactions. ISME J 11: 601–612.

Dimmock, N.J., Easton, A.J., and Leppard, K.N. (2016) Part II: Virus Growth in Cells. In Introduction to Modern Virology. John Wiley & Sons, Inc.

Evans, C., Archer, S., Jacquet, S., and Wilson, W. (2003) Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. Aquat Microb Ecol 30: 207–219.

Falkowski, P.G., Katz, M.E., Knoll, A.H., Quigg, A., Raven, J.A., Schofield, O., and Taylor, F.J.R. (2004) The Evolution of Modern Eukaryotic Phytoplankton. Science 305: 354–360.

Finke, J., Winget, D., Chan, A., and Suttle, C. (2017) Variation in the Genetic Repertoire of Viruses Infecting *Micromonas* pusilla Reflects Horizontal Gene Transfer and Links to Their Environmental Distribution. Viruses 9: 116.

Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399: 541–548.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. Nature 532: 465–470.

Ha, A.D. and Aylward, F.O. (2024) Automated classification of giant virus genomes using a random forest model built on trademark protein families. npj Viruses 2: 9.

Ha, A.D., Moniruzzaman, M., and Aylward, F.O. (2023) Assessing the biogeography of marine giant viruses in four oceanic transects. ISME COMMUN 3: 43.

Heath, S., Knox, K., Vale, P., and Collins, S. (2017) Virus Resistance Is Not Costly in a Marine Alga Evolving under Multiple Environmental Stressors. Viruses 9: 39.

ICTV, I.C. on T. of V. (2023) Virus Taxonomy: 2022 Release.

Karl, D.M. A Sea of Change: Biogeochemical Variability in the North Pacific Subtropical Gyre. 34.

Karl, D.M. and Church, M.J. (2014) Microbial oceanography and the Hawaii Ocean Time-series programme. Nat Rev Microbiol 12: 699–713.

Lennon, J.T., Khatana, S.A.M., Marston, M.F., and Martiny, J.B.H. (2007) Is there a cost of virus resistance in marine cyanobacteria? ISME J 1: 300–312.

Marston, M.F. and Martiny, J.B.H. (2016) Genomic diversification of marine cyanophages into stable ecotypes: Cyanophage diversification into ecotypes. Environ Microbiol 18: 4240–4253.

Martiny, J.B.H., Riemann, L., Marston, M.F., and Middelboe, M. (2014) Antagonistic Coevolution of Marine Planktonic Viruses and Their Hosts. Annu Rev Mar Sci 6: 393–414.

Mayer, J.A. and Taylor, F.J.R. (1979) A virus which lyses the marine nanoflagellate *Micromonas* pusilla. Nature 281: 3.

McKenzie, T. (2018) Kāneʻohe bay and watershed: a review of its hydrogeology, historical and current sources of pollution and methods to quantify pollutant fluxes and sources.

Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., and Grimsley, N. (2010) Marine Prasinovirus Genomes Show Low Evolutionary Divergence and Acquisition of Protein Metabolism Genes by Horizontal Gene Transfer. J Virol 84: 12555–12563.

Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E.V., and Béjà, O. (2017) Novel Abundant Oceanic Viruses of Uncultured Marine Group II Euryarchaeota. Current Biology 27: 1362–1368.

Proctor, L.M. and Fuhrman, J.A. (1990) Viral mortality of marine bacteria and cyanobacteria. Nature 343: 60–62.

Rii, Y., Karl, D., and Church, M. (2016) Temporal and vertical variability in picophytoplankton primary productivity in the North Pacific Subtropical Gyre. Mar Ecol Prog Ser 562: 1–18.

Rosenwasser, S., Ziv, C., Creveld, S.G.V., and Vardi, A. (2016) Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean. Trends in Microbiology 24: 821–832.

Schvarcz, C.R. (2018) Cultivation and Characterization of Viruses Infecting Eukaryotic Phytoplankton from the Tropical North Pacific Ocean.

Suttle, C.A., Chan, A.M., and Cottrell, M.T. (1990) Infection of phytoplankton by viruses and reduction of primary productivity. Nature 347: 467–469.

Thomas, R., Grimsley, N., Escande, M., Subirana, L., Derelle, E., and Moreau, H. (2011) Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations: Viral resistance in Mamiellales. Environmental Microbiology 13: 1412–1420.

Våge, S., Storesund, J.E., and Thingstad, T.F. (2013) Adding a cost of resistance description extends the ability of virus-host model to explain observed patterns in structure and function of pelagic microbial communities: Structuring of microbial communities by viruses. Environ Microbiol 15: 1842–1852.

Weitz, J.S., Hartman, H., and Levin, S.A. (2005) Coevolutionary arms races between bacteria and bacteriophage. Proc Natl Acad Sci USA 102: 9535–9540.

Weynberg, K., Allen, M., and Wilson, W. (2017) Marine Prasinoviruses and Their Tiny Plankton Hosts: A Review. Viruses 9: 43.

Wommack, K.E. and Colwell, R.R. (2000) Virioplankton: Viruses in Aquatic Ecosystems. Microbiol Mol Biol Rev 64: 69–114.

Worden, A.Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M.P., Aerts, A.L., et al. (2009) Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes *Micromonas*. Science 324: 268–272.

Chapter 2: Comparative genomics of four newly sequenced viruses infecting the picoeukaryote *Micromonas*

2.1 ABSTRACT

Viruses that infect algae are an integral part of marine ecosystems, as they alter biogeochemical cycling via cell lysis, manipulate cellular processes of their hosts, and influence biodiversity through mortality, selection, and horizontal gene transfer. The vast majority of viral diversity remains uncultivated, and additional isolates of phytoplankton viruses are needed to establish reference genomes for environmental sequence data, to better characterize the diversity of gene content across viruses, and to better understand viral evolution and host-virus coevolution. Here we introduce four near-complete genomic assemblies of viruses that infect the widespread marine picoeukaryote *Micromonas commoda*. All hosts and virus isolates were obtained from the interior of the North Pacific Subtropical Gyre, a first for viruses infecting the order Mamiellales. Genome length of the new virus isolates range from 205-212 kb, and phylogenetic analysis shows that all four are members of the genus *Prasinovirus*. Three of the viruses form a clade that is adjacent to previously sequenced *Micromonas* viruses, while the fourth novel virus is relatively divergent from previously sequenced prasinoviruses. We explored the gene content of the new viruses in relation to all previously published prasinovirus genomes, as well the genomes of two co-isolated hosts. We identified 61 putative genes not previously found in prasinoviruses, as well as 48 genes that are shared with the hosts and may have been acquired through horizontal gene transfer. Some of the notable genes discovered include a phosphate transporter that is distinct from previously known phycodnavirus phosphate transporters, and a plastid terminal oxidase previously found to be common in cyanophages, both of which are shared with one of the host genomes. By comparing *Micromonas*-, *Ostreococcus*-, and *Bathycoccus*-infecting prasinoviruses we find that ~25% of prasinovirus gene content is significantly correlated with host genus identity, with gene functions suggesting that much of the viral life cycle is differentially adapted to the host genera. Mapping of metagenomic reads from global survey data indicates that one of the new isolates, McV-SA1, is relatively common in multiple ocean basins.

2.2 INTRODUCTION

A significant fraction of marine viral diversity is composed of viruses that infect phytoplankton, the diverse unicellular primary producers that perform the majority of

marine photosynthesis (Suttle, 2007). Viruses that infect phytoplankton affect biogeochemical processes by lysing their hosts, thereby shunting nutrients and energy to smaller microbial cells (Wilhelm and Suttle, 1999), and by shaping phytoplankton production through mortality and metabolic manipulation. In this study we focus on prasinoviruses, which are double-stranded DNA viruses that infect prasinophytes, an ecologically important group of eukaryotic algae. Due to their global ubiquity, relatively high abundance, and amenability to laboratory manipulation, prasinoviruses are a useful model system for understanding the biology and functional consequences of phytoplankton viruses (Moreau et al., 2010; Clerissi et al., 2014; Lopes Dos Santos et al., 2017). In environmental samples, prasinoviruses are typically among the most abundant members of the *Nucleocytoviricota*, which is a highly diverse phylum of large, eukaryote-infecting dsDNA viruses (Farzad et al., 2022; Ha et al., 2023). Prasinoviruses impact marine biogeochemical processes via infection of green algae in the order Mamiellales, which includes the three cosmopolitan genera *Bathycoccus*, *Ostreococcus*, and *Micromonas* (Yau et al., 2015; ICTV, 2023). It should be noted that prasinophytes themselves are a diverse and paraphyletic group (Leliaert et al., 2012), but isolated members of the genus *Prasinovirus* all infect members of the order Mamiellales (Bachy et al., 2021). The Mamiellales are known for their small size, which is exemplified by the species *Ostreococcus tauri*, considered to be the smallest free-living eukaryote at ~0.8 µm cell diameter, while *Bathycoccus* and *Micromonas* are ~2 µm in diameter (Manton and Parke, 1960; Eikrem and Throndsen, 1990; Courties et al., 1994). The Mamiellales are ubiquitous in the sunlit ocean and often major community members in both oligotrophic and eutrophic environments, typically dominating the picoeukaryotic fraction of primary producers under productive conditions (Not et al., 2004; Lopes Dos Santos et al., 2017). Isolates of prasinoviruses have been used to study several topics, such as marine viral gene content (e.g., Finke et al., 2017; Bachy et al., 2021), consequences of host resistance to lytic infection (e.g., Thomas et al., 2012; Heath et al., 2017), and diel changes to the dynamics of viral infection (e.g., Derelle et al., 2015).

In juxtaposition to their diminutive hosts, prasinoviruses are relatively large viruses, with an average genome size of ~192 kbp among published prasinovirus

genomes (Weynberg et al., 2017), which makes prasinovirus genomes approximately 1% of the size of their host genomes. For comparison, the human-associated dsDNA poxviruses have a similar genome size to prasinoviruses at 130-360 kb, which is only ~0.005% the size of the human genome. While there is still much to be understood about how the size and composition of prasinovirus genomes relate to their ecological success, past studies on the comparative genomics of these viruses have described key characteristics of prasinovirus gene content (Moreau et al., 2010; Finke et al., 2017; Bachy et al., 2021). As of this writing 22 prasinovirus genomes have been published, including four *Micromonas* viruses, five *Bathycoccus* viruses, and 12 *Ostreococcus* viruses. Genes shared among all published prasinovirus genomes (i.e., core genes) are largely associated with basic viral functions, such as DNA replication, transcription, and nucleotide metabolism (Moreau et al., 2010). In contrast, non-core gene (i.e., genes present in a clade but not shared by all members) include some involved in cellular functions, such as nutrient acquisition, photosynthesis, and carbohydrate metabolism (Moreau et al., 2010; Finke et al., 2017). However, few prasinovirus genes have been experimentally evaluated for function (Monier et al., 2017). By possessing genes associated with cell function prasinoviruses are potentially able to reshape host metabolism in a way that enhances viral replication, and such gene content may be of particular advantage when resources are scarce or fluctuating.

The characterization of additional viral isolates and their hosts continues to be essential, as isolates contribute to the database of viral reference genomes with known host taxa, while also allowing experimental studies of the ecology and (co)evolution of host-virus systems. Sequencing additional prasinovirus representatives would facilitate a better understanding of the scope of prasinovirus genetic diversity, as well as how genome content has diverged among prasinovirus clades. It would be particularly beneficial to expand the geographic diversity of environments from which isolates have been collected, with prior isolates being largely from the Mediterranean and Atlantic, with some representation from the South Pacific.

In the current work we contribute to the understanding of prasinovirus diversity by introducing and examining four genomic assemblies of *Micromonas commoda* viruses, three isolated from a coastal location (Kāneʻohe Bay, Oʻahu) and one from an

open-ocean location (Station ALOHA, 22°45'N 158°W) near Hawai‘i. Our four isolates represent the first prasinovirus genome assemblies from the interior of the North Pacific Subtropical Gyre. In a metagenomic study, Ha et al. (2023) found that Algavirales, the order to which prasinoviruses belong, dominated giant virus communities in northern latitudes, and were found throughout latitudes of the Atlantic, but were not as abundant in the Pacific. Given that our virus isolates are exclusively from the Pacific, we sought to use similar methodologies to Ha et al. in order to determine if sequences similar to our virus isolates could be found in other ocean basins, or if the Pacific isolates are restricted to their home basin. Additionally, we utilize two recently sequenced genomes of *Micromonas commoda* isolated from Kāne‘ohe Bay (Chapter 4), to compare virus-host gene content. In total our analyses aim to address the following questions:

1. What are the phylogenetic relationships between the four new *Micromonas* viruses and previously sequenced prasinoviruses?

2. How does the gene content of the four new *Micromonas* viruses vary, and how does it compare to the four previously published *Micromonas* viruses, as well as other prasinoviruses? Are there genes in our strains not previously found in *Micromonas* viruses, in prasinoviruses as a whole, or in viruses in general? How do putative gene functions differ between core and non-core genes?

3. Are there consistent differences in gene content among viruses that infect the three Mamiellales genera examined (*Micromonas*, *Ostreococcus*, and *Bathycoccus*)? What gene functions are associated with adapting to infect the different host genera?

4. How does gene content of the new *Micromonas* viruses relate to that of their hosts? What cellular functions are potentially manipulated by host-acquired genes?

5. How common are the newly isolated *Micromonas* virus strains in the global ocean, based on existing metagenomic survey data?

2.3 METHODS

2.3.1 Virus isolation

Four virus strains infecting the marine eukaryote *Micromonas commoda* were examined in this study. Three of the strains, McV-KB2, McV-KB3, and McV-KB4, were isolated from Kāne‘ohe Bay on the windward side of O‘ahu. The fourth strain, McV-SA1, which was previously referred to as MsV-SA1 in Schvarcz (2018), was isolated from the

pelagic research site Station ALOHA (22°45'N 158°W). For simplicity, we will refer to this suite of Hawai‘i *Micromonas commoda* virus strains as “HiMcVs”, short for “Hawai‘i *Micromonas commoda* Viruses.” The four HiMcVs overlap in host range, based on lysis tests with seven *Micromonas* strains isolated from Kāne‘ohe Bay and Station ALOHA, with each viral strain infecting 2-6 *Micromonas* strains (Fig. 2.1). All virus strains are maintained in the UHM Culture Collection via propagation on their original hosts, which were isolated from the same waters as their corresponding virus strains. Full isolation methods are described in Schvarcz (2018). In brief, whole sea water from respective sites was filtered, concentrated, and then added to healthy cell cultures which were subsequently monitored for lysis. If lytic effects were confirmed after multiple transfers to healthy culture, lysates were further purified through several rounds of dilution-to-extinction. The identity of virus-like particles from successful lysates were confirmed via whole-genome sequencing for all four viruses, and with transmission electron microscopy for McV-SA1. Electron micrographs of McV-SA1 established virus particle size to be ~142-160 nm.

In the current study, lysate stocks were maintained through fortnightly transfers of lysate into healthy cultures grown in f/2 -Si medium (Guillard and Ryther, 1962; Guillard, 1975). Once lysis took place, typically within 4-6 days of the initial challenge, lysates were stored at 4°C.

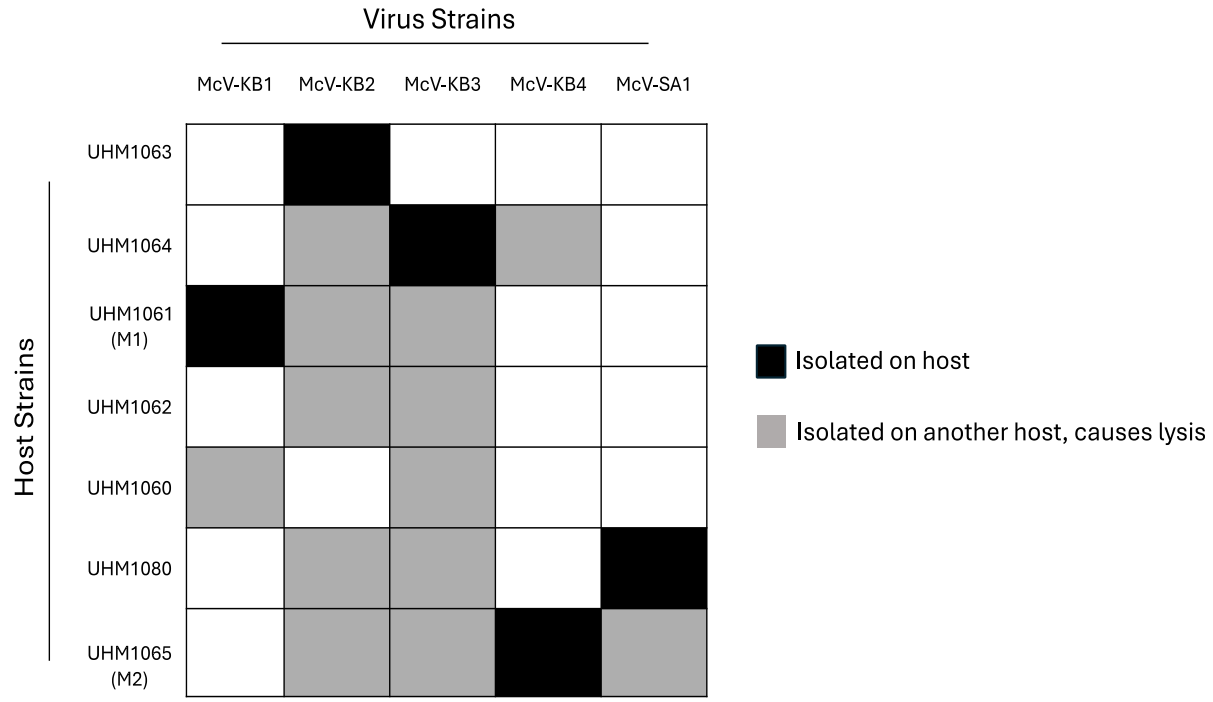


Figure 2.1. Host range of virus strains used in this study. UHM culture collection IDs are listed for 7 *Micromonas* isolates, with shortened IDs in parentheses indicating the two host isolates with sequenced genomes (M1 and M2). Black squares indicate a virus-host pair that were isolated together. Gray squares indicate additional successful lytic infection on non-original hosts.

2.3.2 Whole-genome sequencing

The sequenced and assembled genome of McV-SA1 was provided by Dr. Christopher R. Schvarcz. Purification, extraction, and sequencing protocols for this strain are described in Schvarcz (2018). Briefly, McV-SA1 virions were purified with a CsCl density gradient. DNA was extracted with a Masterpure DNA purification Kit from Epicenter™. Whole genome sequencing was done via PacBio technology at the University of Washington PacBio Sequencing Services facility. Illumina NextSeq sequencing was conducted in tandem at the Georgia Genomics and Bioinformatics Core at the University of Georgia, USA.

Illumina short read technology was used for whole genome sequencing of McV-KB2, McV-KB3, and McV-KB4. We used a combination of filtration and antibiotics to reduce the abundance of contaminating sequences from bacteria and phage present in the host culture. *Micromonas* host cultures were filtered onto a 1 µm Whatman© Nucleopore Track-Etched Membrane polycarbonate filter and then resuspended into

sterile f/2 -Si medium containing broad-spectrum antibiotics (Supplementary Table S2.1). Cultures were transferred three to four times into fresh antibiotic-treated f/2 -Si at 1:100 incolulum:medium. Flow cytometry counts indicated a tenfold reduction of bacteria in treated *Micromonas* cultures. To obtain McV DNA for sequencing, 5 µL of 0.2 µm-filtered lysate was added to 50 mL of the treated host culture. The host cultures were allowed to clear and then filtered through a 0.2 µm pore cellulose acetate syringe filter. Genomic material was then extracted from the filtrate with a Promega Wizard® Genomic DNA Purification kit. Illumina 151 library preparation and sequencing was performed at SeqCenter (formerly Microbial Genome Sequencing Center), located at the University of Pittsburgh, USA.

2.3.3 Genome assembly and annotation

Assembly methods varied between viral strains and were dependent upon sequencing method and the level of contaminating DNA. PacBio reads for Station ALOHA strain McV-SA1 were assembled with Canu v.1.0 and polished with NextSeq data by Christopher R. Schvarcz (Koren et al., 2017; Schvarcz, 2018). The Kāneʻohe Bay virus strains, McV-KB2, McV-KB3, and McV-KB4, contained varying levels of contaminating DNA from phage and bacteria. Two approaches were used to obtain relatively complete assemblies of the Kāneʻohe strains.

The first approach used the default assembler in Geneious 11.1.5 for McV-KB3. Illumina 151 paired-end data was trimmed of adapters (kmer = 27), low quality reads (minimum = 20) and short reads (minimum = 20) using the BBDuk plug-in (https://sourceforge.net/projects/bbmap/) and then normalized with the Geneious built-in tool (default settings) before assembly. Two prasinovirus contigs (147 and 60 kb) were identified via nucleotide BLAST against the standard NCBI databases, and the reads mapped to these contigs were dissolved and re-assembled. This produced a single 205 kb contig. For McV-KB2, the iterative mapping approach using the Geneious assembler did not produce a single contig, and so the metagenome assembly tool metaviralSPAdes (Galaxy Version 3.15.4+galaxy 2, accessed through usegalaxy.org) was tried, and successfully assembled a single 210 kb contig. Finally, for McV-KB4 we were unable to obtain a single contig using either method, but five putative prasinovirus contigs totaling 212 kb were obtained from the Geneious assembler, which we treat as

a draft assembly for use in comparative analyses. We treat the McV-SA1, McV-KB2, and McV-KB4 assemblies as near-complete based on their lengths, which are comparable to known prasinoviruses (Table 2.1), and based on whole genome alignments using progressiveMauve, which indicated high synteny of McV-SA1, McV-KB3, McV-KB4, and the most similar previously sequenced isolated, MpV1 (GenBank NC_014767.1; See section 2.4 Results and Discussion).

Gene prediction was conducted with Prokka v.1.14.5 accessed via Kbase for all four assemblies (Seemann, 2014). Only CDS with a start and stop codon, and with a minimum of 65 amino acids (195 nucleotides), were used in downstream analysis. Functional and structural annotation was performed manually by integrating information from EggNOG mapper (v.2.1.2, evalue≤0.01, minimum 25% nucleotide identity), InterProScan web interface (v.5.66-98.0, default settings), and RefSeq (BLASTp, evalue ≤ 0.001) (Altschul et al., 1990; Quevillon et al., 2005; Cantalapiedra et al., 2021).

2.3.4 Host strain information

The genomes of two *Micromonas* strains in our UHM collection, UHM1061 and UHM1065, referred to here as M1 and M2 for simplicity, were previously sequenced (Chapter 4). M1 and M2 were isolated from surface water of Kāneʻohe Bay and both strains are lysed by both McV-KB2 and McV-KB3 (Fig. 2.1). Additionally, M2 is the original host that McV-KB4 was isolated on, and it continues to be lysed by this virus strain, as well as McV-SA1 (Fig. 2.1). The 18s rRNA sequences of the M1 and M2 cell lines group into a clade with sequences of *Micromonas commoda* and we therefore identify these strains as *Micromonas commoda* (Chapter 3). Whole genome assemblies for M1 and M2 were created with PacBio reads with Canu (v. 1.9), annotated initially with Maker (v3.01.03; Cantarel et al., 2008) and manually corrected. Full sequencing, assembly, and annotation protocols are described in Chapter 4.

2.3.5 Gene content comparisons

We compared the gene content of the four HiMcVs at four levels:

1) Identify the core genome among the four HiMcVs (genes shared by all four viruses), as well as genes that are unique to one or more of these strains.

2) Identify genes in the HiMcV genomes that have orthologs in the genomes of the two sequenced host strains, which may provide insight into horizontal gene transfer.

3) Identify genes in the HiMcV genomes that are not present in previously sequenced prasinovirus genomes, which expand the known functional capacity of prasinoviruses.

4) Quantify whether there are genes that consistently distinguish viruses infecting the three Mamiellales genera (*Micromonas*, *Ostreococcus*, and *Bathycoccus*), which may provide insight into the process of adaptation to different host taxa.

To make these comparisons, we used OrthoFinder (v2.5.5.; Emms and Kelly, 2019) to identify orthologous groups of genes in a dataset containing the four HiMcVs, the two *Micromonas commoda* strains isolated from O‘ahu, and all known prasinovirus genomes published in GenBank. Additionally, we followed Bachy et al. (2021) by including chloroviruses, which form a monophyletic group with prasinoviruses based on marker gene analysis (Bellec et al., 2009), and serve as an outgroup in phylogenetic analyses. Our full genome dataset therefore includes strains that infect the Mamiellales genera of *Micromonas* (n = 8, including the four HiMcVs), *Ostreococcus* (n = 12), and *Bathycoccus* (n = 5); *Paramecium bursaria Chlorella* virus strains (n = 4), which are chloroviruses with robust reference genomes that are closely related to the prasinoviruses (Bachy et al., 2021; ICTV, 2023); and *Micromonas commoda* host strains (n = 2). In total, there were 31 genomes, 25 of which are prasinoviruses, in our OrthoFinder exploration. GenBank Accession numbers are listed in Supplementary Table S2.2.

Using OrthoFinder output data, we merged assigned and unassigned orthogroups with gene information from EggNOG mapper, InterProScan, and RefSeq, to explore functional trends in shared and unshared orthogroups among the four HiMcVs. Orthogroups unique among HiMcV strains were BLASTed against the NCBI nr database to identify orthologs in prasinoviruses that were absent in the refseq_protein database. Six orthogroups that were initially labeled as unique to the HiMcVs were found to have BLAST hits to prasinovirus genes in the NCBI nr database, and therefore we conservatively excluded them from our list of unique HiMcV orthogroups.

Additionally, we used orthogroup gene count data for all prasinovirus genomes in our dataset and performed a linear model analysis with the formula:

Orthogroup gene count per genome ~ host genus

This model quantifies the correlation between the number of putative coding sequences in an orthogroup and the genus of host infected by a viral strain, to identify orthogroups that most strongly differentiate viruses infecting different host genera. Linear models were compared to a null model using Chi-squared likelihood ratio tests in R (v.4.30; Core Team, 2022). P-values were adjusted for the false discovery rate using the p.adjust function in R. Prasinovirus orthogroup count data was also used to create a clustered heatmap using the R pheatmap package (Kolde, 2019).

2.3.6 Species and gene tree construction

We sought to understand how the HiMcV assemblies related phylogenetically to other members of the order Algavirales and accordingly constructed species trees using orthologous genes from the aforementioned published genomes of prasinoviruses and chloroviruses. We used a gene alignment concatenation approach that included all orthogroups shared among all prasinovirus and chlorovirus genomes (n = 26 orthogroups). We separately aligned each orthogroup using MAFFT (v7.450; Katoh and Standley, 2013), accessed through Geneious. Gene alignments were trimmed to eliminate supervenient sequences with Goalign v0.3.7 (https://github.com/evolbioinfo/goalign). If more than one gene copy was present in a genome, the paralog most closely related to orthologs in other genomes was chosen, and then the 26 ortholog alignments were concatenated in Geneious. A phylogeny was estimated with FastTree (v.2.1.2; Price et al., 2010) within Geneious. We also constructed a tree using only polB sequences, in order to compare the gene tree of this common prasinovirus marker gene to our core gene-based species trees. We used FigTree (v.1.4.4; http://tree.bio.ed.ac.uk/software/figtree/) to visualize both gene concatenation and polB trees (Fig. 2.3 & Supplementary Figure S2.1).

Additionally, OrthoFinder used the STAG algorithm to construct rooted species trees using shared orthogroups, similar to our concatenation method (Emms and Kelly, 2018). The STAG algorithm first constructs a species tree using each core orthogroup, based on the distances between the closest pair of genes for each pair of species. It then combines all species trees using a greedy consensus method and calculates support for each bipartition as the proportion of underlying trees that include that partition, and branch lengths as the average branch lengths of each bipartition.

However, the current version of OrthoFinder does not report support values when fewer than 100 shared orthogroups are present, as is the case with our dataset. We included the STAG-generated tree as Supplementary Figure S2.2, as a point of comparison to our concatenated alignment tree.

If sequences in the HiMcV genomes were suspected of being recently acquired from a cellular genome (based on the top BLASTp hit from RefSeq) they were considered for gene tree construction to provide additional evolutionary context. To do this, related gene sequences from cell and virus strains were searched for via BLASTp against the nr database and downloaded, trimmed with Goalign, aligned with MAFFT, and constructed with FastTree. Visualization of gene trees was created in FigTree.

2.3.7 HiMcV detection in metagenomes

To assess the global distribution of the four HiMcVs we utilized publicly available metagenomes, spanning Pacific and Atlantic ocean basins, from the Hawai‘i Ocean Time-series (HOT), the Bermuda Atlantic Time-series Study (BATS), and GEOTRACES (Mende et al., 2017; Biller et al., 2018). Data from HOT include metagenomes from Station ALOHA, the site where McV-SA1 was isolated. Sequencing runs from HOT above 200 meters and filtered onto either 0.2 µm (n = 293) or 0.02 µm (n = 185) filters were included, as these depths and filter sizes are most likely to capture prasinovirus signal (Aylward and Moniruzzaman, 2022; Ha et al., 2023). The GEOTRACES metagenomic dataset consisted of whole seawater filtered onto 0.2 µm filters (n = 490), with samples taken from the GA02 and GA03 transects in the North Atlantic, the GA10 transect in the South Atlantic (off the coast of South Africa), and the GP13 transect in the South Pacific (off the coasts of Australia and New Zealand). Note that the GEOTRACES dataset includes sequences from the BATS station.

We used CoverM v0.6.1 (https://github.com/wwood/CoverM) to search the metagenomic datasets for sequences that mapped onto at least one of the four HiMcV assemblies. Requirements of 95% minimum read identity and 20% minimum covered fraction (Ha et al. 2023), indicated with the flags --min-read-percent-identity and --min-covered-fraction, were used. Relative abundances of each virus (i.e., percent of reads from the metagenome sample) derived from CoverM results were then merged with bioproject metadata to create a map of hits in R statistical software (Fig. 2.7). CoverM

results information, including SRR numbers and metadata, are available in Supplementary Table S2.3.

2.4 RESULTS AND DISCUSSION

2.4.1 Genome assemblies

Four HiMcV draft genomes were assembled, and we posit these assemblies are near-complete based on nucleotide length, the number of predicted genes, and whole-genome alignments comparing these genomes to each other and to the most similar previously sequenced *Micromonas* virus. The assemblies range from 205 to 212 kbp, with the largest genome belonging to McV-KB4 (Table 2.1). To help assess genome completeness we created a Mauve alignment with the published genome of MpV1 (NC_014767.1), the virus most closely related to McV-KB3, McV-KB4 and McV-SA1 (as described in the next section, 2.4.2 Phylogeny). These four genomes exhibited a high degree of synteny, although MpV1 is shorter in total by 20 to 26 kbp (Fig. 2.2A). It appears that MpV1 has one major ~11 kbp inversion, relative to the other three genomes, towards the center of its genome. A Mauve alignment using only McV-KB3, McV-KB4 and McV-SA1 showed that these three genomes have high structural similarity (Fig. 2.2B). Inclusion of McV-KB2 in Mauve alignments resulted in a large number of Locally Collinear Blocks (LCBs), represented as colored blocks in Fig. 2.2C, which indicated that there were substantial differences in genome organization between McV-KB2 and the other HiMcVs. Overall, the organization of the HiMcV genomes is consistent with previous findings that prasinoviruses exhibit a high degree of conservation in genome structure (Moreau et al., 2010), with the notable exception of McV-KB2.

Table 2.1. Characteristics of genome assemblies of four Hawai‘i *Micromonas commoda* virus strains.

Assembly	Assembly size (bp)	# CDS	GC%	Gene density (gene/kbp)
McV-KB2	210,100	242	44.9	1.15
McV-KB3	204,582	271	41.3	1.32
McV-KB4	212,418	272	42.1	1.28
McV-SA1	210,087	270	41.2	1.29

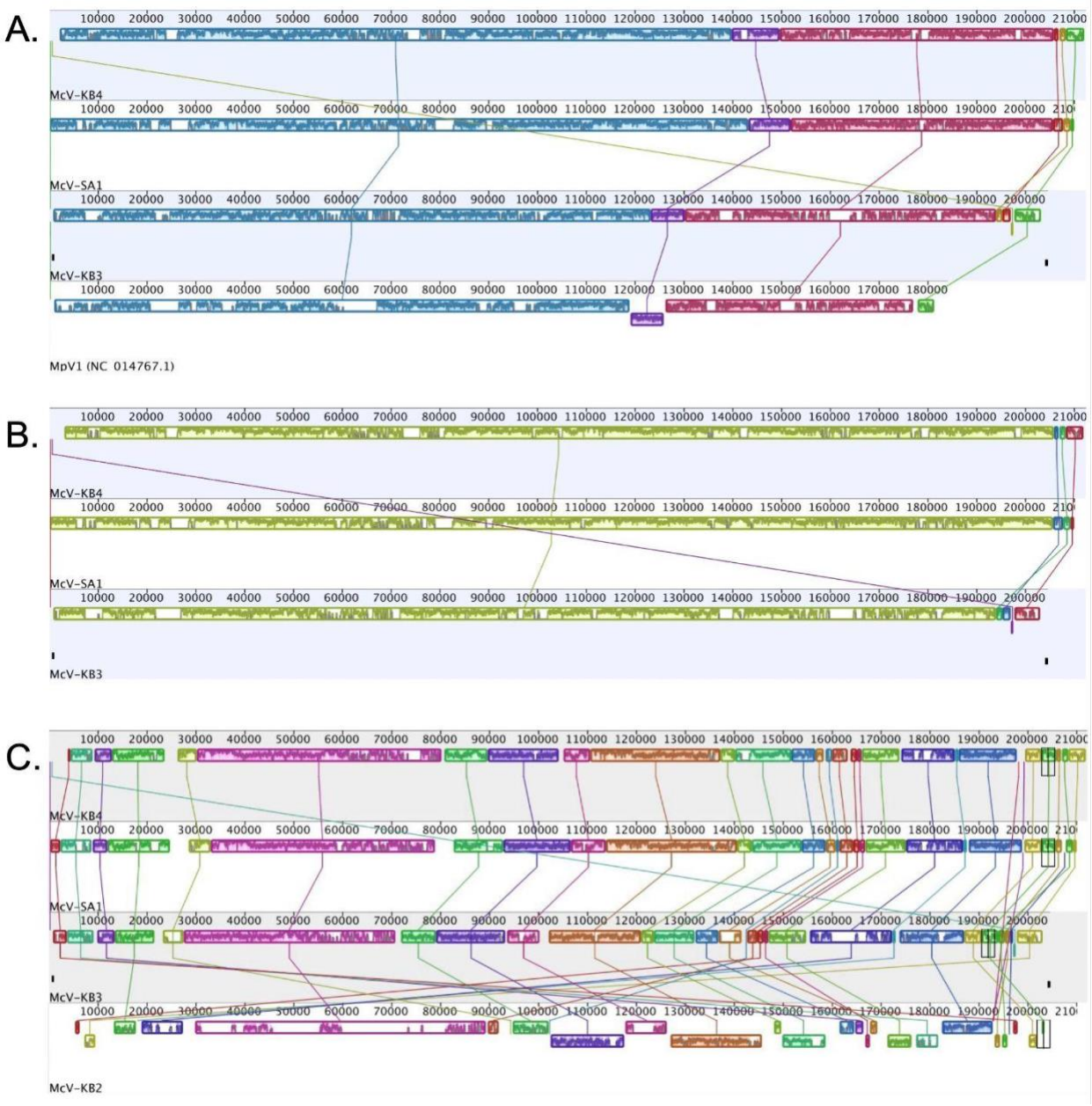


Figure 2.2. Whole genome alignments created with the progressiveMauve algorithm. Colored blocks with corresponding lines represent regions conserved across assemblies. A. An alignment of McV-KB3, McV-KB4, and McV-SA1 with the published genome of *Micromonas pusilla* virus MpV1. B. An alignment of McV-KB3, McV-KB4, and McV-SA1. C. An alignment of all four HiMcV assemblies. Note that the inclusion of McV-KB2 generated a greater number of LCBs.

2.4.2 Phylogeny

The species tree derived from the concatenation of shared prasinovirus and chlorovirus genes is shown in Fig. 2.3. This tree evidences that McV-KB3, McV-KB4, and McV-SA1 group into a clade, with McV-KB4 and McV-SA1 most closely related to each other, and the clade of these three HiMcVs lies within a larger clade containing all

previously published *Micromonas* and *Ostreococcus*-infecting virus genomes. Within this larger clade the *Ostreococcus*-infecting viruses form a monophyletic clade, while the *Micromonas*-infecting viruses are paraphyletic, consistent with previous analyses (Bellec et al., 2009; Bachy et al., 2021). McV-KB2 is quite divergent from the other HiMcVs, as its closest relative is the clade of *Bathycoccus*-infecting viruses, although it diverged from that group soon after their common ancestor arose from the last common ancestor of all the prasinoviruses. The divergence of McV-KB2 and its grouping with *Bathycoccus* viruses is consistent with previous and current findings using a polB phylogeny, as well as the OrthoFinder/STAG-generated species tree (Chapter 3; Supplementary Figures S2.1 & S2.2). Although McV-KB2 is relatively divergent from the other HiMcVs, it should be noted that it nonetheless overlaps in host range with each of the other three viruses (Fig. 2.1). Bachy et al. (2021) presented results suggesting ecotypic divergence among *Bathycoccus* viruses that may reflect their hosts' phylogeny. We did not assess whether *Bathycoccus* strains in culture collections can be infected by the HiMcVs, or whether *Bathycoccus* viruses can infect the *Micromonas commoda* strains from Kāne'ohe Bay. However, as of this writing, no known prasinovirus infects prasinophytes outside of its original host's genus (see Bachy et al., 2021 and references therein).

The STAG-generated tree (Supplementary Figure S2.2) is nearly identical to our concatenated sequence tree, indicating a topology that is robust across multiple approaches to species tree construction. We also generated a polB gene tree that included prasinoviruses and chloroviruses, and its structure was very similar to the core-orthogroup based species trees (Supplementary Figure S2.1). However, the polB tree most notably differed in having poor node support values, as well as placing McV-KB2 outside of the shared clade with *Bathycoccus* viruses.

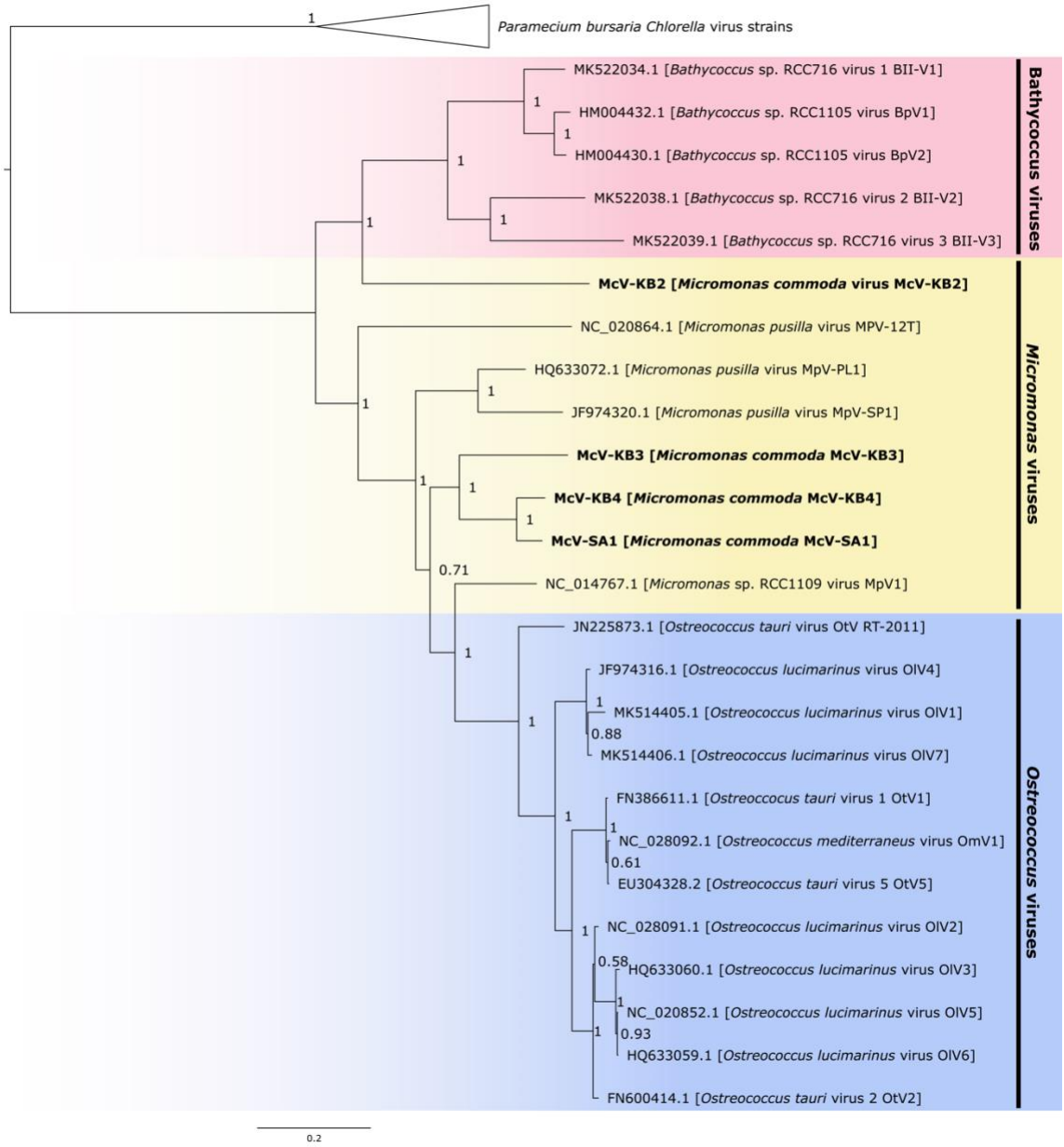


Figure 2.3. Prasinovirus species tree from concatenated amino acid alignments of the 26 orthogroups possessed by all prasinoviruses and chloroviruses in the analysis. Generated with FastTree. Scale bar indicates substitutions per site.

2.4.3 HiMcV core genes

Our OrthoFinder analysis resulted in 344 orthogroups that were present within at least one of our HiMcV genomes (Table 2.2). The four HiMcVs share 152 orthogroups (i.e., the core HiMcV orthogroups), while the other 192 orthogroups are present in three or fewer genomes (Fig. 2.4). Fifty-six of the HiMcV core orthogroups were found in all prasinoviruses, and 117 were found in all *Micromonas* virus genomes. Lastly, there

were 48 orthogroups found in at least one HiMcV that were also found in at least one of the two host genomes, M1 and M2. When comparing HiMcVs to each other McV-KB2 is the most distinct, with 46 unique orthogroups, while McV-KB3 has the second highest number of unique orthogroups (36), both of which are consistent with the phylogenetic distances between the HiMcVs (Fig. 2.3).

Table 2.2. Summary of OrthoFinder results comparing prasinoviruses, chloroviruses, and *Micromonas* hosts M1 and M2.

	No. of Orthogroups
Total prasinovirus	693
Total HiMcV	344
Core HiMcV	152
Core *Ostreococcus* viruses	129
Core *Micromonas* viruses	117
Core *Bathycoccus* viruses	99
Core prasinovirus	56
Core prasinovirus + chlorovirus	26
HiMcV not found in other prasinoviruses	61
HiMcV shared with host	48

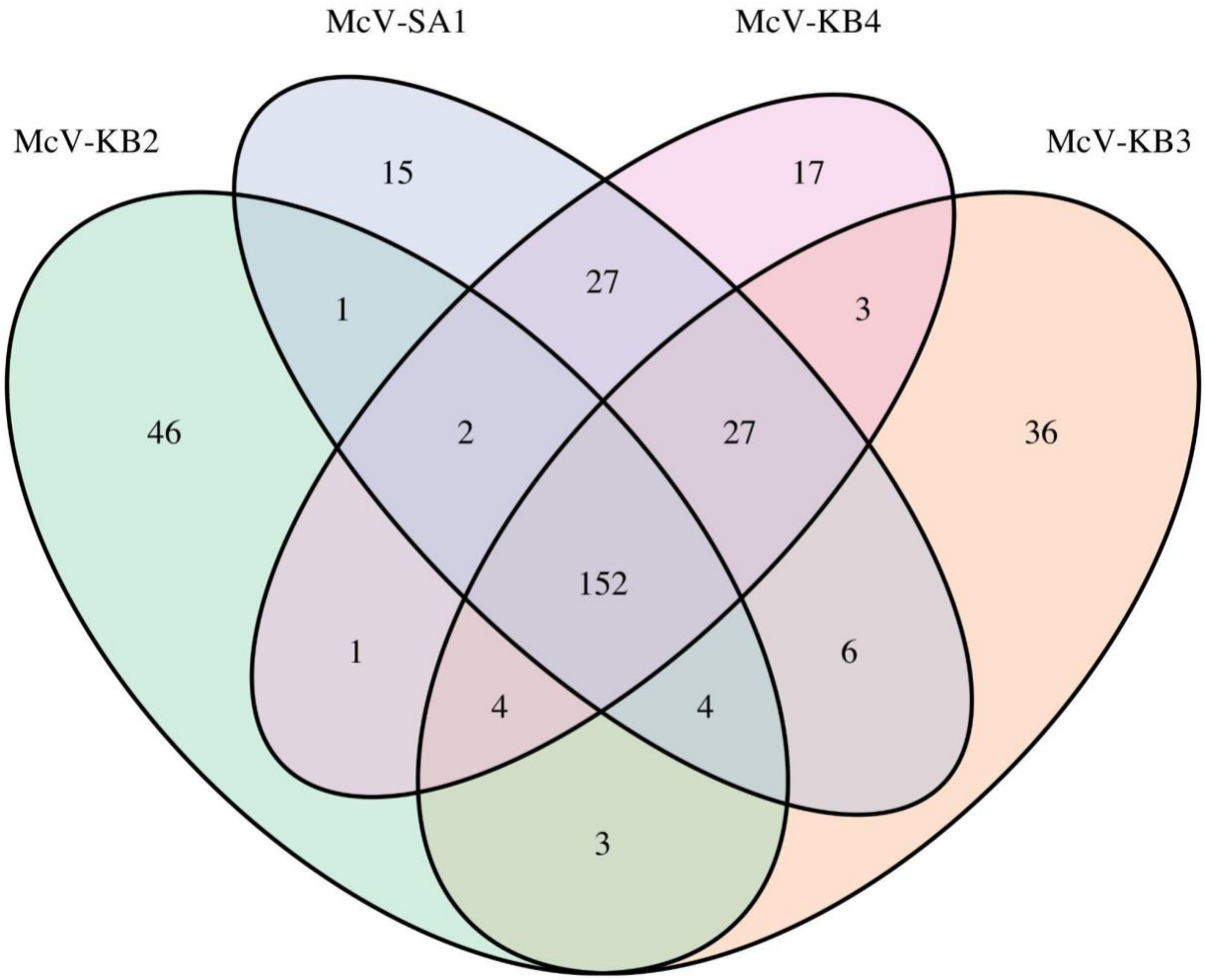


Figure 2.4. Venn diagram of the number of orthogroups shared by and unique to the four HiMCVs.

All 152 core HiMcV orthogroups contain genes that occur in at least one other prasinovirus (Supplementary Table S2.4). The prasinovirus sequences in 127 of these orthogroups were identified as hypothetical proteins in the NCBI refseq_protein database. However, some functional information is available for 95 of these hypothetical proteins from InterProScan member databases. Broadly speaking, core HiMcV orthogroups with putative functions implicate many aspects of the infection cycle, such as virion structure (major capsid protein), genome replication (DNA polymerase, DNA primase), RNA processing and transcription (mRNA capping enzyme, transcription factors, RNAse H), breakdown of host polymers (nucleases, proteases), nucleotide metabolism (dUTPase, dCMP deaminase, thymidylate synthase, ribonucleotide

reductase), carbohydrate metabolism (mannitol dehydrogenase), lipid metabolism (glycerophosphoryl diester phosphodiesterase, phospholipase), and glycosylation (glycogen phosphorylase B, nucleotide-diphospho-sugar transferases). All four HiMcVs possess a cAMP-dependent Kef-type K+ transporter that is also found in *Micromonas* virus strains SP1 and MpV1, a transporter which is known in bacteria to help protect cells from electrophilic compounds (i.e., reducing compounds). Potassium channels are the most common type of membrane transporters in viral genomes, with substantial diversity, likely multiple origins, and generally unknown function (Greiner et al., 2018). In *Paramecium bursaria Chlorella* viruses, a potassium channel depolarizes the cell membrane during infection, preventing infection of the cell by multiple viruses (Greiner et al., 2009). Thus, the Kef-type potassium transporter may play a significant role in HiMcV's infection strategy and merits further study.

Another noteworthy core HiMcV orthogroup is annotated as a chlorophyll a-b binding protein, and which is also present in the assembly for *Micromonas* virus MpV1. This protein is found in the light harvesting complex of photosystem II, and there is growing evidence that it plays an important role in the viruses of eukaryotic phytoplankton, as it has been found in other prasinoviruses, *Chrysochromulina ericina* virus CeV-01B (family Mesomiviridae), and a variety of giant virus metagenome-assembled genomes (Gallot-Lavallée et al., 2017; Moniruzzaman et al., 2020).

A putative PhoH-like phosphate starvation-inducible protein is also among the HiMcV core sequences. This protein has been seen previously in some, but not all, prasinoviruses and is common among marine phages, potentially enhancing viral infection under low-phosphate conditions (Goldsmith et al., 2011; Monier et al., 2012). While genes in this family are common in eukaryotic phytoplankton, it appears that our host genomes do not contain orthologs to the sequences found in our HiMcVs. Previously studied prasinovirus versions of this gene appear to be host-derived, which may mean that the HiMcVs obtained it from other *Micromonas* hosts in the environment (Monier et al., 2012).

A final notable gene shared among all four HiMcVs is an alternative oxidase that also occurs in the *Micromonas commoda* RCC299 genome. Within this orthogroup, McV-KB2, McV-KB3, and McV-SA1 had one ortholog each, while McV-KB4 had two

orthologous sequences, the host strain M1 had one sequence, and M2 had two sequences. Alternative oxidases have been previously found in cyanophages (Puxty et al., 2015), and these enzymes are thought to reduce photodamage to the electron transport chain under stressful conditions, which may be particularly important due to viral inhibition of photosystem I and ferredoxin NADP+ reductase (FNR) (Wang et al., 2023). Our analysis found alternative oxidases in only one other prasinovirus beyond the HiMcVs, *Ostreococcus tauri* virus RT-2011. We were curious to see if the HiMcV orthologs of this gene were more closely related to host or phage sequences, and whether alternative oxidases have been acquired more than once by phytoplankton viruses. Thus, we created an alignment using alternative oxidase sequences from the four HiMcVs, the two Hawai'i *Micromonas* hosts, OtV RT-2011, and cyanobacterial and cyanophage alternative oxidases found through GenBank (see 2.3.6 Species and gene tree construction). Previous work on plant alternative oxidases has found that the mitochondrial alternative oxidase (AOX, also called ubiquinol oxidase) and the plastid alternative oxidase (PTOX, also called plastoquinol terminal oxidase) are often misannotated or annotated in an inconsistent way (Nobre et al., 2016). Therefore, we also included reference PTOX and AOX plant and algal sequences to aid in interpreting the alignment and phylogeny.

The gene tree resulting from our AOX/PTOX alignment is shown in Fig. 2.5. All prasinovirus sequences appear in the same clade, and the closest relatives of this clade are sequences from *Micromonas* isolates, including one of the sequences from M2. This group, containing prasinovirus and *Micromonas* PTOX sequences, is a sister clade to one containing *Bathycoccus* and *Ostreococcus* sequences, although support for the node joining these clades is low (0.21). It should be noted that within the prasinovirus clade there is one CDS from McV-KB4 that diverges before a second and third versions of the gene from McV-KB4, suggesting several duplication events.

Further inspection of the AOX/PTOX phylogeny reveals several important features. First, there is a monophyletic clade containing AOX/ubiquinol oxidase from prasinophytes, other algae, and plants, consistent with an ancient divergence between AOX and PTOX (Nobre et al., 2016). Second, within the PTOX clade there is a relatively early divergence by *Synechococcus/ParaSynechococcus* sequences, placing them in a

different part of the tree from the *Synechococcus* phage sequences, which are more closely related to *Prochlorococcus* phage and *Prochlorococcus* cellular sequences, and which was also noted in a prior analysis (Puxty et al., 2015). Third, after the divergence of *Synechococcus* sequences there is a split between a clade containing the *Prochlorococcus* and cyanophage sequences and a clade containing eukaryote and prasinovirus PTOX sequences as well as other cyanobacterial sequences. Finally, the eukaryote+prasinovirus+cyanobacteria clade contains two main branches, and the two branches possess different paralogs of PTOX from isolates of *Micromonas* (including M2), *Ostreococcus*, and *Bathycoccus*. This split may represent an ancient duplication event resulting in two PTOX copies, one of which was acquired by prasinoviruses. Based on the structure of the phylogeny it may be the case that prasinovirus and cyanophage PTOX genes were acquired through distinct gene transfer events.

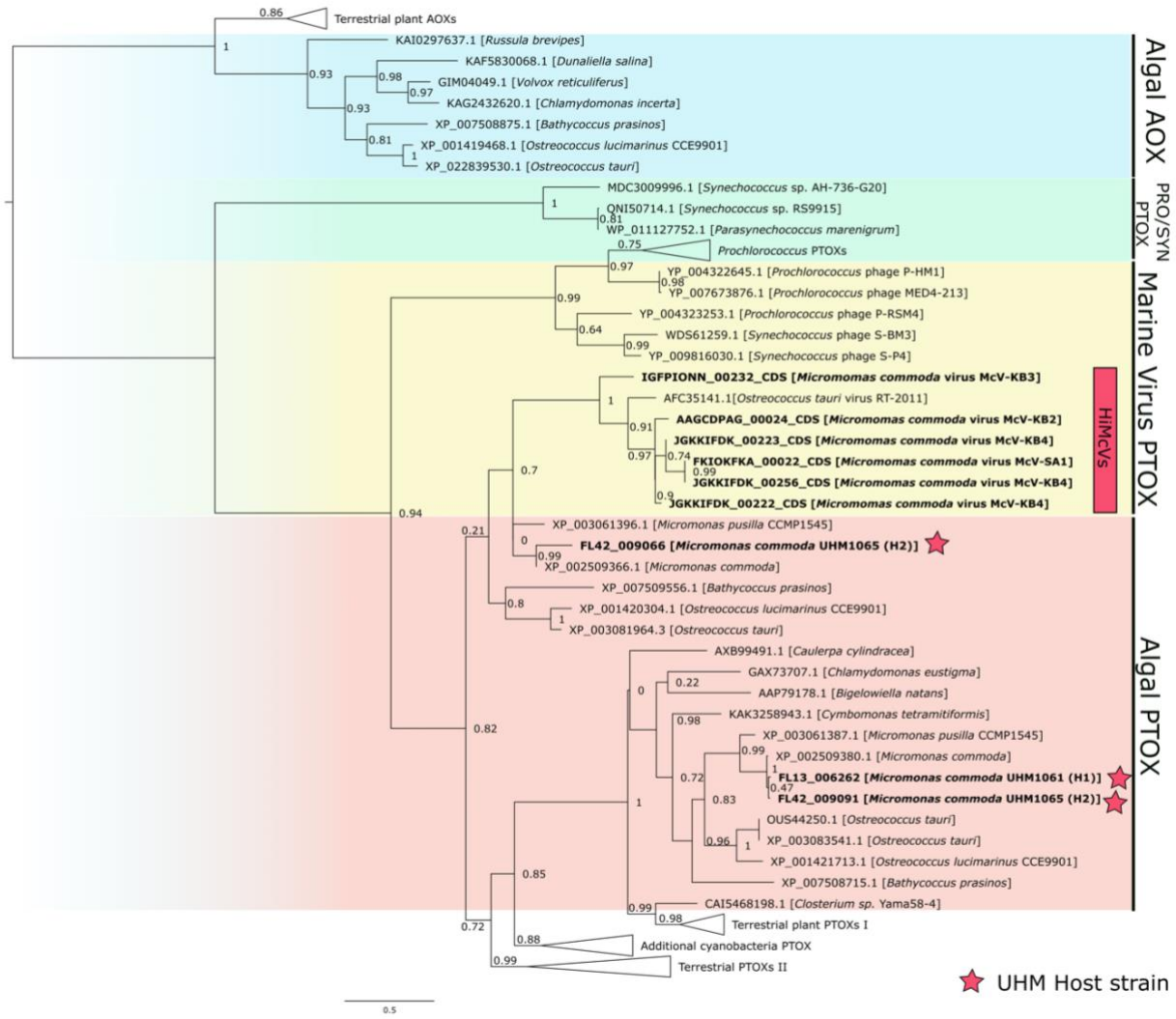


Figure 2.5. Alternative oxidase/plastoquinol terminal oxidase gene tree, including mitochondrial ubiquinol oxidase (AOX) from plants and eukaryotic phytoplankton, and plastoquinol terminal oxidase (PTOX) from plants, eukaryotic phytoplankton, cyanobacteria, prasinoviruses, and cyanophages. Gene sequences were aligned in MAFFT and trimmed with GoAlign, and tree created in FastTree. Node support values reflect FastTree Shimodaira-Hasegawa test values. UHM strains are shown in bold, with red stars by *Micromonas commoda* host cell sequences. The tree was rooted using the split between AOX and PTOX sequences. Scale bar indicates substitutions per site.

2.4.4 Notable non-core and unique HiMcV genes

The 192 non-core HiMcV orthogroups, or those found in only 1-3 of the 4 HiMcVs, often contain sequences orthologous to those previously found in other phycodnaviruses, with the closest refseq_protein and nr database hits generally being hypothetical proteins (Supplementary Table S2.5). However, there are 61 non-core HiMcV orthogroups that represent proteins not previously found in prasinoviruses (verified through a secondary nr BLAST search), a distinction which we will refer to as

“unique” in this writing. All orthogroups unique to HiMcVs are non-core orthogroups, as none contain sequences from all four HiMcVs. Limited functional information is available for the 61 unique orthogroups, as 40 have no BLAST hits against refseq_protein, eight have BLAST hits to hypothetical proteins, and the remaining 13 include many with low-identity (<30% amino acid identity) BLAST hits. Search results from the InterPro databases are comparable, although many of the proteins are predicted to be membrane-bound. One notable unique orthogroup is a phosphate:sodium symporter in McV-SA1 that is also found in both M1 and M2 *Micromonas* host genomes. Top BLAST hits for the McV-SA1 CDS in both nr and refseq_protein are from previously published strains of *Micromonas*, including RCC299, a pelagic strain from the equatorial Pacific, and CCMP1445, a coastal strain from the North Atlantic. This CDS has no orthologs among the other prasinovirus genomes, which may indicate that the McV-SA1 symporter gene has a recent cellular origin via horizontal gene transfer. It should be noted that our data set includes a second orthogroup with prasinovirus phosphate transporter sequences that appear distinct from the McV-SA1 version, as an alignment of the two orthogroups indicated low sequence identity, albeit with scattered matching residues (results not shown). This second orthogroup includes phosphate transporter sequences from BpV1 (HM004432.1), BII-V1 (MK522034.1), OlV6 (HQ633059.1), OlV5 (NC_020852.1), OtV2 (FN600414.1), OlV4 (JF974316.1), OlV2 (NC_028091.1), and OlV1 (MK514405.1), as well as sequences from M1 and M2. This orthogroup corresponds to the prasinovirus phosphate transporters from the PHO4 superfamily identified by Monier et al. (2012) in a comparison of phytoplankton virus phosphate transporters. The PHO4 transporters correspond to InterPro family IPR001204, whereas the McV-SA1 transporter corresponds to InterPro family IPR003841. Therefore, the McV-SA1 phosphate:sodium symporter likely indicates a separate acquisition of a phosphate transporter by prasinoviruses, potentially with different uptake affinity or other physiological differences.

Other unique HiMcV orthogroups with predicted functions include a putative bax inhibitor-1 (McV-KB2), N-6 DNA methylase (McV-KB2), polyamine aminopropyl transferase (McV-KB3), adenosylmethionine decarboxylase (McV-KB3), and glycosyltransferase (McV-SA1). Bax inhibitor-1 is a conserved inhibitor of programmed

cell death, and viruses such as deerpox (Banadyga et al., 2011) and cytomegalovirus (Ma et al., 2012) are known to encode other proteins that suppress bax, thereby countering the elimination of infected cells by apoptosis. Top BLAST hits to the McV-KB2 gene are fungal genes, but with relatively low amino acid identity (30-35%), making it unclear where McV-KB2 may have acquired the gene from. The putative adenosylmethionine decarboxylase and polyamine aminopropyl transferase encoded by McV-KB3 may catalyze linked steps in the synthesis of spermidine from putrescine. Spermidine is a polyamine required for cell growth as well as the replication of many viruses, and enzymes related to spermidine synthesis have been found in a variety of phages and eukaryotic viruses (Li et al., 2023), including the chlorovirus PBCV-1 (Baumann et al., 2007), although the McV-KB3 genes appear to be the first reported occurrence in a prasinovirus isolate. The two McV-KB3 genes have BLAST hits to genes from various bacteria and archaea, although the relatively low amino acid identity (~30-40%) provides little information about the proximate origin of the genes. The other unique orthogroups with putative functions (N-6 DNA methylase and a glycosyltransferase) represent categories of enzymes that are commonly encoded by prasinoviruses, although the uniqueness of these orthogroups indicates these specific genes are not closely related to previously documented prasinovirus genes. A final notable unique orthogroup is a gene found only in McV-KB4 that is orthologous to cyanophage genes of unknown function. This putative CDS has not been seen in other prasinoviruses, and the closest database hit is a hypothetical protein from *Prochlorococcus* phage strain P-SSM2, which infects low-light *Prochlorococcus* ecotypes. The orthogroup may be evidence of gene exchange between cyanophage and a eukaryotic virus.

2.4.5 HiMcV orthogroups shared with their host genomes

A total of 48 HiMcV orthogroups contain genes also found in the M1 and/or M2 host genomes, 27 of which are shared among all 4 HiMcVs. In total M1 shared 27 orthogroups with all four HiMcV strains, with 19 additional orthogroups shared between M1 and at least one other virus (Supplementary Figure S2.3). Results from comparison with M2 were similar, with 24 orthogroups shared between M2 and all four viruses, and with 22 orthogroups shared between M2 and at least one virus (Supplementary Figure

S2.4). The number of shared orthologs is comparable to results from Moreau et al. (2010), in which *Micromonas pusilla* virus MpV1 shared 56 CDS with *Micromonas* sp. strain RCC1109. Forty-five of the HiMcV-host shared orthogroups contain genes found in previously published prasinoviruses, as evidenced by refseq_protein and nr BLAST hits. Some of these orthogroups were described in the section 'HiMcV core genes' (chlorophyll a-b binding protein, PTOX, cAMP-dependent Kef-type K+ transporter). In general the orthogroups shared with hosts are potentially associated with a variety of cellular processes such as protein modification/regulation/processing (N-myristoyltransferase, ubiquitin, cysteine protease, ATP-dependent metalloprotease FtsH), glycosylation (nucleotide-sugar epimerases and transferases), amino acid synthesis (dehydroquinate synthase), nucleotide metabolism (dCMP deaminase, thymidine kinase), transcription regulation (transcription factors), stress response (heat shock protein 70, rhodanese, superoxide dismutase, mannitol dehydrogenase), nucleic acid processing (exonuclease, ribonucleotide reductase, DNA polymerase family X), photosynthesis (PTOX, chlorophyll a/b binding protein), and potentially countering host defenses (methyltransferases) (Supplementary Table S2.6).

There are three HiMcV orthogroups not found in other prasinoviruses that are found in both hosts, which include the aforementioned phosphate:sodium symporter found in McV-SA1, as well as two orthogroups shared with McV-KB2. One of the McV-KB2-host orthogroups contains only hypothetical protein sequences with no hits in NCBI or InterPro databases. The other McV-KB2 host orthogroup contains sequences that are annotated as chlorophyllide a oxygenase (CAO) for M1 and M2. CAO converts chlorophyll a to chlorophyll b, and chlorophyll b is an important accessory pigment in green algae (Jeffrey et al., 2011), which may mean that CAO supports light adsorption during infection by McV-KB2. However, amino acid identity between the McV-KB2 and host sequences is 17.89%, suggesting these sequences may not be truly orthologous. If McV-KB2 indeed encodes for CAO it would be the first virus reported to have this sequence.

2.4.6 Genes differentiating prasinoviruses that infect different host genera

In total there were 693 orthogroups in our analysis that occurred in at least one prasinovirus. Linear models relating copy number of each orthogroup for each prasinovirus isolate to host genus found 170 orthogroups that differed significantly between viruses of the three host genera ($p < 0.05$; Supplementary Table S2.7). Therefore, 25% of prasinovirus orthogroups were significantly associated with host genus identity, suggesting that a substantial portion of the genome is involved in adapting to infect host genera in the same taxonomic order. Twenty-eight orthogroups that differ strongly between host genera ($p < 0.001$) and that also have functional annotations are shown in Table 2.3, to exemplify the diversity of functions that relate viral gene content to host identity. For example, orthogroups that are absent in *Bathycoccus* viruses but present in most or all *Micromonas* and *Ostreococcus* viruses include asparagine synthetase (nitrogen and amino acid metabolism), dCMP deaminase (nucleotide metabolism), DNA polymerase X (potentially for base excision repair, Fernández-García et al., 2017), nucleotide-diphospho-sugar transferases (glycosylation), RNAse H (RNA processing), NTP pyrophosphohydrolase (potentially involved in stress response), and a protein with a rhodanese-like domain (potentially involved in stress response). Orthogroups that are present in most or all *Micromonas* viruses but absent/rare in the other two groups include a glycerophosphodiester phosphodiesterase (lipid metabolism), mannitol dehydrogenase (potentially involved in stress response), ubiquitin, a protein with a zinc finger C2H2-type domain (potential transcription factor), a protein with an integrin alpha domain (potentially used for attachment to the host), and a putative tail fiber protein (potentially used for attachment to the host). Therefore, it may be the case that many stages of the viral life cycle, such as attachment to host receptors and manipulation of host metabolism and defenses, are involved in (co)evolution to infect different host genera.

Table 2.3. Orthogroups that exhibit highly significant differences between viruses infecting different host genera ($p < 0.001$), and which possess putative functional annotations. Reported for each orthogroup is the p-value from a chi-square likelihood ratio test comparing occurrence across host genera (adjusted for false discovery rate), the organism(s) associated with the top refseq_protein database hits, selected annotations from refseq_protein and InterPro, and the proportion of strains infecting a specific host genus that have sequences present in the orthogroup.

Orthogroup	Adjusted p-value	Organisms	Selected Annotations	Proportion *Bathycoccus* - infecting virus strains	Proportion *Micromonas*- infecting virus strains	Proportion *Ostreococcus* - infecting virus strains
OG0000005	7.12E-06	Bathycoccus sp. RCC1105 virus BpV1	Intramolecular chaperone auto-processing domain, Galactose oxidase/kelch, Integrin alpha beta-propellor, FG-GAP repeat	0.80	0.00	0.38
OG0000008	0.0004766 1	*Micromonas* sp. RCC1109 virus MpV1, *Micromonas* pusilla virus SP1, Ancylomarina sp. DW003	Serralysin-like metalloprotease, C-terminal/Tumour necrosis factor-like domain superfamily, C1q domain	0.20	1.00	0.88
OG0000016	5.26E-06	*Micromonas* pusilla virus 12T, *Micromonas* sp. RCC1109 virus MpV1	Coagulation factor 5/8, Concanavalin A-like lectin, Galactose oxidase, Integrin alpha, Serralysin-like metalloprotease, Tumour necrosis factor-like domain superfamily	0.20	0.00	0.88
OG0000135	1.13E-06	*Micromonas* pusilla virus 12T	NFACT, RNA-binding domain	0.00	1.00	0.50
OG0000141	1.65E-19	*Micromonas* sp. RCC1109 virus MpV1,*Micromonas* pusilla virus 12T	Asparagine synthetase	0.00	1.00	0.88

OG0000142	2.52E-05	*Micromonas* pusilla virus 12T, *Micromonas* sp. RCC1109 virus MpV1, *Micromonas* pusilla virus SP1	dCMP deaminase	0.00	0.67	1.00
OG0000143	4.69E-05	Pseudodesulfovibrio sp. SB368, *Micromonas* pusilla virus SP1, *Micromonas* pusilla virus 12T, Ostreococcus tauri virus 1, Chromobacterium amazonense	tail fiber protein	0.00	0.25	0.88
OG0000144	2.15E-05	*Micromonas* sp. RCC1109 virus MpV1, *Micromonas* pusilla virus SP1	cAMP-dependent Kef-type K+ transporter	0.60	0.17	1.00
OG0000149	0.00024067	*Micromonas* pusilla virus SP1	methyltransferase	0.00	1.00	0.63
OG0000157	0	Ostreococcus tauri virus 1, Ostreococcus lucimarinus virus 1, Ostreococcus lucimarinus virus 7, *Micromonas* sp. RCC1109 virus MpV1	NTP pyrophosphohydrolase MazG-related, YvdC	0.00	1.00	1.00
OG0000159	1.04E-08	*Micromonas* sp. RCC1109 virus MpV1	Rhodanese-like domain superfamily	0.00	1.00	0.63
OG0000163	1.65E-19	Ostreococcus mediterraneus virus 1, *Micromonas* sp. RCC1109 virus MpV1, Ostreococcus lucimarinus virus OlV5	Nucleotide-diphospho-sugar transferases	0.00	1.00	0.88
OG0000168	2.52E-05	*Micromonas* pusilla virus 12T, *Micromonas* sp. RCC1109 virus MpV1	DNA polymerase beta-like, N-terminal domain	0.00	0.67	1.00

OG0000174	1.05E-05	*Micromonas* sp. RCC1109 virus MpV1	Holliday junction resolvase, A22, Ribonuclease H-like superfamily	0.00	0.92	0.75
OG0000181	4.07E-07	*Micromonas* pusilla virus 12T, *Micromonas* pusilla virus SP1	endonuclease	1.00	0.25	1.00
OG0000183	1.97E-06	Acinetobacter larvae	lipid A hydroxylase LpxO	0.20	1.00	0.25
OG0000187	1.71E-12	Suillus clintonianus, *Micromonas* pusilla virus 12T	ubiquitin	0.00	0.00	0.88
OG0000213	0.0001721 1	*Micromonas* pusilla virus SP1	cytidyltransferase	0.00	0.83	0.25
OG0000227	5.32E-05	*Micromonas* pusilla virus 12T	Zinc finger C2H2-type	0	0.83	0.75
OG0000251	2.49E-09	*Micromonas* sp. RCC1109 virus MpV1;*Micromonas* pusilla virus SP1	mannitol dehydrogenase	0.00	0.08	0.88
OG0000300	2.49E-09	*Micromonas* sp. RCC1109 virus MpV1;*Micromonas* pusilla virus SP1	Glycerophosphodiester phosphodiesterase domain, PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily	0.00	0.08	0.88
OG0000467	3.26E-05	*Micromonas* pusilla virus 12T	Glucose-6-phosphate dehydrogenase, NAD(P)-binding domain superfamily	0.00	0.00	0.63

We used unsupervised clustering analysis via the R package pheatmap to further understand how gene content varies among the prasinovirus genomes that we analyzed (Fig. 2.6). Consistent with the many differences we found between viruses infecting host genera (Table 2.3), unsupervised clustering largely groups viruses by host genus, with the exception of one *Ostreococcus tauri* virus that occurs in the *Micromonas* virus cluster. In our phylogenetic analysis this strain, *Ostreococcus tauri* virus RT-2011 (JN225873.1), is relatively divergent from the clade containing the other 12 *Ostreococcus* viruses (Fig. 2.3). It is possible that OtV RT-2011 retained gene content similar to *Micromonas* viruses while the other *Ostreococcus* viruses evolved more *Ostreococcus*-specific gene content. Finally, the clustering results again emphasize the uniqueness of McV-KB2 relative to the other HiMcVs, as the other three HiMcVs group together in a cluster, while McV-KB2 is grouped with the genome of MpV-12T, a strain isolated from the coast of The Netherlands.

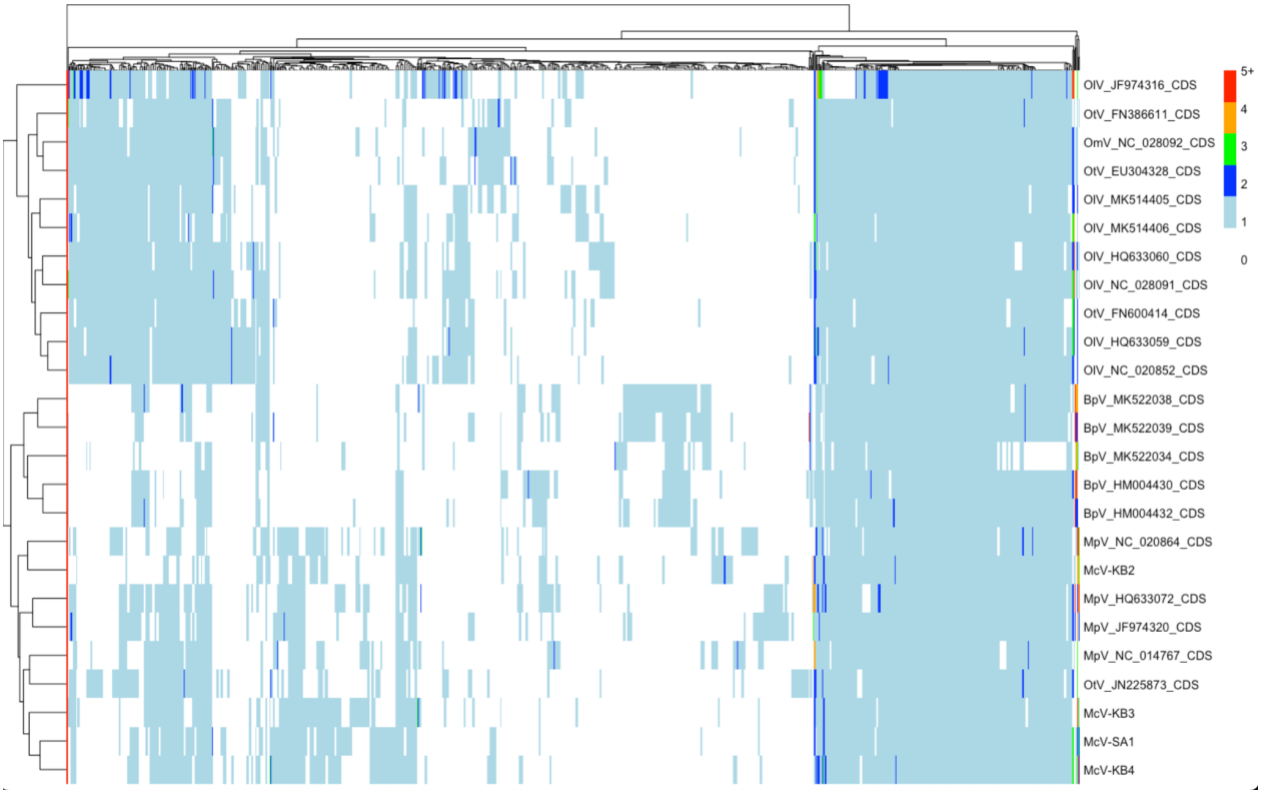


Figure 2.6. Clustered heatmap of 693 prasinovirus orthogroups. Colors in the heatmap represent the number of gene copies in each orthogroup per genome, with warmer colors being higher. The dendrogram on the left-hand side reflects similarity of virus strains based on orthogroup composition, and the dendrogram at the top clusters orthogroups by similarity in patterns of occurrence across strains.

2.4.7 Distribution of HiMcV sequences in the world ocean

Using CoverM we observed 84 instances of metagenomic reads mapping to our HiMcVs from the combined Station ALOHA, BATS, and GEOTRACES datasets (Supplementary Table S2.8). Although our viruses were isolated at, or relatively near, Station ALOHA, only 8 hits occurred at this location, while 76 occurred within the GEOTRACES/BATS dataset. Nevertheless, HOT data did contain the second highest relative abundance of reads mapping to a HiMcV in our dataset (Fig. 2.8). We ran CoverM against a similar number of sequencing runs from each dataset, with 478 runs from Station ALOHA versus 480 from GEOTRACES/BATS. Further emphasizing the lower occurrence of prasinoviruses in the Pacific, we observed no hits from runs in the GEOTRACES transect GP13 in the South Pacific (Fig. 2.8). These results are consistent with those of Ha et al. (2023), who analyzed *Nucleocytoviricota* diversity in the GEOTRACES data, and observed that Algavirales in general were more common in the Atlantic samples. However, the same researchers found that higher viral diversity was captured in longitudinal transects, which were performed in the Atlantic, compared to the latitudinal transects and single-station time series examined in the Pacific. Therefore, it may be possible to find additional sequences similar to the HiMcV viruses by sampling longitudinal transects in the Pacific.

Nearly all HiMcV hits belonged to strain McV-SA1, which is the only HiMcV isolated from the open ocean. The exceptions are hits from GEOTRACES transect GA10 in the South Atlantic, in the highly productive upwelling region near South Africa, which contained the highest relative abundances of HiMcVs of all transects and included McV-KB2 and McV-KB3 hits in addition to McV-SA1 (Fig. 2.8). No sequences similar to McV-KB4 were found in any samples.

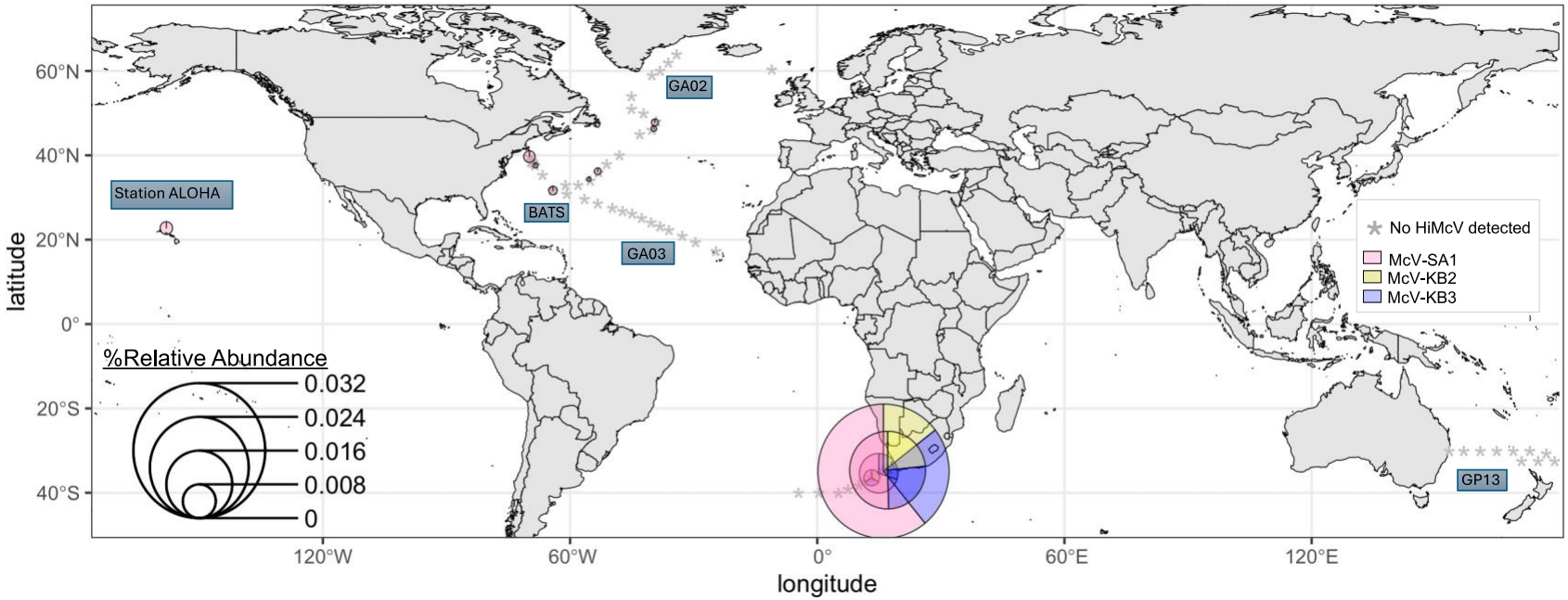


Figure 2.7. Distribution of HiMcV strains in metagenomic samples from GEOTRACES and Station ALOHA. Pie charts represent strain composition at each location, based on mean relative abundance of each strain. The size of each pie chart indicates the summed relative abundance of HiMcV transcripts, averaged over multiple samples at a location. Stations where no HiMcV strains were detected are represented with an asterisk. The smaller pie charts in the North Atlantic and Station ALOHA all contain reads only from McV-SA1.

2.5 CONCLUSIONS

We sequenced four new strains of *Micromonas commoda* viruses and compared them to previously published prasinoviruses. Three of the four viruses were relatively similar in genome structure and gene content, forming their own clade within the prasinoviruses, while one strain (McV-KB2) was relatively divergent from the other HiMcVs and from other previously sequenced prasinoviruses. All four viruses overlap in host range, with each possessing 17-46 orthogroups not found in the other three viruses. This indicates that even within a local environment there is high genetic diversity amongst prasinoviruses that overlap in host range.

Functional traits of the core and specialized HiMcV genes are consistent with gene content of other prasinovirus genomes (Moreau et al., 2010; Finke et al., 2017; Bachy et al., 2021), in which core genes are involved in fundamental processes such as genome replication/transcription and virion structure, but also certain metabolic processes such as photosynthesis, phosphorus starvation, degradation of host polymers, nucleotide metabolism, lipid metabolism, and glycosylation. The flexible pan-genome contains mostly hypothetical genes with some putative metabolic genes that may allow for local adaptation by manipulating host metabolism during infection.

We found 61 orthogroups in HiMcV genomes that are not found in previously sequenced prasinoviruses. Although the majority of these orthogroups are of unknown function, the unique HiMcV orthogroups represent a significant expansion of prasinovirus gene content. The unique orthogroups include a type of phosphate transporter and a type of apoptosis inhibitor that may be novel to viruses in general, and enzymes for spermidine synthesis novel to prasinoviruses. Additionally, we found 48 orthogroups containing HiMcV sequences as well as sequences from one or both hosts, suggesting a substantial history of horizontal gene transfer. The variety of functions encoded by the shared orthogroups suggests that acquiring host genes aids viral manipulation of many host metabolic processes, as well as manipulation of host stress responses and defenses against infection. All HiMcVs appear to carry one or more orthologs to a host PTOX, a sequence previously found in only one other prasinovirus (OtV-RT2011), and which is homologous to the PTOX commonly found in cyanophages. The common occurrence of this gene in disparate virus clades may

indicate an increased need for protection against photodamage for viruses infecting photosynthesizers at low latitudes.

Further emphasizing the strength of coevolutionary host-virus dynamics are the results of our comparison of prasinoviruses infecting *Micromonas*, *Ostreococcus*, and *Bathycoccus*. Viral gene content is strongly tied to host genus identity, and significantly divergent orthogroups contain sequences likely related to diverse functions such as virion attachment, manipulating host stress responses, and metabolism of components needed for virion construction.

Metagenomic data provide evidence that all but one HiMcV strain, McV-SA1, are relatively rare in major ocean basins. Several factors may contribute to the relatively small number of hits against the utilized datasets. Three of the isolates (McV-KB2, McV-KB3, and McV-KB4) were isolated from a coastal site and may be more abundant in coastal locations compared to the primarily open ocean metagenome stations. The highly productive upwelling region near South Africa, surveyed in the GA10 transect, contained the highest abundance and diversity of HiMcVs, consistent with the fact that *Micromonas* is generally more common under nutrient-rich conditions (Not et al., 2004; Lopes Dos Santos et al., 2017). Therefore, further surveys in productive waters may capture more sequences that map to our HiMcVs. In addition, the GEOTRACES metagenomes were created by filtering 100 mL of whole seawater onto 0.2 and 0.02 µm filters, which may be too small of a volume to reliably capture less abundant giant viruses in low-biomass pelagic waters.

Our study shows that isolating and sequencing new viruses within a relatively well-studied clade (the prasinoviruses) can continue to increase our knowledge of marine viral gene content and genome evolution. In addition to sequencing more isolates, future work should focus on verifying putative gene function, as well as evolutionary pathways of genes and their origins. In this way, we can better understand how viruses influence phytoplankton communities and vice versa.

2.6 REFERENCES

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. J Mol Bio 3: 403–410.

Aylward, F.O. and Moniruzzaman, M. (2022) Viral Complexity. Biomolecules 12: 1061.

Bachy, C., Yung, C.C.M., Needham, D.M., Gazitúa, M.C., Roux, S., Limardo, A.J., et al. (2021) Viruses infecting a warm water picoeukaryote shed light on spatial co-occurrence dynamics of marine viruses and their hosts. ISME J 15: 3129–3147.

Banadyga, L., Lam, S.-C., Okamoto, T., Kvansakul, M., Huang, D.C., and Barry, M. (2011) Deerpox Virus Encodes an Inhibitor of Apoptosis That Regulates Bak and Bax. J Virol 85: 1922–1934.

Baumann, S., Sander, A., Gurnon, J.R., Yanai-Balser, G.M., Van Etten, J.L., and Piotrowski, M. (2007) Chlorella viruses contain genes encoding a complete polyamine biosynthetic pathway. Virology 360: 209–217.

Bellec, L., Grimsley, N., Moreau, H., and Desdevises, Y. (2009) Phylogenetic analysis of new Prasinoviruses (*Phycodnaviridae*) that infect the green unicellular algae *Ostreococcus* , *Bathycoccus* and *Micromonas*. Environmental Microbiology Reports 1: 114–123.

Biller, S.J., Berube, P.M., Dooley, K., Williams, M., Satinsky, B.M., Hackl, T., et al. (2018) Marine microbial metagenomes sampled across space and time. Sci Data 5: 180176.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38: 5825–5829.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18: 188–196.

Clerissi, C., Grimsley, N., Ogata, H., Hingamp, P., Poulain, J., and Desdevises, Y. (2014) Unveiling of the Diversity of Prasinoviruses (Phycodnaviridae) in Marine Samples by Using High-Throughput Sequencing Analyses of PCR-Amplified DNA Polymerase and Major Capsid Protein Genes. Appl Environ Microbiol 80: 3150–3160.

Core Team, R. (2022) R: A language and environment for statistical computing.

Courties, C., Vaquer, A., Troussellier, M., Lautier, J., Chrétiennot-Dinet, M.J., Neveux, J., et al. (1994) Smallest eukaryotic organism. Nature 370: 255–255.

Derelle, E., Monier, A., Cooke, R., Worden, A.Z., Grimsley, N.H., and Moreau, H. (2015) Diversity of Viruses Infecting the Green Microalga *Ostreococcus lucimarinus*. J Virol 89: 5812–5821.

Eikrem, W. and Throndsen, J. (1990) The ultrastructure of *Bathycoccus* gen. nov. and B. *prasinos* sp. nov., a non-motile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. Phycologia 29: 344–350.

Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20: 238.

Emms, D.M. and Kelly, S. (2018) STAG: Species Tree Inference from All Genes, Evolutionary Biology.

Farzad, R., Ha, A.D., and Aylward, F.O. (2022) Diversity and genomics of giant viruses in the North Pacific Subtropical Gyre. Front Microbiol 13: 1021923.

Fernández-García, J.L., De Ory, A., Brussaard, C.P.D., and De Vega, M. (2017) *Phaeocystis globosa* Virus DNA Polymerase X: a “Swiss Army knife”, Multifunctional DNA polymerase-lyase-ligase for Base Excision Repair. Sci Rep 7: 6907.

Finke, J., Winget, D., Chan, A., and Suttle, C. (2017) Variation in the Genetic Repertoire of Viruses Infecting *Micromonas pusilla* Reflects Horizontal Gene Transfer and Links to Their Environmental Distribution. Viruses 9: 116.

Gallot-Lavallée, L., Blanc, G., and Claverie, J.-M. (2017) Comparative Genomics of *Chrysochromulina Ericina* Virus and Other Microalga-Infecting Large DNA Viruses Highlights Their Intricate Evolutionary Relationship with the Established Mimiviridae Family. J Virol 91: e00230-17.

Goldsmith, D.B., Crosti, G., Dwivedi, B., McDaniel, L.D., Varsani, A., Suttle, C.A., et al. (2011) Development of phoH as a Novel Signature Gene for Assessing Marine Phage Diversity. Appl Environ Microbiol 77: 7730–7739.

Greiner, T., Frohns, F., Kang, M., Van Etten, J.L., Käsmann, A., Moroni, A., et al. (2009) Chlorella viruses prevent multiple infections by depolarizing the host membrane. Journal of General Virology 90: 2033–2039.

Greiner, T., Moroni, A., Van Etten, J., and Thiel, G. (2018) Genes for Membrane Transport Proteins: Not So Rare in Viruses. Viruses 10: 456.

Guillard, R.R.L. (1975) Culture of Phytoplankton for Feeding Marine Invertebrates. In Culture of Marine Invertebrate Animals. Smith, W.L. and Chanley, M.H. (eds). Boston, MA: Springer US, pp. 29–60.

Guillard, R.R.L. and Ryther, J.H. (1962) STUDIES OF MARINE PLANKTONIC DIATOMS: I. CYCLOTELLA NANA HUSTEDT, AND DETONULA CONFERVACEA (CLEVE) GRAN. Can J Microbiol 8: 229–239.

Ha, A.D., Moniruzzaman, M., and Aylward, F.O. (2023) Assessing the biogeography of marine giant viruses in four oceanic transects. ISME COMMUN 3: 43.

Heath, S., Knox, K., Vale, P., and Collins, S. (2017) Virus Resistance Is Not Costly in a Marine Alga Evolving under Multiple Environmental Stressors. Viruses 9: 39.

ICTV (2023) Virus Taxonomy: 2022 Release.

Jeffrey, S.W., Wright, S.W., and Zapata, M. (2011) Microalgal classes and their signature pigments. In Phytoplankton Pigments. Roy, S., Llewellyn, C.A., Egeland, E.S., and Johnsen, G. (eds). Cambridge University Press, pp. 3–77.

Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30: 772–780.

Kolde, R. (2019) pheatmap: Pretty Heatmaps.

Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., and De Clerck, O. (2012) Phylogeny and Molecular Evolution of the Green Algae. Critical Reviews in Plant Sciences 31: 1–46.

Li, B., Liang, J., Baniasadi, H.R., Phillips, M.A., and Michael, A.J. (2023) Functional polyamine metabolic enzymes and pathways encoded by the virosphere. Proc Natl Acad Sci USA 120: e2214165120.

Lopes Dos Santos, A., Gourvil, P., Tragin, M., Noël, M.-H., Decelle, J., Romac, S., and Vaulot, D. (2017) Diversity and oceanic distribution of prasinophytes clade VII, the dominant group of green algae in oceanic waters. The ISME Journal 11: 512–528.

Ma, J., Edlich, F., Bermejo, G.A., Norris, K.L., Youle, R.J., and Tjandra, N. (2012) Structural mechanism of Bax inhibition by cytomegalovirus protein vMIA. Proc Natl Acad Sci USA 109: 20901–20906.

Manton, I. and Parke, M. (1960) Further observations on small green flagellates with special reference to possible relatives of Chromulina *pusilla* Butcher. J Mar Biol Ass 39: 275–298.

Mende, D.R., Bryant, J.A., Aylward, F.O., Eppley, J.M., Nielsen, T., Karl, D.M., and DeLong, E.F. (2017) Environmental drivers of a microbial genomic transition zone in the ocean's interior. Nat Microbiol 2: 1367–1373.

Monier, A., Chambouvet, A., Milner, D.S., Attah, V., Terrado, R., Lovejoy, C., et al. (2017) Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. Proc Natl Acad Sci USA 114.

Monier, A., Welsh, R.M., Gentemann, C., Weinstock, G., Sodergren, E., Armbrust, E.V., et al. (2012) Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges: Phosphate and commonalities in marine viral-host exchanges. Environmental Microbiology 14: 162–176.

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. Nat Commun 11: 1710.

Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., and Grimsley, N. (2010) Marine Prasinovirus Genomes Show Low Evolutionary Divergence and Acquisition of Protein Metabolism Genes by Horizontal Gene Transfer. J Virol 84: 12555–12563.

Nobre, T., Campos, M.D., Lucic-Mercy, E., and Arnholdt-Schmitt, B. (2016) Misannotation Awareness: A Tale of Two Gene-Groups. Front Plant Sci 7.

Not, F., Latasa, M., Marie, D., Cariou, T., Vaulot, D., and Simon, N. (2004) A Single Species, *Micromonas pusilla* (Prasinophyceae), Dominates the Eukaryotic Picoplankton in the Western English Channel. Appl Environ Microbiol 70: 4064–4072.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5: e9490.

Puxty, R.J., Millard, A.D., Evans, D.J., and Scanlan, D.J. (2015) Shedding new light on viral photosynthesis. Photosynth Res 126: 71–97.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005) InterProScan: protein domains identifier. Nucleic Acids Research 33: W116–W120.

Schvarcz, C.R. (2018) Cultivation and Characterization of Viruses Infecting Eukaryotic Phytoplankton from the Tropical North Pacific Ocean.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068–2069.

Suttle, C.A. (2007) Marine viruses — major players in the global ecosystem. Nat Rev Microbiol 5: 801–812.

Thomas, R., Jacquet, S., Grimsley, N., and Moreau, H. (2012) Strategies and mechanisms of resistance to viruses in photosynthetic aquatic microorganisms. Advances in Oceanography and Limnology 3: 1–15.

Wang, Y., Ferrinho, S., Connaris, H., and Goss, R.J.M. (2023) The Impact of Viral Infection on the Chemistries of the Earth's Most Abundant Photosynthesizes: Metabolically Talented Aquatic Cyanobacteria. Biomolecules 13: 1218.

Weynberg, K., Allen, M., and Wilson, W. (2017) Marine Prasinoviruses and Their Tiny Plankton Hosts: A Review. Viruses 9: 43.

Wilhelm, 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: 8.

Yau, S., Grimsley, N., and Moreau, H. (2015) Molecular ecology of Mamiellales and their viruses in the marine environment. pip 2: 83–89

Chapter 3: Transient, context-dependent fitness costs accompanying viral resistance in isolates of the marine microalga *Micromonas sp.* (class Mamiellophyceae)

3.1 ABSTRACT

Marine microbes are important in biogeochemical cycling, but the nature and magnitude of their contributions are influenced by their associated viruses. In the presence of a lytic virus, cells that have evolved resistance to infection have an obvious fitness advantage over relatives that remain susceptible. However, susceptible cells remain extant in the wild, implying that the evolution of a fitness advantage in one dimension (virus resistance) must be accompanied by a fitness cost in another dimension. Identifying costs of resistance is challenging because fitness is context dependent. We examined the context dependence of fitness costs in isolates of the picophytoplankton genus *Micromonas* and their co-occurring dsDNA viruses using experimental evolution. After generating 88 resistant lineages from two ancestral *Micromonas* strains, each challenged with one of four distinct viral strains, we found resistance led to a 46% decrease in mean growth rate under high irradiance, and a 19% decrease under low. After a year in culture the experimentally selected lines remained resistant, but fitness costs had attenuated. Our results suggest that the cost of resistance in *Micromonas* is dependent on environmental conditions and the duration of population adaptation, illustrating the dynamic nature of fitness costs of viral resistance among marine protists.

3.2 INTRODUCTION

Marine viruses influence oceanic biogeochemical processes through the infection of single-celled organisms such as bacteria, archaea, and protists (Fuhrman, 1999; Suttle, 2007). An important aspect of microbe-virus interactions is the ability of microbes to evolve defenses against viral infection (Breitbart, 2012). Although resistance to infection is commonplace among marine microbes in both laboratory and field settings (Thyrhaug et al., 2003; Waterbury and Valois 1993), the persistence of susceptible hosts suggests a fitness cost associated with viral protections (Lenski, 1988; Lennon et al., 2007). The fitness costs of resistance to infections by lytic viruses are likely to have broad consequences for ecosystems, such as altering the growth rate and productivity of microbial communities and influencing the portion of microbial production that is lysed rather than consumed by predators (Våge et al., 2013).

Much of the early work investigating the evolutionary responses of microscopic marine plankton to viral infections focused on bacteria (see Breitbart, 2012 and references therein), but there is growing interest in how selection by viruses might influence the phenotype of unicellular eukaryotes.

A common metric of fitness is growth rate, which has been observed to decrease in some resistant strains of prokaryotes (Lennon et al., 2007). In eukaryotic algae, Frickel et al. (2016) found a negative correlation between breadth of resistance and growth rate in the freshwater alga *Chlorella variabilis*, using host strains that co-evolved with viruses in a long-term chemostat experiment. In another study, resistant morphotypes of *Emiliana huxleyi* showed a 15%-55% decrease in growth rate compared to ancestral strains (Frada et al., 2017). However, virus resistance was not accompanied by a detectable decline in growth rate in several other studies with various eukaryotic phytoplankton such as *E. huxleyi*, *Ostreococcus tauri*, and *Micromonas* sp. (Thomas et al., 2011; Ruiz, Baudoux, et al., 2017; Ruiz, Oosterhof, et al., 2017).

Quantifying a cost of resistance (COR) is challenging in part because the nature and magnitude of fitness costs may vary with context, such as the physiological condition of the host, and may also depend on the (co)evolutionary history of the host and virus populations. Bohannan et al. (1999) found that fitness costs for *Escherichia coli* varied depending on the phage strain to which cells were resistant and whether resistant cells were grown in a glucose-, trehalose-, or maltose-limiting medium. Previous studies on phytoplankton did not find an effect of resource levels on fitness costs (Lennon et al., 2007; Heath et al., 2017), but it is possible that shifts in COR would emerge under a different, or more severe, resource limitation. Furthermore, the magnitude of fitness costs could be affected by the coevolutionary history of hosts and viruses. Past experiments to measure fitness costs have used host strains and viruses that were isolated from different water masses and/or kept separately in culture for years at a time (Heath et al., 2017; Ruiz, Baudoux, et al., 2017; Ruiz, Oosterhof, et al., 2017). It is possible that hosts and viruses with recent coevolutionary interactions will exhibit a different COR, if a virus is better adapted to its host and/or if a coevolutionary arms race has enhanced mutual adaptation (Thyrhaug et al., 2003; Schwartz and Lindell, 2017). While not specifically examining geography or habitat of cells and

viruses, Lennon et al. (2007) did provide evidence that the magnitude of the reduction in growth rates among resistant *Synechococcus* varied according to the identity of ancestral strains, indicating that the genetic background of a host may influence the cost of adapting to a particular virus.

Another consideration when trying to make sense of coevolutionary dynamics is that the time since acquiring resistance appears to affect the presence of COR. Avrani and Lindell (2015) found that COR attenuated over time in seven strains of the prokaryote *Prochlorococcus* using experimentally selected cultures that were subsequently propagated for 40 months. The authors identified both new genetic mutations and reversions that may have compensated for the initial decline in growth rate.

To better understand the ecological and evolutionary dynamics of eukaryotes and eukaryotic viruses, we used two isolates of the ubiquitous marine eukaryotic alga *Micromonas* and four dsDNA-containing viruses in the genus *Prasinovirus* that were co-isolated from the same coastal waters of Kāneʻohe Bay, Hawaiʻi. *Micromonas* is a ~2 µm diameter flagellated prasinophyte found in all major ocean basins (Cottrell and Suttle, 1995; Thomsen and Buck, 1998; Demory et al., 2018). Previously isolated *Micromonas* dsDNA viruses are lytic viruses (capsid diameter ca. 65 nm) in the family Phycodnaviridae similarly found in all ocean basins (Waters and Chan, 1983; Cottrell and Suttle, 1991; Zingone, 1999; Bellec et al., 2009; Wilson et al., 2009). Using our Kāneʻohe Bay isolates we conducted experimental evolution to generate resistant algal strains, which allowed us to focus on trait changes caused by resistance, controlling for other processes that could influence trait variation. Growth assays under different light conditions were used to test the interactive effects of resistance and resource limitation on fitness.

The host-virus systems used in our study appear to be the first published reports of *Micromonas* and *Micromonas* prasinoviruses isolated from the interior of the North Pacific Subtropical Gyre. We used marker genes of hosts and viruses to construct phylogenetic trees to contextualize our isolates within the framework of published genetic sequences.

3.3 RESULTS

Two *Micromonas* isolates from our culture collection, M1 and M2, were chosen for this study, given their susceptibility to the largest number of isolated virus strains in our collection. Virus isolates, referred to here for simplicity as V1, V2, V3, and V4, were chosen to select for resistant *Micromonas* cells. Strains V2, V3, and V4 correspond to McV-KB2, McV-KB3, and McV-KB4 in Chapter 2. When tested against a suite of seven *Micromonas* isolates in our collection, V1 and V4 had narrower, non-overlapping host ranges, each infecting only 2 of 7 isolates. V2 and V3 had broader, overlapping, but not identical, host ranges, with each infecting 6 of 7 isolates. Prior to selection M1 was susceptible to infection by V1, V2, and V3, and M2 was susceptible to V2, V3, and V4 (Fig.2.1).

3.3.1 Phylogeny of *Micromonas* and virus Isolates

Partial 18S rRNA gene sequences of M1 and M2 revealed 100% nucleotide identity between the two strains. M1 (UHM1061; GenBank accession OQ428650) and M2 (UHM1065; OQ436457), along with another *Micromonas* isolate, UHM1080 (OQ445607), from the oligotrophic Station ALOHA (100 km north of O'ahu, Hawai'i; 22°45'N, 158°W), clustered with members of the species *Micromonas commoda* (Fig. 3.1). Three other published 18S rRNA gene sequences also shared 100% nucleotide identity with M1 and M2, including a sequence from an uncultured eukaryote from the South China Sea (JX188376; Wu et al., 2014) and isolates from pelagic waters in the Sargasso Sea and Pacific Ocean (AY955002; Šlapeta et al., 2006, KU244663; Foulon and Simon, 2016).

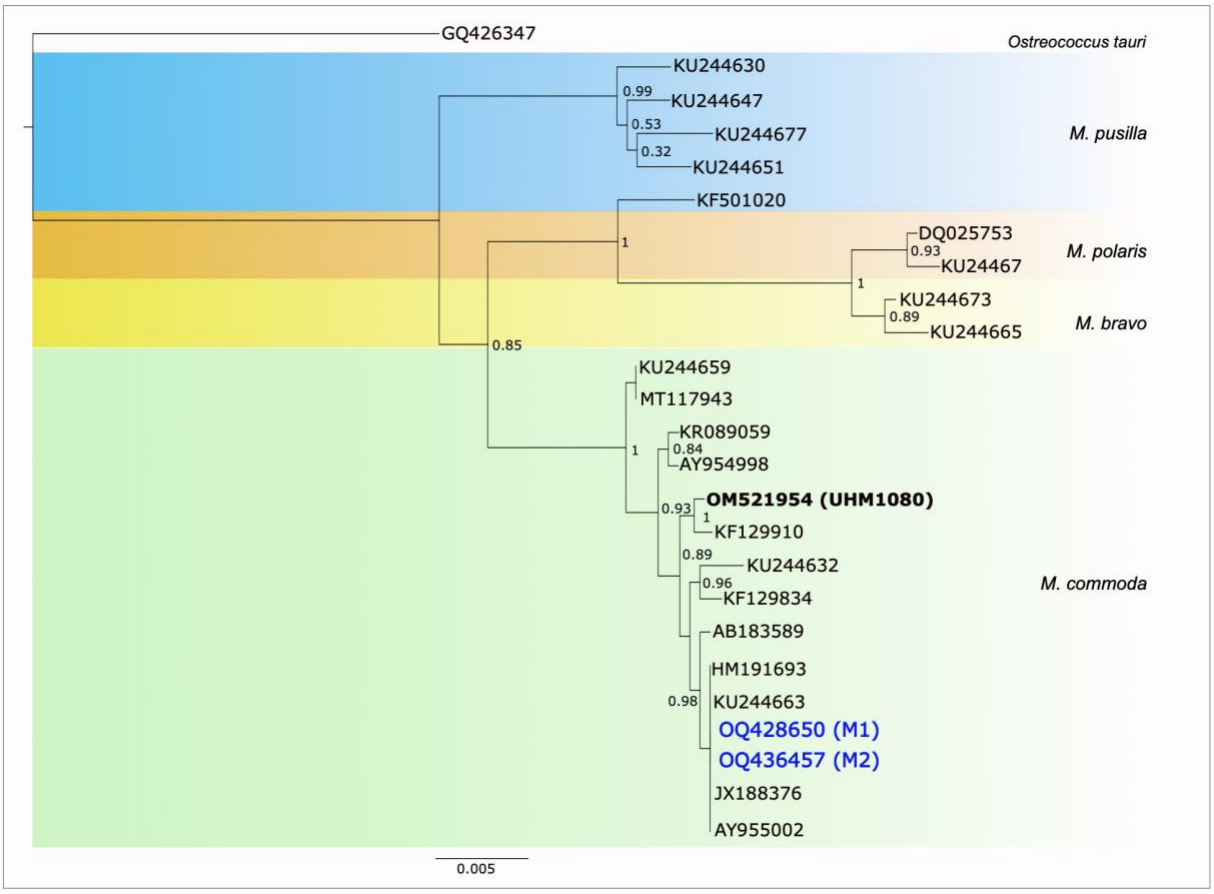


Figure 3.1. Phylogenetic tree of partial 18S rDNA genes of Mamiellales derived from trimmed alignments of 18S rRNA genes from the two *Micromonas* strains used in this study (M1 and M2), a *Micromonas* strain from the pelagic Station ALOHA (UHM1080), and related sequences found using NCBI's BLAST tool. Alignments were created and trimmed with Geneious 11.1 default alignment tool and processed through FastTree using approximately-maximum-likelihood. Node support values reflect FastTree local support values derived from the Shimodaira-Hasegawa test.

The phylogenetic relationships for virus isolates V1, V2, V3, and V4 were more complex than those of the host isolates (Fig. 3.2). The V1 de novo assembly showed evidence of a contaminating sequence. To explore this further, the polB-containing contig from V1 was dissolved to obtain component reads, which were then reassembled with stringent settings. This reassembly resulted in two contigs with relatively equal coverage, one with a polB gene unique to V1 and another contig with 100% nucleotide identity to the polB gene of V4. Thus, V1 may have been contaminated by V4, or the two virus types were co-isolated in the original V1 culture.

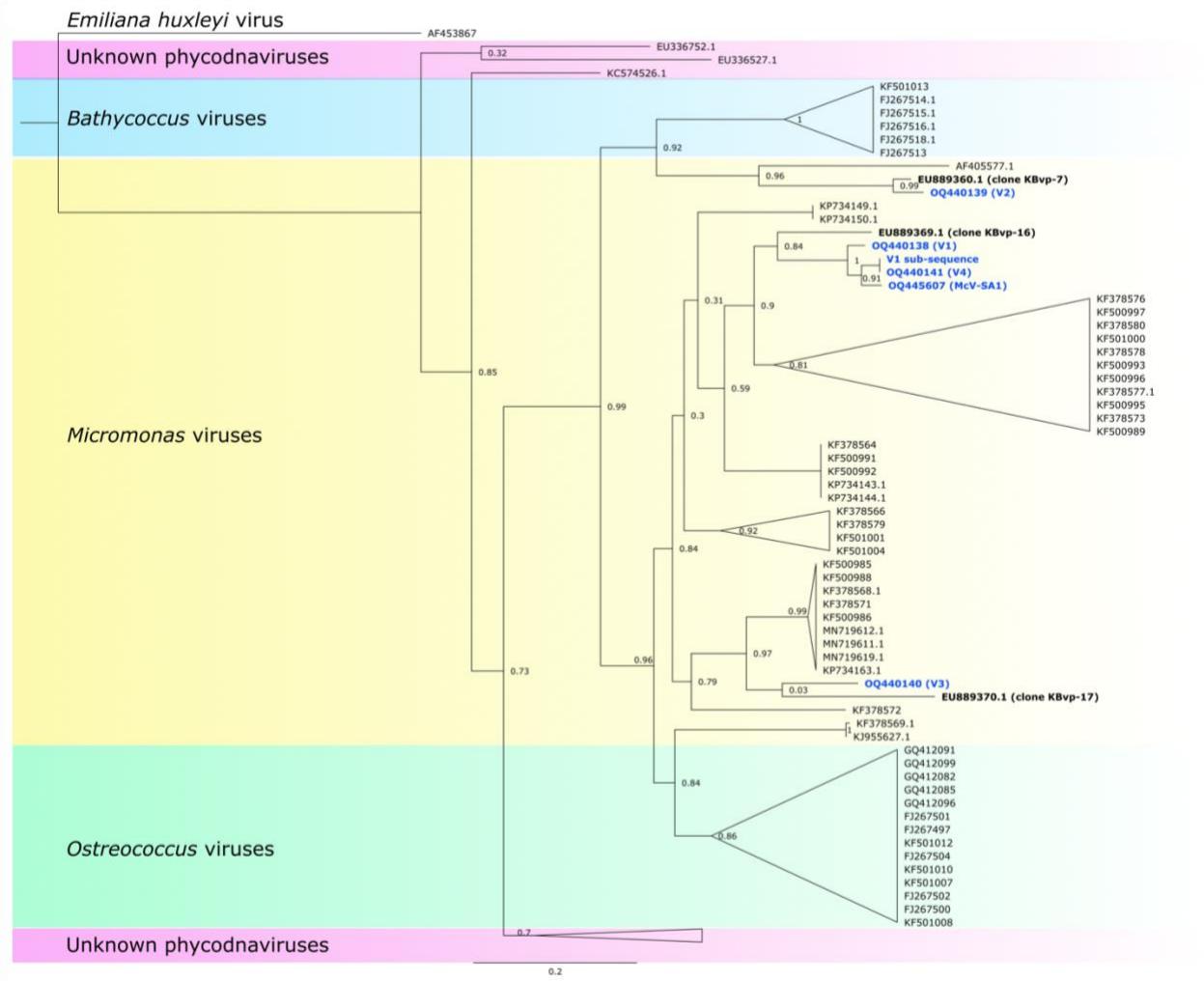


Figure 3.2. Phylogenetic tree of partial polB genes of phycodnaviruses derived from trimmed alignments of polB genes of the four viruses used in this study (V1–V4) from coastal waters of Hawai‘i, a *Micromonas*-infecting virus isolated on UHM1080 (MsV-SA1) from the open ocean waters off the coast of Hawai‘i, and related sequences found through NCBI’s BLAST tool. Alignment correction and tree generation were created using the same methodology described in Fig. 3.1.

The V1 (OQ440138) and V4 (OQ440141) partial polB sequences group into a clade with a previously sequenced *Micromonas* virus in our collection, MsV-SA1, isolated from Station ALOHA on *Micromonas* strain UHM1080 (Schvarcz, 2018). This group also contains a polB sequence (clone KBvp-16) amplified with phycodnavirus-targeted primers from samples collected from Kāne‘ohe Bay (Culley et al., 2009), the location where the viruses used in this study were isolated. V3 (OQ440140) is closely related to another uncultured prasinovirus, KBvp-17, also previously detected in Kāne‘ohe Bay (Culley et al., 2009), but these sequences are grouped into a different clade. Finally, V2 (OQ440139) possesses a polB sequence that is relatively divergent

from other isolated *Micromonas* viruses, with a range of 69.8%-71.3% nucleotide identity with the three other virus strains used in this study. The most closely related sequence to V2 is an uncultured prasinovirus clone, KBvp-7, previously sampled from Kāneʻohe Bay, with 85.7% nucleotide identity (Culley et al., 2009).

3.3.2 Fitness measurements

Our selection experiment yielded 135 cell lines, 88 of which were selected for resistance to infection by one of the viruses, and 47 susceptible lines put through the same dilution-to-extinction process but not exposed to virus (Table 3.1). Amongst the 88 resistant cell lines were 11 to 22 replicates of each cell strain-virus combination. Resulting resistant cell lines were named by concatenating *Micromonas* and virus isolate codes (e.g., there were 11 resistant lines derived from challenging M1 with V1 and these are designated M1V1). Control susceptible cell lines are distinguished from original ancestral strains by appending an 'S' to the original *Micromonas* isolate name (i.e., M1S, M2S). In a study of resistance to viral infection by *Ostreococcus tauri*, Thomas et al. (2011) found continued viral production, perhaps by budding of viruses from intact cells, in a subset of lines that were resistant to lysis. To examine whether viruses were produced by our resistant lines we introduced filtrate from resistant lines to susceptible ancestral cultures, but these showed no signs of lysis, and therefore we ruled out the possibility that the resistant lines maintained a persistent viral infection.

Table 3.1. Host and virus crosses with the resulting number of resistant and susceptible isolates. Susceptible isolates are indicated by the letter "S" at the end of the strain name. Gray cells indicate a host-virus cross that did not result in lytic infection.

M1 Descendants	*M1S*	*M1V1*	*M1V2*	*M1V3*	
# of isolates	23	11	23	10	
M2 Descendants	*M2S*		*M2V2*	*M2V3*	*M2V4*
# of isolates	24		13	18	13

Growth rates from our initial May 2018 experiment, conducted shortly after isolation of descendant cell lines, showed a significant difference between susceptible and resistant cell lines in high light conditions (Fig. 3.3; see Supplementary Figures S3.1 & S3.2 for raw exponential growth curves). Susceptible M1 lines had a mean growth rate of 1.31 d^{-1} and 0.43 d^{-1} under high light and low light, respectively. Susceptible M2 cells had similar growth rates to M1, at 1.41 d^{-1} and 0.34 d^{-1} under high and low light, respectively. Growth rates of resistant cell lines under high light were consistently lower than their susceptible counterparts (for M1: $p < 0.001$, $\chi^2_1 = 13.2$; for M2: $p < 0.001$, $\chi^2_1 = 9.3$), with the percent decrease in mean growth rates among resistant lines ranging from 21% in M2V3 cells to 85% in M1V1 cells.

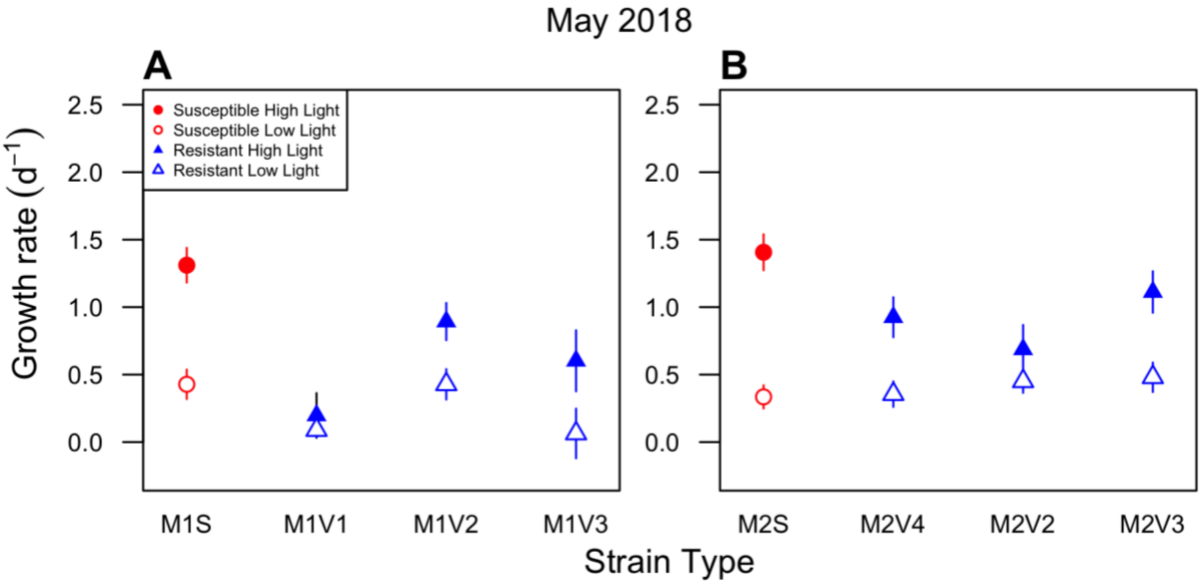


Figure 3.3. Growth rates of susceptible and resistant cell lines in the May 2018 experiment. Panels A and B show results for M1 and M2, respectively. The plotted points are model-estimated means, averaging over multiple replicate experimental lines (n = 11-20; Table 3.1). Closed circles indicate data from the high light treatment, and open circles from the low light treatment. Error bars are 95% confidence intervals. Dotted horizontal lines show 0 and 2 d^{-1} growth rates as a reference.

Under low light the effect of resistance on fitness was more variable, but significant for both M1 and M2 lines (for M1: $p = 0.01$, $\chi^2_1 = 5.5$; for M2: $p = 0.01$, $\chi^2_1 = 5.9$). However, a positive or negative trend in growth rates among resistant lines was not consistent across all host-virus combinations, with M1 resistant cell lines having slower mean growth rates (0.19 d^{-1}) and M2 resistant cell lines having higher rates (0.43 d^{-1}) than susceptible counterparts (M1: 0.42 d^{-1}, M2: 0.33 d-1;Fig. 3.3). Two resistant

lines derived from an M1 ancestor grew at slower mean rates (M1V1: 0.09 d^{-1}, M1V3: 0.06 d^{-1}) than susceptible M1S cell lines (0.42 d^{-1}), but one did not (M1V2: 0.43 d^{-1}). The mean growth rate for one set of resistant lines derived from the M2 ancestor (M2V4: 0.36 d^{-1}) was similar to that of the M2S lines (0.33 d^{-1}), but the mean growth rates for two other resistant lines (M2V2: 0.45 d^{-1}, M2V3: 0.48 d^{-1}) were higher than their susceptible counterparts.

Resistant and susceptible cell lines were maintained in culture for approximately one year with 1% volume transfer into fresh, virus-free, medium every two weeks. M1V1 cell lines went extinct within six months of the aforementioned fitness assay. For cell lines that did survive, we re-examined a subset of these (2 to 3 lines per cross) after one year (September 2019), with new lysis assays and growth experiments. We found that all resistant lines had maintained their resistance, despite being propagated in the absence of viruses, and that susceptible lines had maintained their susceptibility. However, new growth patterns had emerged (Fig.3.4). Mean growth rates of susceptible hosts under high light increased to 1.54 d^{-1} for M1 (previously 1.31 d^{-1}) and 1.83 d^{-1} for M2 (previously 1.41 d^{-1}). An increase in growth rate did not occur at low light for susceptible cell lines. The M1S low light growth rate decreased to 0.37 d^{-1} from 0.42 d^{-1} and M2 maintained its low light growth rate of 0.34 d^{-1}. The strong pattern of decreased fitness of resistant lines under high light seen shortly after selection was not observed 15 months later for host M1 or M2, with no significant differences seen between susceptible and resistant mean growth rates (Fig. 3.4; for M1: $p = 0.74$, $\chi^2_1 = 0.1$; for M2: $p = 0.47$, $\chi^2_1 = 0.5$). When we analyzed the high light subset on its own, the growth rates of resistant lines (M1V2: 1.56 d^{-1}, M1V3: 1.38 d^{-1}; M2V2: 1.69 d^{-1}, M2V3: 2.05 d^{-1}, M2V4: 1.36 d^{-1}) were the same as, or faster than, their susceptible counterparts on average (M1S: 1.54 d^{-1}; M2S: 1.83 d^{-1}). Only one group of resistant lines (M2V4) maintained a fitness cost similar to that initially seen. Notably, M2V2-derived cell lines had the highest mean growth rate of all lines tested, a 13% higher growth rate than M2 susceptible lines ($p = 0.01$, $\chi^2_1 = 6.2$). Under low light, M1-derived resistant lines did not have mean growth rates different from susceptible counterparts ($p = 0.11$, $\chi^2_1 = 2.5$), M2-derived lines did ($p = 0.01$, $\chi^2_1 = 6.5$), with a trend towards increased growth rate.

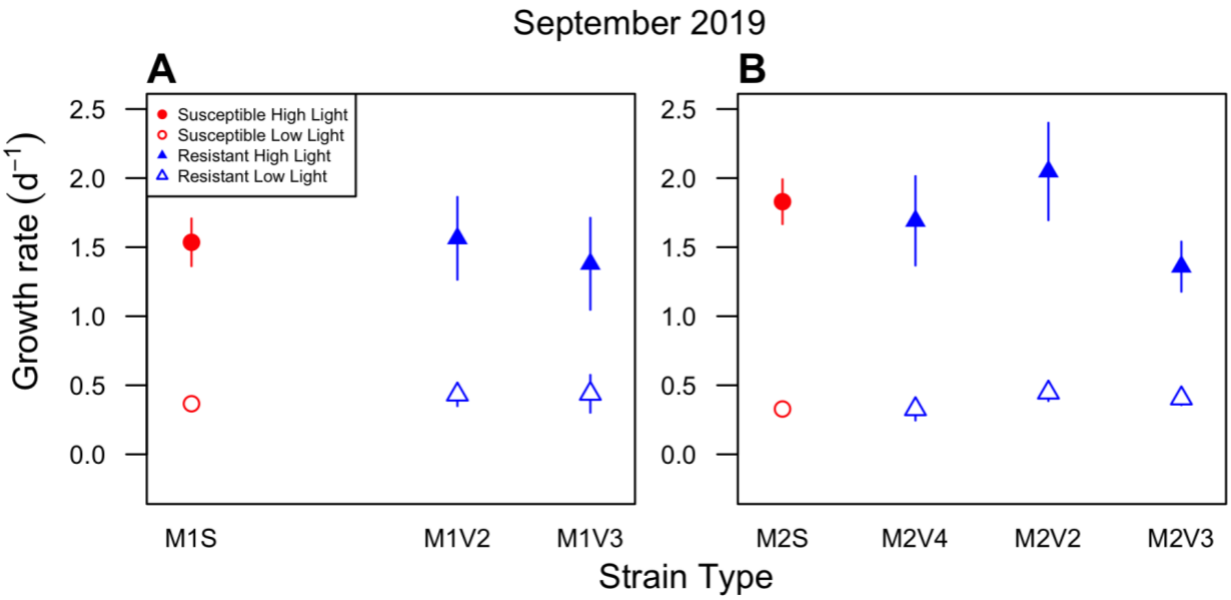


Figure 3.4. Growth rates of susceptible and resistant cell lines in the September 2019 experiment. Panels A and B show results for M1 and M2, respectively. M1V1 cell lines went extinct before this experiment was conducted. The plotted points are model-estimated means, averaging over multiple replicate experimental lines (n = 2-3). Closed circles indicate data from the high light treatment, and open circles from the low light treatment. Error bars are 95% confidence intervals. Dotted horizontal lines show 0 and 2 d^{-1} growth rates as a reference.

3.4 DISCUSSION

The large number of replicate resistant lineages (11-20) generated from each host-virus pair improved the power of our statistical analysis, despite variable responses among virus-host pairs. We detected a consistent fitness cost among cell lines that had been selected for resistance to infection 2–3 months prior to growth assays. This fitness cost took the form of lower growth rates under high light conditions. In contrast, a fitness cost was not consistently present under low light. A subset of resistant cell lines was tested approximately one and a half years later, after verifying continued resistance to population lysis, and the reduction in growth rates under high light disappeared for nearly all observed resistant cell lines. In summary, the presence and magnitude of COR depended on irradiance, identity of host and virus strain, and time since initial selection.

3.4.1 Interplay of resource availability and COR

Our hypothesis, that resource availability would influence the severity of COR, was supported by the results of our May 2018 growth experiments, which showed that the decrease in growth rates for resistant cell lines was larger and more consistent under high light. The mechanism underlying the interaction between light and COR is unknown, but a common mechanism of resistance observed among microbes is the modification or elimination of a cell surface receptor used by the virus for attachment. Viruses in the *Phycodnaviridae* use such receptors to attach to a potential host before inserting their DNA into the cytoplasm (Dunigan et al., 2019). Viral replication then takes place in the nucleus, with virion assembly in the cytoplasm. If the cell surface receptors used for viral attachment are involved in transport of an essential nutrient, their modification via selection for resistance could impair growth rate, as was found in an *Escherichia coli*-phage system (Lenski, 1988). While Thomas et al. (2011) found evidence that this resistance strategy may not be present in the related protist *Ostreococcus*, this mechanism may still occur in our *Micromonas* cell lines, as any adaptations that result in both viral resistance and a reduction in the maximum uptake rate of a limiting nutrient would be consistent with our observations: Under high light, energy is plentiful so nutrients could be limiting and the cost would be evident, but under low light, energy is limiting and a reduced nutrient uptake rate is of less, or no, consequence. This putative mechanism could be further investigated via viral adsorption assays, in which lower viral adsorption rates would suggest a modification in *Micromonas* cell surface receptors. Measuring nutrient uptake kinetics for susceptible and resistant lines under different light levels would also provide insight into nutrient acquisition by resistant cells, however this approach would be convoluted by factors of lowered growth rate as well as a need to test a multitude of limiting nutrients to determine if one or more are affected.

Greater COR associated with high light may also be explained if light levels at 100 µE m^{-2} s^{-1} cause an unforeseen stress that interacts with an underlying mechanism of resistance. However, our “high” irradiance of 100 µE m^{-2} s^{-1} is not out of the ordinary for surface waters of Kāneʻohe Bay, the habitat from which these strains were isolated. A study of daily light levels at 2 m depth in the bay over a 14-month period indicated

that Photosynthetically Active Radiation (PAR) ranged from 0.04-1810 µE m^{-2} s^{-1}, with an average daily mean of 85.74 µE m^{-2} s^{-1}(Ritson-Williams and Gates, 2016). Additional experiments using a greater number of light levels would help establish the light level for optimum growth and examine the interplay between light stress and resistance.

Our results differed from some previous work by researchers who studied the interplay between resource availability and viral immunity-associated fitness costs. Heath et al. (2017) tested varying light, phosphate, salinity, and temperature levels for resistant lines of *Ostreococcus tauri*, an alga closely related to *Micromonas*. Heath et al. did not find a correlation between resource level and cost of resistance. One difference among our experiments is the magnitude of resource limitation. Heath et al. compared growth under irradiances of 60 vs. 85 µE m^{-2} s^{-1}, which may not shift the factors limiting growth as much as 10 vs. 100 µE m^{-2} s^{-1}. It is also possible that the different results in these studies are simply attributable to the differing physiology among species, which may lead to different COR responses.

3.4.2 Fitness costs attenuated over time

Our resistant cultures lost their cost of resistance, after fifteen months of propagation in the absence of virus, even though they maintained resistance to population lysis. Because the experimentally evolved cell lines were propagated in the absence of viruses it is possible that genetic diversity in the cultures arose during the interim period, and susceptible subpopulations could have been present in the resistant cultures. However, our results indicate that cells with the resistance phenotype continued to dominate the populations, because dominance of a susceptible subpopulation would have led to a detectable decline in population density in response to viral inoculation, which was not observed for any of the resistant cell lines. The loss of a fitness cost could be a result of compensatory mutations, which has been reported in comparable experiments using the prokaryotes *Escherichia coli* (Lenski, 1988) and *Prochlorococcus* (Avrani and Lindell, 2015).

Frickel et al. (2016) conducted a three-month coevolution experiment with the freshwater green alga *Chlorella variabilis* and its chlorovirus, finding arms race dynamics in which the breadth of host resistance increased over time, while average host growth rate declined over time. These results suggest that continued selective

pressure from coevolving viruses may maintain or increase fitness costs, although the results do not rule out growth rate depression being a short-term (i.e., lasting a few months) consequence, comparable to our results. In contrast, researchers using resistant lines that were previously cultured for many years have reported little evidence of depression in growth rates, suggesting that more time allows for the accumulation of compensatory mutations (Heath et al., 2017; Ruiz, Baudoux, et al., 2017; Ruiz, Oosterhof, et al., 2017).

3.4.3 Environmental implications

Based on our observation that COR was greater under high light conditions, we hypothesize that a higher relative abundance of susceptible cells will occur in natural high light environments, where the cost to maintain resistance is greater. Given the rapid decline in irradiance with depth in the ocean, depth may play a role in determining the frequency of viral immunity among *Micromonas* populations. In situations where photosynthesis extends below the mixed layer (e.g., shallow mixed layers in relatively clear waters), a comparison of resistance among *Micromonas* populations within the "high" light mixed layer with "low" light closer to the base of the euphotic zone, would be worth examining. In addition, if a greater fraction of the population is resistant under lower irradiance, this could mean that a greater fraction of primary production is consumed by zooplankton, rather than being lysed by viruses. Consumption of phytoplankton production by zooplankton transfers energy and nutrients to larger organisms in higher trophic levels, while viral lysis shunts organic matter to the dissolved pool consumed by smaller microbes, and therefore the prevalence of resistance in natural populations may be important for ecosystem size structure and associated differences in carbon export (Wilhelm and Suttle, 1999).

3.4.4 Approaches to measuring fitness Costs

Ruiz et al. (2017) used diverse *Micromonas* and virus isolates from established cultures, rather than experimental evolution, to test for a fitness cost of viral resistance, by asking whether natural variation in the breadth of resistance among the hosts correlated with fitness metrics. The authors did not observe a correlation between growth rate and the two metrics of resistance used in their study, one metric being defined as the breadth of resistance of a host strain and the other being the viral burst

size propagating from a host strain. Those results may be consistent with the findings from our current study, as the lack of a fitness cost in Ruiz et al. (2017) may reflect compensatory evolution in long-term culture collections after isolation, or in natural populations. It is also possible that a lack of coevolutionary history between the host and virus isolates, or natural variation in host growth rate due other causes, obscured the effect of resistance.

Our study measured fitness as population growth rate, and it is possible that the temporal dynamics of fitness costs would differ if another measure of fitness was utilized. For example, Thomas et al. (2011) used experimental evolution to study COR in *Ostreococcus tauri* and found no difference in growth rates between resistant and susceptible lines. However, they did find that their resistant cell line declined in direct competition with its susceptible ancestor, which implies that relevant aspects of fitness were not captured by growth rates in batch culture. A follow-up to our work could use competition assays, ideally using interspecific competitors in addition to intraspecific competitors, to see whether competitive differences are maintained over time even as differences in growth rate attenuate.

3.5 MATERIALS AND METHODS

3.5.1 *Micromonas* and MicV isolates

The UHM phytoplankton collection contains seven *Micromonas* sp. strains. Six of these, including the two used to generate cell lines in this study, were isolated by Schvarcz (2018) from the surface waters (> 2 m) of Kāneʻohe Bay, Oʻahu, Hawaiʻi (21°27 N, 157°48 W) in 2011. Additionally, a *Micromonas* sp. strain was isolated from surface waters (< 25 m) at the oligotrophic site Station ALOHA, ~100 km north of O'ahu (22°45 N, 158°00 W) in 2012. A dsDNA virus strain was isolated on each one of the seven *Micromonas* sp. strains using water from the same location of each host, in 2011 for Kāneʻohe virus strains and in 2015 for the Station ALOHA virus strain. Full descriptions of cell and virus strains can be found in Schvarcz (2018). Each of the seven cell lines exhibit different cross-infection dynamics when challenged separately with the seven virus strains. Two *Micromonas* strains were chosen for this study out of the seven strains in our collection, given their amenability to culturing and susceptibility to multiple isolated virus strains (each susceptible to 3 of the 7 virus strains in our

collection), which allowed for the selection of resistance against multiple virus strains (Fig. 2.1). These two *Micromonas* sp. strains were isolated from Kāneʻohe Bay with original isolate identifiers of KB-FL13 and KB-FL42 (Schvarcz 2018) and were assigned University of Hawaiʻi at Mānoa (UHM) culture collection identifiers of UHM1061 and UHM1065. For simplicity in the presentation of this paper these cell lines are referred to as M1 and M2, respectively. The *Micromonas* cell line originating from Station ALOHA, UHM1080 (referred to as AL-FL30 in Schvarcz, 2018), was not included in phenotypic assay but was included in phylogenetic analysis described below.

Cell cultures were maintained in f/2 -Si medium (Guillard and Ryther, 1962; Guillard, 1975) at 25 °C with an irradiance of ~100 µE m-2 s-1 provided by Philips F17T8/TL841 17-watt fluorescent bulbs. The irradiance level of 100 µE m^{-2} s^{-1} is typically near the optimum for growth of marine phytoplankton (Edwards et al., 2015). Cultures were transferred approximately every two weeks by diluting to one percent in fresh medium. Four viruses with the UHM culture collection IDs McV-KB1 , McV-KB2, McV-KB3, and McV-KB4, (corresponding to viruses isolated on hosts KB-FL13, KB-FL22, KB-FL28 and KB-FL42 in Schvarcz, 2018) were used in the selection process for resistant hosts. For simplicity of presentation, McV-KB1 , McV-KB2, McV-KB3, and McV-KB4 will be referred to here as V1, V2, V3, and V4, respectively. The gene sequences of a fifth virus, MsV-SA1, isolated from waters of Station ALOHA on the cell line UHM1080, was used in phylogenetic analysis. Viral stocks were maintained by challenging exponentially growing cultures of respective isolated host strains approximately every two weeks. Lysates were filtered with a 0.2 µm syringe filter once algal cultures were visually cleared. The resulting filtrate was then stored at ~4 °C.

3.5.2 Phylogenetic Analysis of Host and Virus Isolates

Gene phylogenies were constructed using the 18S rRNA gene for *Micromonas* isolates, and the DNA polymerase gene Beta (polB) for the viruses. Genomes of M1 and M2 were sequenced with PacBio Sequel II technology at the University of Delaware DNA Sequencing and Genotyping Center. Genome assembly of cellular DNA was conducted at the University of Delaware Bioinformatics Data Science Core using Canu ver 1.9 (Koren et al., 2017) and 18S rRNA gene sequences were identified in genome

assemblies using the Basic Local Alignment Search Tool (Altschul et al., 1990) against *Micromonas* sp. RCC299's 18s rRNA partial gene sequence (HM191693).

Phylogenetic analysis of M1 and M2 18S rRNA genes was carried out using closely related sequences found via nucleotide BLAST against GenBank (Sayers et al., 2022). Using Geneious 11.1.5, full and near-full gene sequences were aligned. The resulting 1,771 bp alignment was used to generate a phylogenetic tree using FastTree 2.1.12 (Price et al., 2010) in Geneious using default settings.

The genomes of the four viruses were sequenced with Illumina paired-end 151bp technology at the Microbial Genomic Sequencing Center (MiGs) at the University of Pittsburgh. Partial de novo assemblies were created in Geneious 11.1.5. A phylogenetic tree was constructed for the B-family DNA polymerase (polB) gene, which is commonly used as marker gene for Phycodnaviridae (Chen and Suttle, 1995). The polB genes in our sequences were identified by BLAST search using published *Micromonas* virus sequences as the query. Differences in base-call quality among reads for the putative polB gene of V1 indicated the presence of two strains. The contig containing the polB sequence was dissolved to acquire and re-assembled with high stringency to resolve contaminating sequences. To construct a phylogeny, polB sequences similar to each of our four viruses were found through a BLASTn search against the NCBI nucleotide database. We included 108 published polB sequences, as well as a polB sequence from a virus isolated on UHM1080, to create a 453 bp alignment using MAFFT accessed through Geneious. FastTree was then used to construct a phylogeny in Geneious using default settings.

3.5.3 Establishing susceptible and resistant host strains

We were unable to grow *Micromonas* on solid media and used dilution-to-extinction procedure in liquid medium to isolate resistant cell lines. To isolate strains grown from single resistant cells, 3.5 mL susceptible M1 and M2 cultures with cell density of 106 cells mL^{-1} were serially diluted 10-, 10^2-, and 10^3- fold. For each dilution, 2 mL aliquots were transferred to all the wells of a 24-well untreated CELLSTAR® suspension culture plate and then challenged during exponential growth with 500 µL fresh, undiluted, viral lysate (~107 infectious particles mL^{-1}). We combined dilution and viral inoculation to increase the probability that each resistant line would be derived from

a single cell, and that different resistant lines would potentially represent different genotypes. Given that the highest dilution in our series was 103, and our lysate had 107 infectious particles mL^{-1}, the lowest concentration of infection particles was 104 infectious particles mL^{-1}. As our host organisms are motile flagellates, encounter rates of cells and viruses would provide sufficient viral pressure.

Plates were incubated (25 °C, ~100 µE m^{-2} s^{-1}, 12:12 light:dark cycle) and routinely examined at 400× with an inverted microscope until resistant mutants grew to a density detectable by microscopy (≥ 103 cells mL-1), taking approximately 3-4 weeks. Approximately 10% of wells from the highest dilution plate of each host-virus strain combination contained detectable cells, which were used to establish new 3.5 mL cultures. Once in exponential growth, the putative resistant lines were re-challenged with viral lysate and monitored for signs of lysis to verify resistance. Susceptible ancestors were put through the same dilution and pipetting steps, without the addition of lysate, in order to control for phenotypic evolution arising from the genetic bottlenecks and selection associated with these steps. As with the resistant strains, cells growing back at the highest dilution were used to establish experimental cultures. The resulting susceptible strains were named by appending “S” to the ancestral host code name. Multiple strains (from 10 to 24) of each susceptible or resistant strain type were generated (Table 3.1).

Thomas et al. (2011) reported viral budding in cultures of *Ostreococcus tauri* that had been selected for resistance, suggesting that a shift from lytic to chronic infection could underlie the resistant phenotype. To confirm that resistant lines in our experiment were truly resistant and not under chronic infection, resistant cell cultures were filtered (0.2 µm) and the resulting filtrate was added to susceptible host cultures. These cultures were then monitored for reductions in cell density over the course of two weeks.

3.5.4 High and low light growth experiment

Growth experiments were conducted over three to four weeks, during which 135 cell lines (Table 3.1) were grown in untreated 24-well CELLSTAR® suspension plates (2.5 mL culture volume). All descendant cell lines were grown in both low light (~10 µE m^{-2} s^{-1}) and high light (~100 µE m^{-2} s^{-1}), with duplicates of every line in each light condition. All cultures were grown in the same Percival AL36L4 growth chamber with

cool white lights mounted above individual shelves adjusted to either high or low light conditions.

Daily growth data was collected for each well via fluorescence readings in a plate reader (TECAN Spark). Preliminary experiments found that fluorescence values correlate strongly with hemocytometer-based cell counts of *Micromonas*. Cultures were transferred into a new plate with fresh medium on a semi-weekly basis to maintain exponential phase growth. Maintaining cultures in exponential growth throughout the experiment allowed us to generate multiple exponential growth curves for each cell line. We obtained approximately 520 growth curves which were used to estimate the growth rates for each experimental isolate. A linear regression was fit to exponential phase data using R software in order to extract exponential growth rates (R Core Team, 2022). We used a mixed model to determine if growth rates were affected by susceptibility (resistant vs. susceptible) and light level (high vs. low) for each experimental sample with R package glmmTMB (Brooks et al., 2012) using the following terms:

growth rate ~ susceptibility + light + susceptibility:light + (1|plate) + (1|isolate)

where growth rate (the response variable) is modeled as a function of the terms to the right of the '~' (the predictor variables). Here, 'susceptibility:light' is an interaction term that accounts for any interactive effects of light and susceptibility. The terms (1|plate) and (1|isolate) are random effects that account for variation in growth among replicate plates and variation in growth among isolates that is not explained by susceptibility.

An initial growth experiment was conducted in May 2018, two to three months after the ancestral cultures were challenged to generate resistant cell lines. A follow-up experiment was conducted in September 2019 to observe if COR changed after fifteen months of cultivation (~220 generations). A randomized selection of cell lines from each host-virus cross were used in the September 2019 experiment due to resource restraints (n = 12). Between growth experiments cell lines were maintained at 25 °C, ~100 µE m^{-2} s^{-1}, and passaged approximately every two weeks. The September 2019 experiment did not include resistant lines from M1V1 challenges as these lines went extinct in long-term culture. Resistance status was verified in tandem with the second

growth experiment by introducing fresh viral lysate to resistant lines and monitoring for a decline in cell density.

3.6 REFERENCES

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. J Mol Bio 3: 403–410.

Avrani, S. and Lindell, D. (2015) Convergent evolution toward an improved growth rate and a reduced resistance range in *Prochlorococcus* strains resistant to phage. Proc Natl Acad Sci USA 112.

Bellec, L., Grimsley, N., Moreau, H., and Desdevises, Y. (2009) Phylogenetic analysis of new Prasinoviruses (*Phycodnaviridae*) that infect the green unicellular algae *Ostreococcus* , *Bathycoccus* and *Micromonas*. Environmental Microbiology Reports 1: 114–123.

Bohannan, B.J.M., Travisano, M., and Lenski, R.E. (1999) Epistatic Interactions Can Lower the Cost of Resistance to Multiple Consumers. Evolution 53: 292.

Breitbart, M. (2012) Marine Viruses: Truth or Dare. Annu Rev Mar Sci 4: 425–448.

Brooks, M.L., Fleishman, E., Brown, L.R., Lehman, P.W., Werner, I., Scholz, N., et al. (2012) Life Histories, Salinity Zones, and Sublethal Contributions of Contaminants to Pelagic Fish Declines Illustrated with a Case Study of San Francisco Estuary, California, USA. Estuaries and Coasts 35: 603–621.

Chen, F. and Suttle, C.A. (1995) Amplification of DNA polymerase gene fragments from viruses infecting microalgae. Appl Environ Microbiol 61: 1274–1278.

Core Team, R. (2022) R: A language and environment for statistical computing.

Cottrell, M. and Suttle, C. (1991) Wide-spread occurrence and clonal variation in viruses which cause lysis of a cosmopolitan, eukaryotic marine phytoplankter *Micromonas pusilla*. Mar Ecol Prog Ser 78: 1–9.

Cottrell, M.T. and Suttle, C.A. (1995) Genetic Diversity of Algal Viruses Which Lyse the Photosynthetic Picoflagellate *Micromonas pusilla* (Prasinophyceae). Appl Environ Microbiol 61: 3088–3091.

Culley, A.I., Asuncion, B.F., and Steward, G.F. (2009) Detection of inteins among diverse DNA polymerase genes of uncultivated members of the Phycodnaviridae. ISME J 3: 409–418.

Demory, D., Baudoux, A.-C., Monier, A., Simon, N., Six, C., Ge, P., et al. (2018) Picoeukaryotes of the *Micromonas* genus: sentinels of a warming ocean. ISME J 13: 132–146.

Dunigan, D.D., Al-Sammak, M., Al-Ameeli, Z., Agarkova, I.V., DeLong, J.P., and Van Etten, J.L. (2019) Chloroviruses Lure Hosts through Long-Distance Chemical Signaling. J Virol 93: e01688-18.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A., and Litchman, E. (2015) Light and growth in marine phytoplankton: allometric, taxonomic, and environmental variation: Light and growth in marine phytoplankton. Limnol Oceanogr 60: 540–552.

Foulon, E. and Simon, N. (2016) *Micromonas commoda* 18S ribosomal RNA gene, partial sequence.

Frada, M.J., Rosenwasser, S., Ben-Dor, S., Shemi, A., Sabanay, H., and Vardi, A. (2017) Morphological switch to a resistant subpopulation in response to viral infection in the bloom-forming coccolithophore *Emiliania huxleyi*. PLoS Pathog 13: e1006775.

Frickel, J., Sieber, M., and Becks, L. (2016) Eco-evolutionary dynamics in a coevolving host-virus system. Ecol Lett 19: 450–459.

Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399: 541–548.

Guillard, R.R.L. (1975) Culture of Phytoplankton for Feeding Marine Invertebrates. In Culture of Marine Invertebrate Animals. Smith, W.L. and Chanley, M.H. (eds). Boston, MA: Springer US, pp. 29–60.

Guillard, R.R.L. and Ryther, J.H. (1962) STUDIES OF MARINE PLANKTONIC DIATOMS: I. CYCLOTELLA NANA HUSTEDT, AND DETONULA CONFERVACEA (CLEVE) GRAN. Can J Microbiol 8: 229–239.

Heath, S., Knox, K., Vale, P., and Collins, S. (2017) Virus Resistance Is Not Costly in a Marine Alga Evolving under Multiple Environmental Stressors. Viruses 9: 39.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. Genome Res 27: 722–736.

Lennon, J.T., Khatana, S.A.M., Marston, M.F., and Martiny, J.B.H. (2007) Is there a cost of virus resistance in marine cyanobacteria? ISME J 1: 300–312.

Lenski, R.E. (1988) Experimental Studies of Pleiotropy and Epistasis in *Escherichia coli*. II. Compensation for Maldaptive Effects Associated with Resistance to Virus T4. Evolution 42: 433.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5: e9490.

Ritson-Williams, R. and Gates, R.D. (2016) Kaneohe Bay Light Data 2014 And 2015.

Ruiz, E., Baudoux, A.-C., Simon, N., Sandaa, R.-A., Thingstad, T.F., and Pagarete, A. (2017) *Micromonas* versus virus: New experimental insights challenge viral impact: *Micromonas* - MicV empirical estimations. Environ Microbiol 19: 2068–2076.

Ruiz, E., Oosterhof, M., Sandaa, R.-A., Larsen, A., and Pagarete, A. (2017) Emerging Interaction Patterns in the *Emiliania huxleyi*-EhV System. Viruses 9: 61.

Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., et al. (2022) Database resources of the national center for biotechnology information. Nucleic Acids Research 50: D20–D26.

Schvarcz, C.R. (2018) Cultivation and Characterization of Viruses Infecting Eukaryotic Phytoplankton from the Tropical North Pacific Ocean.

Schwartz, D.A. and Lindell, D. (2017) Genetic hurdles limit the arms race between *Prochlorococcus* and the T7-like podoviruses infecting them. ISME J 11: 1836–1851.

Šlapeta, J., López-García, P., and Moreira, D. (2006) Global Dispersal and Ancient Cryptic Species in the Smallest Marine Eukaryotes. Molecular Biology and Evolution 23: 23–29.

Suttle, C.A. (2007) Marine viruses — major players in the global ecosystem. Nat Rev Microbiol 5: 801–812.

Thomas, R., Grimsley, N., Escande, M., Subirana, L., Derelle, E., and Moreau, H. (2011) Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations: Viral resistance in Mamiellales. Environmental Microbiology 13: 1412–1420.

Thomsen, H.A. and Buck, K.R. (1998) Nanoflagellates of the central California waters: taxonomy, biogeography and abundance of primitive, green flagellates (Pedinophyceae, Prasinophyceae). Deep Sea Research Part II: Topical Studies in Oceanography 45: 1687–1707.

Thyrhaug, R., Larsen, A., Thingstad, T., and Bratbak, G. (2003) Stable coexistence in marine algal host-virus systems. Mar Ecol Prog Ser 254: 27–35.

Våge, S., Storesund, J.E., and Thingstad, T.F. (2013) Adding a cost of resistance description extends the ability of virus-host model to explain observed patterns in structure and function of pelagic microbial communities: Structuring of microbial communities by viruses. Environ Microbiol 15: 1842–1852.

Waterbury, J.B. and Valois, F.W. (1993) Resistance to Co-Occurring Phages Enables Marine *Synechococcus* Communities To Coexist with Cyanophages Abundant in Seawater. 7.

Waters, R.E. and Chan, A.T. (1983) *Micromonas pusilla* Virus: the Virus Growth Cycle and Associated Physiological Events Within the Host Cells; Host Range Mutation. J gen Virol 63: 199–206.

Wilhelm, 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: 8.

Wilson, W.H., Van Etten, J.L., and Allen, M.J. (2009) The Phycodnaviridae: The Story of How Tiny Giants Rule the World. In Lesser Known Large dsDNA Viruses. Current Topics in Microbiology and Immunology. Van Etten, J.L. (ed). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–42.

Wu, W., Huang, B., and Zhong, C. (2014) Photosynthetic picoeukaryote assemblages in the South China Sea from the Pearl River estuary to the SEATS station. Aquat Microb Ecol 71: 271–284.

Zingone, A. (1999) Seasonal dynamics in the abundance of *Micromonas pusilla* (Prasinophyceae) and its viruses in the Gulf of Naples (Mediterranean Sea). Journal of Plankton Research 21: 2143–2159.

Chapter 4: Genetic signatures of resistance to viral infection in the marine picoeukaryote *Micromonas*

4.1 ABSTRACT

Phytoplankton resistant to viral infection are commonplace in both environmental and laboratory settings, having broad implications for biogeochemical cycling in the marine environment. Experimental selection for resistant phenotypes of microbes has provided insights into the consequences of viral immunity via biological assays. However, the genomic basis of resistance has only been explored in a handful of marine microbes. Previous work suggests that the genetic elements underlying viral immunity may include smaller polymorphisms as well as large-scale structural variants within the genomes of marine phytoplankton. Here, we used short- and long- read sequencing technology to compare genomes of resistant and susceptible cell lines of the ubiquitous marine eukaryote, *Micromonas*, and found potential genetic signatures of resistance in the form of small polymorphisms (<50bp) as well as potential structural variants (>50bp). The most compelling small polymorphisms amongst resistant cell lines occurred in genes associated with surface proteins and DNA transcription. These small polymorphisms appeared to be dispersed throughout each genome. In contrast, only one out of eight putative structural variants implicated any genes. Moreover, structural variants tended to concentrate in hypervariable outlier chromosomes.

4.2 INTRODUCTION

Viruses of phytoplankton, including those that infect prokaryotic and eukaryotic algae, are abundant and regulate algal populations largely through mortality (Suttle, 2007; Breitbart, 2012). However, resistance to viral infection in marine phytoplankton is commonly observed in natural environments and is readily evolved in experimental settings (Thingstad, 2000; Thyrhaug et al., 2003). While resistance allows cells to avoid lysis, immune phenotypes may incur a loss of fitness in terms of growth rate (Lennon et al., 2007; Thomas et al., 2011). Past researchers have used phenotypic and genetic assays to identify putative mechanisms that confer resistance in microbes, such as modification of cell-surface proteins, life stage selection, changes to membrane-bound glycosphingolipids, and programmed cell death (Bohannan and Lenski, 2000; Frada et al., 2008; Bidle, 2016; Schleyer et al., 2023). While previous methodologies often relied on phenotypic observations, Next Generation Sequencing now allows for the investigation of the genomic basis using a dataset of dozens to hundreds of genomes.

Genomic comparisons of resistant and susceptible cell lines are becoming more common, but two studies highlight the current state of research. Avrani et al. (2011) utilized short-read Illumina technology to examine small-scale variants in 27 strains of the marine cyanobacterium *Prochlorococcus*. Each strain was experimentally evolved to establish resistance to one of five podoviruses. The genomes of these resistant strains were compared to eight susceptible control genomes. The researchers found that mutated genes in resistant genomes were associated with changes to cell surface receptors, a resistance mechanism implicated in heterotrophic bacteria (Bohannan et al., 1999). Furthermore, these variants tended to concentrate around hypervariable genomic islands, suggesting that coevolution with viruses drives rapid evolution of these genes.

In eukaryotic algae, large-scale structural variants (SVs) may also play a key role in viral immunity. Yau et al. (2018) used pulse field gel electrophoresis to identify signatures of chromosomal rearrangements within the genomes of experimentally evolved *Ostreococcus tauri* that were resistant to *Ostreococcus tauri* virus (OtV). A follow-up study by this research group utilized Illumina HiSeq and optical mapping to further characterize changes in resistant *O. tauri* genomes (Yau et al., 2020). The results of both studies indicated that the rearrangements involved the Small Outlier Chromosome (SOC), one of two low-GC hypervariable chromosomes possessed by *Ostreococcus*, as well as related species in the order of Mamiellales. Additionally, Blanc-Mathieu et al. (2017) found a positive correlation between resistance to viral infection and the size of the same hypervariable outlier chromosome in *O. tauri*.

The subject of the current study is the prasinophyte *Micromonas*, which is a relative of *Ostreococcus*, in the order Mamiellales. Prasinophytes are a group of small single-celled eukaryotes found in all major ocean basins. Along with their viruses, this group of phytoplankton has provided considerable insights into marine viral ecology (Mayer and Taylor, 1979; Cottrell and Suttle, 1995a, 1995b; Thomas et al., 2011; Finke et al., 2017). The Mamiellales are of particular interest given their potential role in dominating eukaryotic fractions in the sunlit marine environments as the oceans warm (Demory et al., 2018). Among the picoeukaryotes, *Micromonas* has been observed to dominate its size class in relatively productive waters (Not et al., 2004). The genomes of

two strains of this alga were previously characterized by Worden et al. (2009), who identified hypervariable regions concentrated in two outlier chromosomes.

Viruses infecting the *Micromonas* genus are found in all major ocean basins (Cottrell and Suttle, 1995). More recently, metagenome-assembled genomes potentially representing the *Micromonas* dsDNA virus were commonly found in oligotrophic gyres, implying that this virus may play a larger ecological role than previously thought (Ha et al., 2023).

Previously, we used two strains of *Micromonas* and four strains of double-stranded DNA virus (*Prasinovirus*) to select for 88 resistant cell lines, in order to measure potential trade-offs associated with resistance. Continuing our assessment of resistant *Micromonas*, here we sequenced the genomes of 20 cell lines, including both resistant and susceptible representatives, to perform an exploratory analysis of putative genetic signatures associated with resistance. In this current work we attempted to resolve both small-scale polymorphisms and large-scale structural variants associated with resistance to lytic infection in *Micromonas* by utilizing, respectively, short-read Illumina sequencing and long-read PacBio sequencing.

Our main questions are:

I. Whether specific variants or genes are commonly associated with immunity.

II. Whether different host strains obtain different variants in response to selection by the same viral strain.

III. Whether different viral strains select for similar or different mutations in the same host strain.

Our exploratory dataset may help provide insight to these questions using an ecologically relevant host-virus system.

4.3 METHODS

4.3.1 Host-virus systems

The strains used in this study are two *Micromonas* cell lines, M1 and M2, corresponding to UHM1061 and UHM1065 in the UH Mānoa culture collection, and four double-stranded DNA viruses, V1, V2, V3, V4, corresponding to McV-KB1, McV-KB2, McV-KB3, and McV-KB4 in the UH Mānoa culture collection (Fig. 2.1). Host and virus

strains were isolated from Kāneʻohe Bay in 2011 (Schvarcz, 2018). M1 and M2 were chosen because they grew easily in culture and could be lysed by several (n=3 and 4, respectively) virus strains in our collection. The four virus strains in this study were chosen either because they were isolated on M1 and M2 (V1 and V4, respectively) or because they infect a relatively large number of *Micromonas* strains in our collection, including M1 and M2 (V2 and V3). Overlapping host ranges allowed us to assess whether a particular virus strain selected for similar mutations in multiple host strains, and whether different viruses selected for similar mutations in the same host strain.

4.3.2 Generation of resistant and susceptible cell lines

Full isolation methods for resistant and susceptible cell lines are described in Chapter 3. Briefly, to generate resistant cell lines, a dilution series with 10-fold increments of *Micromonas* cultures was created with the goal of isolating single cells at high dilution steps. The same volume of undiluted viral lysate was added to each experimental dilution step to coerce resistant cell lines to outcompete susceptible *Micromonas* counterparts. Cultures were monitored for evidence of re-growth over the course of 4-5 weeks. Resistant lines derived from wells that regrew were challenged again with fresh lysate to confirm resistance. Susceptible cell lines were generated through the same dilution-to-extinction process, but without addition of viral lysate, as a control to account for selection and genetic drift that may occur during the experiment. Fresh, sterile media was added to susceptible controls in place of lysate. Our methodology produced 10-20 cell lines for each of six host-virus crosses. Lines produced with this methodology were named after the host and virus strain used to select for resistant lines plus a unique identifier, e.g., M1V1a indicates cell line ‘a’ of those that descended from M1 and are resistant to lysis from the V1 virus strain. Susceptible descendants were amended with the letter ‘S’, e.g., M1Sa (See Tables 4.1 and 4.2).

4.3.3 Reference genome assemblies and genome annotation

Reference genome assemblies were created for M1 and M2 using sequence data from PacBio long-read technology. We used antibiotic treatments followed by a Percoll density gradient to isolate *Micromonas* cells, to reduce contamination with DNA from bacteria present in algal cultures. Cultures of each cell line were transferred three

times consecutively into fresh f/2 -Si media containing 1% of a stock antibiotic mixture (Guillard and Ryther, 1962; Guillard, 1975). The recipe for the antibiotic mixture can be found in Supplementary Table S2.1. Cells were inoculated at a ratio of 1:100 and allowed to grow until they reached approximately 106 cells mL^{-1} before the next transfer, corresponding to late-stage exponential growth. After the third transfer, two liters of culture were concentrated to a 50 mL volume through serial centrifugation (5,000 x g, 20 min at 4°C) and decanting of supernatant. The resulting concentrated 50 mL volume was added to a 100 mL Percoll density step gradient created following Price et al. (1978). The goal of this method is to separate a band of *Micromonas* cells from contaminating bacteria based on differences in cell densities by using liquids of varying specific mass. Once a clean band of *Micromonas* cells was identified, it was extracted using a high-gauge syringe. Flow cytometry was used to establish that the resulting ratio of *Micromonas*:bacteria was > 100:1. Purified cells were extracted from the gradient, pelleted by centrifugation, and frozen at -20ºC. Frozen samples were shipped to the University of Delaware Sequencing and Genotyping Center for high molecular weight DNA extraction using a CTAB method (Supplementary Material S4.1). Library preparation and sequencing was performed with PacBio Sequel II SMRT cells.

We also sequenced total RNA of M1 and M2 grown under a variety of conditions, to generate a library of transcripts to aid in gene calling. Total RNA was generated from M1 and M2 cell cultures by sampling 10 mL of exponential phase culture (~106 cells mL^{-1}) every four hours for 24 hours (to capture diel variation in gene expression). An additional culture sample was given a heat shock treatment to stimulate stress response, in which 10 mL aliquots were placed in a 30ºC water bath for 30 minutes. Samples for RNA extraction were syringe filtered onto 25 mm 1 µm polycarbonate filters (STERLITECH®) and then frozen immediately in liquid nitrogen and stored at -80ºC. Filters were then thawed over ice and extraction of total RNA took place within a week of sampling using the ZymoBIOMICS RNA Miniprep Kit. Extracted RNA was sent to University of Delaware Sequencing and Genotyping Center for Illumina sequencing.

PacBio consensus long read (CLR) data was assembled using Canu (ver. 1.9) in PacBio-raw CLR assembly mode, providing an estimated genome size parameter of 22 Mbp (Koren et al., 2017). Resulting contigs were further polished to remove remaining

InDel errors by iterative rounds of mapping CLR reads to reference contigs using BLASR (ver 5.3.3 with default parameters except: maxMatch=30, minSubreadLength=750, minAlnLength=750, minPctSimilarity=70, minPctAccuracy=70, hitPolicy=randombest), followed by error correction using Arrow (Pacific Biosciences GenomicConsensus ver 2.3.3 with default parameters except: minCoverage=5, minConfidence=40, coverage=120) until a stable reference was obtained (4 iterations) (Chaisson and Tesler, 2012). Closing of circular/organellar elements was performed using Circlator (v1.5.5) and additional manual finishing was performed including manual assessment and scaffolding/overlap of adjacent contigs and resolution/dereplication of haplotype bubbles. Chromosome assignments were manually made using alignment to reference genomes with ProgressiveMauve (v2.4.0) (Worden et al., 2009; Genbank accession CP001323).

Annotation was performed with Maker (v3.01.03). A custom repeat library was generated using RepeatModeler (v2.0.1) with RepeatScout (v1.0.6) and TRF (v4.0.9). These repeats and repeats for order Mamiellales (CONS-Dfam_3.1-rb20170127) were identified by RMBLAST (v2.10.0) in RepeatMasker (v4.1.0) and masked for gene model annotation. Genome-specific *ab initio* gene calls by GeneMarkES (v 2.5p) and SNAP (v2006.07.28) were used to train Augustus (v3.3.3) gene models using e-training scripts (BUSCO v4.0.2).

Illumina RNAseq data was quality trimmed with TrimGalore! (v0.6.5) using cutadapt (v3.3) and mapped to draft genomes with STAR (v2.7.9a) using the two pass method and was the basis for a genome-guided transcriptome assembly using Trinity (v2.13.2). Trinity transcripts and primary CDS and protein sequences annotated in *Micromonas pusilla* assemblies RCC299_229_v3.0 and CCMP1545_v3.0, and additional protein sequences extracted from Uniprot for order Mamiellales were mapped to the genome as evidence and used to assess support for *ab initio* gene models. Noncoding RNA were identified using tRNAscan-SE (v2.0.5) and rnammer (v1.2), functional annotations were made by using BLASTp (v 2.12.0) against a swissprot database (v2021.04), and additional annotations were applied using InterProScan (v5.53-87.0).

Phylogenetic analysis of M1 and M2 18s rRNA marker genes is reported in Chapter 3. In summary, a 1771 bp trimmed alignment was used with FastTree 2.1.12 to assess phylogenetic relationships with published prasinophyte sequences.

4.3.4 Sequencing and mapping of descendant genomes

To investigate small-scale genetic changes associated with viral resistance (i.e., variants less than 50 bp in length), a selection of 17 resistant and susceptible descendants of M1 and M2 were sequenced with Illumina 151 technology at the University of Delaware Sequencing and Genotyping Center (Table 4.1). Two cell lines representing susceptible descendants from each ancestral host strain were included, in addition to representatives of each host strain resistant to different virus strains. Cells destined for Illumina short-read sequencing were treated with antibiotics to reduce bacterial contaminants using the methods described in the previous section. Notably, two out of three sequences from M1 cell lines resistant to infection by V1 (M1V1 cell lines) and all M1 sequences resistant to V3 (M1V3 cell line) appeared to be derived from genotypes not closely related to the reference genomes, based on the number of variants called for each cell line (~100,000+ variants as opposed to ~100-200 variants for all other cell lines). Remapping of these outlier sequences to the M2 reference genome resulted in similar variant call numbers, suggesting that these lines were not the result of contamination of M1 cultures with M2 cells. Thus, we excluded all M1V3 cell lines and the two outlier M1V1 cell lines from further analysis. Finally, as depicted in Fig. 2.1, M1 cannot be lysed by V4 and M2 cannot be lysed by V1 and therefore no cell lines were created to represent these crosses.

Illumina short reads were error corrected and read coverage was normalized with built-in Geneious v11.1 tools. Corrected reads were quality trimmed with Geneious BBDuk plugin with a minimum Phred score of 20 (Bushnell 2014). Reads were mapped to reference assemblies using the Geneious default mapper.

Table 4.1**.** Cell lines sequenced with Illumina. Note that certain host-virus crosses are not represented in this table due to lack of lytic infection (i.e., M1V4, M2V1) or due to exclusion of outlier sequences (i.e., M1V3).

Susceptible cell lines	Resistant to V1	Resistant to V2	Resistant to V3	Resistant to V4
M1Sa, M1Sb, M2Sa, M2Sb	M1V1a	M1V2a, M1V2b, M1V2c, M2V2a, M2V2b, M2V2c	M2V3a, M2V3b, M2V3c	M2V4a, M2V4b, M2V4c

4.3.5 Variant calling and filtering

The Geneious default variant caller was used on each descendant genome to find variants less than 50 bp long that occurred in at least 80% of reads. Variants in regions above and below 2 SD of mean coverage were excluded from further analysis. In order to identify candidate variants associated with resistance, we filtered variant calls into three nested categories. This filtering was intended to enrich for variants that were selected for during the experiment and could causally underlie resistant phenotype(s).

1) Phenotype-specific (PS) variants: We excluded variants that occurred in both susceptible and resistant cell lines. Our rationale is that these shared variants were likely selected prior to the start of the experiment, as the cell lines evolved to live in laboratory conditions, or were potentially selected during the experiment for the same reason. It is also possible that variants shared by all susceptible and resistant lines correspond to errors in the reference genomes. In either case they are not likely to be associated with resistance to infection. After filtering, we were left with variants found only in susceptible or resistant phenotypes, but not both. We categorize such variants as phenotype-specific or PS.

2) PS Nonsynonymous variants (PSNS): Phenotype-specific variants were further binned into synonymous and nonsynonymous categories. Particular attention was paid to genes with phenotype-specific nonsynonymous (PSNS) variants, as the change in protein sequence means we can conclude with more certainty that these variants may have caused some phenotypic change. PSNS candidate genes were functionally categorized into clusters of orthologous groups (COGs) with EggNOG mapper (default settings). PSNS variants were used to generate non-metric

multidimensional scaling (NMDS) plot using metaMDS in R vegan package, to visualize associations between host phenotype, ancestral host identity, virus strain used to select for resistance, and functional categories of genes with PSNS variants (R Core Team, 2022; Oksanen, J et al., 2022).

3) Pervasive PSNS variants: Phenotype-specific nonsynonymous variants that were found in more than one genome were labeled “pervasive.” The occurrence of a specific variant in multiple resistant cell lines increases the likelihood that this variant underlies resistant phenotypes. Genes with pervasive variants were searched for in the literature to determine if previous studies implicated them in conjunction with viral infection or immunity.

4.3.6 PacBio sequencing and structural variant analysis

Six descendant cell lines from ancestral strain M1 were sequenced with PacBio CLR technology at the University of Delaware Sequencing and Genotyping Center. We chose to focus on M1 cell lines due to their relatively large decrease in growth rate associated with resistance, as described in Chapter 3. The six M1 cell lines were selected for PacBio sequencing based on their ability to grow to a density of $>10^6$ cells mL^{-1} after serial antibiotic treatment (Table 4.2). Reference mapping and detection of structural variants in six M1 descendant genomes were carried out via the PacBio Structural Variant calling and analysis tool (PBSV; https://github.com/PacificBiosciences/pbsv). Structural variant calls were filtered to include only 1/1 genotypes, a remnant of the diploid assumption in the PBSV algorithm but interpreted as a dominance of the alternative variant for our haploid organism. These high frequency variants occur in >80% of reads at each position.

For functional categorization, translated sequences of genes in regions of structural variant calls were submitted to EggNOG (minimum e-value: 0.01, minimum percent identity: 25%) and InterProScan (default settings). BLAST searches against NCBI databases were also used to further elucidate potential gene functions when necessary.

Table 4.2. M1 cell lines sequenced with PacBio.

Susceptible	Resistant to V1	Resistant to V2	Resistant to V3
M1Sa	M1V1a, M1V1b	M1V2b, M1V2c	M1V3a

4.4 RESULTS AND DISCUSSION

4.4.1 Assembly notes and phylogenetics

The genomes of M1 and M2 are relatively similar in size at 20.7 Mbp and 21.01 Mbp, respectively (Table 4.3). The gene density of M1 (440.24 genes/Mbp) was lower than M2 (521.48 genes/Mbp) and GC content was higher (69.2% compared to 64%). The M1 genome assembly resulted in 26 contigs, 21 of which were assigned to 17 chromosomes. Five contigs, ranging from 11kbp to 60kbp, could not be assigned to a chromosome. The M2 assembly produced 17 contigs, all of which represent a complete chromosome. Phylogenetic analysis using 18s rRNA marker genes place our *Micromonas* strains in a clade with published *Micromonas commoda* sequences (Fig. 3.1).

Table 4.3. Information on PacBio-derived reference genome assemblies.

Strain	Genome Size (bp)	#Genes	GC%
M1 (UHM1061)	20,718,608	9,113	69.2
M2 (UHM1065)	21,012,695	10,951	64.0

4.4.2 Small-scale variants in Illumina-derived genomes

Amongst the 17 genomes sequenced with Illumina, 307 phenotype-specific variants were found (refer to Supplementary Figures S4.1 and S4.2 for information regarding all variants detected). Of the phenotype-specific variants, 182 were nonsynonymous. Amongst resistant cell lines, an average of 89.42% of the phenotype-specific variants were nonsynonymous, and in seven cell lines all phenotype-specific variants were nonsynonymous (Fig. 4.1). The percentage of nonsynonymous variants within the phenotype-specific bin among susceptible cell lines was dissimilar between M1 and M2 descendants. Susceptible M1 nonsynonymous mutations were nearly all phenotype-specific whereas susceptible M2 cell lines either had no phenotype-specific

variants (M2Sa) or had nearly the same proportion of shared nonsynonymous mutations as phenotype-specific nonsynonymous mutations (M2Sb).

Based on these observations, it appears that phenotype-specific variants in resistant cell lines were more likely to cause changes to amino acids, compared to shared variants or variants associated with the susceptible phenotype. The enrichment in nonsynonymous mutations among variants specific to resistant cell lines suggests that these variants may be largely the result our selection experiment, as opposed to genetic drift.

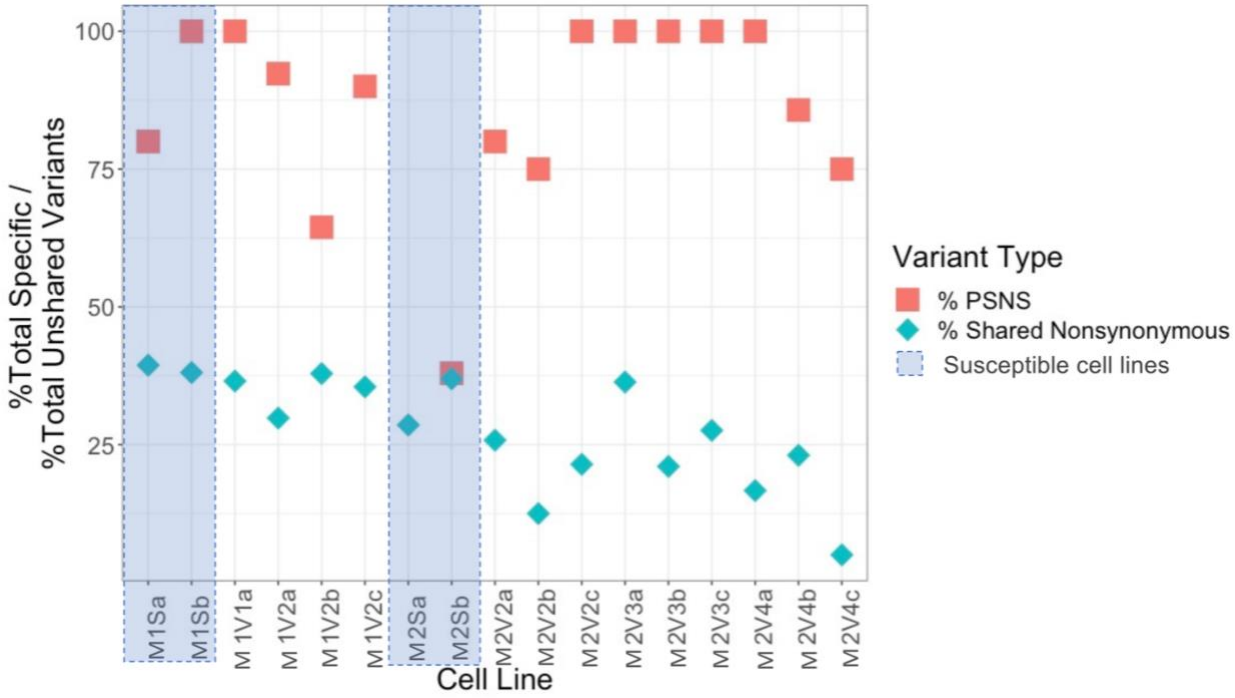


Figure 4.1. Proportions of nonsynonymous mutations amongst phenotype-specific (pink squares) and shared (green diamonds) variants within each cell line. Data for susceptible cell lines are highlighted in blue boxes. In all resistant cell lines, the great majority of phenotype-specific variants are nonsynonymous.

The susceptible cell line M2Sb contains the largest number of variants at 183 total variants, with 52 PSNS variants (Supplementary Figure S4.2). The next largest number of variants amongst cell lines are 140 variants in two M1V2 cell lines with 24 and 29 PSNS variants in each. The larger number of variants could indicate that the M2Sb cell line deviated from the ancestral line before the selection event.

There does not appear to be specific genomic regions with concentrations of PSNS variants nor a universal pattern among outlier hypervariable chromosomes (Figs. 4.2 and 4.3). Rather, some variants may be shared among a handful of genomes.

However, an examination of pervasive PSNS variants does provide insight as to what genes tend to be affected in multiple genomes.

Figure 4.2. Line plots of the genomes of M1 derived cell lines demonstrating placement of PSNS variants within each genome. Light blue lines represent individual genomes with red triangles signifying the position of a PSNS variant.

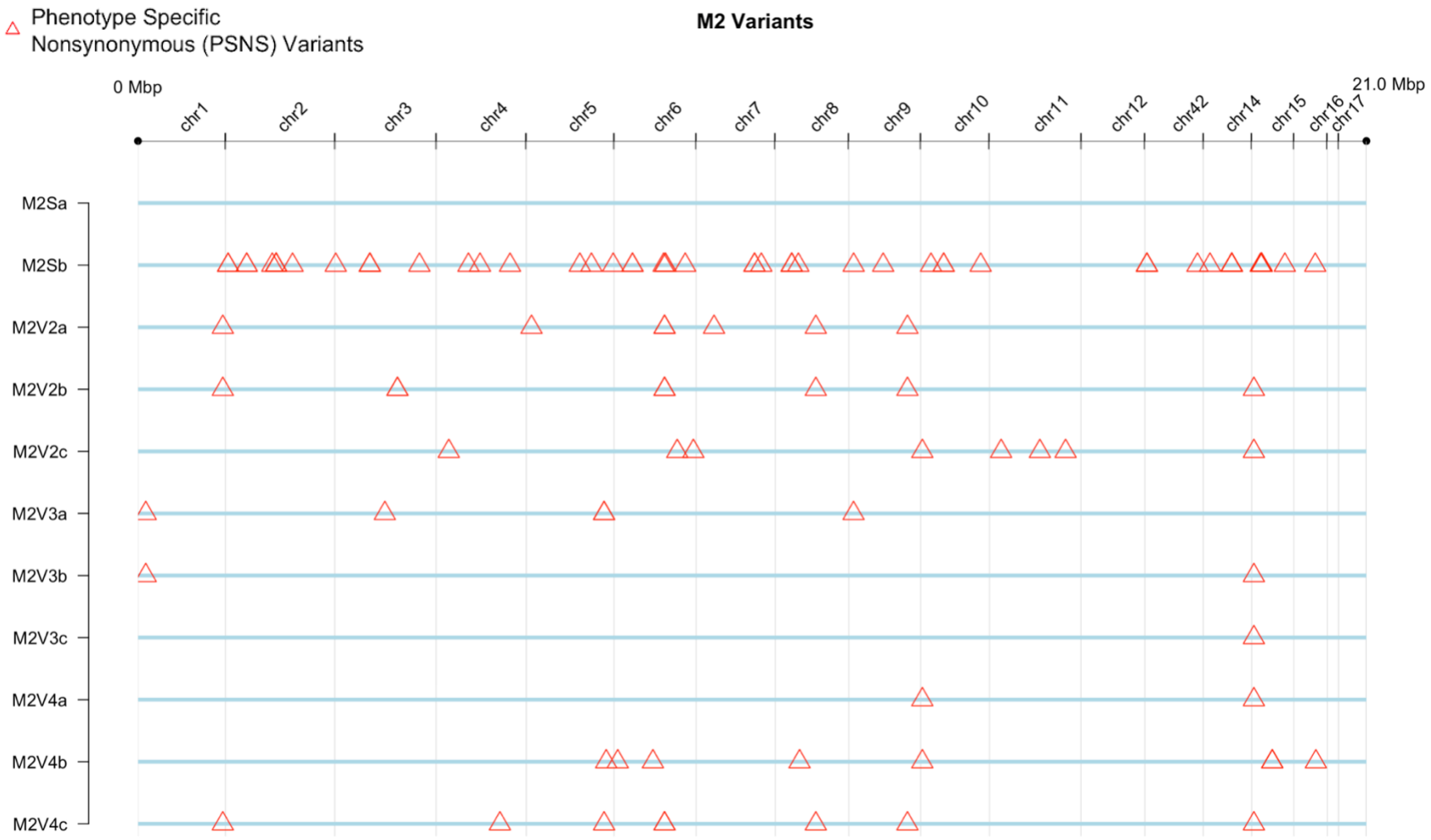


Figure 4.3. Line plot of the genomes of M2 derived cell lines demonstrating placement of PSNS variants within each genome. Light blue lines represent individual genomes with red triangles signifying the position of a PSNS variant.

Non-metric multidimensional scaling (NMDS) of the functional composition of PSNS variants (based on COG categories) indicates two main clusters of resistant cell lines, corresponding to ancestral host identity (Fig. 4.4). Within these clusters there is some evidence of sub-groupings of cell lines based on the identity of the virus strain used for selection. In particular, the M2V3 lines group together, while there is overlap among M2V2 and M2V4 lines. Between the two major clusters of resistant cell lines is a triangular “stripe” of PSNS variants belonging to susceptible cell lines. Collectively these clusters suggest that M1 resistant cell lines possess variants in different gene functional categories than variants in resistant M2 cell lines, and that susceptible cell lines also tend to possess a distinct variant pattern, although there is substantial functional variation within each of the clusters.

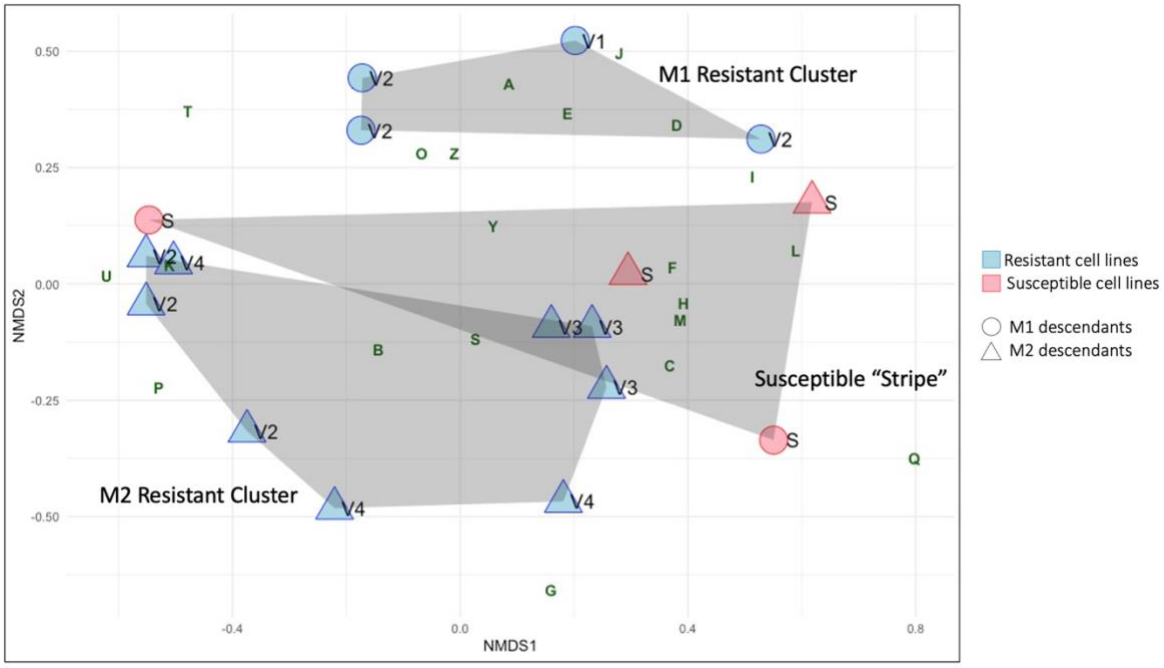


Figure 4.4. NMDS ordination of the composition of PSNS variants within each genome, categorized using the COG categories of the genes containing each variant. Light blue and light pink symbols represent variants from resistant and susceptible genomes, respectively. Circles are associated with M1 cell lines and triangles indicate M2 cell lines. Three gray convex hulls encompass clusters of PSNS variants belonging to M1 resistant cell lines, M2 resistant cell lines, and susceptible cell lines of both ancestral strains.

To better understand the potential function of variants underlying resistance we focused on predicted functions of genes containing PSNS variants that occurred in multiple resistant genomes. Thirty-four PSNS variants occurred in multiple genomes,

and we refer to these variants as “pervasive.” All pervasive variants were found in the genomes of resistant cell lines (i.e., no PSNS variant occurred in more than one susceptible genome), and zero PSNS variants occurred in M2Sa. The most compelling patterns among pervasive variants come from our M2 cell lines, which had a larger sample size (n=10 genomes) as well as a smaller number of PSNS pervasive variants (n=7) (Table 4.4).

The most frequent pervasive variant was a deletion in the gene TMCO4, which codes for the transmembrane and coiled-coil domain-containing protein 4. The variant associated with this gene was detected in six of the eight resistant M2 cell lines. This deletion is the only PSNS variant to occur in M2V3c, and one of only two PSNS variants in M2V3b and M2V4a. Little is known about the function of the protein that TMCO4 codes for, but a homologous gene in yeast, MIL1, was shown to encode for a peripheral membrane protein localized to the Golgi apparatus/early endosome, and is involved in recruiting proteins to the membrane to regulate vesicle formation and transport. The function of MIL1 could indicate that TMCO4 is involved with viral attachment or entry in *Micromonas*, if TMCO4 is localized to the cell membrane. Alternatively, if TMCO4 is localized to an interior membrane it could be required for trafficking of the virus within the cell. Mutations in TMCO4 have been associated with diseases in higher organisms, suggesting TMCO4 is part of a stress response or immunity cascade (Sirchia and Luparello, 2007; Hauser et al., 2014; Lassen et al., 2016). A relevant example to our work is the upregulation of TMCO4 expression in blue-winged teal super shedders of the avian influenza virus which directly correlates this gene with response to viral infection.

Only one pervasive variant in M2 cell lines occurred in a hypervariable outlier chromosome: a single nucleotide polymorphism in the gene FGT1, which is located in chromosome 1 (the Big Outlier Chromosome). FGT1 encodes for Protein FORGETTER 1, which is associated with cellular response to heat shock (Brzezinka et al., 2016). While FGT1 has not been implicated in host defense, changes to this gene in *Pseudomonas* bacteria has been associated with its virulence in *Nicotiana* plants, which implicates this gene in host-pathogen dynamics (Taguchi et al., 2009).

The five other pervasive variants in the M2 dataset appear to be associated with gene expression, cell surface structures, and cellular secretion. These five variants occur together in the same three genomes and include a hypothetical protein in the Lipase class 3 family, a hypothetical protein with FG-GAP repeat regions, a putative calcium-transporting ATPase 11 gene (ACA11), and two variants in a Histone-lysine N-methyltransferase gene (ASHH2). Lipase class 3 enzymes, also known as triacylglycerol lipases, aid in lipid metabolism. One enzyme from this class has been indicated in resistance mechanisms in maize against sugarcane mosaic virus (Xu et al., 2022). The hypothetical protein with an FG-GAP repeat is of particular interest as FG-GAP repeat regions occur in alpha integrins and are associated with host-pathogen interactions (Sun et al., 2014). Furthermore, integrins are used as cellular receptors by a diverse range of viruses (Hussein et al., 2015). Lastly, the two pervasive variants in ASHH2, a deletion and a substitution, suggests that methylation may be of particular importance in viral resistance mechanisms. Genes encoding for similar methylating proteins were differentially expressed in a species of grape during viral infection (Aquea et al., 2011).

Table 4.4. Pervasive PSNS Variants in M2 descendant cell lines.

Variant	Position	Gene Name	COG Category	Genomes Affected	Polymorphism Type	Host Virus Cross
- GGTGC CGAGG G	chr9 37601	Transmembra ne and coiled-coil domain-containing protein 4 (TMCO4)	S	M2V2b; M2V2c; M2V3b; M2V3c; M2V4a; M2V4c	Deletion	M2V2; M2V3; M2V4
(G)7 -> (G)8	chr9 29472	hypothetical protein - in Lipase 3 Protein Family	G	M2V2c; M2V4a; M2V4b	Insertion (tandem repeat)	M2V2; M2V4
A -> G	chr9 1026624	hypothetical protein - contains FG-GAP repeat region	S	M2V2a; M2V2b; M2V4c	SNP (transition)	M2V2; M2V4
C -> G	chr8 709397	Putative calcium-transporting ATPase 11, plasma membrane-type (ACA11)	P	M2V2a; M2V2b; M2V4c	SNP (transversion)	M2V2; M2V4
-C	chr6 883245	Histone-lysine N-methyltransfe rase ASHH2 (ASHH2)	U	M2V2a; M2V2b; M2V4c	Deletion	M2V2; M2V4
CCCG -> GGTT	chr6 883240	Histone-lysine N-methyltransfe rase ASHH2 (ASHH2)	U	M2V2a; M2V2b; M2V4c	Substitution	M2V2; M2V4
G -> A	chr1 1484544	Protein FORGETTER 1 (FGT1)	K, T	M2V2a; M2V2b; M2V4c	SNP (transition)	M2V2; M2V4

Results from M1 cell lines were less straightforward. There were a large number of pervasive variants (26), occurring among a relatively small sample size of six cell lines, which makes it challenging to associate particular variants with the resistance phenotype (Table 4.5). However, it is noteworthy that there is no overlap between genes in the M1 and M2 PSNS variant lists, which may be evidence that these closely related host strains achieved resistance through disparate mechanisms.

It is notable that two variants occurred in three out of the four M1 resistant cell lines. Both variants are insertions of tandem repeats, the first of which occurs in the cell division cycle protein 27 homolog B (CDC27B) gene, which is involved in mitotic regulation. Lack of CDC27B expression is associated with pathogen defense in *Nicotiana* plants (Kudo et al., 2007). The second variant occurs in a hypothetical protein in the integrin beta subunit protein family, which is often associated with cell signaling pathways, and as discussed previously, integrins are often used at virus receptors. Aside from possibly acting as receptors for viruses, there may be additional relevant functions of this gene. Two of the top five NCBI BLASTn hits for this gene sequence are carbonic anhydrase genes in the diatoms *Conticribra weissflogii* (53% pairwise identity, e-value 4e-118) and *Thalassiosira oceanica* (57% pairwise identity, e-value 9e-65). The carbonic anhydrase enzyme contributes to pH homeostasis and additionally allows for the uptake of dissolved inorganic carbon via the carbon concentration mechanism in *T.* weissflogii, allowing this diatom to continue photosynthesis in low CO2 conditions. Perhaps inorganic carbon plays an important role in homeostatic pathways associated with viral resistance among our cell lines, as viral infection can disrupt cellular processes even in the absence of lysis (Alterio et al., 2015).

There are twenty-four variants in the M1 dataset that only occur in two genomes. The genes in this dataset do not appear to follow any overarching patterns, as some are tied to membrane trafficking (e.g., QKY), DNA transcription (e.g., SBNO1), or cellular metabolism (e.g., APT1). Of note, none of the variants in the M1 dataset occur in either of the hypervariable outlier chromosomes.

Table 4.5. Pervasive PSNS Variants in M1 descendant cell lines.

Variant	Position	Gene Name	COG Category	Genomes Affected	Polymorphism Type	Host Virus Cross
(G)7 -> (G)8	Chr3 4121190	Cell division cycle protein 27 homolog B (CDC27B)	D (Cell cycle control and mitosis)	M1V1a; M1V2b; M1V2c	Insertion (tandem repeat)	M1V1; M1V2
(C)7 -> (C)8	Chr9 13472815	hypothetical protein - Integrin Beta Subunit Protein Family	-	M1V1a; M1V2a; M1V2b	Insertion (tandem repeat)	M1V1; M1V2
G -> C	Chr2 2228921	Argininosuccinate lyase (ARG7)	E (Amino acid metabolism)	M1V2a; M1V2c	SNP (transversion)	M1V2
G -> A	Chr2 3104185	Protein QUIRKY (QKY)	N (Cell motility), A (RNA processing and modification)	M1V2a; M1V2c	SNP (transition)	M1V2
G -> T	Chr2 3875044	Protein ABERRANT POLLEN TRANSMISSION 1 (APT1)	E (Amino Acid metabolism and transport)	M1V2a; M1V2c	SNP (transversion)	M1V2
G -> A	Chr2 3941544	Polyprotein of EF-Ts, chloroplastic (PETs)	J (Translation)	M1V2a; M1V2c	SNP (transition)	M1V2
G -> C	Chr3 4923948	Guanine nucleotide exchange factor SPIKE 1 (SPK1)	T (Signal Transduction)	M1V2a; M1V2c	SNP (transversion)	M1V2
C -> G	Chr6 8907405	Protein strawberry notch homolog 1 (Sbno1)	K (Transcription) , T (Signal Transduction)	M1V2a; M1V2c	SNP (transversion)	M1V2
A -> G	Chr6 9403926	AP-2 complex subunit alpha-2 (ALPHAC-AD)	U (Intracellular trafficking and secretion)	M1V2a; M1V2c	SNP (transition)	M1V2

(G)6 -> (G)7	Chr6 9598691	hypothetical protein - Integrin Beta Subunit Family	- Phyre2: Transferase	M1V1a; M1V2a	Insertion (tandem repeat)	M1V1; M1V2
(G)6 -> (G)5	Chr4 6042953	Exosome component 10 (EXOSC10)	J (Translation)	M1V1a; M1V2a	Deletion (tandem repeat)	M1V1; M1V2
-CCTG	Chr4 6325625	E3 ubiquitin-protein ligase Ufd4 (Ufd4)	O (Post-translational modification, protein turnover, chaperone functions)	M1V2a; M1V2c	Deletion	M1V2
(C)7 -> (C)8	Chr4 7042109	Regulator of nonsense transcripts UPF2 (UPF2)	A (RNA processing and modification)	M1V1a; M1V2a	Insertion (tandem repeat)	M1V1; M1V2
(C)8 -> (C)9	Chr8 11838302	Pentatricopeptide repeat-containing protein At5g55840 (At5g55840)	B (Chromatin Structure and dynamics)	M1V1a; M1V2a	Insertion (tandem repeat)	M1V1; M1V2
(G)5 -> (G)6	Chr8 12336093	Sperm flagellar protein 2 (SPEF2)	S (Function Unknown)	M1V1a; M1V2a	Insertion (tandem repeat)	M1V1; M1V2
C -> T	Chr9 12440294	U-box domain-containing protein 1 (PUB1)	-	M1V2a; M1V2c	SNP (transition)	M1V2
G -> C	Chr9 12476697	hypothetical protein FL13_004203	N (Cell motility), A (RNA processing and modification) Isomerase	M1V2a; M1V2b	SNP (transversion)	M1V2
(G)7 -> (G)8	Chr10 14076228	Nuclear pore complex protein NUP54 (NUP54)	U (Intracellular trafficking and secretion), Y (Nuclear structure)	M1V2a; M1V2c	Insertion (tandem repeat)	M1V2

(G)5 -> (G)6	Chr11 14972117	WD repeat-containing protein 20 (WDR20)	S (Function Unknown)	M1V2a; M1V2b	Insertion (tandem repeat)	M1V2
- CAGCG CAT	Chr11 15238246	hypothetical protein FL13_005489	N (Cell motility), A (RNA processing and modification) Phyre2: Isomerase	M1V2a; M1V2c	Deletion	M1V2
(GA)2 -> (GA)3	Chr12 16471012	Partial mRNA	A (RNA processing and modification)	M1V2a; M1V2c	Insertion (tandem repeat)	M1V2
C -> T	Chr14 18445304	Putative RNA-binding protein Luc7-like 1 (Luc7l)	A (RNA processing and modification)	M1V2a; M1V2c	SNP (transition)	M1V2
G -> C	Chr14 18587869	Carotene epsilon-monooxygenase, chloroplastic (CYP97C1)	Q (Secondary Structure)	M1V2a; M1V2c	SNP (transversion)	M1V2
G -> C	Chr14 18870947	Dynein axonemal heavy chain 7 (DNAH7)	Z (Cytoskeleton)	M1V2a; M1V2c	SNP (transversion)	M1V2
(G)7 -> (G)8	Chr7 10862216	Mitogen-activated protein kinase 4b (MPK4b)	T (Signal Transduction)	M1V2b; M1V2c	Insertion (tandem repeat)	M1V2
TC -> GG	Chr5 8673474	Leucine-rich repeat-containing protein 74A (Lrrc74a)	S (Function Unknown)	M1V2a; M1V2b	Substitution	M1V2
T -> C	Chr13 17966288	Cilia- and flagella-associated protein 74 (CFAP74)	Z (Cytoskeleton)	M1V2a; M1V2c	SNP (transition)	M1V2

4.4.3 Structural variants

Eight high frequency structural variants, in which >80% of reads supported the variant, were detected by the PBSV algorithm in our six PacBio-derived, mapped assemblies of M1 descendant genomes (Table 4.6). Six out of the eight variants occur in the outlier chromosomes (Chr1, the big outlier chromosome [BOC], and Chr17, the small outlier chromosome [SOC]) in resistant cell lines. Five of these take the form of deletions in chromosome 1. Two variants, a ~1.2 Mbp deletion in the BOC and a ~100 kbp inversion in the SOC, occur together in two genomes, M1V1a and M1V3a. The only high frequency structural variant in our susceptible cell line is a translocation in a non-outlier chromosome that was not present in any of the resistant cell lines. These observations are consistent with an association between structural variants in outlier chromosomes and viral resistance, but the inclusion of only one susceptible line, which has its own structural variant in a non-outlier chromosome, makes it difficult to draw firm conclusions.

Table 4.6. High frequency structural variant calls in M1 descendant genomes. Breakends are listed where a breakpoint is undetected (i.e. an imprecise structural variant call). “+”/”-” indicates the presence/absence of a putative structural variant.

Breakend/ Breakpoint	Variant Type	M1Sa	M1V1a	M1V1b	M1V2b	M1V2c	M1V3a
chr1 1,396,761; chr1 139,679	Deletion	-	+	-	-	-	+
chr17 199,518; chr17 37,167	Inversion	-	+	-	-	-	+
chr1 419,098; chr1 419,609	Deletion	-	-	-	-	+	-
chr1 704,527; chr1 704,739	Deletion	-	-	+	-	-	-
chr1 704,833; chr1 704,943	Deletion	-	+	-	-	-	-
chr1 1,496,912; chr1 1,497,045	Deletion	-	-	-	-	+	-
chr3 927,716	Insertion	-	-	-	-	+	-
chr10 749,897; chr11 415,047; chr10 750,027; chr11 415,027	Translocation	+	-	-	-	-	-

Seven of the eight structural variants detected occur in intergenic regions. However, a ~160 kbp inversion in the small outlier chromosome (SOC) found in two resistant genomes (M1V1a, M1V3a) was detected in a region containing 19 predicted genes (highlighted in light blue in Table 4.7). For context, the SOC is a ~200 kbp chromosome containing 27 predicted genes, 14 of which are hypothetical. Genes in this chromosome with functional annotations appear to largely involve carbohydrate metabolism, cell wall biogenesis, and post-translational modification. However, we were unable to assign any functional categorization to 15 of the genes within the region of the inversion. None of the 19 genes in the region of the inversion have small-scale variants discussed in previous sections. This may indicate that small scale mutations and structural variants could work in parallel to confer resistance. Furthermore, structural variants may affect gene expression in nearby genomic regions even if the coding sequences of those genes are unaffected.

Table 4.7. Genes in M1 chromosome 17, the small outlier chromosome. Genes within the region of a major 160 kbp inversion are highlighted in light blue.

Position	Gene Name	COG Category
Chr17 7340	hypothetical protein FL13_008944	-
Chr17 14870	Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase (GNE)	G (Carbohydrate Metabolism and Transport), K (Transcription)
Chr17 19358	CMP-sialic acid transporter 1 (At5g41760)	G (Carbohydrate Metabolism and Transport)
Chr17 29550	Glutamine-dependent NAD (Os07g0167100)	H (Coenzyme metabolism)
Chr17 32477	Uncharacterized protein MJ1065 (MJ1065)	M (Cell wall/membrane/envelope biogenesis)
Chr17 34372	hypothetical protein FL13_008949	T (Signal Transduction)
Chr17 37730	hypothetical protein FL13_008950	-
Chr17 43175	hypothetical protein FL13_008951	-
Chr17 46626	hypothetical protein FL13_008952	-
Chr17 83322	SURP and G-patch domain-containing protein 1-like protein (At3g52120)	O (Post-translational modification, protein

		turnover, chaperone functions)
Chr17 93752	Putative DNA (FV3-083R)	-
Chr17 95283	Putative DNA (FV3-083R)	-
Chr17 104017	hypothetical protein FL13_008955	-
Chr17 112380	hypothetical protein FL13_008956	-
Chr17 114969	hypothetical protein FL13_008957	-
Chr17 127557	hypothetical protein FL13_008958	S (Function Unknown)
Chr17 136787	Probable BsuMI modification methylase subunit YdiP (ydiP)	S (Function Unknown)
Chr17 145589	hypothetical protein FL13_008960	-
Chr17 177314	hypothetical protein FL13_008961	S (Function Unknown)
Chr17 182002	hypothetical protein FL13_008962	-
Chr17 184123	GDP-mannose 3,5-epimerase (At5g28840)	G (Carbohydrate Metabolism and Transport), M (Cell wall/membrane/envelope biogenesis)
Chr17 184690	GDP-mannose 3,5-epimerase (At5g28840)	G (Carbohydrate Metabolism and Transport), M (Cell wall/membrane/envelope biogenesis)
Chr17 194291	hypothetical protein FL13_008964	S (Function Unknown)
Chr17 195682	Putative DNA (FV3-083R)	S (Function Unknown)
Chr17 199029	GDP-mannose 6-dehydrogenase (algD)	M (Cell wall/membrane/envelope biogenesis)
Chr17 206607	GDP-mannose 3,5-epimerase (At5g28840)	G (Carbohydrate Metabolism and Transport), M (Cell wall/membrane/envelope biogenesis)
Chr17 208819	hypothetical protein FL13_008968	-

4.5 CONCLUSIONS

Constraints to sample size makes it challenging to derive clear overarching conclusions regarding genetic signatures of resistance in *Micromonas*. However, there are several lines of evidence suggesting that some of the variants we have identified are worthy of further study regarding their role in infection and resistance.

4.5.1 Characterization of small-scale variants

Hundreds of small-scale variants resulted from our analysis of 20 Illumina-derived genomes, but a much smaller number of variants corresponded to phenotype-specific nonsynonymous changes that occurred in more than one resistant cell line (7 variants in M2 lines and 26 in M1 lines). The ten M2-derived cell lines provided the most promising results, and the seven pervasive PSNS variants associated with resistance, occurred in genes with reasonable putative ties to viral infection and/or defense. The most pervasive variant was found in TMCO4, a transmembrane protein that could function in viral attachment/entry or intracellular trafficking. This gene has also been implicated in a study of host response to avian influenza virus. Further study is needed to explore this gene's involvement in viral infection in phytoplankton via gene knockout and/or transcriptomic studies. Aside from TMCO4, a suite of pervasive genes in the M2 dataset were also tied to putative host defense function, including changes to stress response (FGT1), DNA methylation (ASHH2), and potential viral receptors (hypothetical protein in Lipase class 3, and a hypothetical alpha integrin).

The pervasive PSNS variants in the M1 dataset were more challenging to interpret. Two pervasive variants occurred in three of the four resistant M1 cell lines, one in the CDC27B gene and one in a gene for a hypothetical beta integrin. Both involve protein families that have some evidence of interacting with viruses and may play a role in defense through changes in the cell cycle, a strategy seen in *E. huxleyi*, or changes to an integrin which may affect viral attachment (Frada et al., 2008; Hussein et al., 2015). The rest of the 22 pervasive variants in the M1 dataset have varying function and it is difficult to discern an overarching pattern in this group of genes. However, this set of genes can serve as a starting point in further investigations regarding genetic signatures of resistance in protists.

Hypervariable regions of genomes have been implicated in viral resistance in marine microorganisms (Avrani et al., 2011; Yau et al., 2016). It is noteworthy that only one pervasive PSNS variant, in the M2 dataset, occurred in a gene in an outlier chromosome. This is inconsistent with the results of Avrani et al.'s study with *Prochlorococcus*, in which the majority of resistance-associated variants are concentrated in hypervariable regions of the genome. However, we did not investigate whether there are additional hypervariable regions outside of the two previously identified outlier chromosomes, which would be an important subject for future studies.

4.5.2 Preliminary insights into structural variants

Despite our small sample size of six cell lines sequenced with PacBio, it is interesting to see that structural variants were consistently called within the outlier chromosomes of resistant cell lines. Outlier chromosomes are hypervariable by definition, suggesting elevated mutation rates in these regions, and therefore SV occurrence may be frequent in these chromosomes regardless of a selection pressure such as viral infection. Additionally, the lack of SV detection in chromosome 1 or 17 within the genome of the susceptible cell line is consistent with previous work showing that resistance is tied to outlier chromosome structure in a related prasinophyte-virus system. Specifically, changes to the small outlier chromosome were found by Yau et al. to be related to resistance. These researchers found that the SOC changed significantly in size depending on the susceptibility of *Ostreococcus mediterraneus* cell lines. Sequencing data revealed a ~50 kb deletion of the SOC in a *O. mediterraneus* cell line that recently switched from a resistant to a susceptible phenotype. The region of deletion was largely made up of repeats, containing seven genes of mixed functional attributes that were differentially expressed compared to cell lines with other phenotypic attributes.

Additional methods would bolster structural variant detection in *Micromonas*, which may include optical mapping and pulsed field gel electrophoresis and could provide further data on SVs in hard to detect regions, such as those with a large number of repeats. Additionally, a transcriptomic approach would provide insights into whether SVs affect gene expression.

4.5.3 Future directions

Future studies should focus on sequencing cell lines at various points before and after isolation of resistant phenotypes. Previous work suggests that genetic signatures of resistance can change over time (Avrani and Lindell, 2015). A transcriptomic approach alongside comparative genomics will further bring potential mechanisms of resistance into focus. Variant calling at both small and large scales should continue, as both small and structural variants appear to be associated with resistance to viral immunity.

4.6 REFERENCES

Alterio, V., Langella, E., De Simone, G., and Monti, S. (2015) Cadmium-Containing Carbonic Anhydrase CDCA1 in Marine Diatom Thalassiosira weissflogii. Marine Drugs 13: 1688–1697.

Aquea, F., Vega, A., Timmermann, T., Poupin, M.J., and Arce-Johnson, P. (2011) Genome-wide analysis of the SET DOMAIN GROUP family in Grapevine. Plant Cell Rep 30: 1087–1097.

Avrani, S. and Lindell, D. (2015) Convergent evolution toward an improved growth rate and a reduced resistance range in *Prochlorococcus* strains resistant to phage. Proc Natl Acad Sci USA 112.

Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D. (2011) Genomic island variability facilitates *Prochlorococcus*–virus coexistence. Nature 474: 604–608.

Bidle, K.D. (2016) Programmed Cell Death in Unicellular Phytoplankton. Current Biology 26: R594–R607.

Blanc-Mathieu, R., Krasovec, M., Hebrard, M., Yau, S., Desgranges, E., Martin, J., et al. (2017) Population genomics of picophytoplankton unveils novel chromosome hypervariability. Sci Adv 3: e1700239.

Bohannan, B.J.M. and Lenski, R.E. (2000) Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol Letters 3: 362–377.

Bohannan, B.J.M., Travisano, M., and Lenski, R.E. (1999) Epistatic Interactions Can Lower the Cost of Resistance to Multiple Consumers. Evolution 53: 292.

Breitbart, M. (2012) Marine Viruses: Truth or Dare. Annu Rev Mar Sci 4: 425–448.

Brzezinka, K., Altmann, S., Czesnick, H., Nicolas, P., Gorka, M., Benke, E., et al. (2016) Arabidopsis FORGETTER1 mediates stress-induced chromatin memory through nucleosome remodeling. eLife 5: e17061.

Bushnell, Brian (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner.

Chaisson, M.J. and Tesler, G. (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13: 238.

Core Team, R. (2022) R: A language and environment for statistical computing.

Cottrell, M.T. and Suttle, C.A. (1995) Genetic Diversity of Algal Viruses Which Lyse the Photosynthetic Picoflagellate *Micromonas pusilla* (Prasinophyceae). Appl Environ Microbiol 61: 3088–3091.

Demory, D., Baudoux, A.-C., Monier, A., Simon, N., Six, C., Ge, P., et al. (2018) Picoeukaryotes of the *Micromonas* genus: sentinels of a warming ocean. ISME J 13: 132–146.

Finke, J., Winget, D., Chan, A., and Suttle, C. (2017) Variation in the Genetic Repertoire of Viruses Infecting *Micromonas pusilla* Reflects Horizontal Gene Transfer and Links to Their Environmental Distribution. Viruses 9: 116.

Frada, M., Probert, I., Allen, M.J., Wilson, W.H., and de Vargas, C. (2008) The “Cheshire Cat” escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. Proc Natl Acad Sci USA 105: 15944–15949.

Guillard, R.R.L. (1975) Culture of Phytoplankton for Feeding Marine Invertebrates. In Culture of Marine Invertebrate Animals. Smith, W.L. and Chanley, M.H. (eds). Boston, MA: Springer US, pp. 29–60.

Guillard, R.R.L. and Ryther, J.H. (1962) STUDIES OF MARINE PLANKTONIC DIATOMS: I. CYCLOTELLA NANA HUSTEDT, AND DETONULA CONFERVACEA (CLEVE) GRAN. Can J Microbiol 8: 229–239.

Ha, A.D., Moniruzzaman, M., and Aylward, F.O. (2023) Assessing the biogeography of marine giant viruses in four oceanic transects. ISME COMMUN 3: 43.

Hauser, Michael, A., Ashley-Koch, Allison E, Qin, Xuejun, Strickland, Shelby, Liu, Yutao, Girkin, Christopher A., et al. (2014) Rare Genetic Variants are Associated with POAG in Populations of African Ancestry.

Hussein, H.A.M., Walker, L.R., Abdel-Raouf, U.M., Desouky, S.A., Montasser, A.K.M., and Akula, S.M. (2015) Beyond RGD: virus interactions with integrins. Arch Virol 160: 2669–2681.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. Genome Res 27: 722–736.

Kudo, C., Suzuki, T., Fukuoka, S., Asai, S., Suenaga, H., Sasabe, M., et al. (2007) Suppression of Cdc27B expression induces plant defence responses. Mol Plant Pathol 8: 365–373.

Lassen, K.G., McKenzie, C.I., Mari, M., Murano, T., Begun, J., Baxt, L.A., et al. (2016) Genetic Coding Variant in GPR65 Alters Lysosomal pH and Links Lysosomal Dysfunction with Colitis Risk. Immunity 44: 1392–1405.

Lennon, J.T., Khatana, S.A.M., Marston, M.F., and Martiny, J.B.H. (2007) Is there a cost of virus resistance in marine cyanobacteria? ISME J 1: 300–312.

Mayer, J.A. and Taylor, F.J.R. (1979) A virus which lyses the marine nanoflagellate *Micromonas pusilla*. Nature 281: 3.

Not, F., Latasa, M., Marie, D., Cariou, T., Vaulot, D., and Simon, N. (2004) A Single Species, *Micromonas pusilla* (Prasinophyceae), Dominates the Eukaryotic Picoplankton in the Western English Channel. Appl Environ Microbiol 70: 4064–4072.

Oksanen, J, Simpson, G, Blanchet, F, Kindt, R, Legendre, P, Minchin, P, et al. (2022) vegan: Community Ecology Package.

Price, C.A., Reardon, E.M., and Guillard, R.R.L. (1978) Collection of dinoflagellates and other marine microalgae by centrifugation in density gradients of a modified silica sol 1. Limnol Oceanogr 23: 548–553.

Schleyer, G., Kuhlisch, C., Ziv, C., Ben-Dor, S., Malitsky, S., Schatz, D., and Vardi, A. (2023) Lipid biomarkers for algal resistance to viral infection in the ocean. Proc Natl Acad Sci USA 120: e2217121120.

Schvarcz, C.R. (2018) Cultivation and Characterization of Viruses Infecting Eukaryotic Phytoplankton from the Tropical North Pacific Ocean.

Sirchia, R. and Luparello, C. (2007) Mid-region parathyroid hormone-related protein (PTHrP) and gene expression of MDA-MB231 breast cancer cells. bchm 388: 457–465.

Sun, Z., Li, S., Li, F., and Xiang, J. (2014) Bioinformatic Prediction of WSSV-Host Protein-Protein Interaction. BioMed Research International 2014: 1–9.

Suttle, C.A. (2007) Marine viruses — major players in the global ecosystem. Nat Rev Microbiol 5: 801–812.

Taguchi, F., Suzuki, T., Takeuchi, K., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2009) Glycosylation of flagellin from Pseudomonas syringae pv. tabaci 6605 contributes to evasion of host tobacco plant surveillance system. Physiological and Molecular Plant Pathology 74: 11–17.

Thingstad, T.F. (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45: 1320–1328.

Thomas, R., Grimsley, N., Escande, M., Subirana, L., Derelle, E., and Moreau, H. (2011) Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations: Viral resistance in Mamiellales. Environmental Microbiology 13: 1412–1420.

Thyrhaug, R., Larsen, A., Thingstad, T., and Bratbak, G. (2003) Stable coexistence in marine algal host-virus systems. Mar Ecol Prog Ser 254: 27–35.

Worden, A.Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M.P., Aerts, A.L., et al. (2009) Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes *Micromonas*. Science 324: 268–272.

Xu, X.-J., Geng, C., Jiang, S.-Y., Zhu, Q., Yan, Z.-Y., Tian, Y.-P., and Li, X.-D. (2022) A maize triacylglycerol lipase inhibits sugarcane mosaic virus infection. Plant Physiology 189: 754–771.

Yau, S., Hemon, C., Derelle, E., Moreau, H., Piganeau, G., and Grimsley, N. (2016) A Viral Immunity Chromosome in the Marine Picoeukaryote, *Ostreococcus tauri*. PLoS Pathog 12: e1005965.

Yau, S., Krasovec, M., Benites, L.F., Rombauts, S., Groussin, M., Vancaester, E., et al. (2020) Virus-host coexistence in phytoplankton through the genomic lens. Sci Adv 6: eaay2587.

Chapter 5: Conclusions

This dissertation is a story in three parts. First, we introduced four virus strains with novel qualities not previously seen in their clade. Next, we saw how selection against lytic infection from these virus strains impacted the fitness of their hosts. Then, we examined how the selection for specific phenotypes of these hosts potentially influenced host genetics. While one can appreciate the cleanness of a triptych, there is much work to be done in all dimensions surrounding this work.

5.1 CHAPTER 2

A major conclusion from this chapter is that virus strains isolated from the same environment, with overlapping host ranges, can exhibit a surprising amount of diversity. Among the four Hawaiian *Micromonas commoda* virus strains (HiMcVs), three separate phylogenetic analyses resulted in three HiMcVs strains grouping together within a clade containing all other published *Micromonas* viruses, while the fourth appeared at quite a distance away from other *Micromonas* viruses. The arguably singular HiMcV, McV-KB2, would continue to be the odd one out in whole genome alignments and gene content comparisons. Furthermore, within the three HiMcVs that grouped together, there were still dozens of genes that were possessed by only one or two of the three viruses. The genes in this pan-genome have the potential to better our understanding of adaptation to seemingly small environmental variation. Unfortunately, characterization of the function of the genes of *Micromonas* viruses, and marine viruses more broadly, are still lacking. I believe systematically evaluating hypothetical proteins is the next frontier of marine viral ecology, further emphasizing the need for isolates accessible in culture. My prediction is that the majority of these putative genes are involved in host recognition and may be involved in viral attachment and entry. Missing from this chapter is the mapping of orthogroups in the HiMcV genomes. Perhaps unique orthogroups occur in hypervariable regions such that virus strains can quickly evolve to overcome host defenses.

Forty-eight orthogroups among the HiMcVs can be found in the cellular genomes of *Micromonas* isolated from Hawaiian waters, labeled M1 and M2 throughout this dissertation. Functional and structural annotations for these orthogroups are more abundant, highlighting the disparity between cellular and viral gene characterization. Among these host-associated orthogroups are genes involved with cellular metabolism,

stress response, and host defense against viral infection. It would be interesting to compare the infection dynamics and gene content of our HiMcVs and their hosts, both isolated from oligotrophic regions, to prasinoviruses and hosts from highly productive areas, because nutrient availability may affect host-virus coevolution. If *Micromonas* viruses from different environments partially overlap in host range, then host range comparisons and competition experiments, in which virus strains are pitted against each other to infect hosts from different environments, would provide important insight into the genomic basis for cross-infection networks. A major conclusion from Chapter 2 is that approximately a quarter of prasinovirus gene content is correlated with host genus, reflecting coevolutionary patterns of marine protists and their associated viruses.

The last portion of Chapter 2 is dedicated to the search for sequences similar to HiMcVs in global metagenomes. We found that our HiMcV strain isolated from the pelagic waters, McV-SA1, was found throughout the Atlantic and at Station ALOHA. The metagenomes used in our search are largely from oligotrophic waters and it is logical to expect our pelagic strain to have greater presence in these datasets. Two strains, McV-KB3 and the singular McV-KB2, were also found, but only in the productive region of the Atlantic coast of South Africa. Oddly, McV-KB4, which is highly similar to McV-SA1 and McV-KB3 based on phylogenetic analysis, gene content, and whole genome alignment, was not found by using CoverM software to search through published metagenomes. This could mean that even small genetic differences can have a large impact on viral selection in different oceanic regions and/or that rare viruses are difficult to sample. Metagenomic sampling of HiMcVs could be expanded in two ways, first by sampling in a wider range of environments with different nutrient profiles, as well as incorporating methodology specifically targeting larger viruses. In particular, the GEOTRACES metagenomes were generated from 100 mL of sea water samples from each station. This is an astonishingly small volume of water for virus work, even for amplicon sequencing. My predecessor and colleague, Dr. Christopher R. Schvarcz, used 20 to 400 L of sea water to obtain viable isolates of viruses through tangential flow filtration. While the methodologies for procuring metagenomic data and isolates are not completely comparable, the scale at which Dr. Schvarcz conducted his work should

provide an indicator as to what volumes may be more appropriate for marine eukaryotic viruses.

Expansion of metagenomic datasets should also be accompanied by the continued isolation and sequencing of virus strains. This chapter shows that even a small sample size of four viruses provides insight into not only gene diversity, but biogeography of the viruses of prasinophytes, which are ecologically relevant primary producers. While the isolation of host-virus systems is a non-trivial task, my predecessor's work shows that it is possible.

5.2 CHAPTER 3

Chapter 3 examined the effects of HiMcVs on the fitness of Hawaiian *Micromonas commoda* strains. The relatively large number of resistant strains, resulting from six combinations of host and virus strain challenge experiments, resulted in a robust dataset with clear patterns of fitness under different conditions. Prior to our work, researchers were only able to detect fitness costs associated with resistance in marine protists on an inconsistent basis. The clear patterns of cost of resistance found in my work make this chapter the most fulfilling experiment that I have conducted in my scientific career.

The identities of the ancestral host strain and the virus strain used to select for resistant cell lines played a key role in the magnitude of fitness costs. Resistant descendants of M1 cell lines experienced the largest depression in growth rates, with cell lines resulting from the M1V1 potentially going extinct because of this cost. The connection between ancestral strain identity and the cost of resistance could be further explored in the future using the rest of the algal isolates in the UHM culture collection.

For me, however, the most predominant conclusion of this chapter is how the magnitude of fitness costs attenuates under lower light. As a follow-up, I attempted to isolate *Micromonas* cell and virus strains from different depths at Station ALOHA in 2019. For context, all of the UHM *Micromonas* strains are from high-light surface waters. Prasinophyte ecotypes correlated to depth have been reported, with a specific study showing that low-light *Micromonas* strains contain the novel pigment Chl $C_{CS\text{-}170}$ (Jeffery 1989). I wanted to see if specific growth patterns would emerge among *Micromonas* strains from different depths at Station ALOHA based on the results of my

growth experiments and the knowledge that different prasinophytes exist at different depth strata. Unfortunately, my isolation efforts were hampered by a steep learning curve associated with establishing clean algal cultures and, ultimately, ended in a mass extinction of putative isolates when an incubator overheated after a power outage. With more time and resources, I would relaunch this project to test if (1) resistance in low-light *Micromonas* strains came with different dynamics regarding fitness costs and (2) the characteristics of low-light strains were different based on pigment and genetic analysis. Metagenomic analysis may capture some of this population variability, but low-light strains from the base of the euphotic zone would need to be established as references.

There is still quite a bit of work to be done even without establishing new isolates. Growing resistant cell lines at more than two light levels would provide a better understanding of how irradiance relates to fitness cost. Furthermore, if, as alluded to in the conclusions of Chapter 3, nutrient uptake plays a role in the interaction of fitness cost and light, then nutrient deprivation experiments could provide corroborating data to support this hypothesis. To that end, I did attempt to grow *Micromonas* in phosphate- and nitrogen-poor media. These experiments failed as the picoeukaryotes continued to grow uninhibited by my attempts to deprive them of nutrients. Whether using filtered sea water or artificial sea water, I could not get media with a low enough nutrient content. Nutrient uptake kinetic experiments would potentially provide better data and obviate the issue.

Arguably, a more interesting exploration of cost of resistance would be direct competition between susceptible and resistant cell lines under different resource regimes. For a more ecologically relevant analysis, such a competition experiment would benefit from including more algal species as competitors. More holistically, the inclusion of zooplankton grazers and microfaunal filter feeders would provide meaningful context for the radiating effects of marine viruses on the larger ecosystem scale, while also increasing our understanding of the shaping of phytoplankton communities from different selection pressures.

As of now, the resistant cell lines I established did provide insights into the putative mechanisms of resistance in *Micromonas* and their potential impacts in the

marine ecosystem. Potential mechanisms of resistance were explored through comparative genomics in chapter 4.

5.3 CHAPTER 4

Admittedly, while this chapter did provide some interesting insights, it was the one with the least resolved conclusions. Our analysis was hampered by a small sample size, specifically for descendants of M1. For my small-scale variant analysis, I successfully mapped sequences of six M1 descendants to their ancestor's genome. Of those six mapped assemblies, two represented susceptible descendants and the three of the four resistant cell lines sequences resulted from the same host-virus combination (Table 4.1). In comparison, I was successful in mapping Illumina sequences of 11 M2 descendants, two of which were susceptible, and nine resistant cell lines resulting from three separate host-virus strain challenges. For large-scale structural analysis, I was able to successfully sequence six M1 cell lines, only one of which was susceptible (Table 4.2). With a larger and more diverse data set, the M2 Illumina data provided the most clarity in putative effects of viral resistance on the genomes of *Micromonas* strains.

Despite issues associated with a small sample size, there were promising results that provided a path forward in shaping future experiments that could better characterize genetic effects of viral resistance. Nonsynonymous variants found in only resistant genomes were observed throughout both M1 and M2 descendants. Interestingly, the genes that these variants affected were quite different between the two data sets, reflecting observations in Chapter 3 in which M1 and M2 descendants experienced dissimilar fitness costs. Virus strains are hyper-specific to host strains, often only infecting only a handful of genotypes from the same species. It would then stand to reason that different hosts would require unique genetic strategies in achieving resistance to each virus. Three M1 descendants and three M2 descendants were resistant to the same virus, V2, with no shared variants between these two groups of descendants, further emphasizing the importance of host identity in resistance response and, in a more holistic view, the survival of individual viral strains.

5.4 SYNTHESIS

In Chapter 2, we examined genomic sequences of four *Micromonas* virus strains. A tidy story would conclude with a clear link between the gene content of these four strains and the genes with resistance variants in M1 and M2 cell lines. Drawing such connections is necessarily speculative at this point, but there are common variants in resistant cell genomes that are noteworthy for possible ties to viral processes indicated by HiMcV gene annotations.

One pervasive PSNS variant in M2 lines occurred in the FGT1 gene, which is known to promote heat stress-induced gene expression in *Arabidopsis* (Brzezinka *et al.*, 2016). The HiMcV genomes contain heat-shock protein 70 (Hsp70), which has a direct ortholog in M1 and M2, and which is commonly found in other members of virus phylum *Nucleocytoviricota* (Ha *et al.*, 2021). Hsp70 maintains protein function under stressful conditions and is also a direct suppressor of apoptosis (Kennedy *et al.*, 2014), and this protein can both inhibit and enhance viral replication in various viruses (Manzoor *et al.*, 2014). The occurrence of a heat shock protein gene in HiMcV annotations, along with a variant in a gene that modulates responses to heat stress in resistance cell lines, could be evidence of the importance of stress response mechanisms in host-virus interactions and coevolution in *Micromonas*-virus systems.

Expanding this hypothesis, there are additional genes among the resistant host variants and the HiMcV genomes that are involved in cellular stress response. M2 has a pervasive PSNS variant in a putative calcium-transporting ATPase, ACA11, which is a gene implicated in programmed cell death (Ren *et al.*, 2021). In a cell unaffected by viral attempts at gene suppression, programmed cell death could be a response to infection that prevents additional members of the host population from becoming infected (Verburg *et al.*, 2022). Viruses may inhibit this cellular process in order to keep hosts alive long enough to fully exploit them for virus reproduction. Another pervasive variant occurs in the Histone-lysine N-methyltransferase gene ASHH2. In *Arabidopsis thaliana*, ASHH2, along with other histone methyltransferases, have been associated with stress responses related to infection by *Pseudomonas* bacteria (De-La-Peña *et al.*, 2012; Nunez-Vazquez *et al.*, 2022). Lastly, the pervasive PSNS variant found in the largest number of resistant *Micromonas* genomes occurs within the transmembrane and coiled-

coil domain-containing protein 4 (TMCO4) gene. This gene is involved in stress response and immunity cascades in humans (Sirchia and Luparello, 2007; Hauser et al., 2014; Lassen et al., 2016). Mutations in a homologous gene in yeast, MIL1, increases yeast sensitivity to the antidepressant sertraline (Whitfield *et al.*, 2016), which may in turn trigger autophagy via overaccumulation of sertraline (Chen *et al.*, 2012), again implicating programmed cell death as a means of viral control.

In HiMcVs genomes, additional cell-derived stress response genes (beyond Hsp70) include bax inhibitor-1, rhodanese, superoxide dismutase, and mannitol dehydrogenase. The bax inhibitor, which has not been previously found in protistan viruses, is particularly noteworthy in this context. Bax-mediated apoptosis is an antiviral defense activated by various DNA and RNA viruses infecting humans and other mammals, and many viruses encode proteins that inhibit bax in order to replicate (Chattopadhyay *et al.*, 2011; Verburg *et al.*, 2022).

In addition to stress responses and programmed cell death, another key locus of host-virus interactions is viral attachment and entry. One pervasive PSNS resistance variant occurs in an M2 hypothetical protein that contains FG-GAP repeats (Table 4.4). FG-GAP repeats are found in extracellular structures and are closely associated with alpha-integrins, which are important for ligand binding, and which are associated with intercellular interaction, pathogen recognition, and immune response. Therefore, the host FG-GAP repeat gene with a resistance variant could be a viral receptor. FG-GAP repeats also occur in twelve HiMcV genes, including the major capsid protein, integrins, and intramolecular chaperones, all of which are likely involved with protein-protein interaction. It may be that genes encoding for proteins with FG-GAP repeats, such as integrins, co-evolve, as hosts respond to viral pressure by changing cell surface structure through mutations, and viruses overcome host defenses by changing attachment structures found on their capsids (Martiny *et al.*, 2014). The pervasive resistance variant in TMCO4 may also be involved in viral attachment or entry, as this is a transmembrane gene known to be involved in endocytosis in yeast (Attwood and Schiöth, 2020).

Revisiting the results of Chapter 3, in which depression of growth rates of resistant cell lines was exacerbated under high light, we hypothesized that the

mechanism of resistance in our *M. commoda* cell lines may be the modification of cell surface receptors. Our examination of the functional groups of genes with pervasive variants supports this hypothesis as a possibility. However, we also asserted that the modification of receptors could affect nutrient uptake, should said receptors be involved in nutrient transport. Unfortunately for our tidy story, the pervasive PSNS variants in Chapter 4 did not occur in genes encoding for known nutrient transporters. It seems more reasonable to speculate that a higher fitness cost under high light may results from a more active stress response system, given the suite of stress response genes affected by resistance and found among predicted viral genes. It is not clear why an energy-demanding stress response would be more costly under high irradiance, when energy is more plentiful, than under low irradiance, unless the stress response itself is only activated when sufficient photon energy is available. Alternatively, it may be that the stress response takes cellular energy away nutrient uptake, which would therefore put resistant cell lines at a larger disadvantage under high light, rather than at low light, because the demand for nutrient uptake is higher under high light.

5.5 DISCUSSION

A major takeaway from this work is the importance of studying well-characterized host and virus systems in the laboratory environment. However, in the future a significant effort must also be made to understand the functions of potentially important genes we have identified within these systems, in order to provide a more resolved image of how host-virus interactions and evolution work under different environmental conditions. Only then can we form better predictive models that can provide essential information about biogeochemical cycling in a changing global climate.

5.6 REFERENCES

Attwood, M.M. and Schiöth, H.B. (2020) Classification of Trispanins: A Diverse Group of Proteins That Function in Membrane Synthesis and Transport Mechanisms. *Front Cell Dev Biol* **7**: 386.

Brzezinka, K., Altmann, S., Czesnick, H., Nicolas, P., Gorka, M., Benke, E., et al. (2016) Arabidopsis FORGETTER1 mediates stress-induced chromatin memory through nucleosome remodeling. *eLife* **5**: e17061.

Chattopadhyay, S., Yamashita, M., Zhang, Y., and Sen, G.C. (2011) The IRF-3/Bax-Mediated Apoptotic Pathway, Activated by Viral Cytoplasmic RNA and DNA, Inhibits Virus Replication. *J Virol* **85**: 3708–3716.

Chen, J., Korostyshevsky, D., Lee, S., and Perlstein, E.O. (2012) Accumulation of an Antidepressant in Vesiculogenic Membranes of Yeast Cells Triggers Autophagy. *PLoS ONE* **7**: e34024.

De-La-Peña, C., Rangel-Cano, A., and Alvarez-Venegas, R. (2012) Regulation of disease-responsive genes mediated by epigenetic factors: interaction of Arabidopsis–Pseudomonas. *Molecular Plant Pathology* **13**: 388–398.

Ha, A.D., Moniruzzaman, M., and Aylward, F.O. (2021) High Transcriptional Activity and Diverse Functional Repertoires of Hundreds of Giant Viruses in a Coastal Marine System. *mSystems* **6**: e00293-21.

Hauser, Michael, A., Ashley-Koch, Allison E, Qin, Xuejun, Strickland, Shelby, Liu, Yutao, Girkin, Christopher A., et al. (2014) Rare Genetic Variants are Associated with POAG in Populations of African Ancestry.

Jeffrey, S. W. (1989) Chlorophyll *C* Pigments And Their Distribution In The Chromophyte Algae, in J C Green, B S C Leadbeater, and W L Diver (eds), *The Chromophyte Algae: Problems and Perspectives* (Oxford, 1990; online edn, Oxford Academic, 31 Oct. 2023)

Kennedy, D., Jäger, R., Mosser, D.D., and Samali, A. (2014) Regulation of apoptosis by heat shock proteins. *IUBMB Life* **66**: 327–338.

Lassen, K.G., McKenzie, C.I., Mari, M., Murano, T., Begun, J., Baxt, L.A., et al. (2016) Genetic Coding Variant in GPR65 Alters Lysosomal pH and Links Lysosomal Dysfunction with Colitis Risk. Immunity 44: 1392–1405.

Manzoor, R., Kuroda, K., Yoshida, R., Tsuda, Y., Fujikura, D., Miyamoto, H., et al. (2014) Heat Shock Protein 70 Modulates Influenza A Virus Polymerase Activity. *Journal of Biological Chemistry* **289**: 7599–7614.

Martiny, J.B.H., Riemann, L., Marston, M.F., and Middelboe, M. (2014) Antagonistic Coevolution of Marine Planktonic Viruses and Their Hosts. *Annu Rev Mar Sci* **6**: 393–414.

Nunez-Vazquez, R., Desvoyes, B., and Gutierrez, C. (2022) Histone variants and modifications during abiotic stress response. *Front Plant Sci* **13**: 984702.

Ren, H., Zhao, X., Li, W., Hussain, J., Qi, G., and Liu, S. (2021) Calcium Signaling in Plant Programmed Cell Death. *Cells* **10**: 1089.

Sirchia, R. and Luparello, C. (2007) Mid-region parathyroid hormone-related protein (PTHrP) and gene expression of MDA-MB231 breast cancer cells. bchm 388: 457–465.

Verburg, S.G., Lelievre, R.M., Westerveld, M.J., Inkol, J.M., Sun, Y.L., and Workenhe, S.T. (2022) Viral-mediated activation and inhibition of programmed cell death. *PLoS Pathog* **18**: e1010718.

Whitfield, S.T., Burston, H.E., Bean, B.D.M., Raghuram, N., Maldonado-Báez, L., Davey, M., et al. (2016) The alternate AP-1 adaptor subunit Apm2 interacts with the Mil1 regulatory protein and confers differential cargo sorting. *MBoC* **27**: 588–598.

Appendix

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

Supplementary Table S2.1. Antibiotic recipe used to clean *Micromonas* culture of bacteria and associated phage. This recipe was developed by colleagues at Observatoire océanologique de Banyuls-sur-Mer, France.

Antibiotic name	Final concentration ($\mu g\ mL^{-1}$)	Mass (g) in 1000X stock (10mL)
Ampicillin	50	0.5
Gentamicin	50	0.5
Kanamycin	20	0.2
Neomycin	100	1

Supplementary Table S2.2. Strain information for prasinovirus and chlorovirus strains used in OrthoFinder and phylogenetic analysis.

Accession	Strain Name	Strain Abbreviation	Host Genus	Authors
HM004430	*Bathycoccus* sp. RCC1105 virus	BpV2	*Bathycoccus*	Moreau et al. (2010)
HM004432	*Bathycoccus* sp. RCC1105 virus	BpV1	*Bathycoccus*	Moreau et al. (2010)
MK522034	*Bathycoccus* sp. RCC716 virus 1	BII-V1	*Bathycoccus*	Bachy et al. (2019)
MK522038	*Bathycoccus* sp. RCC716 virus 2	BII-V2	*Bathycoccus*	Bachy et al. (2019)
MK522039	*Bathycoccus* sp. RCC716 virus 3	BII-V3	*Bathycoccus*	Bachy et al. (2019)
HQ633072	*Micromonas pusilla* virus PL1	MpV-PL1	*Micromonas*	Henn et al. (2010)
JF974320	*Micromonas pusilla* virus SP1	MpV-SP1	*Micromonas*	Henn et al. (2010)
NC_014767	*Micromonas* sp. RCC1109 virus	MpV1	*Micromonas*	Moreau et al. (2010)
NC_020864	*Micromonas pusilla* virus 12T	MpV-12T	*Micromonas*	Henn et al. (2010)
HQ633059	*Ostreococcus lucimarinus* virus 6	OlV6	*Ostreococcus*	Henn et al. (2010)
HQ633060	*Ostreococcus lucimarinus* virus 3	OlV3	*Ostreococcus*	Henn et al. (2010)
JF974316	*Ostreococcus lucimarinus* virus 4	OlV4	*Ostreococcus*	Henn et al. (2010)

MK514405	*Ostreococcus lucimarinus* virus 1	OlV1	*Ostreococcus*	Zimmerman et al. (2019)
MK514406	*Ostreococcus lucimarinus* virus 7	OlV7	*Ostreococcus*	Zimmerman et al. (2019)
NC_020852	*Ostreococcus lucimarinus* virus 5	OlV5	*Ostreococcus*	Henn et al. (2010)
NC_028091	*Ostreococcus lucimarinus* virus 2	OlV2	*Ostreococcus*	Derelle et al. (2015)
NC_028092	*Ostreococcus mediterraneus* virus 1	OmV1	*Ostreococcus*	Derelle et al. (2015)
EU304328	*Ostreococcus tauri* virus 5	OtV5	*Ostreococcus*	Derelle et al. (2015)
FN386611	*Ostreococcus tauri* virus 1	OtV1	*Ostreococcus*	Weynberg et al. (2009)
FN600414	*Ostreococcus tauri* virus 2	OtV2	*Ostreococcus*	Weynberg et al. (2009
JN225873	*Ostreococcus tauri* virus RT-2011	OtV-RT2011	*Ostreococcus*	Thomas et al. (2011)
DQ491002	*Paramecium bursaria Chlorella* virus NY2A	NY2A	*Paramecium bursaria Chlorella*	Van Etten et al. (2006)
DQ491003	*Paramecium bursaria Chlorella* virus AR158	AR158	*Paramecium bursaria Chlorella*	Van Etten et al. (2006)
DQ890022	*Paramecium bursaria Chlorella* virus FR483	FR483	*Paramecium bursaria Chlorella*	Fitzgerald et al. (2007)
NC_000852	*Paramecium bursaria Chlorella* virus 1	PbCV1	*Paramecium bursaria Chlorella*	Yanai-Balser et al. (2010)

Supplementary Table S2.3. Full metagenomic dataset searched with CoverM, including SRR accession numbers and metadata. The corresponding spreadsheet can be found online here and through the attached xlsx file.

Supplementary Table S2.4. Orthogroups found in all four HiMcVs (i.e., core HiMcV orthogroups). Table includes top hits from refseq_protein BLAST, information from InterPro member databases, and HiMcVs putative gene IDs. The corresponding spreadsheet can be found here and through the attached xlsx file.

Supplementary Table S2.5. Orthogroups not shared by all four HiMcVs (i.e., non-core HiMcV orthogroups), including those not found in other prasinoviruses. Table includes top hits from refseq_protein BLAST, information from InterPro member databases, and HiMcVs putative gene IDs. Unique orthogroups are highlighted in yellow. The corresponding spreadsheet can be found here and through the attached xlsx file.

Supplementary Table S2.6. Orthogroups shared between HiMcVs and *Micromonas* hosts M1 and M2. Table includes top hits for HiMcVs from refseq_protein BLAST, information from InterPro member databases, HiMcVs putative gene IDs, as well as the numbers of host and HiMcV strains with sequences present in each orthogroup. The corresponding spreadsheet can be found here and through the attached xlsx file.

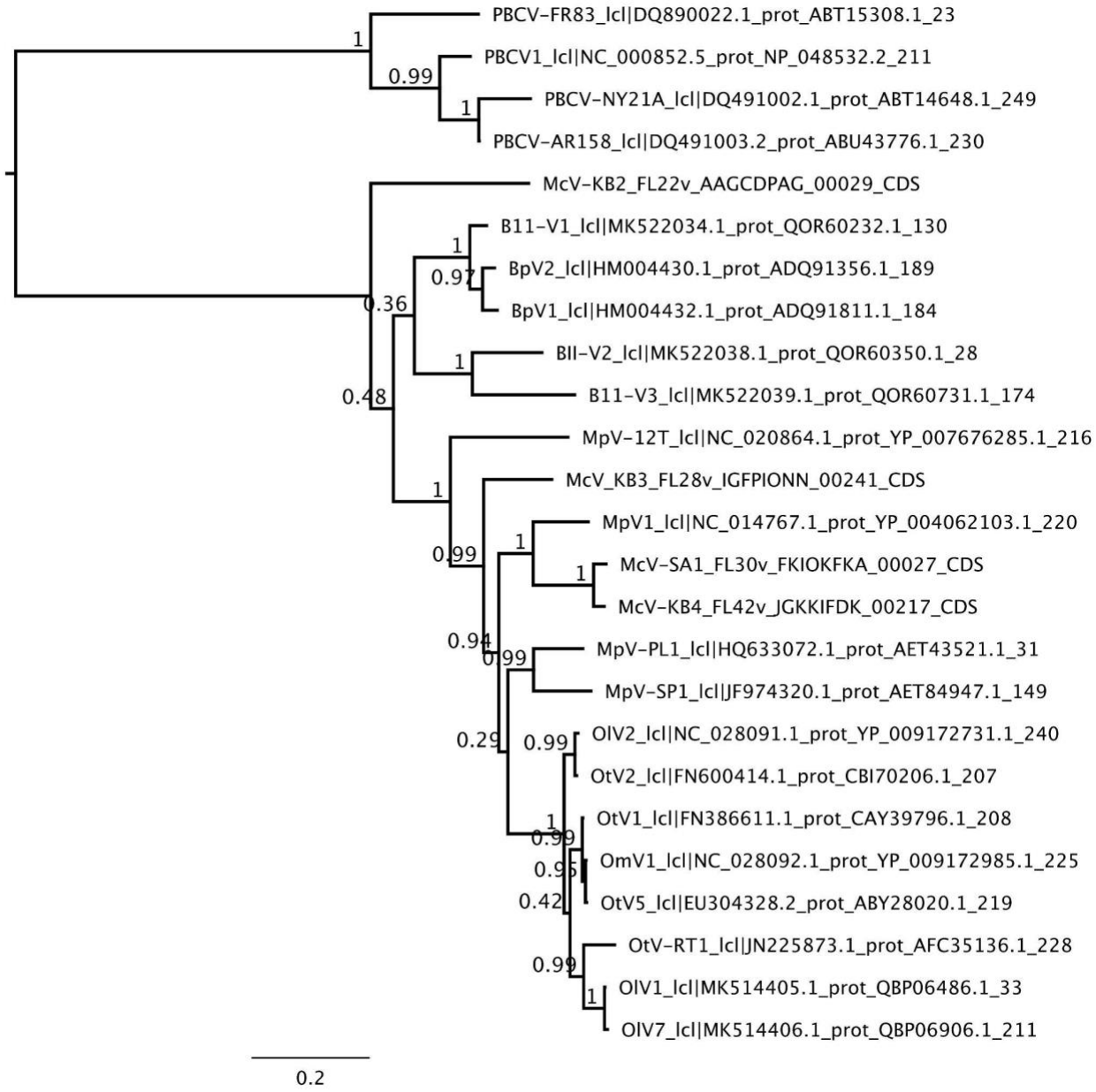
Supplementary Table S2.7. Comparison of orthogroup occurrence across viruses infecting different host genera. Columns for each prasinovirus strain contain sequence count data for each orthogroup (i.e., values > 1 indicate multiple paralogs per strain). Raw p-values, p-values adjusted for false discovery rates, and sequence annotation are included. The corresponding spreadsheet can be found here and through the attached xlsx file.

Supplementary Table S2.8. Metagenome samples containing reads that mapped successfully to HiMcV assemblies, using the CoverM criteria of 95% nucleotide identity and 20% cover. Table include strain name of virus, GenBank NCBI SRA accessions, percent of unmapped reads from each run, relative abundance of reads mapping to HiMcV assembly, and the name of the metagenomic data set. The ALOHA/BATs and BGT dataset are from Biller et al. (2018), and the Mende data set is from Mende et al. (2017).

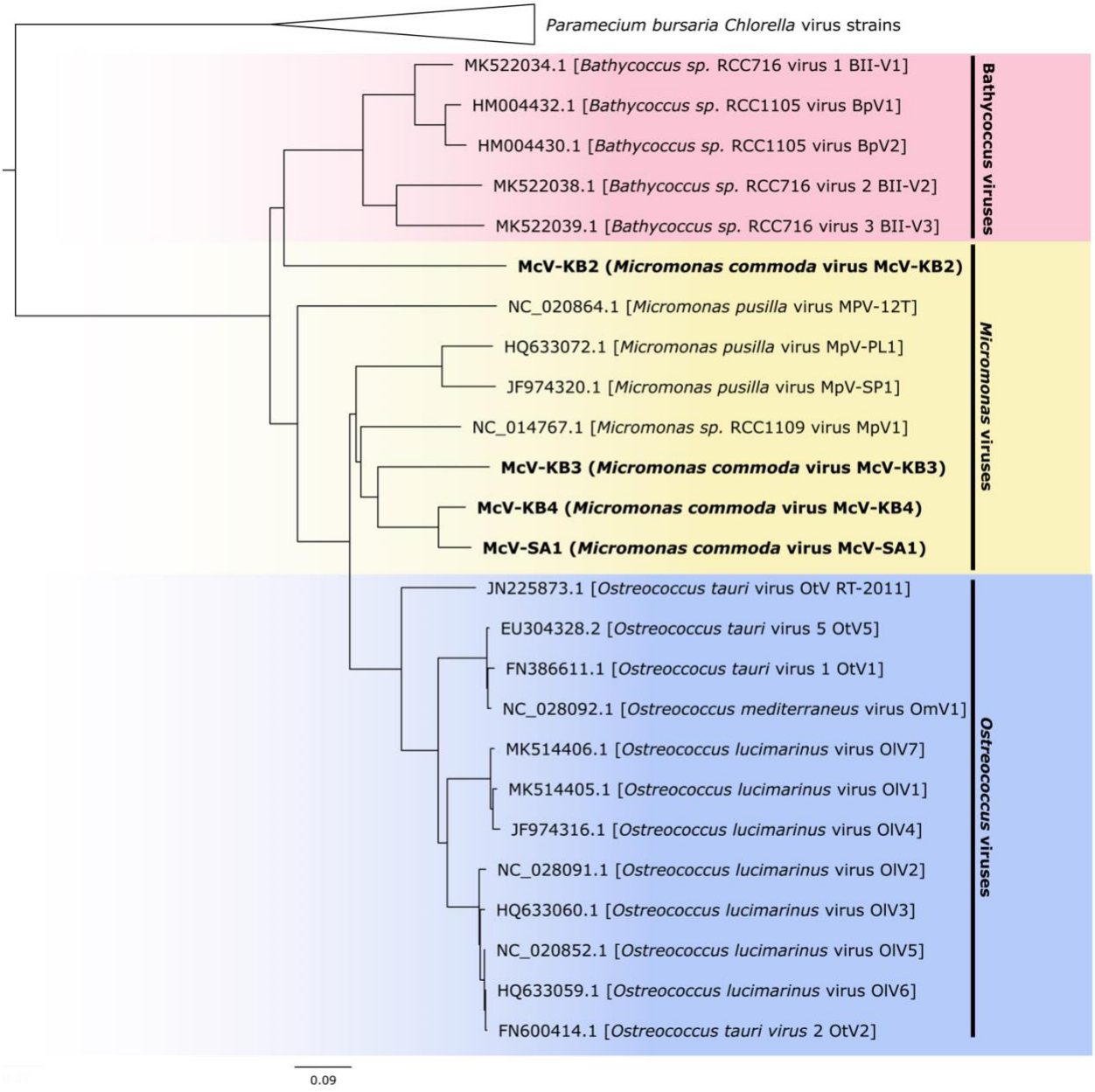
Genome	SRR	unmapped(%)	RelativeAbundance(%)	DataSet
McVSA1	SRR5720231	99.99596	0.00404624	ALOHA/BATS
McVSA1	SRR5720232	99.999	0.00100338	ALOHA/BATS
McVSA1	SRR5720236	99.99674	0.00325706	ALOHA/BATS
McVSA1	SRR5720237	99.99721	0.00279628	ALOHA/BATS
McVSA1	SRR5720238	99.99823	0.00177218	ALOHA/BATS
McVSA1	SRR5720249	99.99895	0.00104743	ALOHA/BATS
McVSA1	SRR5720250	99.99886	0.00114006	ALOHA/BATS
McVSA1	SRR5720254	99.99887	0.00113518	ALOHA/BATS
McVSA1	SRR5720256	99.999146	0.00085947	ALOHA/BATS
McVSA1	SRR5720259	99.99889	0.00110498	ALOHA/BATS
McVKB2	SRR5788033	99.9834	0.00304499	BGT
McVKB3	SRR5788033	99.9834	0.00474699	BGT
McVSA1	SRR5788033	99.9834	0.00880504	BGT
McVKB2	SRR5788075	99.98971	0.00126892	BGT
McVKB3	SRR5788075	99.98971	0.00126892	BGT
McVSA1	SRR5788075	99.98971	0.00126892	BGT
McVSA1	SRR5788089	99.99724	0.00276746	BGT
McVSA1	SRR5788109	99.99846	0.00154367	BGT
McVKB3	SRR5788026	99.98261	0.00454977	BGT
McVSA1	SRR5788026	99.98261	0.0128368	BGT
McVKB3	SRR5788027	99.98899	0.00335733	BGT
McVSA1	SRR5788027	99.98899	0.0076518	BGT
McVKB3	SRR5788028	99.994484	0.00177133	BGT
McVSA1	SRR5788028	99.994484	0.00374259	BGT
McVKB3	SRR5788030	99.99456	0.00183901	BGT
McVSA1	SRR5788030	99.99456	0.00360023	BGT
McVKB3	SRR5788031	99.99267	0.0025355	BGT
McVSA1	SRR5788031	99.99267	0.00479299	BGT
McVKB2	SRR5788032	99.984024	0.0026116	BGT
McVKB3	SRR5788032	99.984024	0.00448607	BGT
McVSA1	SRR5788032	99.984024	0.00887816	BGT
McVKB2	SRR5788130	99.973625	0.0019791	BGT
McVKB3	SRR5788130	99.973625	0.00782666	BGT
McVSA1	SRR5788130	99.973625	0.01656745	BGT
McVKB2	SRR5788131	99.98399	0.00956761	BGT

McVKB3	SRR5788131	99.98399	0.0028609	BGT
McVSA1	SRR5788131	99.98399	0.00357325	BGT
McVKB3	SRR5788136	99.98759	0.00442311	BGT
McVSA1	SRR5788136	99.98759	0.00798864	BGT
McVKB3	SRR5788137	99.98548	0.00462654	BGT
McVSA1	SRR5788137	99.98548	0.00989475	BGT
McVKB2	SRR5788207	99.970116	0.00395817	BGT
McVKB3	SRR5788207	99.970116	0.00713887	BGT
McVSA1	SRR5788207	99.970116	0.01878805	BGT
McVKB2	SRR5788208	99.96208	0.00564524	BGT
McVKB3	SRR5788208	99.96208	0.00944147	BGT
McVSA1	SRR5788208	99.96208	0.02283623	BGT
McVKB2	SRR5788209	99.96813	0.00430875	BGT
McVKB3	SRR5788209	99.96813	0.00766576	BGT
McVSA1	SRR5788209	99.96813	0.01989256	BGT
McVKB2	SRR5788210	99.97286	0.00374107	BGT
McVKB3	SRR5788210	99.97286	0.00671979	BGT
McVSA1	SRR5788210	99.97286	0.01668005	BGT
McVKB2	SRR5788211	99.96458	0.00538717	BGT
McVKB3	SRR5788211	99.96458	0.00913204	BGT
McVSA1	SRR5788211	99.96458	0.02090218	BGT
McVKB2	SRR5788212	99.97119	0.00460325	BGT
McVKB3	SRR5788212	99.97119	0.00693099	BGT
McVSA1	SRR5788212	99.97119	0.01727143	BGT
McVSA1	SRR5788229	99.99854	0.00146091	BGT
McVSA1	SRR5788230	99.998886	0.001113	BGT
McVSA1	SRR5788233	99.99854	0.00145187	BGT
McVSA1	SRR5788283	99.99851	0.00148699	BGT
McVSA1	SRR5788282	99.99712	0.00287704	BGT
McVSA1	SRR5788281	99.996635	0.00336963	BGT
McVSA1	SRR5788284	99.99817	0.00182955	BGT
McVSA1	SRR5788285	99.99818	0.00182364	BGT
McVSA1	SRR5788286	99.99813	0.00186652	BGT
McVSA1	SRR5788288	99.998695	0.00130245	BGT
McVSA1	SRR5788318	99.9988	0.00119912	BGT
McVSA1	SRR5788374	99.99895	0.00105563	BGT
McVSA1	SRR5788430	99.99832	0.00168226	BGT
McVSA1	SRR5788431	99.99872	0.00127943	BGT
McVSA1	SRR5788436	99.99881	0.00118995	BGT
McVSA1	SRR5788435	99.998566	0.00142894	BGT
McVSA1	SRR5788429	99.998055	0.00194779	BGT

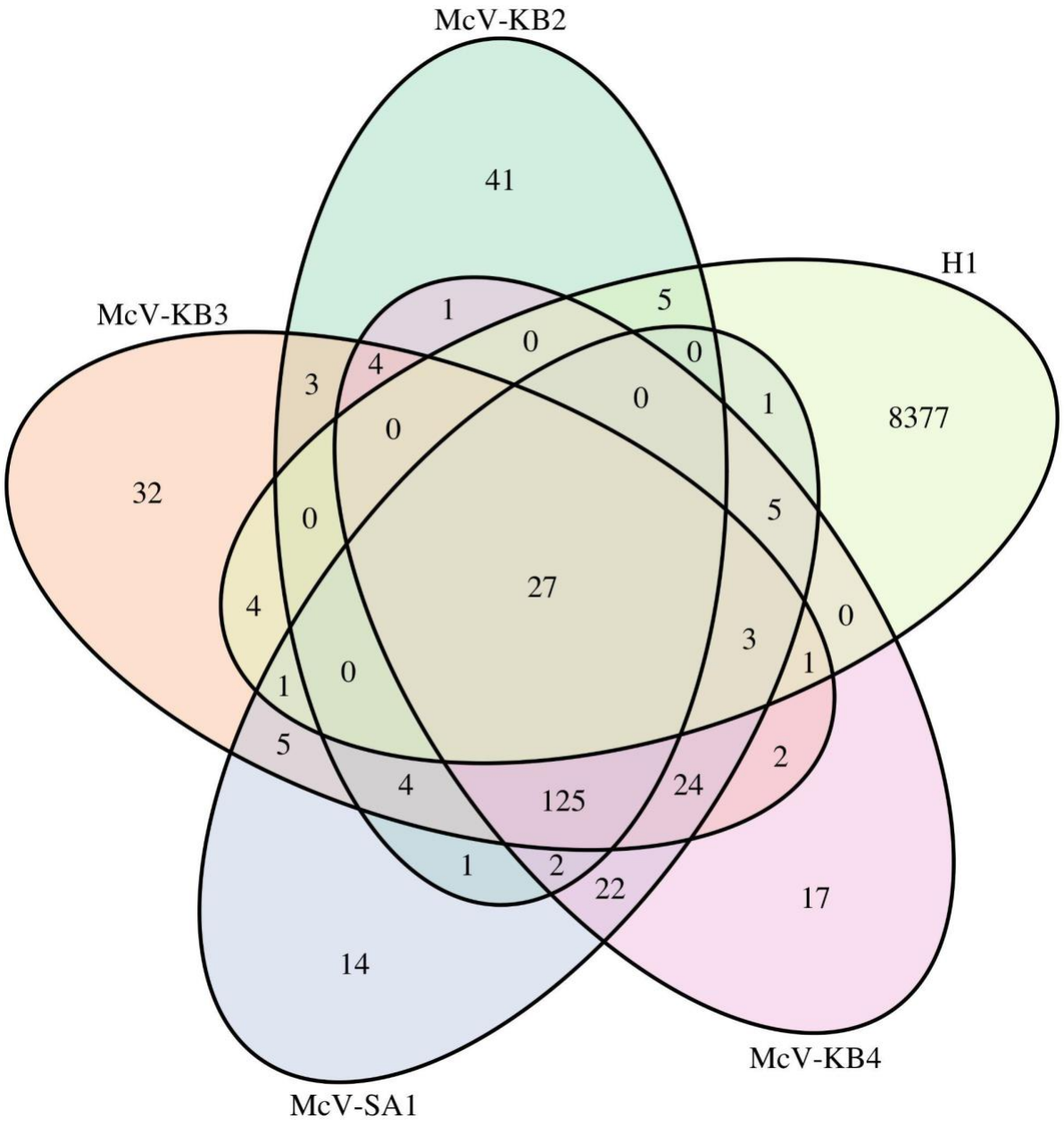
McVSA1	SRR9178106	99.99787	0.00212924	Mende
McVSA1	SRR9178213	99.998184	0.00181834	Mende
McVSA1	SRR9178292	99.99803	0.00196652	Mende
McVSA1	SRR9178368	99.99813	0.00186345	Mende
McVSA1	SRR9178358	99.998146	0.00185231	Mende
McVSA1	SRR9178335	99.997986	0.00201276	Mende
McVSA1	SRR9178205	99.99348	0.00652092	Mende
McVSA1	SRR9178320	99.994	0.00599536	Mende



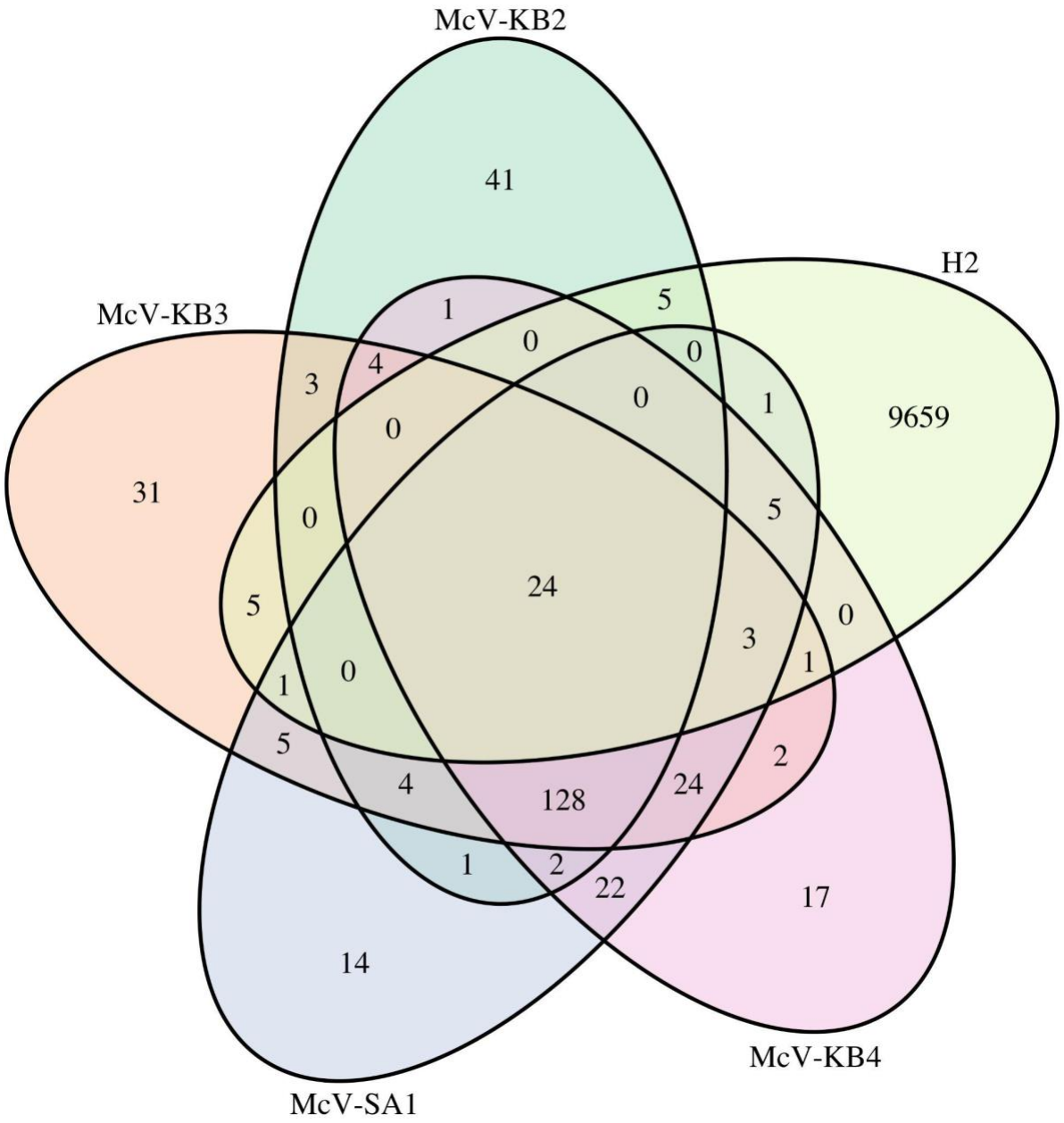
Supplementary Figure S2.1. Prasinovirus and chlorovirus species tree based on the polB orthogroup. Tree was created using FastTree, scale bar represents substitutions per site.



Supplementary Figure S2.2. STAG-generated species tree using the 26 orthogroups that are possessed by all prasinoviruses and chloroviruses genomes used in our OrthoFinder analysis. STAG bipartition support values are not available for datasets with fewer than 100 shared orthogroups. Scale bar indicates substitutions per site.

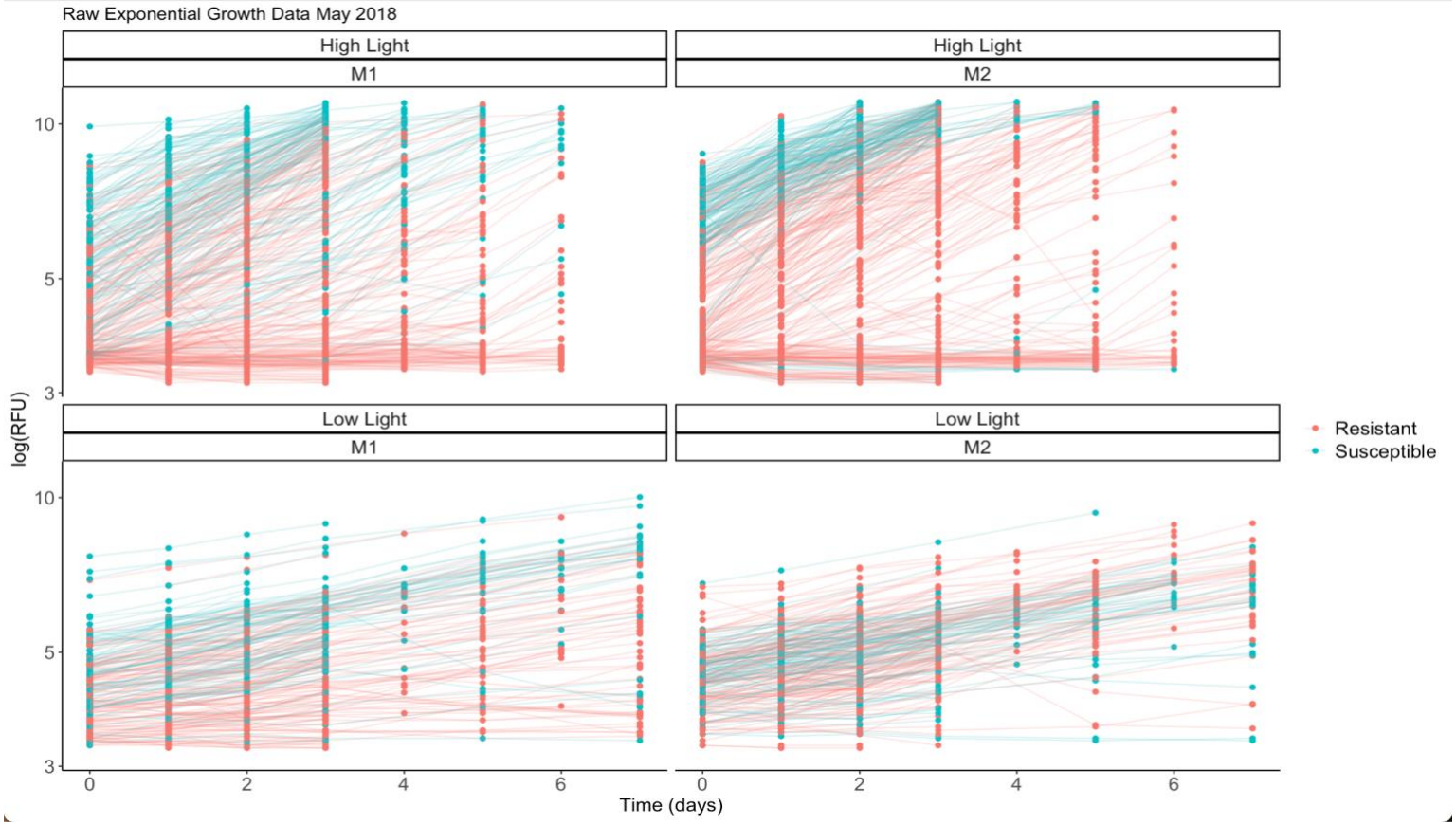


Supplementary Figure S2.3. Venn Diagram of the number of orthogroups shared between the four HiMcVs and host M1.

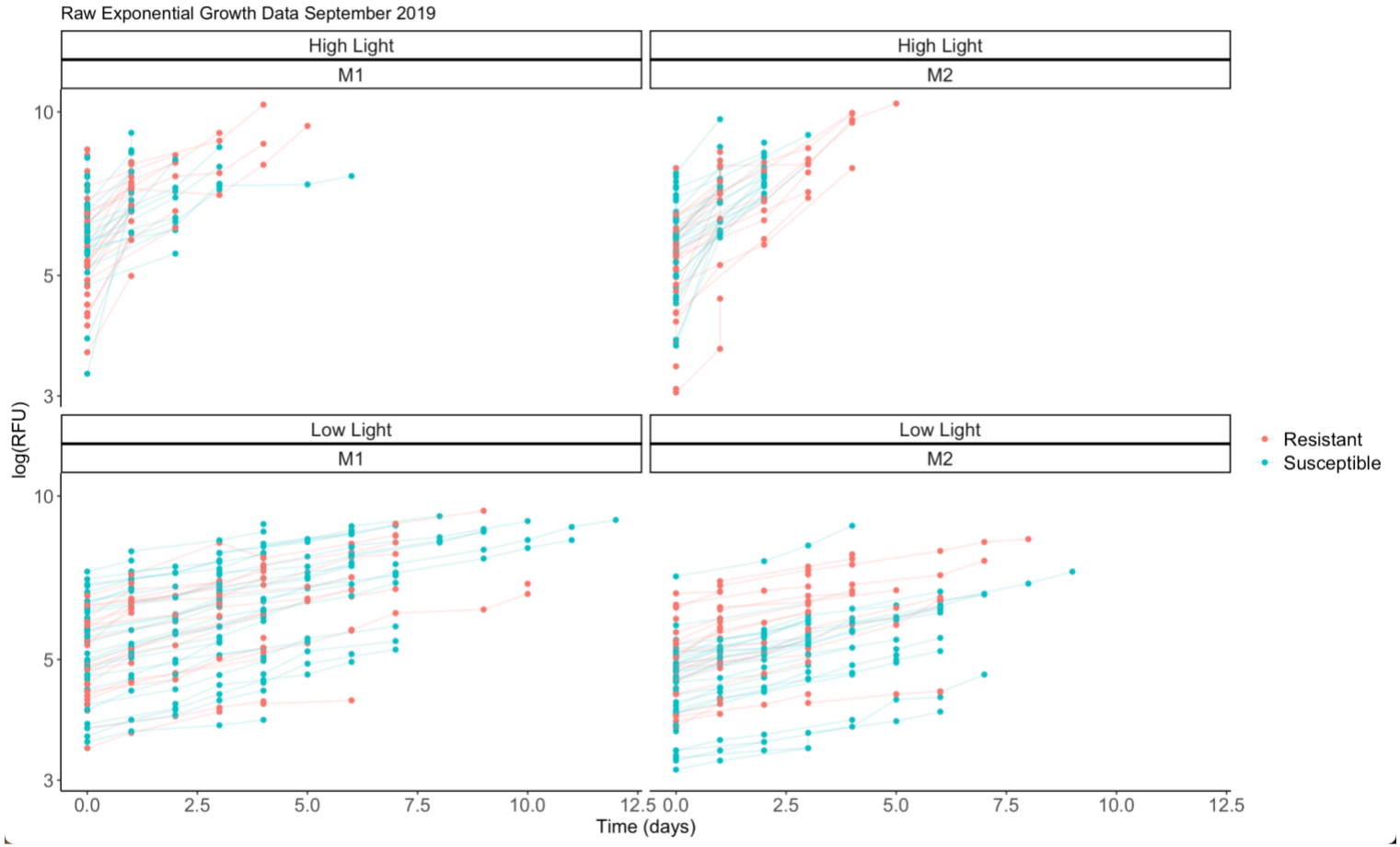


Supplementary Figure S2.4. Venn Diagram of the number of orthogroups shared between the four HiMcVs and host M2.

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

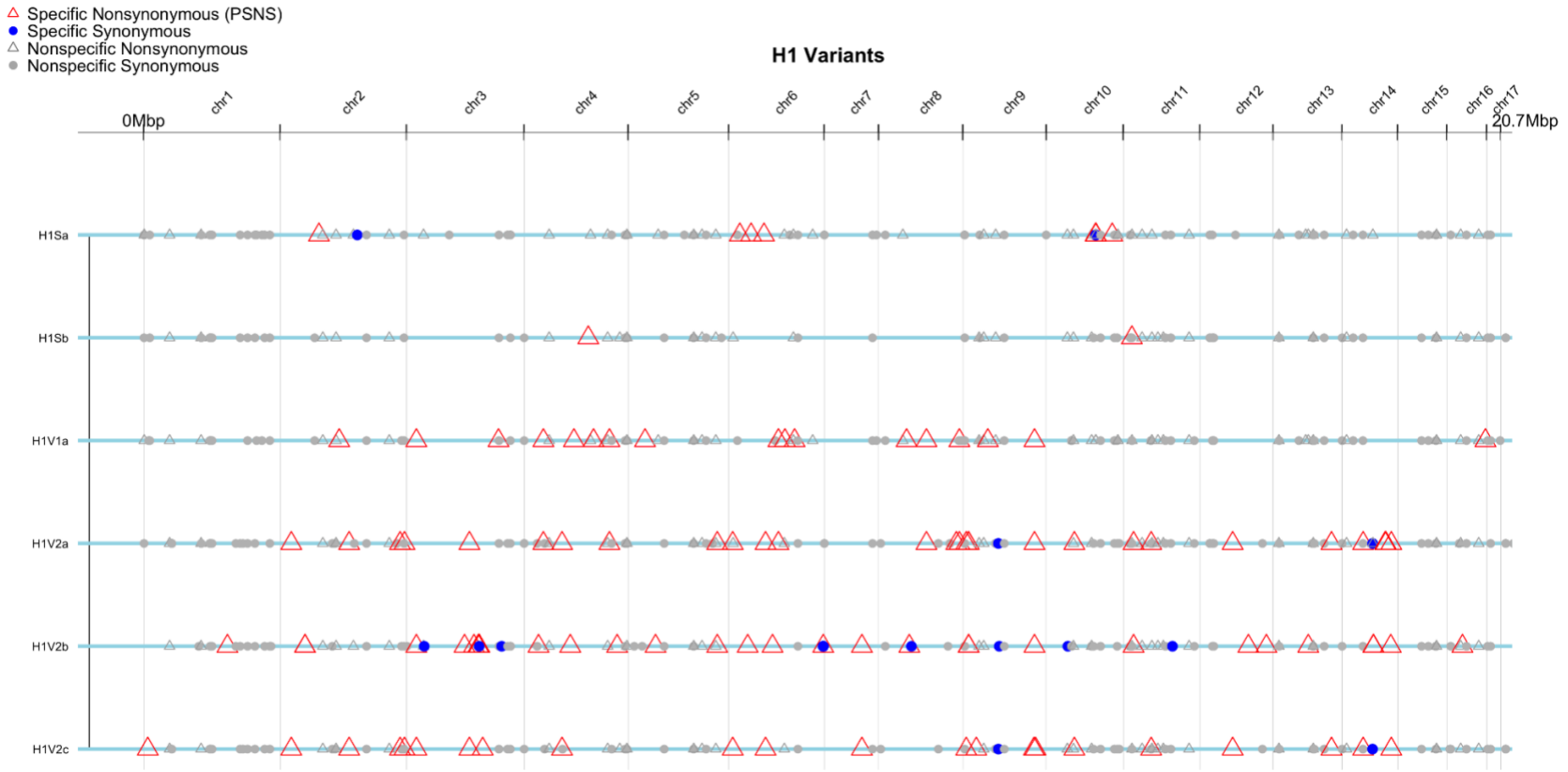


Supplementary Figure S3.1. Exponential growth curves from May 2018. All data used to estimate exponential growth rate of each cell line (88 resistant, 47 susceptible, with duplicates) are represented. Duplicates of 88 resistant and 47 susceptible cell lines per transfer are shown. High light observations encompass three transfers, with low light accounting for two. Resistant cell lines are in pink and susceptible cell lines are in blue. The y-axis, representing raw fluorescence units (RFU), is on a log scale, and the x-axis is time since the start of exponential growth phase, with both lag phase and stationary phase timepoints excluded from each growth curve.

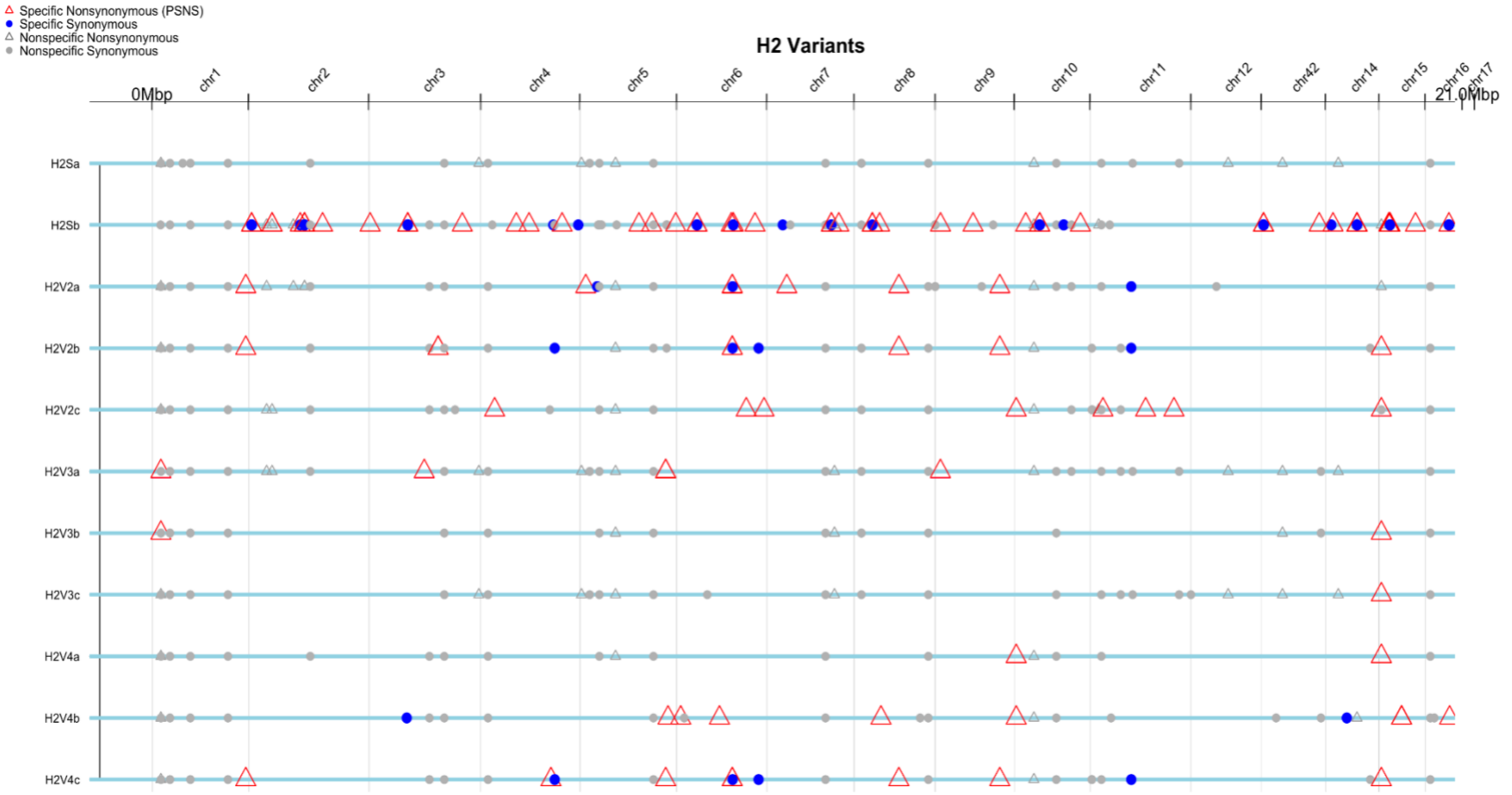


Supplementary Figure S3.2. Exponential growth curves from September 2019 growth experiments. All data used to estimate exponential growth rates of each cell line is represented. Resistant cell lines are in pink and susceptible cell lines are in blue. Duplicates of 12 resistant and 14 susceptible cell lines per transfer are shown. High light observations encompass three transfers, with low light accounting for two. The y-axis, representing raw fluorescence units (RFU), is on a log scale, and the x-axis is time since the start of exponential growth phase, with both lag phase and stationary phase timepoints excluded from each growth curve.

SUPPLEMENTARY MATERIAL FOR CHAPTER 4



Supplementary Figure S4.1. All variants found in M1 cell lines, categorized based whether variant was found in both resistant and susceptible cell lines (“Nonspecific”), only found in either resistant or susceptible cell line (“Specific”), and whether synonymous or nonsynonymous.



Supplementary Figure S4.2. All variants found in M2 cell lines, categorized based whether variant was found in both resistant and susceptible cell lines (“Nonspecific”), only found in either resistant or susceptible cell line (“Specific”), and whether synonymous or nonsynonymous.

Supplementary Material S4.1

CTAB extraction Protocol

Authored by
Erin L. Bernberg, PhD.
Senior Scientist
University of Delaware Sequencing and Genotyping Center

*Use wide bore tips

1) Mechanically homogenize pellet in 1ml of CTAB extraction buffer
2) Dilute with 5 additional ml of CTAB extraction buffer
3) Aliquot into 2m tubes and heat to 55C for 10-15 minutes. Invert tubes every 5 minutes. DO NOT VORTEX EVER!
4) Add 20mg/ml RNAseA to a final conc. of 100ug/ml and incubate at 37C for 30 min. Invert tubes every 5-10 minutes.
5) Centrifuge at 13K RPM for 10 min at RT to remove debris
6) Save supernatant and extract with same volume of phenol:chloroform:IAA (25:24:1, pH6.6) - spin at 4C for 10 minutes
7) Place top layer in a new tube and add 1/10 volume of prewarmed (55C) CTAB/NaCl Buffer and mix well by inversion
8) Extract with same volume of phenol:chloroform:IAA- spin at 4C for 10 minutes
9) Place the top layer in a new tube and extract with chloroform:IAA (24:1)
10) Place the top layer in a new tube and add 0.8 volume of isopropanol- place at -80C overnight to precipitate
11) Centrifuge at 13K RPM for 30 minutes at 4C to pellet
12) Wash the pellet with 70% EtOH and spin for 5 min
13) Wash the pellet with 100% EtOH and spin for 5 min
14) Air dry the pellet and resuspend in warmed (37C) Tris (50-100ul)

DNA extraction buffer:

2% w/v CTAB	2g
100 mM Tris pH8	10ml from IM
20mM EDTA	4 ml from 0.5M
1.4M NaCl	28 ml from 5M
1% PVP40	1g
Q.S.to	100 ml

Add 2% beta mercaptoethanol to the amount of extraction buffer needed before each isolation- Add fresh each time

CTAB/NaCl Buffer:

2% w/v CTAB	10g
0.7M NaCl	4.1g
Q.S. to	100 ml