RICE UNIVERSITY

Detecting and interpreting viral dynamics in marine invertebrate holobionts

By

Alex J. Veglia

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Adrienne Simoes Correa (May 3, 2023 10:04 CDT)

Adrienne Simoes Correa
Assistant Professor of BioSciences

Michael Gustin
Professor of BioSciences

Julia Saltz
Associate Professor of BioSciences

Todd Treangen
Assistant Professor of Computer Science

HOUSTON, TEXAS

April 2023

ABSTRACT

Detecting and interpreting viral dynamics in marine invertebrate holobionts

By

Alex J. Veglia

Metazoan holobionts are comprised of an animal host and symbiotic microorganisms, which may include dinoflagellates, bacteria, archaea, and fungi. The community composition of these symbionts, as well as their interactions with each other and their host, can impact emergent holobiont phenotypes, such as stress tolerance or disease resilience. Viruses are also components of metazoan holobionts, with diverse viruses likely targeting all other holobiont members as their respective hosts. Deciphering the dynamics of viral infections and their impact on holobionts is challenging due to the high mutation rates and variable genome content of viruses, and the limited bioinformatic resources available for this group. This dissertation aims to expand our understanding of viral diversity and dynamics within a foundational marine invertebrate holobiont – stony corals and their microorganisms – and advance environmental virology through bioinformatic tool development. In my first chapter, endogenous viral elements (EVEs) are leveraged to investigate the ecology of an enigmatic RNA virus ('dinoRNAV') that is commonly detected from coral holobionts. I generate evidence that dinoRNAVs target the dinoflagellate endosymbionts (Family Symbiodiniaceae) of corals as their hosts, and that this interaction has likely occurred for >160 million years. The presence of dinoRNAV genomic elements in Symbiodiniaceae genomes is hypothesized to contribute to dinoflagellate antiviral capacity. In my second chapter, I present an accessible bioinformatic program, vAMPirus, that ameliorates challenges related to standardization in the analysis of viral amplicon sequence data. vAMPirus can improve the comparability and reproducibility of viral amplicon sequencing studies by (1) focusing analyses on non-dataset dependent amplicon sequence variants and their unique amino acid translations ('aminotypes'); and (2) providing a freely accessible central repository for vAMPirus-generated files that environmental virologists can use to replicate analyses. Test cases demonstrating the efficacy of vAMPirus based on previously published DNA virus datasets are provided, as well as results from the application of vAMPirus to a novel RNA virus dataset. My third chapter investigates the roles of viruses in the etiology of stony coral tissue loss disease (SCTLD), which has caused coral mortality across the wider Caribbean. Comparative viromics

suggests that virus community diversity in corals in a SCTLD-transmission experiment was driven mainly by dominant Symbiodiniaceae type, followed by coral species, and to a lesser-extent, fragment health state. This suggests that resident viruses (viruses present in the coral prior to SCTLD infection) are correlated with SCTLD etiology via opportunistic infections. Our findings did not implicate a single virus lineage as the causative agent of SCTLD. Nine novel genomes of Symbiodiniaceae-infecting filamentous RNA viruses were detected; the diversity and prevalence of filamentous viruses in Caribbean coral holobionts has likely been underestimated due to the difficulties associated with generating sequence data for RNA viruses from coral tissues. Altogether, this dissertation lays the groundwork for the development of lineage-specific molecular and bioinformatic approaches for the viruses of marine invertebrate holobionts and can ultimately improve our understanding of how viruses influence the health and resilience of coral holobionts, as well as support the development of virus-based tools (e.g., diagnostics, viral therapy) for reef management practices.

ACKNOWLEDGMENTS

My PhD experience at Rice University has been incredibly formative, serving as a catalyst for profound personal and scientific growth. In my time at Rice, I have made meaningful connections with a diverse group of supportive and enthusiastic collaborators whom I thank for their inspiration which has nurtured my passion for the sciences. I want to first thank wholeheartedly my advisor Dr. Adrienne Correa, who has been nothing less than extremely supportive and has been instrumental in my growth as a scientist. Adrienne has encouraged me always to follow my curiosity and develop a research direction that I was passionate about. Adriennes mentorship will have lasting impacts on how I conduct myself as a scientist and a mentor of future scientists. Entering my PhD at Rice, I was facing a significant case of imposter syndrome that manifested itself as lack of interest in my research and writers' block. Since day 1, Adrienne has helped me regain the confidence in myself and my abilities that allowed me to break out of that case imposter syndrome. Without Adriennes mentorship, I would not have an opportunity to fulfill my goal of returning to Puerto Rico to continue my research and educational efforts. In addition, my long-time Correa Lab mates and friends, Lauren Howe-Kerr and Carsten Grupstra, were also extremely important for my development throughout my PhD. Carsten and Lauren were always willing to help me in my research and provide crucial advice for navigating my PhD. I also want to thank all other current and former members of the Correa Lab who have provided guidance and support at various stages of my PhD journey. This includes, but is certainly not limited to, Amanda Shore, Sam Coy, Jenn Drummond, Joyah Watkins, Kara Titus, Carly Karrick, Taylor Walker, and Clark Hamor.

I would like to also express my sincere appreciation to my previous mentors Dr. Nikoloas Schizas and Dr. Marcia Marston for their invaluable mentorship throughout my academic career. Each instilled in me the confidence and skills I often called upon in my PhD and without them I would not be where I am now. I am also immensely thankful to my committee members Julia Saltz, Michaeal Gustin, and Todd Treangen for their unwavering support and guidance during my PhD. Moreover, I am profoundly appreciative of all my collaborators who have been instrumental in the success of my work. I am grateful for their valuable insights, constructive feedback, and willingness to collaborate, which have truly elevated the quality and impact of our collective efforts.

Furthermore, I would like to express my heartfelt gratitude to my family and friends, who have played an indispensable role in shaping the person I have become. I am deeply grateful to my parents for providing me with the opportunity to follow my childhood dream of being a marine biologist. Their guidance and constant belief in me have created the foundation of who I have developed into today. Thank you to all my other family members, brothers, cousins, aunts, uncles, and grandparents for being a source of happiness and stability throughout my life that allowed me to always remember what life is truly about even when I was stressed in graduate school. I want to also thank all my life-long friends made throughout my life, from SBHS, to RWU, to UPRM, all of you have filled my life with happiness and memories that I cherish greatly. Finally, I want to especially express my deepest and most profound gratitude to my partner Joyah Watkins for being there with me step by step in the final months of my PhD. Whenever my motivation or confidence wavered, Joyah was there to support me and express her overwhelming belief in me which drove me until the final moments. Joyah's relentless support and willingness to help me attain my dream career opportunity have made a permanent mark on my success, and I am forever grateful for her impact on my life and work.

TABLE OF CONTENTS

ABSTRACT ..... **ii**

ACKNOWLEDGMENTS ..... **iv**

TABLE OF CONTENTS ..... **vi**

LIST OF FIGURES ..... **viii**

LIST OF TABLES ..... **ix**

INTRODUCTION ..... **x**

Chapter 1 ..... **1**

Abstract 2

Introduction 3

Methods 7

Results and Discussion 11

Conclusions 29

Acknowledgements and Contributions 29

References 31

Chapter 2 ..... **40**

Abstract 41

1 Introduction 42

2 Materials and methods 46

3 Validating the vAMPirus workflow with published double-stranded DNA (dsDNA) virus datasets 55

4 Applying vAMPirus to study a novel environmental RNA virus dataset 60

5 Discussion 63

6 Acknowledgments 66

7 References 68

8 Authors' Contributions 77

Chapter 3 ..... **78**

Abstract 79

Introduction 81

Materials and Methods 85

Results 92
Discussion 106
Acknowledgements 113
References 114
CONCLUDING REMARKS 121
CHAPTER 1 APPENDIX 127
Chapter 1 Supplementary Tables and Figures 127
CHAPTER 2 APPENDIX 135
Chapter 2 Supplementary Methods Tables and Figures 135
CHAPTER 3 APPENDIX 143
Chapter 3 Supplementary Figures 143
INTRODUCTION AND CONCLUDING REMARKS REFERENCES 147

LIST OF FIGURES

1.1……14
1.2……16
1.3……20
1.4……24
1.5……25
2.1……45
2.2……48
2.3……58
2.4……59
2.5……62
3.1……96
3.2……97
3.3……99
3.4……101
3.5……102
3.6……103
3.7……105
3.8……106

LIST OF TABLES

1.1……17
2.1……57
2.2……60
3.1……86
3.2……90
3.3……95
3.4……99
3.5……101
3.6……102
3.7……104
3.8……105

INTRODUCTION

All multicellular organisms harbor microbial symbionts (e.g., protists, viruses, bacteria, archaea, fungi), living and evolving collectively as a single biological unit termed the "holobiont" (Bosch and McFall-Ngai 2011, Bordenstein and Theis 2015). The gain or loss of symbiotic partners, or alterations to interactions among existing holobiont members (e.g., metazoan-symbiont-symbiont), can influence emergent holobiont phenotypic traits and/or fitness (Bordenstein and Theis 2015; Pita et al. 2016, Hussa and Goodrich-Blair 2013). Changes in holobiont traits may drive ecological niche expansion through mechanisms like nutrient acquisition from a novel food source (Kleiner et al. 2012) or enhanced tolerance of environmental stressors (e.g., temperature stress) (Perry et al 2010). For example, when the grass host *Dichanthelium lanuginosum* is infected by the fungal endosymbiont *Curvularia protuberata*, it gains thermal tolerance, but only if the fungus is also infected with Curvularia thermal tolerance virus (Marquez et al., 2007). As a result, *D. lanuginosum* holobionts with this fungus-virus interaction can survive in geothermally heated soils at temperatures up to 65 °C, while grass holobionts lacking this interaction cannot (Marquez et al., 2007). Therefore, understanding and predicting the holobionts responses, resistance, and resilience to environmental stressors require considering the interactions among all symbiotic partners.

Viruses, collectively, infect all members of the holobiont and the impact of these infections on holobiont functioning can range from positive (mutualistic) to neutral (commensalistic) to negative (parasitic) (Vega Thurber et al 2017; Correa et al 2021; Bao & Roossinck, 2013; Grasis, 2017; Roossinck, 2011, 2015; Sweet & Bythell, 2017). Within metazoan tissues, viral lysis of specific microorganisms can shape and maintain symbiont communities, as well as contribute to the host's innate immune system by eliminating potentially

parasitic or disease-causing taxa (Vega Thurber et al. 2017; Neil and Cadwell 2018; Grasis 2017; Quistad et al 2017). However, viral predation may cause the loss of biologically important symbiotic members, resulting in dysbiosis (microbial imbalance) and a decline in holobiont fitness and/or health (Pita et al., 2016; Quistad et al., 2017). Virus infection can also be a vector of horizontal gene exchange, an evolutionary mechanism that can mediate microbiome-driven fitness differences between holobionts (Garcia-Lopez et al 2019; Wang et al 2021). For example, in the fruit fly, *Drosophila melanogaster*, fitness differences induced by a horizontally transferred gene in fly microbiomes influenced adaptation of fly holobionts to a thiamine poor diet (Wang et al 2021).

Scleractinian (stony) corals are complex, non-model marine invertebrate holobionts that are the foundation (habitat-forming) species of highly productive marine ecosystems like coral reefs. Coral reefs are massive subaquatic structures built over time through the deposition of calcium carbonate by stony corals and are home to more than 25% of all marine eukaryotes (Moberg & Folke, 1999). These habitats are vulnerable to biotic (e.g., disease outbreaks) and abiotic (e.g., temperature anomalies) stressors and have faced significant declines in ecosystem health and productivity globally. The proper health and functioning of coral reefs essentially depend on the stability of the coral holobiont and the interactions therein. The most well-characterized interaction in the coral holobiont is that between the host animal and dinoflagellates of the Family Symbiodiniaceae, which can translocate up to 95% of the sugars they photosynthesize to their hosts (LaJeunesse et al., 2018; Martinez et al., 2022). Environmental stress events like periods of abnormally high temperature on the reef can cause the breakdown of this interaction and result in the loss or expulsion of Symbiodiniaceae from the coral tissues, a phenomenon known as bleaching (Hayes & Bush, 1990; Brown, 1997).

Bleaching is increasingly a major source of coral mortality since the 1980s (Brown, 1997; Jackson et al., 2014; van Woesik & Kratochwill, 2022) and, by reducing coral populations, has substantial ecosystem-wide consequences (e.g., phase shifts to algal dominance). Environmentally-driven stress events like bleaching, alongside disease outbreaks and human pressures like overfishing, have led to a 50% reduction in global coral coverage, 60% reduction in global fish catch per unit effort, and an overall 50% decline in coral reefs' capacity to provide ecosystem services (e.g., shore line protection) worldwide (Eddy et al., 2021). The roles of viruses in the resilience and adaptability of coral holobionts in the face of major environmental and anthropogenic threats are still not well understood (Vega Thurber et al., 2017).

Investigating virus functions and impact in coral and other marine invertebrate holobionts can provide unique clues into viral roles in the early evolution of animals (Collins et al 2005). In the past two decades, foundational research on stony corals, sponges, and anemones has revealed the high abundance and diversity of prokaryotic viruses (such as tailed phages in the order *Caudovirales*) and eukaryotic viruses (including members of *Megavirales*, *Herpesvirales*, *Chitovirales*, *Reovirales*, *Picornavirales*, and *Pimascovirales*) in both healthy and diseased marine invertebrate tissues (Bettarel et al., 2013; Buerger et al., 2018; Cárdenas et al., 2020; Correa et al., 2013, 2016; Davy & Patten, 2007; Davy et al., 2006; Grupstra, Howe-Kerr et al., 2022; Jahn et al., 2019, 2021; Laffy et al., 2018; Leruste et al., 2012; Marhaver et al., 2008; Messyasz et al., 2020; Montalvo-Proaño et al., 2017; Nguyen et al., 2021; Nguyen-Kim et al., 2015; Pascelli et al., 2018, 2020; Patten et al., 2008; Rusanova et al., 2022; Soffer et al., 2014, 2015; Veglia et al., 2021, 2022; Weynberg et al., 2014, 2015, 2017; Wilson et al., 2005; Wilson & Chapman 2001; Wood-Charlson et al., 2015). Nonetheless, our understanding of the complete viral diversity associated with these coral reef holobionts is still in its infancy (Ambalavanan et

al., 2021; Thurber et al., 2017). In addition, we are just beginning to explore the prevalence, dynamics, and target hosts of single virus lineages commonly associated with coral reef invertebrates (Buerger et al., 2019; C. G. B. Grupstra et al., 2022; Howe-Kerr et al., 2023; Jahn et al., 2021; Montalvo-Proaño et al., 2017; Planes et al., 2019; Thurber et al., 2017). Consequently, although some viral taxa may be relatively well resolved from sequence datasets, empirical evidence linking these taxa to their hosts or functional role within coral reef holobionts is sparse (reviewed in Thurber et al., 2017 and Ambalavanan et al., 2021). In stony corals, virus lineages that target the inhabiting Symbiodiniaceae are of great interest to the field as it is thought virus infection can contribute to coral bleaching or impact symbiont thermal resilience (Correa et al., 2016; Grupstra et al., 2022; Levin et al., 2017; Messyasz et al., 2020; Thurber et al., 2017).

DinoRNAVs are a lineage of positive-sense single-stranded RNA viruses that are hypothesized to infect Symbiodiniaceae in coral tissues. Genomic evidence of dinoRNAVs has been reported in coral holobiont tissue samples (Correa et al., 2013; Grupstra et al., 2022; Howe-Kerr, 2022; Montalvo-Proaño et al., 2017; Weynberg et al., 2014, 2017) as well as Symbiodiniaceae cultures (Levin et al. 2017). Studies have shown that an increase in temperature can lead to an increase in dinoRNAV infection productivity, suggesting that their activity may have a significant impact on the coral-dinoflagellate symbiosis (Grupstra et al., 2022; Howe-Kerr et al., 2023; Levin et al., 2017). However, the lack of dinoRNAV isolates in culture hinders the field's ability to experimentally confirm that Symbiodiniaceae is its target host. In my first chapter, I employ an *in-silico* approach to confirm the host of dinoRNAV by leveraging endogenous viral elements (EVEs) – virus-sourced DNA within host genomes (Aiewsakun & Katzourakis, 2015). This study demonstrates that dinoflagellate genomes contain dinoRNAV

elements, whereas cnidarian genomes do not, strongly suggesting that dinoflagellates (and not cnidarians) are the target hosts of dinoRNAVs. We also uncovered evidence that indicates that the dinoRNAV genes and sometime whole genomes were integrated into dinoflagellate genomes during a major period of retrotransposon activity that occurred prior to the diversification of Symbiodiniaceae (>160 million years ago; (González-Pech et al., 2021; LaJeunesse et al., 2018; Liu et al., 2018; Stephens et al., 2020). These findings open new avenues for research investigating the potential contributions of these EVEs to the Symbiodiniaceae innate immune system.

In addition to using *in silico* analysis of EVEs, another useful approach for understanding the dynamics and diversity of specific viral lineages on reefs is amplicon sequencing analysis. Over the last two decades, the application of amplicon sequencing to investigate viruses in diverse systems has become more frequent (e.g., 16 peer-reviewed publications in 1998 compared to 127 in 2021 based on a Web of Science search of 'virus amplicon sequencing', November 2022). However, accessible bioinformatic pipelines developed to help virologists, especially those studying non-model viruses, determine the most appropriate analytical approach is currently lacking. Given this, the field of virology will greatly benefit from accessible bioinformatics pipelines and resources tailored to the analysis of virus amplicon sequencing data. In my second chapter, I introduce and describe a new bioinformatic tool, vAMPirus (github.com/Aveglia/vAMPirus), that enables reproducible "raw reads-to-results" analyses of virus community diversity and dynamics from amplicon data. The vAMPirus program is composed of open-source software, well suited to address cross-study comparability issues in the field (e.g., see: vAMPirus Analysis Repository (zenodo.org/communities/vampirusrepo/)), and designed to be an adaptable community-driven tool. I validate the function of vAMPirus and

illustrate its utility by recapitulating results from previously published studies. I also applied vAMPirus to investigate reservoirs of dinoRNAV diversity on the reef. vAMPirus has the potential to become the field standard for analyses of virus amplicon sequencing data and could support viral research efforts similarly to the way QIIME2 (Bolyen et al., 2019) and mothur (Schloss 2020) facilitate the analysis of bacterial/archaeal 16S rRNA amplicon sequence datasets.

In the past two decades, the Caribbean has experienced frequent outbreaks of coral diseases, posing a serious threat to the region's coral reefs (Weil & Rogers, 2011). Determining the roles of viruses, if any, is critical to effective disease management strategies (e.g., anti-microbial vs anti-viral medications) on reefs. The most recent coral disease outbreak is that of the stony coral tissue loss disease (SCTLD) which was first reported in Florida in 2014 (Aeby et al., 2019; Brandt et al., 2021). SCTLD affects 22 (of 45) Caribbean corals species and since 2014, has spread across the Caribbean causing significant damage to local coral populations and reef ecosystems (Aeby et al., 2019; Precht et al., 2016; Walton et al., 2018). The causative agent of SCTLD is still unknown, although recent hypotheses suggest filamentous RNA viruses as potential contributors based on transmission electron microscopy images (Work et al., 2021). Work et al. (2021) posits that filamentous RNA viruses infect the symbiotic dinoflagellates (Family Symbiodiniaceae) of corals, inducing the expression of a dinoflagellate toxin, which causes the observed tissue wasting in the coral host. More investigation is needed to validate or refute the proposed contribution of viruses to SCTLD etiology. In my final chapter, I explore metatranscriptomes generated from SCTLD-affected and unaffected corals sourced from a disease transmission experiment and *in situ* sample collection to characterize the virus diversity significantly associated with SCTLD infections across five different coral species. The results

produced from this chapter present evidence of a cumulative contribution of resident coral virus lineages to SCTLD etiology via opportunistic infections, with limited evidence suggesting a single virus lineage as the SCTLD pathogen. Results also reveal increased production of virus orders hypothesized to infect Symbiodiniaceae (e.g., Algavirales, Durnavirales, Patatavirales) and coral (e.g., Picornavirales) suggesting that these opportunistic infections occur in both the symbiont and the coral itself. These findings provide crucial information regarding the viral groups with potential to contribute to SCTLD pathology, thereby facilitating more effective future investigations into the viral roles in SCTLD and further refining our search for the etiological agent(s).

In conclusion, this dissertation offers novel insight into the diversity of reef holobiont-associated viruses and their potential roles in disease, while also presenting innovative methods that contribute to the advancement of coral reef virology, as well as the broader field of virology. The findings presented in the following three chapters can lead to improved management of coral reef habitat and its foundational organisms, as well as more informed coral reef virology studies in the future. Specifically, the results described indicate 1) dinoRNAVs infect Symbiodiniaceae and dinoRNAV EVEs could influence Symbiodiniaceae antiviral capabilities, 2) vAMPirus is an effective tool for the field and aminotyping is a promise approach for increasing signal-to-noise ratio in amplicon sequencing-based investigations of virus diversity and dynamics, and 3) several resident virus lineages are differentially productive in SCTLD-affected coral holobionts, and merit future investigations inside and outside the context of SCTLD.

Chapter 1

Endogenous viral elements reveal associations between a non-retroviral RNA virus and symbiotic dinoflagellate genomes

Alex J. Veglia*, Kalia S.I. Bistolas, Christian R, Voolstra, Benjamin C. C. Hume, Hans-Joachim Ruscheweyh, Serge Planes, Denis Allemand, Emilie Boissin, Patrick Wincker, Julie Poulain, Clémentine Moulin, Guillaume Bourdin, Guillaume Iwankow, Sarah Romac, Sylvain Agostini, Bernard Banaigs, Emmanuel Boss, Chris Bowler, Colomban de Varga, Eric Douville, Michel Flores, Didier Forcioli, Paola Furla, Pierre Galand, Eric Gilson, Fabien Lombard, Stéphane Pesant, Stéphanie Reynaud, Matthew B. Sullivan, Shinichi Sunagawa, Olivier Thomas, Romain Troublé, Didier Zoccola, Adrienne M.S. Correa, and Rebecca L. Vega Thurber

This chapter is in press at *Communications Biology*: https://doi.org/10.1101/2022.04.11.487905

Abstract

Endogenous viral elements (EVEs) offer insight into the evolutionary histories and hosts of contemporary viruses. This study leveraged DNA metagenomics and genomics to detect and infer the host of a non-retroviral dinoflagellate-infecting +ssRNA virus (dinoRNAV) common in coral reefs. As part of the Tara Pacific Expedition, this study surveyed 269 newly sequenced cnidarians and their resident symbiotic dinoflagellates (Symbiodiniaceae), associated metabarcodes, and publicly available metagenomes, revealing 178 dinoRNAV EVEs, predominantly among hydrocoral-dinoflagellate metagenomes. Putative associations between Symbiodiniaceae and dinoRNAV EVEs were corroborated by the characterization of dinoRNAV-like sequences in 17 of 18 scaffold-scale and one chromosome-scale dinoflagellate genome assembly, flanked by characteristically cellular sequences and in proximity to retroelements, suggesting potential mechanisms of integration. EVEs were not detected in dinoflagellate-free (aposymbiotic) cnidarian genome assemblies, including stony corals, hydrocorals, jellyfish, or seawater. The pervasive nature of dinoRNAV EVEs within dinoflagellate genomes (especially *Symbiodinium*), as well as their inconsistent within-genome distribution and fragmented nature, suggest ancestral or recurrent integration of this virus with variable conservation. Broadly, these findings illustrate how +ssRNA viruses may obscure their genomes as members of nested symbioses, with implications for host evolution, exaptation, and immunity in the context of reef health and disease.

Introduction

Endogenous viral elements, or "EVEs," arise when whole or fragmented viral genomes are incorporated into host cell germlines. Once integrated, EVEs may propagate across successive host generations, potentially becoming fixed in a population through natural selection or drift (Johnson 2015, 2019). Therefore, the presence and content of EVEs can provide clues into the evolutionary relationships among host species and shed light on ancient and modern virus-host interactions (Johnson 2010). To date, most EVEs described in metazoan and plant genomes are retroviral, as this viral group must integrate their genome (as a provirus) into the genome of the host to replicate. Retroviruses thus possess and encode all of the molecular machinery (e.g., reverse transcriptases, integrases) required to integrate autonomously (Stoye 2012). Remarkably, however, sequences from viruses that do not encode reverse transcriptases or exploit integration as a component of an obligate replication strategy – even viruses with no DNA stage – have also recently been detected as EVEs in diverse eukaryotic genomes (Gallot-Lavallée & Blanc 2017, Flynn and Moreau 2019, Horie et al. 2010, Katzourakis & Gifford 2010, Chiba et al. 2011, Chu et al. 2014, Kojima et al. 2021). These non-retroviral RNA EVEs have been reported in hosts ranging from unicellular algae to chiropteran (bat) genomes (Ballinger et al. 2012, Tromas et al. 2014, Palantini et al. 2017, Wang et al. 2014, Jebb et al. 2020, Moniruzzaman et al. 2020, Skirmuntt et al. 2020). Though the mechanisms behind non-retroviral integration continue to be explored, viral sequences may be introduced via nonhomologous recombination and repair, through interactions with host-provisioned integrases and reverse transcriptases supplied on mobile elements (e.g., retrotransposons), or by utilizing co-infecting viruses (Horie et al. 2010, Flynn & Moreau 2019).

Endogenization of any viral sequence (including non-retroviral EVEs) may have positive, neutral or negative effects on a host (Roossinck 2011, Harrison & Brockhurst 2017, Correa et al.

2021). While many EVEs are functionally defective or deleterious and ultimately removed from a population via purifying selection, retained EVEs may remodel the genomic architecture of their hosts or introduce sources of genetic innovation later co-opted for host function (i.e. exaptation; Jern & Coffin 2008, Oliveira et al. 2008). Such 'domesticated' EVEs can be co-opted by hosts and utilized as regulatory elements, transcription factors, or functional proteins with purposes ranging from organism development to synaptic plasticity in the mammalian brain (Feschotte & Gilbert 2012, Frank & Feschotte 2017, Mortelmans et al., 2016, Sofuku & Honda 2017, Takahashi et al. 2019). In particular, non-retroviral EVEs potentially serve as antiviral prototypes that help hosts combat infection by exogenous viruses currently circulating in the population (Witfield et al. 2017, Ter Horst 2019, Palantini et al. 2017, Suzuki et al. 2020). Mechanisms underpinning EVE-derived immunity can include cell receptor interference, nucleic acid sequence recognition (e.g., RNAi), or even replication sabotage through production of faulty virus proteins from EVEs (Aswad and Katzourakis 2012). If expressed, EVEs may have a significant influence on the health, physiology and/or behavior of their hosts in natural and experimental systems (Parker & Brisson 2019, Suzuki et al. 2020, Wilson et al. 2001).

Investigating the distribution, sequence identity, and function of EVEs can yield insight into virus-host interactions across generations. EVEs catalogue a subset of the viruses that a host lineage has encountered and can link homologous extant viruses to contemporary hosts or known disease states (Holmes 2011, Suzuki et al. 2020). Because integrated elements may accrue mutations at a slower rate than exogenous viral genomes (Aiewsakun and Katzourakis, 2015, Flynn & Moreau 2019), EVEs can fill gaps in virus-host networks and act as synapomorphies, indicating the minimum time that a virus may have interacted with a host. As 'genomic fossils', EVEs have helped paleovirologists date the minimum origin of *Circoviridae (Belyi et al, 2010),*

Hepadnaviridae (Gilbert & Feschotte, 2010), Bornaviridae (Kawasaki et al, 2021), Flaviviridae (Li et al, 2022), Lentiviridae (Cui & Holmes, 2012; Keckesova et al, 2009), and *Spumaviridae (Katzourakis et al, 2009)* infections within metazoans (Feschotte & Gilbert 2012, Johnson 2019, Aiewsakun & Katzourakis 2015, Katzourakis 2013, Patel et al, 2011, and reviewed by Barreat and Katzourakis in 2022).

Coral holobionts – the cnidarian animal and its resident microbial assemblage, including dinoflagellates in the family Symbiodiniaceae, bacteria, archaea, fungi, and viruses – are an ecologically and economically valuable, multipartite non-model system (Knowlton & Rohwer, 2003, Matthews et al. 2020). Symbiodiniaceae are key obligate nutritional symbionts of corals and support their hosts in the construction of reef frameworks (LaJeunesse et al. 2018). However, environmental stress can break down coral-Symbiodiniaceae partnerships, resulting in bleaching – the mass loss of Symbiodiniaceae cells (Glynn 1996). Some bleaching signs (paling of a coral colony) are hypothesized to also result from viral lysis of Symbiodiniaceae (van Oppen et al. 2009, Correa et al. 2016, Vega Thurber et al. 2017, Messyasz et al. 2020, Correa et al. 2021, Grupstra et al. 2022), but direct evidence supporting this hypothesis remains limited. Overall, the role of viruses in coral colony health and disease requires further examination.

Non-retroviral +ssRNA dinoRNAV sequences were first reported in stony corals based on five metatranscriptomic sequences and corroborated by Symbiodiniaceae EST libraries (Correa et al. 2013). Subsequent studies indicated that similar +ssRNA viruses are commonly detected in coral RNA viromes and metatranscriptomes, as well as via targeted amplicon assays (Weynberg et al. 2014, Levin et al. 2017, Montalvo-Proano et al. 2017, Grupstra et al. 2022). These viruses exhibit synteny and significant homology to *Heterocapsa circularisquama* RNA virus (HcRNAV; Levin et al. 2017), the sole recognized representative of the genus

Dinornavirus and a known pathogen of free-living dinoflagellates (Nagasaki et al. 2005). Both HcRNAV and dinoRNAV sequences detected in coral holobiont tissues contain two ORFs – a Major Capsid Protein (*MCP)* and RNA dependent RNA polymerase (*RdRp*). Furthermore, icosahedral virus-like particle (VLP) arrays resembling HcRNAV (but with 40% smaller individual particle diameters) have been imaged in the Symbiodiniaceae-dense coral gastrodermis tissue and in Symbiodiniaceae themselves (Lawrence et al. 2014). Levin et al. (2017) assembled the 5.2kb genome of a putative dinoRNAV from a poly(A)-selected metatranscriptome generated from cultured *Symbiodinium*. The assembly contained a 5' dinoflagellate spliced leader ("dinoSL"; Zhang et al. 2013) — a component of >95% of Symbiodiniaceae mRNAs, speculated to illustrate molecular mimicry — and exhibited >1000-fold higher expression in a thermosensitive *Cladocopium C1* population relative to a thermotolerant population of this Symbiodiniaceae strain at ambient temperatures (27□C, Levin et al. 2017, LaJeunesse et al. 2018). Together, the findings from these studies suggest that Symbiodiniaceae are target hosts of reef-associated dinoRNAVs.

This study (1) systematically searched for putative endogenized dinoRNAVs in metagenomes from *in situ* (symbiotic) coral colonies and seawater, as well as in available genomes of Symbiodiniaceae and aposymbiotic (symbiont-free) cnidarians, (2) investigated the evolutionary relationship of putative dinoRNAV EVEs to exogenous reef-associated dinoRNAV sequences, and (3) made preliminary inferences regarding the distribution and possible function of these dinoRNAV EVEs based on their detection, prevalence, and genomic context.

Methods

Identification and computational validation of dinoRNAV EVEs leveraging meta'omics

The Tara Pacific Expedition (2016-2018) sampled coral reefs to investigate reef health and ecology using multiple methods, including amplicon sequencing and metagenomics (see Planes et al. 2019 and doi:10.5281/zenodo.4068293 for coral reef sampling and processing methods). In this study, we explored metagenomes generated from hydrocorals (n=60 *Millepora*), stony corals (n=108 *Porites*, n=101 *Pocillopora*) sampled from 11 islands (three replicate sites per island) across the South Pacific Ocean during the Tara Pacific Expedition for dinoRNAV EVEs (Figure 1, Supplemental Table 1A, 1B; Pesant et al. 2020). Amplicon libraries of the dinoflagellate Internal Transcribed Spacer 2 (ITS2) gene fragment were sequenced in tandem with the metagenomes, to characterize the dominant Symbiodiniaceae harbored by hydrozoan and stony coral colonies (Hume et al. 2020).

To confirm that these dinoRNAV EVE sequences were affiliated with coral holobionts and reduce the possibility that they are technical artifacts, publicly available metagenome libraries were analyzed (Supplemental Table 1B). These additional libraries included 120 assembled pelagic water samples presumed to include pelagic dinoflagellate sequences from the Tara Oceans dataset (2009-2013; Pesant et al. 2015) and 30 MiSeq metagenomes from unfractionated samples of the stony coral genus *Acropora*, which were processed and sequenced via a different pipeline (Supplemental Table 1B, Supplemental Figure 1). Publicly accessible transcriptomes from nine Symbiodiniaceae assemblies (Supplemental Table 1B) were also queried to determine if dinoRNAV-like sequences were present in poly(A)-selected dinoflagellate transcriptomes and resembled EVEs in terms of proximal gene composition and presence of a characteristic pre-mRNA spliced leader (dinoSL) sequence (as in Levin et al,

2017). Details regarding the collection of samples, generation of metagenomes and associated Symbiodiniaceae amplicon libraries, and associated bioinformatic analyses are provided in Supplemental Figure 1).

Metagenomic and transcriptomic scaffolds were annotated against a curated database of dinoRNAV-like sequences (Supplemental Table 2) via BLASTx (e-value $<1x10^{-5}$; Altschul et al. 1990, see Supplemental Figure 1 for workflow). Alignments to the custom database with a bit score <50 and percent shared amino acid identity <30% were excluded from further analysis. A length penalty was not imposed during this step due to the limited length of assembled scaffolds (average N50=3341±127 nt across all queried libraries). Open reading frames (ORFs) from selected scaffolds were called via Prodigal (v.2.6.3; Hyatt et al. 2010) and annotated against the NCBI-nr database (e-value<0.001; DIAMOND v.2.0.6; Buchfink et al. 2015) to confirm homology to dinoRNAVs and to identify adjacent dinoflagellate sequences (e-value$<1x10^{-5}$, bit≥50). In the absence of complete ORFs (potentially due to the limited size of scaffolds, partial integrations, etc.), homology was confirmed through comparison of the initial alignments to the curated database and 300nt of upstream/downstream flanking sequences (bedtools v.2.30.0; Quinlan et al, 2010) against the NCBI-nr database (e-value<0.001; DIAMOND v.2.0.6; Buchfink et al. 2015). This served as further curation and verification, as EVEs can exist in fragmented or degraded states. Non-normalized quality-controlled reads were mapped via bbmap (v.38.84; Bushnell et al. 2017), and putative EVEs were assessed for uniform read coverage across scaffolds, reducing the probability of chimeric assembly. RNA secondary structure was predicted via mfold (v.3.5; Zuker et al 2003).

DinoRNAV EVEs in dinoflagellate and aposymbiotic cnidarian genomes

Publicly available dinoflagellate and aposymbiotic (dinoflagellate-free) cnidarian genome

assemblies were queried to resolve the putative host(s) of dinoRNAVs, to assess homology among detected dinoRNAVs within coral holobionts, and to compare genes proximal to dinoRNAV EVEs in different host species/strains. A chromosome-scale dinoflagellate genome assembly generated from a *Symbiodinium microadriaticum* culture (Accession: GSE152150, Nand et al. 2021), and scaffold-scale genome assemblies were examined for dinoRNAV EVEs (Supplemental Table 1B, Supplemental Figure 1). Scaffold-scale genome assemblies were from the closely related families Symbiodiniaceae and Suessiaceae, and included representatives from the genera *Symbiodinium* (n=9), *Breviolum* (n=1), *Cladocopium* (n=3), *Durusdinium* (n=1), *Fugacium* (n=2), and *Polarella* (n=2), as well as 25 aposymbiotic cnidarian genome assemblies, including the stony coral genera *Acropora* (n=13), *Astreopora* (n=1), *Galaxea* (n=1), *Montastraea* (n=1), *Montipora* (n=3), *Orbicella* (n=1), *Pocillopora* (n=2), *Porites* (n=1), and *Stylophora* (n=1), and the jellyfish *Clytia* (n=1; Figure 2, Supplemental Table 1B). Genome completeness and quality were assessed via BUSCO (v3; Simão et al. 2015) with the Eukaryota dataset and QUAST (v5.0.2; Gurevich et al. 2013), respectively. Scaffolds/chromosomes containing putative dinoRNAV EVEs were identified by aligning sequences to the protein version of the Reference Viral DataBase (RVDB v.19; Bigot et al. 2019) using DIAMOND BLASTx (v0.9.30; Buchfink et al. 2015). The same exclusion criteria were maintained for alignments of metagenomic scaffolds, also omitting alignments <100 amino acids. Regions of dinoflagellate genomes exhibiting similarity to the *MCP* or *RdRp* of reef-associated dinoRNAV reference genomes (Levin et al. 2017) or other closely related +ssRNA viruses (Supplemental Table 2) were extracted and re-aligned to the NCBI-nr database to further confirm viral homology.

We tested the relationship between the number of identified dinoRNAV EVE-containing

scaffolds, dinoflagellate genera, and genome quality metrics using a linear model. Model selection was performed with an F-test (package car, v.3.0-12) and assumptions were visually checked. Pairwise comparisons between genera were conducted using the package emmeans (v.1.7.2). Putative whole dinoRNAV-like genomes within scaffolds were identified based on the presence of *MCP* and *RdRp*-like sequences on the same scaffold no further than 1.5 Kbp apart (Table 1; Supplemental Figure 2). IRESPred (Kolekar et al., 2016) was utilized to identify internal ribosomal entry sites (IRES) with default parameters on putative dinoRNAV EVE with whole sequence integrations.

ORFs were predicted and annotated from dinoRNAV EVE-containing scaffolds and all dinoflagellate chromosomes using Prodigal (Hyatt et al. 2010) and MAKER2 annotation pipeline (Holt and Yandell 2011) with the AUGUSTUS gene prediction software (Stanke et al. 2006). Translated ORFs were then aligned to a hybrid database containing the UniProt/Swiss-Prot database and protein version of RVDB (v.19; DIAMOND-BLASTp). ORFs on putative dinoRNAV EVE-containing scaffolds and chromosomes were further annotated using InterProScan (v5.48-83.0, Pfam analysis with default parameters) to identify sequences proximal to putative dinoRNAV integrations. Presence of dinoflagellate spliced leaders ("dinoSLs") within 500nt of dinoRNAV EVEs was examined using BLASTn (default parameters except word size=9), as described by Gonzalez-Pech et al. (2021).

Phylogenetic analysis of dinoRNAV EVEs

Amino acid-based phylogenetic trees were generated with dinoRNAV EVE ORFs (*MCP* and *RdRp*) from scaffold-scale genomic assemblies, metagenomes, transcriptomes, and sequences from exogenous and closely related +ssRNA reference viruses (Supplemental Table 1A,B, Supplemental Table 2). Sequences were aligned using the best fit algorithm determined by

MAFFT (v7.464; Katoh and Standley 2013) and reviewed and trimmed manually in MEGA (v7; Kumar et al. 2016). Maximum-likelihood trees were generated with IQTREE2 (Minh et al. 2020) using the model determined by ModelFinder (Kaylaanamoorthy et al. 2017) and 50,000 parametric bootstraps (Hoang et al. 2018) with nearest neighbor interchange optimization. ORFs from the chromosome-level assembly for S. microadriaticum culture CCMP2467 were not included in the phylogeny in order to avoid redundancy with those from the analogous scaffold-level assembly. To calculate dN/dS, ORFs were aligned in Clustal Omega (v.1.2.4), refined in MUSCLE (v.3.6), before using pal2nal (v.14) for codon-based nucleic acid alignment. Evolutionary trajectory was then assessed via CODEML (PAML package, v.4.10.5).

Results and Discussion

Evidence of Endogenized dinoRNAVs in Coral Holobiont Metagenomes

Putative dinoRNAV EVEs were detected in metagenomes generated from 42 cnidarian holobionts out of 269 sampled across the South Pacific Ocean. The majority of endogenized dinoRNAVs were identified in hydrocoral metagenomes (*Millepora* spp.; 70.5%, n=105) which predominantly harbored *Symbiodinium* dinoflagellates but EVE-like sequences were also observed in scleractinian coral metagenomes (*Pocillopora* spp.; 29.5%, n=15.) which predominantly harbored *Cladocopium* and *Durusdinium* dinoflagellates (Figure 1A,C). No dinoRNAV-like sequences were detected among *Porites* spp. metagenomes (Figure 1, Figure 2). Hydrocoral metagenomes were sequenced at equivalent depths as scleractinian corals and had comparable levels of annotation (Supplemental Figure 3, Supplemental Table 3); thus, higher dinoRNAV EVE prevalence in hydrocoral libraries was likely not a result of methodological bias. Of the 11 evaluated South Pacific islands, dinoRNAV EVEs were identified in samples from eight (Guam, Gambier, Moorea, Cook, Niue, Malpelo, Coïba, and Las Perlas), spanning 18

unique sites (Figure 1B,D). Among *Pocillopora* spp. metagenomes, putative dinoRNAV EVEs were only identified on the Central American coast (CAMR, Coastal Pacific Longhurst Province) and were absent in Melanesia, Micronesia, and Polynesia; at these latter sites, dinoRNAVs were largely found in *Millepora* hydrocoral metagenomes. Importantly, endogenized dinoRNAV open reading frames (ORFs) appeared to be immediately adjacent to ORFs identified as dinoflagellate (typically Symbiodiniaceae) genes—they were not proximal to coral genes or those of other cellular organisms abundant in these metagenomes (Supplemental Table 4).

We examined the Symbiodiniaceae ITS2 profiles (Hume et al. 2020) associated with each metagenome and found that putative dinoRNAV EVEs were primarily associated with *Symbiodinium*, *Cladocopium*, and *Durusdinium,* which exhibited variation on both host and regional scales (Figure 1C). DinoRNAV EVEs were more common in *Symbiodinium*-dominated cnidarians ($F_{2,1044}=25.8$, $p<0.0001$, nested ANOVA; Supplemental Figure 4) relative to cnidarians hosting other Symbiodiniaceae genera, regardless of host. This suggested that dinoRNAV integration may be particularly recurrent or conserved within the genus *Symbiodinium* (Figure 1).

To determine if these putative viral integrations were specific to cnidarian holobiont metagenomes and ensure that they were not artifacts of shared sample processing and sequencing procedures of the Tara Pacific pipeline, we also analyzed seawater metagenomes and publicly available metagenomes from the stony coral-dinoflagellate holobiont, *Acropora* spp. (Supplemental Table 1B). Examination of 120 Tara Oceans pelagic seawater metagenomes (Pesant et al. 2015) yielded no sequences sharing homology to dinoRNAVs. The concentration of Symbiodiniaceae cells within cnidarian tissues is significantly higher than that of the

surrounding seawater (Littman et al. 2008, Scheufen et al. 2017, Fujise et al. 2021, Grupstra et al. 2021). On average,1.46± 0.08% of assembled contigs in seawater metagenomes were annotated as Symbiodiniaceae. Thus, lack of detection of dinoRNAV-like sequences from seawater metagenomes is likely due to reduced genomic signal of Symbiodiniaceae in the water column, rather than a lack of EVEs associated with Symbiodiniaceae lineages in seawater. However, it also must be noted that these Tara Oceans seawater metagenomes were not collected concurrently with coral samples (Sunagawa et al, 2015). Analysis of the 30 non-Tara *Acropora* holobiont metagenomes identified 29 more putative dinoRNAV EVEs (Figure 2). These dinoRNAV EVEs were again neighboring dinoflagellate ORFs. While the Caribbean *Acropora* metagenomes analyzed contained too few reads to resolve the dominant Symbiodiniaceae present, earlier studies of the same coral colonies identified *Symbiodinium* spp. as the primary symbiont present (Muller et al. 2018).

The identification of endogenized dinoRNAV-like sequences in cnidarian holobiont metagenomes, combined with the proximity of dinoRNAV-like ORFs to dinoflagellate-like sequences across metagenomes harboring diverse dinoflagellate consortia, collectively indicate that dinoRNAV EVEs are widespread among Symbiodiniaceae genera (Figure 2 cyan dots).

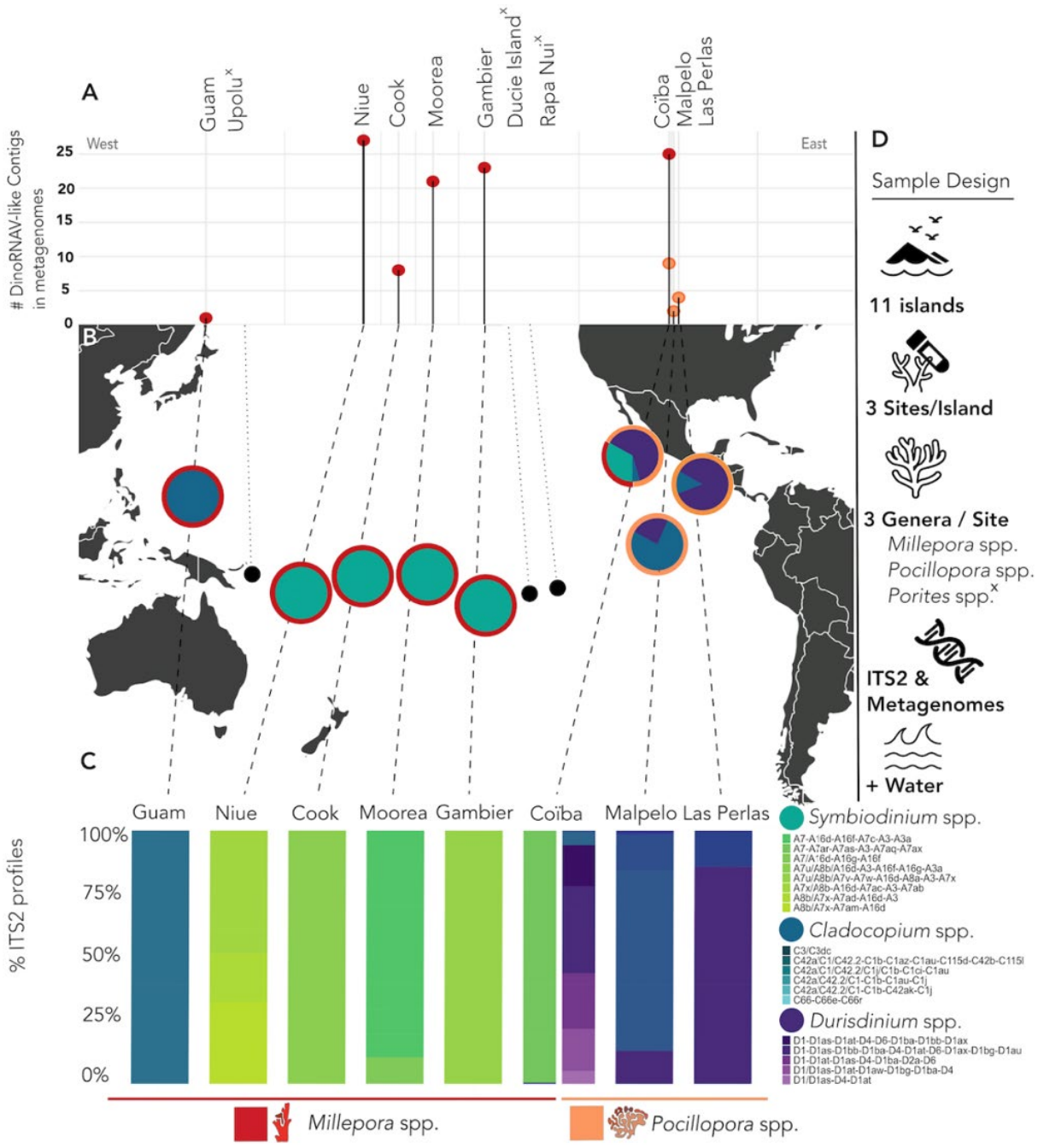


Figure 1. Islands and species (cnidarian and dinoflagellate) correlating with dinoRNAV EVE-like sequence detection among Tara Pacific metagenomes. (A) Count of scaffolds with putative endogenized dinoRNAV-like sequences among Tara Pacific metagenomes, grouped by island and spaced longitudinally by location sampled. (B) Sampling sites of Tara Pacific metagenomes explored for endogenized dinoRNAV-like sequences in this study. Internal circles indicate dominant Symbiodiniaceae genera based on ITS2 type profiles, outer ring denotes coral host(s) sampled at each island. (C) Symbiodiniaceae ITS2 type profile metabarcoding as delineated via Symportal (Hume et al, 2019) within island and host. (D) Sample design of Tara Pacific libraries queried for dinoRNAV EVEs. [x] and black circles on map indicate island locations or species where no dinoRNAV-like sequences were detected.

Endogenized DinoRNAVs Detected in Symbiodiniaceae Genomes

To further test the hypothesis that dinoRNAVs on reefs infect dinoflagellate symbionts and not cnidarians, we examined 18 scaffold-scale genome assemblies representing the dinoflagellate families Symbiodiniaceae and Suessiaceae as well as 25 cnidarian genomes spanning 10 genera (Supplemental Table 1B; Figure 2; Table 1). Alignments revealed no evidence of endogenized dinoRNAVs in any of the 151,782 aposymbiotic (dinoflagellate-free) cnidarian scaffolds. In contrast, the same approach uncovered 351 (of 593,433) dinoflagellate scaffolds with evidence of endogenized dinoRNAVs (Figure 2; Table 1). The identified 351 dinoRNAV EVE-containing scaffolds were observed across 17 of the 18 dinoflagellate genome assemblies (Table 1). DinoRNAV EVEs were also observed in two assemblies from the free-living dinoflagellate genus, *Polarella* (family Suessiaceae), which is closely related to the family Symbiodiniaceae, and served as an outgroup in this study (Janouškovec et al. 2017; Stephens et al. 2020). Interestingly, assemblies belonging to *Symbiodinium*, the most ancestral Symbiodiniaceae genus (LaJeunesse et al. 2018), contained a higher number of scaffolds with putative dinoRNAV EVEs ($\bar{x}$=28.11, stdev=10.7) relative to assemblies of other Symbiodiniaceae genera ($\bar{x}$=8.71, stdev=11; Figure 2 cyan dots; Table 1). This result may clarify why observations of dinoRNAV-like ORFs were more common in metagenomes dominated by *Symbiodinium* (Figure 1C). The dinoflagellate genome assembly with no detected dinoRNAV EVEs belonged to a relatively incomplete assembly of *Cladocopium* C15, which had the second lowest N50 and lowest BUSCO completeness score of all genomes examined (completeness 11.6%, relative to the average 24.54%; Table 1, Supplemental Table 5). The lower coverage/completeness of the *Cladocopium* C15 assembly indicates a reduced window into this genome. It is therefore possible that when a more complete assembly is generated, dinoRNAV

EVE-like sequences will be detectable from this dinoflagellate. However, a linear model suggested that there was no relationship between dinoRNAV EVE detection and assembly statistics (i.e. query length, N50, or completeness; see Supplemental Table 6 for linear model output). Instead, dinoflagellate genus was the strongest predictor of dinoRNAV detection in a genome (LM results: Genus F= 5.74, p = 0.012) and dinoRNAV detections were significantly higher in *Symbiodinium* than *Cladocopium* genomes (pairwise estimated difference= -27.77 ± 5.91, p = 0.01; Supplemental Table 7). Furthermore, since we were unable to detect dinoRNAV EVEs in *Porites* metagenomes – a coral species primarily harboring *Cladocopium* C15 symbionts – we hypothesize that dinoRNAV endogenization was either less common in this lineage of Symbiodiniaceae or integrations have been lost over evolutionary time (Tan et al, 2020, Qin et al, 2019).

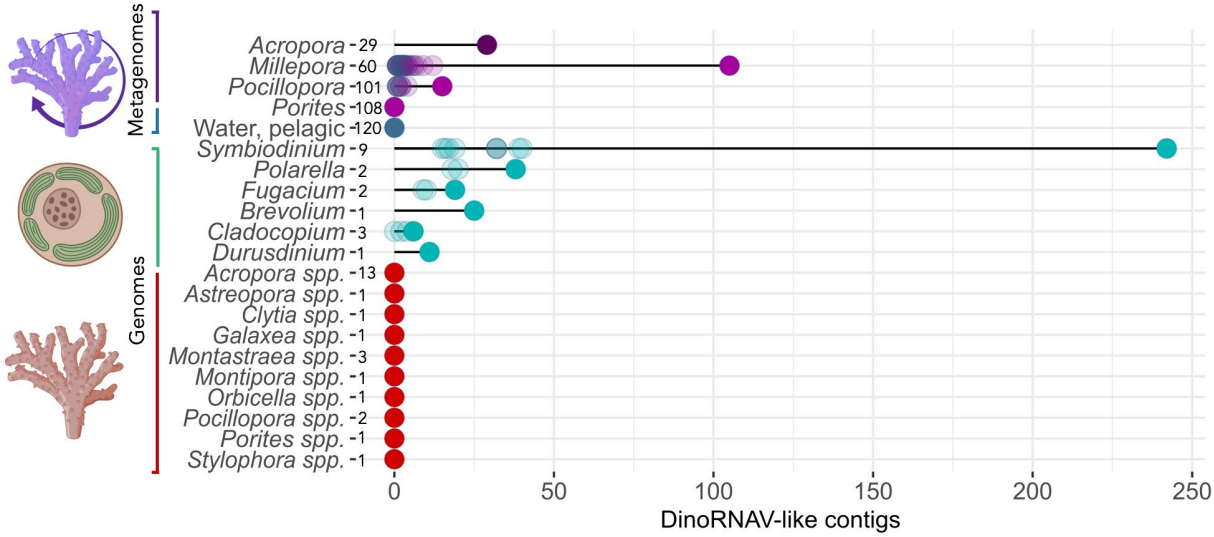


Figure 2. Total quantity of putative endogenized dinoRNAV EVEs identified, broadly organized by sample source (metagenome or genome), and number of libraries or assemblies queried (numbers follow a dash to the right of source name). Opaque circles denote the sum total of dinoRNAV EVE-like sequences identified from each source, while transparent circles denote individual counts of putative dinoRNAV EVEs per library.

	Dinoflagellate Species (strain)	Total # Scaffolds	Host	Location	BUSCO score	dinoRNAV EVE ORFs on scaffolds		
						RdRp	*MCP*	Both
Symbiodiniaceae	*Symbiodinium linucheae* (CCMP2456) [1]	37,772	*Plexaura homamalla*	Bermuda	21.8%	39	0	1
	Symbiodinium microadriaticum (04-503SCI.03) [1]	57,558	*Orbicella faveolata*	Florida, USA	41.6%	30	1	3
	Symbiodinium microadriaticum (CassKB8) [1]	67,937	*Cassiopea sp.*	Hawaii, USA	73.3%	29	1	3
	Symbiodinium microadriaticum (CCMP2467) [2]	9,688	*Stylophora pistillata*	Red Sea	15.6%	29	1	3
	Symbiodinium natans (CCMP2548) [1]**	2,855	N/A (Isolated from seawater)	Hawaii, USA	15.5%	14	1	3
	Symbiodinium necroappetens (CCMP2469) [1]*	104,583	*Condylactis gigantea*	Jamaica	22.8%	37	4	2
	Symbiodinium pilosum (CCMP2461) [1]**	48,302	*Zoanthus sociatus*	Jamaica	19.8%	15	0	0
	Symbiodinium sp. A5 (formerly *S. tridacnidorum*) [1]	6,245	*Heliofungia actiniformis*	Australia	21.1%	17	1	0
	Symbiodinium tridacnidorum (sh18 A3 Y106) [3]	16,176	*Tridacna crocea*	Japan	19.8%	20	1	0
	Brevolium minutum (Mf1.05b) [4]	21,899	*Orbicella faveolata*	Florida, USA	14.2%	21	3	1
	Cladocopium C15 [5]	34,589	*Porites lutea*	Australia	11.6%	0	0	0
	Cladocopium sp. *C1acro* (formerly *C. goreaui* [6]	41,289	*Acropora tenuis*	Australia	27.7%	4	0	0
	Cladocopium sp C92 (Y103) [3]	6,686	*Fragum* sp.	Japan	19.5%	2	0	0
	Durusdinium trenchii [7]	19,593	*Favia speciosa*	Japan	28.7%	10	1	0
	Fugacium kawagutti (CS156 CCMP2468) [6]	16,959	N/A (Free-living)	Hawaii, USA	8.3%	8	1	0
	Fugacium kawagutti (CCMP2468) [8]	30,040	N/A (Free-living)	Hawaii, USA	17.9%	9	1	0
						n=314 scaffolds with DinoRNAV EVE-like sequences		
Suessiaceae	*Polarella glacialis* (CCMP1383) [9] **	33,494	N/A (Free-living, isolated from seawater)	Antarctica	20.8%	20	0	0
	Polarella glacialis (CCMP2088) [9] **	37,768	N/A (Free-living, isolated from seawater)	Arctic	21.8%	18	0	0
						n=38 total scaffolds with DinoRNAV EVE-like sequences		

Table 1. DinoRNAV EVE-like detections from representative Symbiodiniaceae and Suessiaceae dinoflagellate scaffold-level genome assemblies, as well as the host species and location of isolation for each dinoflagellate. Assembly coverage and completeness are measured via BUSCO score (% completeness, or %C; Simão et al, 2015). * Indicates species with documented opportunistic life history; ** Indicates species with documented free-living life history per principal species description. Total counts of dinoflagellate scaffolds in genomes queried with individual endogenized dinoRNAV ORFs (*RdRp, MCP*) or both ORFs nearby one another are provided. *RdRp* = RNA-dependent RNA polymerase; *MCP* = major capsid protein. [1] Gonzalez-Pech et al. 2021, [2] Aranda et al. 2016, [3] Shoguchi et al. 2018, [4] Shoguchi et al. 2013, [5] Robbins et al. 2019, [6] Liu et al. 2018, [7] Shoguchi et al. 2020, [8] Lin et al 2015, [9] Stephens et al. 2020. Further genome citations (including accession numbers) and BUSCO completion metrics can be found in Supplemental Table 5.

Incomplete ORFs and Possible Duplications Indicate Endogenization of DinoRNAVs

The repeated observation of putative dinoRNAV EVEs in dinoflagellate scaffolds and contigs from metagenomes and genomes suggests these sequences are either (1) conserved sequence artifacts of Symbiodiniaceae-dinoRNAV interactions, and/or (2) evidence of highly prevalent dinoflagellate viruses, commonly integrated and propagated via their single-celled hosts. If the observed dinoRNAV-like sequences represent active infections capable of generating virions during egress, we would, at minimum, expect essential ORFs associated with replication (RNA-dependent RNA polymerase, *RdRp*) and virion structure (Major Capsid Protein, *MCP*) to be endogenized on the same scaffold. We would additionally expect to observe overall conservation of ORF length/composition (with a lack of internal stop codons or significant deletions) when aligning the dinoRNAV-like sequences detected here with known exogenous dinoRNAV sequences.

However, both DIAMOND and gene prediction analyses generally depicted dinoRNAV-like ORFs in isolation on separate scaffolds. While 28 MCP and 73 RdRp dinoRNAV ORFs were annotated, both ORFs were present on a Symbiodiniaceae scaffold – potentially representing whole dinoRNAV genome integrations – in only 14 instances. Thirteen of these 14 were from *Symbiodinium* genomes, whereas one scaffold was from *Breviolum minutum*, a member of the second most ancestral dinoflagellate genus (Table 1; LaJeunesse et al. 2018). To assess the conservation of putative dinoRNAV EVE sequence length/composition, we aligned

the genomic and single ORF EVEs to reference exogenous dinoRNAV sequences. The reference genome for reef-associated dinoRNAVs is ~5 Kbp long and contains a 1,071 bp noncoding region between ORFs, with a 124-nucleotide internal ribosomal binding site (Levin et al. 2017). In this study, for 13 of the scaffolds in which dinoRNAV ORFs were detected, the putative noncoding region between the *MCP* and *RdRp* EVEs ranged from ~200-800 bp (except for a scaffold belonging to *S. linucheae* CCMP2456, which contained a ~79 kbp noncoding region, and was excluded in further alignments). No internal ribosomal binding sites were detected within the putative dinoRNAV EVEs identified in dinoflagellate genomes. A nucleotide-based alignment to Levin et al.'s (2017) reference dinoRNAV genome indicated that the putative dinoRNAV EVEs presented here contained substantial insertions and/or deletions (Supplemental Figure 2). Translated exogenous dinoRNAV *MCP* ORFs are reported to be ~358 aa in length (Levin et al. 2017; Figure 3 top sequences), but dinoRNAV-like *MCP* sequences recovered in this study ranged from 116-605aa in length. Furthermore, comparisons of these endogenous *MCPs* to exogenous reference sequences revealed internal stop codons and overall low similarity (Figure 3). Amino acid-based alignment of endogenous dinoRNAV MCPs to metatranscriptome- and amplicon-generated exogenous reference sequences (Levin et al. 2017, Montalvo-Proaño et al. 2017) revealed that indels and regions of low similarity were observed between three conserved regions across both endogenous and exogenous MCP sequences (red boxes in Figure 3).

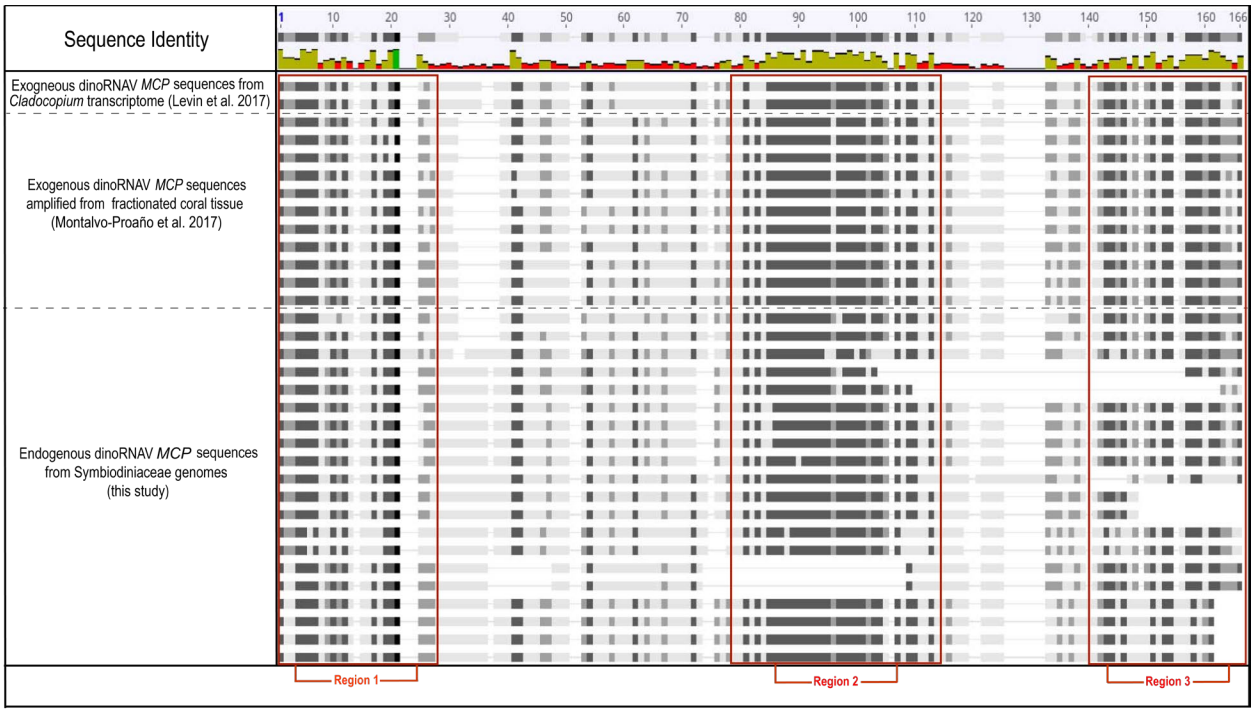


Figure 3. Amino acid alignment including major capsid protein (*MCP*) reference sequences from exogenous dinoRNAV-like +ssRNA viruses, as well as putatively endogenous dinoRNAV *MCP* sequences. Exogenous reference sequences include: 1. Symbiodiniaceae +ssRNA virus *MCP* ORFs recovered from a *Cladocopium* sp. transcriptome (Levin et al, 2017), and 2. dinoRNAV *MCP* amplicons from fractionated coral tissue (Montalvo-Proaño et al. 2017). Conserved regions observed in exogenous and putatively endogenous sequences are labeled as Regions 1-3.

Interestingly, multiple whole dinoRNAV integrations were sometimes observed in a single dinoflagellate genome. For example, genome assemblies of four different *S. microadriacticum* strains contained two or three whole dinoRNAV EVEs each (Table 1; Figure 2). Pairwise alignments measuring shared nucleotide identity of whole dinoRNAV EVEs across Symbiodiniaceae scaffolds revealed that the *S. microadriaticum* genomes and the *S. necroappetens* genome share two whole genome dinoRNAV EVEs (provisionally dinoRNAV-A and dinoRNAV-B; Supplemental Figure 2; Clustal-Omega; Sievers et al. 2011). *S. microadriaticum* dinoRNAV-B was identical in all strains and shared 97% identity with the *S. necroappetens* dinoRNAV-B, yet proximal genes varied (Supplemental Table 8). Importantly, the inconsistent composition and fragmented nature of both the genomic and single ORF dinoRNAV EVEs reported here supports the hypothesis that these sequences are not capable of

generating replicative virions and are best interpreted as multiple integrations of dinoRNAVs into a host genome.

A Potential Mechanism for dinoRNAV Endogenization: Host-Provisioned Retroelements

To assess if general genomic "neighborhoods" are conserved across dinoRNAV integrations (e.g., site location and synteny) and to better understand the genes proximal to EVEs on Symbiodiniaceae genomes, a chromosome-scale *Symbiodinium microadriaticum* genome assembly was evaluated (Figure 4). The highest quality dinoflagellate genome assembly currently available revealed dinoRNAV-like ORFs on 18 of 94 chromosomes, with at least one *RdRp* on each, and some with multiple (two with n=2 *RdRps*, three with n=3 *RdRps*). On three of the chromosomes (# 30, 35, and 74), there were predicted ORFs annotated as dinoRNAV *MCPs* in close proximity to a *RdRp* ORF (separated by noncoding regions 319-656nt), indicative of a potential full-length dinoRNAV genome integration. These results corroborate detections of multiple genomic dinoRNAV EVEs in scaffold-scale assemblies of *Symbiodinium microadriaticum* genomes (Supplemental Figure 2). The higher-resolution *S. microadriaticum* chromosome-level assembly facilitated the identification of an additional dinoRNAV genomic EVE (n=4 for chromosome-level vs. n=3 for scaffold-level, Supplemental Figure 2), two of which were identified on Chromosome 74 and were separated by 2,501 nucleotides. Of note, Nand et al. (2021) reported a decreasing abundance and expression of genes towards the center of chromosomes (past ~2Mpb of a telomere), where there was an increase in repetitive elements; this is where 26 of 29 putative dinoRNAV EVEs were identified in the chromosome-level assembly. Furthermore, ORFs neighboring integrations often varied widely, both in proximity and predicted function, from collagen and RNA binding protein to reverse transcriptase and non-LTR retrotransposable elements.These ORFs potentially contributed to the endogenization of

dinoRNAV via mechanisms such as retroposition (Figure 4, Supplemental Table 8).

Retroposition through host-provisioned retroelements is one proposed mechanism of non-retroviral RNA virus integration into eukaryotic genomes (Horie et al, 2010, Flynn & Moreau 2019). An indicator of such gene retrotransposition in these dinoflagellate genomes is the nearby presence of a relict dinoflagellate spliced leader ("dinoSL"), a 22nt sequence located at the 5' end of mRNAs (Zhang et al. 2007, Lidie and van Dolah 2007, Song et al. 2017, González-Pech et al. 2021). Such a sequence flanks the *RdRp* gene on some extant dinoRNAVs (Levin et al. 2017). We detected dinoSLs within 500bp of 23.1% (six of 26) endogenized *RdRp* ORFs on *S. micoradriaticum* chromosomes, providing support for retrotransposition of these viral elements into Symbiodiniaceae genomes (Supplemental Table 9). DinoRNAV gene retrotransposition may be facilitated by any of three major orders of retroelements associated with Symbiodiniaceae, including long terminal repeat (LTR) retrotransposons, short interspersed retroelements (SINEs), and long interspersed retroelements (LINEs; Elbarbary et al. 2016, González-Pech et al. 2021, Nand et al. 2021, Mita and Boeke 2016). Evidence suggests that these LINEs are common and non-active remnants of an ancient proliferation of LINE retrotransposons that preceded the diversification of Suessiales (Liu et al. 2018, González-Pech et al. 2021, Stephens et al. 2020). *Symbiodinium* contain more LINEs relative to other Symbiodiniaceae genera, comprising 74.10-171.31 Mbp of *Symbiodinium* genomes, relative to an average of 7.48 Mbp of the genomes of in other genera, indicating the loss of these retroelements across speciation events (González-Pech et al. 2021, Nand et al. 2021). The loss of LINEs in more recently derived Symbiodiniaceae genera coincides with a decrease in dinoRNAV EVE detection in these genomes (Table 1). Conversely, the genomes of *Polarella,* the psychrophilic and free-living outgroup from which Symbiodiniaceae diversified ~160 million years ago, are LINE-rich and generally have

comparable numbers of dinoRNAV EVEs to *Symbiodinium* (Table1; Janouškovec et al. 2017, Stephens et al. 2020; González-Pech et al. 2021; LaJeunesse et al. 2018). Together, this suggests that LINE activity during speciation may have facilitated dinoRNAV integration and the resulting EVEs may constitute dinornavirus "fossils." This may explain their degree of sequence fragmentation and relatively low sequence similarity to modern extant dinoRNAVs (Figure 3).

LINE-mediated retroposition is further supported by the observation of a LINE reverse transcriptase homolog ~17 kbp upstream of a *RdRp* EVE with a relict dinoSL on chromosome 45 (Supplemental Table 8) and a LINE retroelement 95 bp downstream of an EVE recovered from a *Pocillopora* metagenome (Figure 4). Additionally, ~40% of annotated ORFs (35 of 88 annotated proteins) proximal to dinoRNAV ORFs on *S. microadriaticum* chromosomes were similar to non-LTR retrotransposable elements seen in other eukaryotic genomes (Figure 4, Supplemental table 6), sometimes <300bp 5' upstream (Supplemental Table 8, Supplemental Figure 5). Collectively, these findings implicate host provisioned retroelements, such as LINEs, as facilitators of dinoRNAV gene integration.

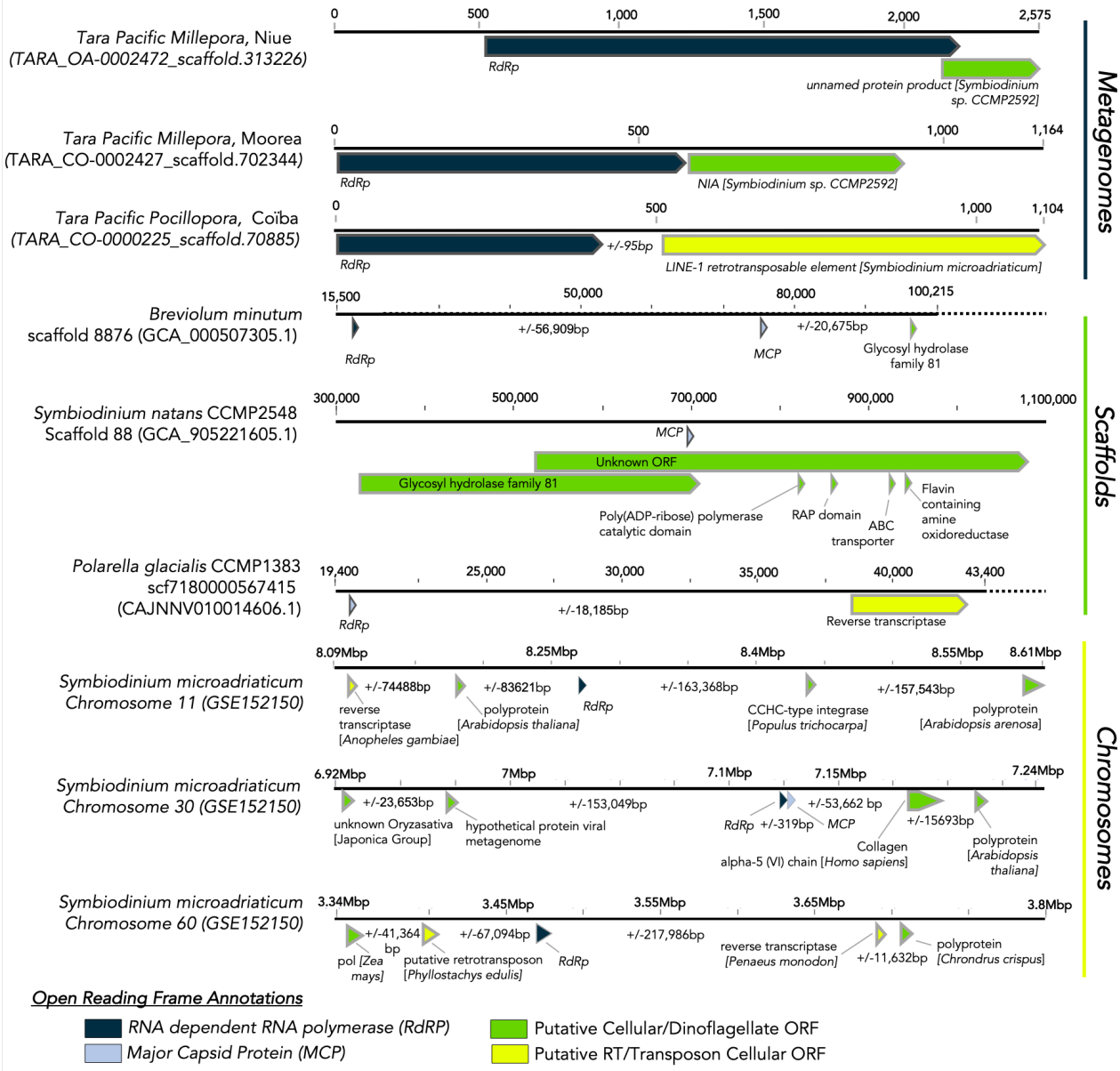


Figure 4. Representative scaffolds and chromosome fragments containing putative dinoRNAV EVEs (*MCP*, light blue; *RdRp*, navy blue ORFs with complete description in Supplemental table 10). Open reading frame (ORF) color broadly indicates cellular versus putative +ssRNA viral homology; yellow and some green (e.g., integrases, polyproteins) ORFs may be exploited mechanisms for viral integration. (+/-) base pair values represent sequence lengths between ORFs.

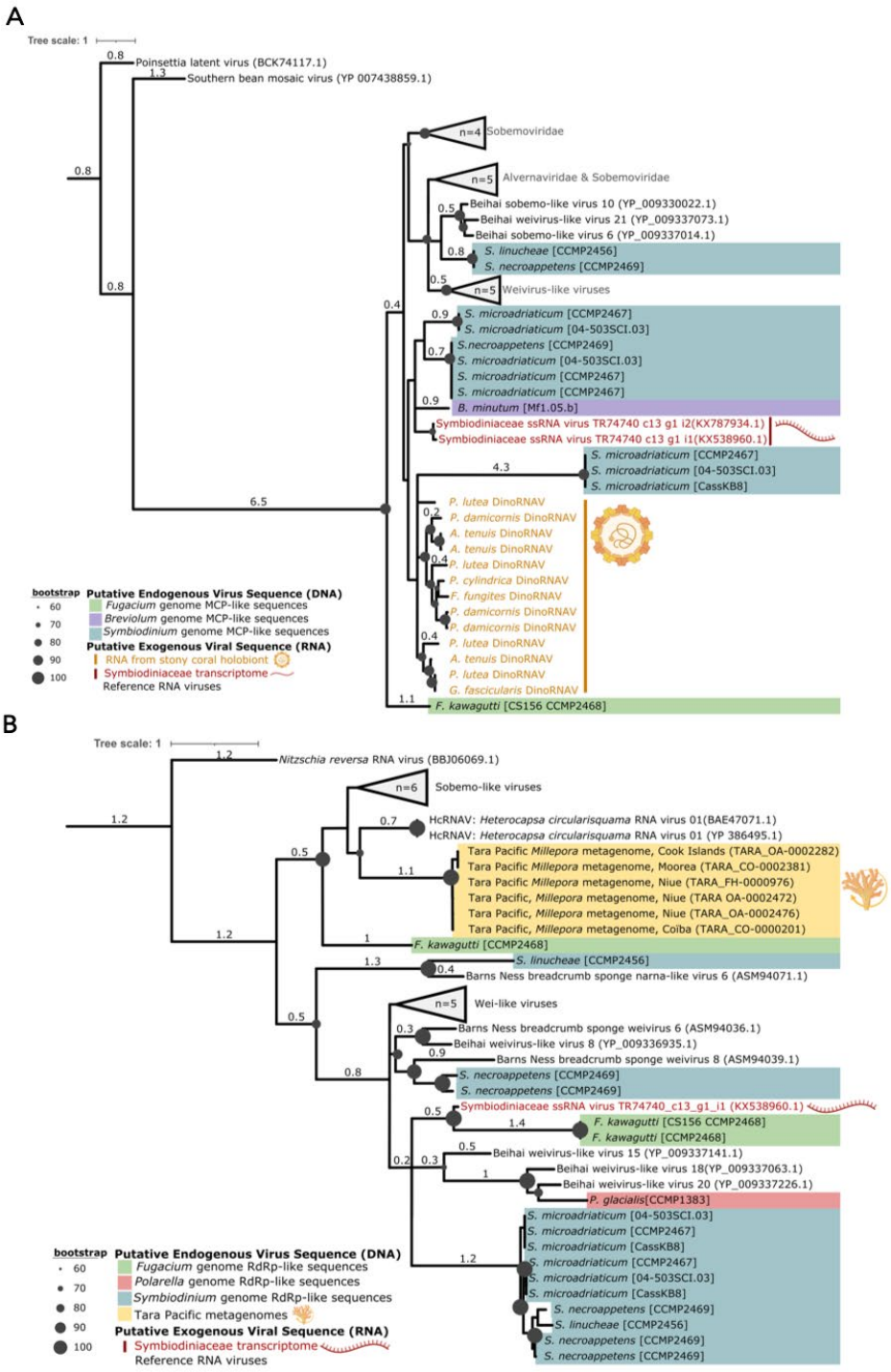


Figure 5. Phylogenies of dinoRNAV Major Capsid Protein (*MCP,* **A**) and RNA-dependent RNA polymerase (*RdRp,* **B**) ORFs recovered from host metagenomes, transcriptomes, polished genomes, and extant +ssRNA reference viruses from amplicon libraries (A only). Both trees include *Dinornavirus* reference sequences. (A) Maximum-likelihood tree of *MCP* amino acid sequences generated with a LG+F+G4 substitution model and 50,000 parametric bootstraps, illustrating the similarity of putative dinoRNAV EVEs (this study) to extant dinoRNAVs from stony coral colonies. (B) Maximum-likelihood tree of *RdRp* amino acid sequences generated with a Blosum62+G4 substitution model and 50,000 parametric bootstraps, demonstrating the similarity of metagenomic dinoRNAV EVE *RdRps* to *RdRps* of the sole recognized *Dinornavirus*, *Heterocapsa circularisquama* RNA virus (HcRNAV), as well as alignment to each other. Trees were visualized in iTOL.

DinoRNAV EVEs Show Homology to Extant Exogenous Viruses

Modern, exogenous dinoRNAVs (Order: *Sobelivirales*) are highly divergent and hypothesized to form chronic infections within dinoflagellate hosts (Correa et al. 2013; Levin et al. 2017; Montalvo-Proaño et al. 2017; Grupstra et al. 2022). This chronic infection strategy likely provides opportunities for retroelement-driven endogenization into host genomes. Because many EVEs evolve at the rate of the host genome, rather than at the much faster rate of exogenous +ssRNA viral genomes, EVEs can serve as a snapshot of viral ancestry (Holmes et al. 2010). We compared translated dinoRNAV EVEs to exogenous dinoRNAVs and other *Dinornavirus* taxa to assess the conservation of EVEs, the potential for host utilization of these elements, and their relatedness to contemporary dinoRNAVs. We found that amino acid translations of endogenous dinoRNAV *MCP* sequences contained conserved motifs observed in the exogenous *MCP* sequences (e.g. Regions 1-3 in Figure 3), yet the associated phylogeny was highly polyphyletic along inferred ancestral nodes (Figure 5A). Endogenous MCP ORFs also appear to be evolving under neutral selection (dN/dS=0.958).

Endogenized dinoRNAV MCP form their own clades within the MCP tree, each closely related to specific clades consisting of extant dinoRNAVs or environmental (i.e. unclassified) sobeliviruses with similar conserved motifs. The majority of dinoRNAV MCP EVEs shared similarity to extant *MCP*s identified from unfractionated stony coral holobionts via amplicon sequencing (Montalvo-Proaño et al. 2017); these sequences formed an independent, disorganized clade (Figure 5A clade containing yellow and blue sequences), relative to those recovered from dinoflagellate transcriptomes or those of other invertebrate hosts. Likewise, dinoRNAV *RdRp* EVEs identified via metagenomics appear most similar to HcRNAV, the defining member of family Alvernaviridae and a protist pathogen, further supporting the affiliation of this EVE with

a dinoflagellate host. *MCP* and *RdRp* ORFs putatively derived from the same dinoflagellate genomes often shared clades (clades containing multiple blue or green sequences in Figure 5A, B), perhaps indicative of duplications within genomes or multiple integration events of particular dinoRNAV lineages within host genera. The detection of putative dinoRNAV *RdRp* ORFs within *Polarella* genomes is therefore indicative of either the antiquity of dinoRNAV-dinoflagellate interactions and/or a propensity for recent dinoRNAV integration across Dinophyceae families. However, the exclusion of the *P. glacialis* dinoRNAV-*RdRp* from *RdRps* of other dinoflagellate clades (pink, Figure 5B) further illustrates the congruence between EVEs and their host genomes. Overall, the evident homology to contemporary *Dinornaviruses* support these integrations as Alvernaviridae within order *Sobelivirales.*

The expression and functional potential of endogenized dinoRNAV elements (if any) remains unclear. With no isolated Symbiodiniaceae-infecting dinoRNAV strains available, investigation into EVE functionality is limited to *in silico* approaches. Sequence data mining efforts identified RNA sequences either sharing sequence similarity with dinoRNAVs, or containing whole dinoRNAV-like ORFs that also annotated as dinoflagellate transcripts (i.e. with cellular ORFs or sequence similarity) in seven out of nine publicly accessible dinoflagellate transcriptomes (Supplemental Table 11). Additionally, two transcripts from an exogenous dinoRNAV infection identified in *Cladocopium* transcriptomes carried *MCP* ORFs of +ssRNA viral sequences (‘TR74740_c13-g1_i1’ and ‘TR74740_c13-g1_i2’, Levin et al. 2017, red text in Figure 5A) and form a clade with putative *Symbiodinium* dinoRNAV EVEs (Figure 5). Likewise, the *RdRp* ORF of ‘TR74740_c13-g1_i1’ and the *RdRp* of ‘GAKY01194223.1’— a transcript derived from a cultured *Symbiodinium microadriaticum* A1 transcriptome—shared some areas of similarity to putative endogenous dinoRNAVs (Figure 5B; Levin et al. 2017, Baumgarten et al.

2013). Importantly, both RNA transcripts also shared features characteristic of dinoflagellates, such as a 5' dinoSL (Zhang et al. 2013) or dinoflagellate sequence space flanking the dinoRNAV itself (Baumgarten et al. 2013). Furthermore, 'TR74740_c13-g1_i1' appeared to be in the top 0.03% of expressed transcripts at under certain thermal conditions, and GAKY01194223.1 appeared to exhibit moderately differential expression at the extremes of temperature and ionic stress in a cultured host (Levin et al. 2017, Baumgarten et al. 2013).

While viral *RdRps* have been leveraged by eukaryotes in multiple pathways (Lipardi and Paterson 2010), the apparent fragmentation of the putative dinoRNAV EVEs *in silico* may indicate a role in triggering antiviral mechanisms within their hosts (Blair et al. 2020, Suzuki et al. 2020). Given that the *Symbiodinium* genome contains all core RNAi protein machinery, including Argonaute and Dicer, and that GAKY01194223.1 folds into several hairpins (ΔG = -142.5kcal/mol; Supplemental Figure 6 examples), Symbiodiniaceae may use the putative EVE ncRNA identified here to develop host immunity against extant, exogenous dinoRNAVs. Furthermore, Symbiodiniaceae harboring dinoRNAV EVEs also contained numerous non-retroviral EVEs of other viral families (Supplemental Table 9, Figure 7) in close proximity, such as *Herpesviridae, Baculoviridae, Poxviridae, Iridoviridae, Phycodnaviridae, Pandoraviridae* and *Pithoviridae*, ssDNA viruses of the family *Shotokuvirae*, -ssRNA viruses from the family *Rhabdoviridae* and +ssRNA viruses from the family *Coronaviridae* (Supplemental Figure 7). Metagenomes corroborate findings of similar *RdRps* from these viral families (Supplemental Figure 7). This provides support for host-mediated integration (e.g. retrotransposition) as a means of defense for single celled organisms, though further research is needed (Yan and Chen 2012).

Conclusions

Over recent decades, endogenous viral elements (EVEs) have enabled investigators to better understand the evolutionary history of viruses ("paleovirology") in diverse terrestrial systems, uncovering ancient and modern virus-host interactions. Our study further demonstrates how *in silico* identification of EVEs can provide ecological context for enigmatic viral genomes in non-model, multipartite systems such as coral holobionts, impacting how we study coral reefs and their viral consortia. Here, we detected heritable integrations of multiple putative dinoRNAV genes in Symbiodiniaceae scaffolds from cnidarian metagenomes, as well as in diverse genomes of cultured Symbiodiniaceae; no integrations were detected from seawater metagenomes nor diverse aposymbiotic cnidarian genomes. The apparent pervasive nature of dinoRNAV-like sequences among dinoflagellate genomes (especially the genus *Symbiodinium*) suggests widespread and recurrent/ancestral integration of these EVEs. We propose that host-provisioned mechanisms drive dinoRNAV integration into single-celled dinoflagellate genomes as EVEs. The findings presented in this study further validate the dinoRNAV-Symbiodiniaceae virus-host pair, enhancing our understanding of ecologically and economically important cnidarian holobionts and opening the door to examining the role of EVEs in reef health.

Acknowledgements and Contributions

Special thanks to the Tara Ocean Foundation, the R/V Tara crew and the Tara Pacific Expedition Participants (https://doi.org/10.5281/zenodo.3777760). Thank you also to Carsten G.B. Grupstra and Clark Hamor for their input regarding analyses. We are keen to thank the commitment of the following institutions for their financial and scientific support that made this unique Tara Pacific Expedition possible: CNRS, PSL, CSM, EPHE, Genoscope, CEA, Inserm, Université Côte d'Azur, ANR, agnès b., UNESCO-IOC, the Veolia Foundation, the Prince

Albert II de Monaco Foundation, Région Bretagne, Billerudkorsnas, AmerisourceBergen Company, Lorient Agglomération, Oceans by Disney, L'Oréal, Biotherm, France Collectivités, Fonds Français pour l'Environnement Mondial (FFEM), Etienne Bourgois, and the Tara Ocean Foundation teams. Tara Pacific would not exist without the continuous support of the participating institutes. This research is further supported by NSF OCE #2145472 to AMSC, NSF DOB Grant 2025457 to RLVT, and with additional support from NSF PRFB #1907184 to KSIB.

References

Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology*, *479–480*, 26–37. https://doi.org/10.1016/j.virol.2015.02.011

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

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V. B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., LaJeunesse, T. C., & Voolstra, C. R. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports*, *6*(1), 39734. https://doi.org/10.1038/srep39734

Aswad, S., Katzourakis, A. (2012). Paleovirology and virally derived immunity. Trends Ecol Evol, 27(11):627-36. https://doi.org/10.1016/j.tree.2012.07.007

Ballinger, M. J., Bruenn, J. A., & Taylor, D. J. (2012). Phylogeny, integration and expression of sigma virus-like genes in *Drosophila*. *Molecular Phylogenetics and Evolution*, *65*(1), 251–258. https://doi.org/10.1016/j.ympev.2012.06.008

Barreat, J.G.N., Katzourakis, A. (2022). Paleovirology of the DNA viruses of eukaryotes. Trends Microbiol. 30(3):281-292. https://doi.org/10.1016/j.tim.2021.07.004

Baumgarten, S., Bayer, T., Aranda, M., Liew, Y. J., Carr, A., Micklem, G., & Voolstra, C. R. (2013). Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. *BMC Genomics*, *14*(1), 704. https://doi.org/10.1186/1471-2164-14-704

Belyi, V.A., Levine, A.J., Skalka, A.M. (2010). Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. J Virol. 84(23):12458-62. https://doi.org/10.1128/JVI.01789-10.

Bigot, T., Temmam, S., Pérot, P., & Eloit, M. (2020). RVDB-prot, a reference viral protein database and its HMM profiles. *F1000 Research*, *8*, 530. https://doi.org/10.12688/f1000research.18776.2

Blair, C. D., Olson, K. E., & Bonizzoni, M. (2020). The Widespread occurrence and potential biological roles of endogenous viral elements in insect genomes. *Current Issues in Molecular Biology*, 13–30. https://doi.org/10.21775/cimb.034.013

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE*, *12*(10), e0185056. https://doi.org/10.1371/journal.pone.0185056

Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., & Suzuki, N. (2011). Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLOS Pathogens*, *7*(7), e1002146. https://doi.org/10.1371/journal.ppat.1002146

Chu, H., Jo, Y., & Cho, W. K. (2014). Evolution of endogenous non-retroviral genes integrated into plant genomes. *Current Plant Biology*, *1*, 55–59. https://doi.org/10.1016/j.cpb.2014.07.002

Correa, A. M. S., Ainsworth, T. D., Rosales, S. M., Thurber, A. R., Butler, C. R., & Vega Thurber, R. L. (2016). Viral outbreak in corals associated with an *in situ* bleaching event: atypical herpes-like viruses and a new Megavirus infecting *Symbiodinium*. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00127

Correa, A. M. S., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., & Weitz, J. S. (2021). Revisiting the rules of life for viruses of microorganisms. *Nature Reviews Microbiology*, *19*(8), 501–513. https://doi.org/10.1038/s41579-021-00530-x

Correa, A. M. S., Welsh, R. M., & Vega Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal*, *7*(1), 13–27. https://doi.org/10.1038/ismej.2012.75

Cui, J., Holmes, E.C. (2012). Endogenous lentiviruses in the ferret genome. J Virol. 86(6):3383-5. https://doi.org/10.1128/JVI.06652-11.

Feschotte, C., & Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics*, *13*(4), 283–296. https://doi.org/10.1038/nrg3199

Flynn, P. J., & Moreau, C. S. (2019). Assessing the diversity of endogenous viruses throughout ant genomes. *Frontiers in Microbiology*, *10*, 1139. https://doi.org/10.3389/fmicb.2019.01139

Frank, J. A., & Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Current Opinion in Virology*, *25*, 81–89. https://doi.org/10.1016/j.coviro.2017.07.021

Fujise, L., Suggett, D. J., Stat, M., Kahlke, T., Bunce, M., Gardner, S. G., Goyen, S., Woodcock, S., Ralph, P. J., Seymour, J. R., Siboni, N., & Nitschke, M. R. (2021). Unlocking the phylogenetic diversity, primary habitats, and abundances of free-living Symbiodiniaceae on a coral reef. *Molecular Ecology*, *30*(1), 343–360. https://doi.org/10.1111/mec.15719

Gallot-Lavallée, L., & Blanc, G. (2017). A Glimpse of Nucleo-Cytoplasmic Large DNA Virus biodiversity through the eukaryotic genomics window. *Viruses*, *9*(1), 17. https://doi.org/10.3390/v9010017

Gilbert, C. and Feschotte, C. (2010). Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol.8(9):e1000495. https://doi.org/10.1371/journal.pbio.1000495

Glynn, P. W. (1996). Coral reef bleaching: facts, hypotheses and implications. *Global Change Biology*, *2*(6), 495–509. https://doi.org/10.1111/j.1365-2486.1996.tb00063.x

González-Pech, R. A., Chen, Y., Stephens, T. G., Shah, S., Mohamed, A. R., Lagorce, R., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2019). *Genomes of Symbiodiniaceae reveal extensive sequence divergence but conserved functions at family and genus levels* [Preprint]. Genomics. https://doi.org/10.1101/800482

González-Pech, R. A., Stephens, T. G., Chen, Y., Mohamed, A. R., Cheng, Y., Shah, S., Dougan, K. E., Fortuin, M. D. A., Lagorce, R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2021). Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*. *BMC Biology*, *19*(1), 73. https://doi.org/10.1186/s12915-021-00994-6

Grupstra, C. G. B., Rabbitt, K. M., Howe-Kerr, L. I., & Correa, A. M. S. (2021). Fish predation on corals promotes the dispersal of coral symbionts. *Animal Microbiome*, *3*(1), 25. https://doi.org/10.1186/s42523-021-00086-4

Grupstra, C. G. B., L. I. Howe-Kerr, A. J. Veglia, R. L. Bryant, S. R. Coy, P. L. Blackwelder and A. M. S. Correa (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *The ISME Journal*. 10.1038/s41396-022-01194-y

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Harrison, E., & Brockhurst, M. A. (2017). Ecological and evolutionary benefits of temperate phage: what does or doesn't kill you makes you stronger. *BioEssays*, *39*(12), 1700112. https://doi.org/10.1002/bies.201700112

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Holmes, E. C. (2011). The evolution of endogenous viral elements. *Cell Host & Microbe*, *10*(4), 368–377. https://doi.org/10.1016/j.chom.2011.09.002

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491. https://doi.org/10.1186/1471-2105-12-491

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J. M., & Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature*, *463*(7277), 84–87. https://doi.org/10.1038/nature08695

Hume, B. C. C., Smith, E. G., Ziegler, M., Warrington, H. J. M., Burt, J. A., LaJeunesse, T. C., Wiedenmann, J., & Voolstra, C. R. (2019). SymPortal: A novel analytical framework and platform for coral algal symbiont next-generation sequencing *ITS2* profiling. *Molecular Ecology Resources*, *19*(4), 1063–1080. https://doi.org/10.1111/1755-0998.13004

Hume, Benjamin C.C., Poulain, Julie, Pesant, Stéphane, Belser, Caroline, Ruscheweyh, Hans-Joachim, Moulin, Clémentine, Boissin, Emilie, Bourdin, Guillaume, Iwankow, Guillaume, Romac, Sarah, Agostini, Sylvain, Banaigs, Bernard, Boss, Emmanuel, Bowler, Chris, de Vargas, Colomban, Douville, Eric, Flores, J. Michel, Forcioli, Didier, Furla, Paola, … Voolstra, Christian R. (2020). Tara Pacific metabarcoding sequencing (16S, 18S, ITS2) reference & replication tables version 1 (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4073035

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

Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., Bright, K. J., Imanian, B., Strom, S. L., Delwiche, C. F., Waller, R. F., Fensome, R. A., Leander, B. S., Rohwer, F. L., & Saldarriaga, J. F. (2017). Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences*, *114*(2), E171–E180. https://doi.org/10.1073/pnas.1614842114

Jebb, D., Huang, Z., Pippel, M., Hughes, G. M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L. S., Skirmuntt, E. C., Katzourakis, A., Burkitt-Gray, L., Ray, D. A., Sullivan, K. A. M., Roscito, J. G., Kirilenko, B. M., Dávalos, L. M., Corthals, A. P., Power, M. L., Jones, G., … Teeling, E. C. (2020). Six reference-quality genomes reveal evolution of bat adaptations. *Nature*, *583*(7817), 578–584. https://doi.org/10.1038/s41586-020-2486-3

Jern, P., & Coffin, J. M. (2008). Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics*, *42*(1), 709–732. https://doi.org/10.1146/annurev.genet.42.110807.091501

Johnson, W. E. (2010). Endless Forms Most Viral. *PLoS Genetics*, *6*(11), e1001210. https://doi.org/10.1371/journal.pgen.1001210

Johnson, W. E. (2015). Endogenous retroviruses in the genomics era. *Annual Review of Virology*, *2*(1), 135–159. https://doi.org/10.1146/annurev-virology-100114-054945

Johnson, W. E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses. *Nature Reviews Microbiology*, *17*(6), 355–370. https://doi.org/10.1038/s41579-019-0189-2

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Katzourakis, A., Gifford, R.J., Tristem, M., Gilbert, M.T., Pybus, O.G. (2009) Macroevolution of complex retroviruses. Science. 18;325(5947):1512. https://doi.org/10.1126/science.1174149.

Katzourakis, A. (2013). Paleovirology: inferring viral evolution from host genome sequence data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1626), 20120493. https://doi.org/10.1098/rstb.2012.0493

Katzourakis, A., & Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLOS Genetics*, *6*(11), e1001191. https://doi.org/10.1371/journal.pgen.1001191

Kawasaki, J., Kojima, S., Mukai, Y., Tomonaga, K., Horie, M. (2021). 100-My history of bornavirus infections hidden in vertebrate genomes. Proc Natl Acad Sci U S A. 118(20):e2026235118. https://doi.org/10.1073/pnas.2026235118.

Keckesova, Z., Ylinen, L.M., Towers, G.J., Gifford, R.J., Katzourakis, A. (2009) Identification of a RELIK orthologue in the European hare (Lepus europaeus) reveals a minimum age of 12 million years for the lagomorph lentiviruses. Virology.5;384(1):7-11. https://doi.org/10.1016/j.virol.2008.10.045.

Knowlton, N., & Rohwer, F. (2003). Multispecies Microbial Mutualisms on Coral Reefs: The Host as a Habitat. *The American Naturalist*, *162*(S4), S51–S62. https://doi.org/10.1086/378684

Kojima, S., Yoshikawa, K., Ito, J., Nakagawa, S., Parrish, N. F., Horie, M., Kawano, S., & Tomonaga, K. (2021). Virus-like insertions with sequence signatures similar to those of endogenous nonretroviral RNA viruses in the human genome. *Proceedings of the National Academy of Sciences*, *118*(5). https://doi.org/10.1073/pnas.2010758118

Kolekar, P., Pataskar, A., Kulkarni-Kale, U. et al. (2016). IRESPred: Web Server for Prediction of Cellular and Viral Internal Ribosome Entry Site (IRES). *Sci Rep*, 6:27436. https://doi.org/10.1038/srep27436

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic revision of symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Current Biology*, *28*(16), 2570-2580.e6. https://doi.org/10.1016/j.cub.2018.07.008

Lawrence, S. A., Davy, J. E., Aeby, G. S., Wilson, W. H., & Davy, S. K. (2014). Quantification of virus-like particles suggests viral infection in corals affected by *Porites* tissue loss. *Coral Reefs*, *33*(3), 687–691. https://doi.org/10.1007/s00338-014-1168-8

Levin, R. A., Voolstra, C. R., Weynberg, K. D., & van Oppen, M. J. H. (2017). Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *The ISME Journal*,

11(3), 808–812. https://doi.org/10.1038/ismej.2016.154

Li, Y.Q., Bletsa, M., Zisi, Z., Boonen, I. Gryseels, S., Kafetzopoulou, L., Webster, J.P., Catalano, S., Pybus, O.G., Van de Perre, F., Li, H.T., Li, Y.Y., Li, Y.C., Abramov, A., Lymberakis, P., Lemey, P., Lequime S. (2022). Discovery of Flaviviridae-derived endogenous viral elements in shrew genomes provide novel insights into Pestivirus ancient history. bioRxiv. 02.11.480044. https://doi.org/10.1101/2022.02.11.480044

Lidie, K.B., van Dolah, F.M. (2007). Spliced leader RNA-mediated trans-splicing in a dinoflagellate, Karenia brevis. J Eukaryot Microbiol. 54(5):427-35. https://doi.org/10.1111/j.1550-7408.2007.00282.x

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z., Cai, M., Zhuang, Y., Shi, X., Lin, L., Wang, L., Wang, Z., Liu, X., Yu, S., Zeng, P., … Morse, D. (2015). The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, *350*(6261), 691–694. https://doi.org/10.1126/science.aad0408

Lipardi, C., & Paterson, B. M. (2009). Identification of an RNA-dependent RNA polymerase in *Drosophila* involved in RNAi and transposon suppression. *Proceedings of the National Academy of Sciences*, *106*(37), 15645–15650. https://doi.org/10.1073/pnas.0904984106

Littman, R. A., van Oppen, M. J. H., & Willis, B. L. (2008). Methods for sampling free-living *Symbiodinium* (zooxanthellae) and their distribution and abundance at Lizard Island (Great Barrier Reef). *Journal of Experimental Marine Biology and Ecology*, *364*(1), 48–53. https://doi.org/10.1016/j.jembe.2008.06.034

Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., Cooke, I., Aranda, M., Bourne, D. G., Forêt, S., Miller, D. J., van Oppen, M. J. H., Voolstra, C. R., Ragan, M. A., & Chan, C. X. (2018). *Symbiodinium* genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Communications Biology*, *1*(1), 95. https://doi.org/10.1038/s42003-018-0098-3

Matthews, J. L., Raina, J., Kahlke, T., Seymour, J. R., Oppen, M. J. H., & Suggett, D. J. (2020). Symbiodiniaceae-bacteria interactions: rethinking metabolite exchange in reef-building corals as multi-partner metabolic networks. *Environmental Microbiology*, *22*(5), 1675–1687. https://doi.org/10.1111/1462-2920.14918

Messyasz, A., Rosales, S. M., Mueller, R. S., Sawyer, T., Correa, A. M. S., Thurber, A. R., & Vega Thurber, R. (2020). Coral bleaching phenotypes associated with differential abundances of Nucleocytoplasmic Large DNA Viruses. *Frontiers in Marine Science*, *7*, 555474. https://doi.org/10.3389/fmars.2020.555474

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mita, P., & Boeke, J. D. (2016). How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development*, *37*, 90–100. https://doi.org/10.1016/j.gde.2016.01.001

Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., & Aylward, F. O. (2020). Widespread endogenization of giant viruses shapes genomes of green algae. *Nature*, *588*(7836), 141–145. https://doi.org/10.1038/s41586-020-2924-2

Montalvo-Proaño, J., Buerger, P., Weynberg, K. D., & van Oppen, M. J. H. (2017). A PCR-Based Assay targeting the major capsid protein gene of a Dinorna-Like ssRNA virus that

infects coral photosymbionts. *Frontiers in Microbiology*, *8*, 1665. https://doi.org/10.3389/fmicb.2017.01665

Mortelmans, K., Wang-Johanning, F., Johanning, G. L. (2016). The role of human endogenous retroviruses in brain development and function. *APMIS.* 124, 105–115. doi: 10.1111/apm.12495. PMID: 26818265.

Muller, E. M., Bartels, E., & Baums, I. B. (2018). Bleaching causes loss of disease resistance within the threatened coral species Acropora cervicornis. *ELife*, *7*, e35066. https://doi.org/10.7554/eLife.35066

Nagasaki, K., Shirai, Y., Takao, Y., Mizumoto, H., Nishida, K., & Tomaru, Y. (2005). Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate *Heterocapsa circularisquama*. *Applied and Environmental Microbiology*, *71*(12), 8888–8894. https://doi.org/10.1128/AEM.71.12.8888-8894.2005

Nand, A., Zhan, Y., Salazar, O. R., Aranda, M., Voolstra, C. R., & Dekker, J. (2021). Genetic and spatial organization of the unusual chromosomes of the dinoflagellate *Symbiodinium microadriaticum*. *Nature Genetics*, *53*(5), 618–629. https://doi.org/10.1038/s41588-021-00841-y

Oliveira, N. M., Satija, H., Kouwenhoven, I. A., & Eiden, M. V. (2007). Changes in viral protein function that accompany retroviral endogenization. *Proceedings of the National Academy of Sciences*, *104*(44), 17506–17511. https://doi.org/10.1073/pnas.0704313104

Palatini, U., Miesen, P., Carballar-Lejarazu, R., Ometto, L., Rizzo, E., Tu, Z., van Rij, R. P., & Bonizzoni, M. (2017). Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. *BMC Genomics*, *18*(1), 512. https://doi.org/10.1186/s12864-017-3903-3

Parker, B. J., & Brisson, J. A. (2019). A laterally transferred viral gene modifies aphid wing plasticity. *Current Biology*, *29*(12), 2098-2103.e5. https://doi.org/10.1016/j.cub.2019.05.041

Patel, M. R., Emerman, M., & Malik, H. S. (2011). Paleovirology—ghosts and gifts of viruses past. *Current Opinion in Virology*, *1*(4), 304–309. https://doi.org/10.1016/j.coviro.2011.06.007

Pesant, Stéphane, Lombard, Fabien, Bourdin, Guillaume, Poulain, Julie, Petit, Emmanuelle, Boss, Emmanuel, Cassar, Nicolas, Cohen, Natalie R., Dimier, Céline, Douville, Eric, Flores, J. Michel, Gorsky, Gabriel, Hume, Benjamin C.C., John, Seth G., Kelly, Rachel L., Lin, Yajuan, Marie, Dominique, Pedrotti, Maria-Luiza, Pujo Pay, Mireille, … Allemand, Denis. (2020). *Tara Pacific samples provenance and environmental context - version 1*. Zenodo. https://doi.org/10.5281/ZENODO.4068293

Qin, Z., Yu, K., Chen, B., Wang, Y., Liang, J., Luo, W., Xu, L., Huang, X. (2019). Diversity of Symbiodiniaceae in 15 Coral Species From the Southern South China Sea: Potential Relationship With Coral Thermal Adaptability. Front Microbiol. 10:2343. https://doi.org/10.3389/fmicb.2019.02343

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Robbins, S. J., ReFuGe2020 Consortium, Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A. U., Ying, H., Baker, A., Bell, S. C., Morrow, K. M., Ragan, M. A., Miller, D. J., Forêt, S., Voolstra, C. R., Tyson, G. W., & Bourne, D. G. (2019). A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nature Microbiology*, *4*(12),

2090–2100. https://doi.org/10.1038/s41564-019-0532-4
Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nature Reviews Microbiology*, *9*(2), 99–108. https://doi.org/10.1038/nrmicro2491
Scheufen, T., Iglesias-Prieto, R., & Enríquez, S. (2017). Changes in the number of symbionts and *Symbiodinium* cell pigmentation modulate differentially coral light absorption and photosynthetic performance. *Frontiers in Marine Science*, *4*, 309. https://doi.org/10.3389/fmars.2017.00309
Shoguchi, E., Beedessee, G., Hisata, K., Tada, I., Narisoko, H., Satoh, N., Kawachi, M., & Shinzato, C. (2021). A new dinoflagellate genome illuminates a conserved gene cluster involved in sunscreen biosynthesis. *Genome Biology and Evolution*, *13*(2), evaa235. https://doi.org/10.1093/gbe/evaa235
Shoguchi, E., Beedessee, G., Tada, I., Hisata, K., Kawashima, T., Takeuchi, T., Arakaki, N., Fujie, M., Koyanagi, R., Roy, M. C., Kawachi, M., Hidaka, M., Satoh, N., & Shinzato, C. (2018). Two divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC Genomics*, *19*(1), 458. https://doi.org/10.1186/s12864-018-4857-9
Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., … Satoh, N. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*, *23*(15), 1399–1408. https://doi.org/10.1016/j.cub.2013.05.062
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539. https://doi.org/10.1038/msb.2011.75
Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351
Skirmuntt, E. C., Escalera-Zamudio, M., Teeling, E. C., Smith, A., & Katzourakis, A. (2020). The potential role of Endogenous Viral Elements in the evolution of bats as reservoirs for zoonotic viruses. *Annual Review of Virology*, *7*(1), 103–119. https://doi.org/10.1146/annurev-virology-092818-015613
Sofuku, K., & Honda, T. (2018). *Influence of Endogenous Viral Sequences on Gene Expression*. IntechOpen. https://doi.org/10.5772/intechopen.71864
Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(Web Server), W435–W439. https://doi.org/10.1093/nar/gkl200
Stephens, T. G., González-Pech, R. A., Cheng, Y., Mohamed, A. R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2020). Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated single-exon genes with adaptive functions. *BMC Biology*, *18*(1), 56. https://doi.org/10.1186/s12915-020-00782-8
Stoye, J. P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nature Reviews Microbiology*, *10*(6), 395–406. https://doi.org/10.1038/nrmicro2783
Suzuki, Y., Baidaliuk, A., Miesen, P., Frangeul, L., Crist, A. B., Merkling, S. H., Fontaine, A., Lequime, S., Moltini-Conclois, I., Blanc, H., van Rij, R. P., Lambrechts, L., & Saleh, M.-

C. (2020). Non-retroviral Endogenous Viral Element limits cognate virus replication in *Aedes aegypti* ovaries. *Current Biology*, *30*(18), 3495-3506.e6. https://doi.org/10.1016/j.cub.2020.06.057

Takahashi, H., Fukuhara, T., Kitazawa, H., & Kormelink, R. (2019). Virus latency and the impact on plants. *Frontiers in Microbiology*, *10*, 2764. https://doi.org/10.3389/fmicb.2019.02764

Tan, Y.T.R., Wainwright, B.J., Afiq-Rosli, L., Ip, Y.C.A., Lee, J.N., Nguyen, N.T.H., Pointing, S.B., Huang, D. (2020). Endosymbiont diversity and community structure in Porites lutea from Southeast Asia are driven by a suite of environmental variables. Symbiosis 80, 269–277. https://doi.org/10.1007/s13199-020-00671-2

Tara Oceans Consortium Coordinators, Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., & Searson, S. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, *2*(1), 150023. https://doi.org/10.1038/sdata.2015.23

ter Horst, A. M., Nigg, J. C., Dekker, F. M., & Falk, B. W. (n.d.). Endogenous Viral Elements are widespread in arthropod genomes and commonly give rise to PIWI-interacting RNAs. *Journal of Virology*, *93*(6), e02124-18. https://doi.org/10.1128/JVI.02124-18

Vega Thurber, R., Payet, J. P., Thurber, A. R., & Correa, A. M. S. (2017). Virus–host interactions and their roles in coral reef health and disease. *Nature Reviews Microbiology*, *15*(4), 205–216. https://doi.org/10.1038/nrmicro.2016.176

Tromas, N., Zwart, M. P., Forment, J., & Elena, S. F. (2014). Shrinkage of genome size in a plant RNA virus upon transfer of an essential viral gene into the host genome. *Genome Biology and Evolution*, *6*(3), 538–550. https://doi.org/10.1093/gbe/evu036

van Oppen, M. J. H., Leong, J.-A., & Gates, R. D. (2009). Coral-virus interactions: A double-edged sword? *Symbiosis*, *47*(1), 1–8. https://doi.org/10.1007/BF03179964

Wang, L., Wu, S., Liu, T., Sun, J., Chi, S., Liu, C., Li, X., Yin, J., Wang, X., & Yu, J. (2014). Endogenous viral elements in algal genomes. *Acta Oceanologica Sinica*, *33*(2), 102–107. https://doi.org/10.1007/s13131-014-0446-1

Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A., & van Oppen, M. J. H. (2014). Generating viral metagenomes from the coral holobiont. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00206

Whitfield, Z. J., Dolan, P. T., Kunitomi, M., Tassetto, M., Seetin, M. G., Oh, S., Heiner, C., Paxinos, E., & Andino, R. (2017). The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Current Biology*, *27*(22), 3511-3519.e7. https://doi.org/10.1016/j.cub.2017.09.067

Wilson, W., Francis, I., Ryan, K., & Davy, S. (2001). Temperature induction of viruses in symbiotic dinoflagellates. *Aquatic Microbial Ecology*, *25*, 99–102. https://doi.org/10.3354/ame025099

Wood-Charlson, E. M., Weynberg, K. D., Suttle, C. A., Roux, S., & van Oppen, M. J. H. (2015). Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise: Methodological biases in coral viromics. *Environmental Microbiology*, *17*(10), 3440–3449. https://doi.org/10.1111/1462-2920.12803

Yan, N., Chen, Z. (2012) Intrinsic antiviral immunity. *Nat Immunol,* 13, 214–222. https://doi.org/10.1038/ni.2229

Zhang, H., Hou, Y., Miranda, L., Lin., S. (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci.* 104 (11) 4618-4623.

https://doi.org/10.1073/pnas.070025810
Zhang, H., Zhuang, Y., Gill, J., & Lin, S. (2013). Proof that dinoflagellate Spliced Leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium kawagutii* as a case study. *Protist*, *164*(4), 510–527. https://doi.org/10.1016/j.protis.2013.04.002
Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595

Chapter 2

vAMPirus: A versatile amplicon processing and analysis program for studying viruses

Alex J. Veglia[1], Ramón E. Rivera-Vicéns[2,3], Carsten G.B. Grupstra[1,4], Lauren I. Howe-Kerr[1], and Adrienne M.S. Correa[1]

This chapter is in revision at *Molecular Ecology Resources* and currently available as a preprint: https://doi.org/10.22541/au.167592965.50010700/v1

Abstract

Amplicon sequencing is an effective and increasingly applied method for studying viral communities in the environment. Here, we present vAMPirus, a user-friendly, comprehensive, and versatile DNA and RNA virus amplicon sequence analysis program, designed to support investigators in exploring virus amplicon sequencing data and running informed, reproducible analyses. vAMPirus intakes raw virus amplicon libraries and, by default, performs nucleotide- and protein-based analyses to produce results such as sequence abundance information, taxonomic classifications, phylogenies, and community diversity metrics. The vAMPirus pipelines additionally include optional approaches that can increase the biological signal-to-noise ratio in results by leveraging tools not yet commonly applied to virus amplicon data analyses. In this paper, we validate the vAMPirus analytical framework and illustrate its implementation into the general virus amplicon sequencing workflow by recapitulating findings from two previously published double-stranded DNA virus datasets. As a case study, we also apply the program to explore the diversity and distribution of a coral reef-associated RNA virus. vAMPirus is incorporated with the Nextflow workflow manager, offering straightforward scalability, standardization, and communication of virus lineage-specific analyses. The vAMPirus framework itself is also designed to be adaptable; community-driven analytical standards will continue to be incorporated as the field advances. vAMPirus supports researchers in revealing patterns of virus diversity and population dynamics in nature, while promoting study reproducibility and comparability.

1 Introduction

From the human gut to sediments in the deep ocean, viruses are abundant, diverse, and shape the systems they inhabit (Breitbart et al., 2018; A. M. S. Correa et al., 2021; Suttle, 2007). The advent of high-throughput sequencing (HTS) techniques like amplicon sequencing has transformed the field of virology, illuminating the currently unculturable virosphere (Labadie et al., 2020; Metcalf et al., 1995; Paez-Espino et al., 2017; Zayed et al., 2022) and helping identify the impacts of viruses on ecosystem and host function (Braga et al., 2020; Breitbart et al., 2018; Thurber et al., 2017; Uyaguari-Diaz et al., 2016). Amplicon sequencing is a targeted, polymerase chain reaction (PCR)-based HTS approach that allows deep characterization of genetic variants within virus populations (Short et al. 2010). The targeted nature of amplicon sequencing reduces the economic and computational investment required for spatiotemporal investigations of virus communities at ecologically relevant scales (see Finke & Suttle, 2019; Frantzen & Holo, 2019; Grupstra et al., 2022; Gustavsen & Suttle, 2021; Howe-Kerr et al., 2022; Montalvo-Proaño et al., 2017). The number of studies leveraging virus amplicon sequencing has increased rapidly over the last two decades (e.g., 16 peer-reviewed publications in 1998 compared to 127 in 2021 based on a Web of Science search of ‘virus amplicon sequencing’, November 2022).

The general virus amplicon sequencing workflow includes: 1. Extraction of virus nucleic acid (DNA or RNA), 2. PCR amplification of virus marker gene or transcript, 3. HTS of virus marker gene amplicons, and 4. Bioinformatic analysis of sequencing data (Short et al., 2010; Figure 1). The effective analysis and interpretation of amplicon sequencing data relies on biologically accurate binning of marker gene sequences into taxonomically or ecologically distinct units. Recognizing viral taxa or ecotypes, however, can be challenging. For example, non-model viruses have limited baseline information available to inform the selection of

clustering thresholds. Other viruses, such as RNA viruses, have error-prone polymerases and produce quasispecies, a population structure consisting of large numbers of variant genomes (Domingo & Perales, 2019) that may not be easily resolved by the same clustering percentage. Amplicon sequence variants (ASVs) are a promising non-clustering-based approach for virus amplicon analyses that offers high precision and biological accuracy as error-derived sequence variants are removed during ASV generation (Callahan et al., 2017; Edgar, 2016b). In addition, since the identity of an ASV is not specific to a given dataset (as identity can be in clustering of marker gene sequences into *de novo* OTUs based on a percent identity value, Callahan et al., 2017), ASVs and their unique translations ('aminotypes', see Grupstra et al., 2022) can be compared directly among studies (Callahan et al., 2017).

To promote the standardization, reproducibility and cross-comparison of DNA and RNA virus amplicon sequence analyses, we developed the automated bioinformatics tool, vAMPirus (github.com/Aveglia/vAMPirus). vAMPirus intakes raw (unprocessed) virus amplicon libraries, performs all read processing and diversity analysis steps, and produces reports detailing results (e.g., relative abundance plots, community diversity metrics) with interactive figures and tables. vAMPirus supports initial explorations of viral amplicon sequence datasets via a 'DataCheck' pipeline, which generates an HTML report with information on data quality and sequence diversity. Results from the exploratory DataCheck pipeline can then be used to optimize parameters in the read processing or ASV generation steps within the vAMPirus 'Analyze' pipeline; this can improve the signal-to-noise ratio in downstream analyses. vAMPirus is integrated with the Nextflow workflow manager, which uses a configuration file that can be shared among investigators, facilitating the standardization and dissemination of virus amplicon sequence analyses across projects and research groups. To that end, we also created the

vAMPirus Analysis Repository (https://zenodo.org/communities/vampirusrepo/) to act as a central location for all published vAMPirus analyses. vAMPirus is intended to be accessible to researchers with a range of bioinformatics experience levels, and includes substantial help documentation with step-by-step instructions for running the tool (https://github.com/Aveglia/vAMPirus/blob/master/docs/). By facilitating the standardization of viral lineage-specific analyses and increasing the signal-to-noise ratio in community diversity analyses, vAMPirus will enhance the effectiveness of virus amplicon studies and lead to a more developed understanding the global virosphere.

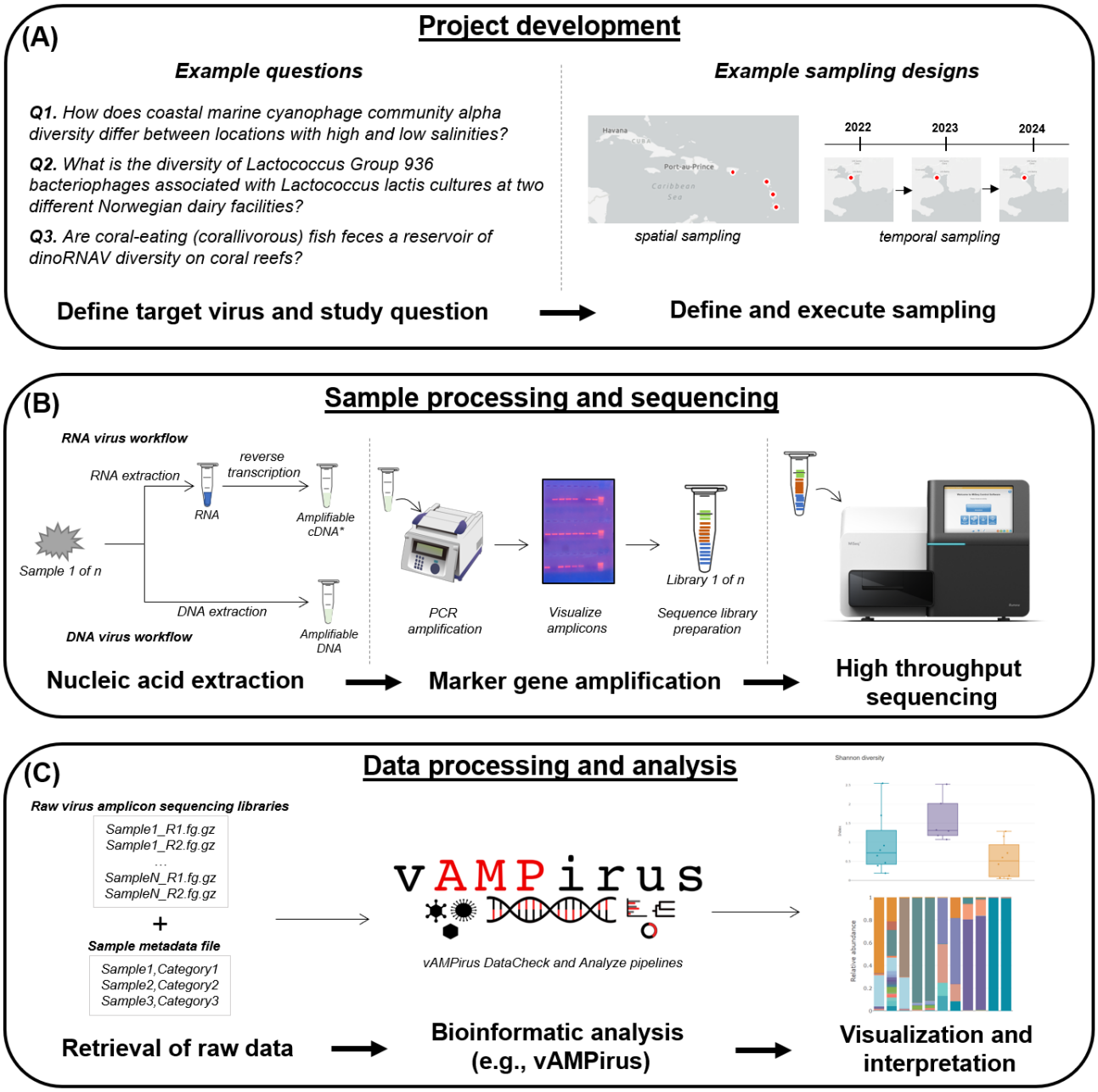


Figure 1. General workflow of virus amplicon sequencing projects (bolded text). *cDNA = complementary DNA.

2 Materials and methods

This section outlines the pipelines and processes therein that comprise vAMPirus v2.1.0 (Figure 2). For a more detailed explanation of vAMPirus processes and output, see the manual stored in the vAMPirus GitHub repository (github.com/Aveglia/vAMPirus/).

2.1 vAMPirus implementation and configuration

2.1.1 Overview of vAMPirus execution

vAMPirus is composed of three main components that are recommended to be deployed sequentially: 1. A startup script to install dependencies and databases for taxonomy processes, 2. A 'DataCheck' pipeline that provides users with detailed information on data quality and diversity to inform subsequent analysis, and 3. An 'Analyze' pipeline that runs a comprehensive biology-focused analysis of the data using specified parameters and program options. vAMPirus is incorporated with Nextflow, a scientific workflow manager that allows easy configuration and deployment of the program using Conda, Docker, Singularity or cloud systems like Amazon Web Services (Di Tommaso et al., 2017). Nextflow natively communicates with scheduling managers like SLURM, PBS, or Torque, making it easy to run vAMPirus on high-performance computing clusters or on a local laptop or workstation. vAMPirus analyses can be configured using the Nextflow configuration file to promote efficient utilization of computing resources and reduce run times. Real-time monitoring and remote launching of vAMPirus analyses can be done using Nextflow Tower with no alterations to the vAMPirus script.

2.1.1 vAMPirus startup script

A startup script written in BASH is provided within the vAMPirus installation directory that will automatically install dependencies and prepare the vampirus.config file for use. Users can deploy this script to download the Nextflow workflow manager and Conda package management system if these programs are not already installed/accessible on the computer

system. The script can also be directed to download one or more protein/taxonomy databases to be used in vAMPirus taxonomy processes. Available databases include: 1. The proteic version of the Reference Virus DataBase (RVDB, Bigot et al., 2020), 2. NCBI virus protein RefSeq database (Brister et al., 2015), and 3. Complete NCBI NR protein database (O'Leary et al., 2016). If directed to do so, the startup script will also download the NCBI Taxonomy Database (Schoch et al., 2020) and last common ancestor (LCA) information for sequences curated within the RVDB (Bigot et al., 2020). The script then edits the vAMPirus configuration file with the updated paths to any downloaded databases and to the vAMPirus installation directory. Lastly, text documents that include general next steps for the user and commands to test the installation are printed in the vAMPirus directory. If test analyses complete successfully, the user then updates the configuration file with project-specific parameters (e.g., project name, database for taxonomy inference, primer sequence information, number of allocated threads, working memory, scheduling manager) prior to running vAMPirus on a dataset.

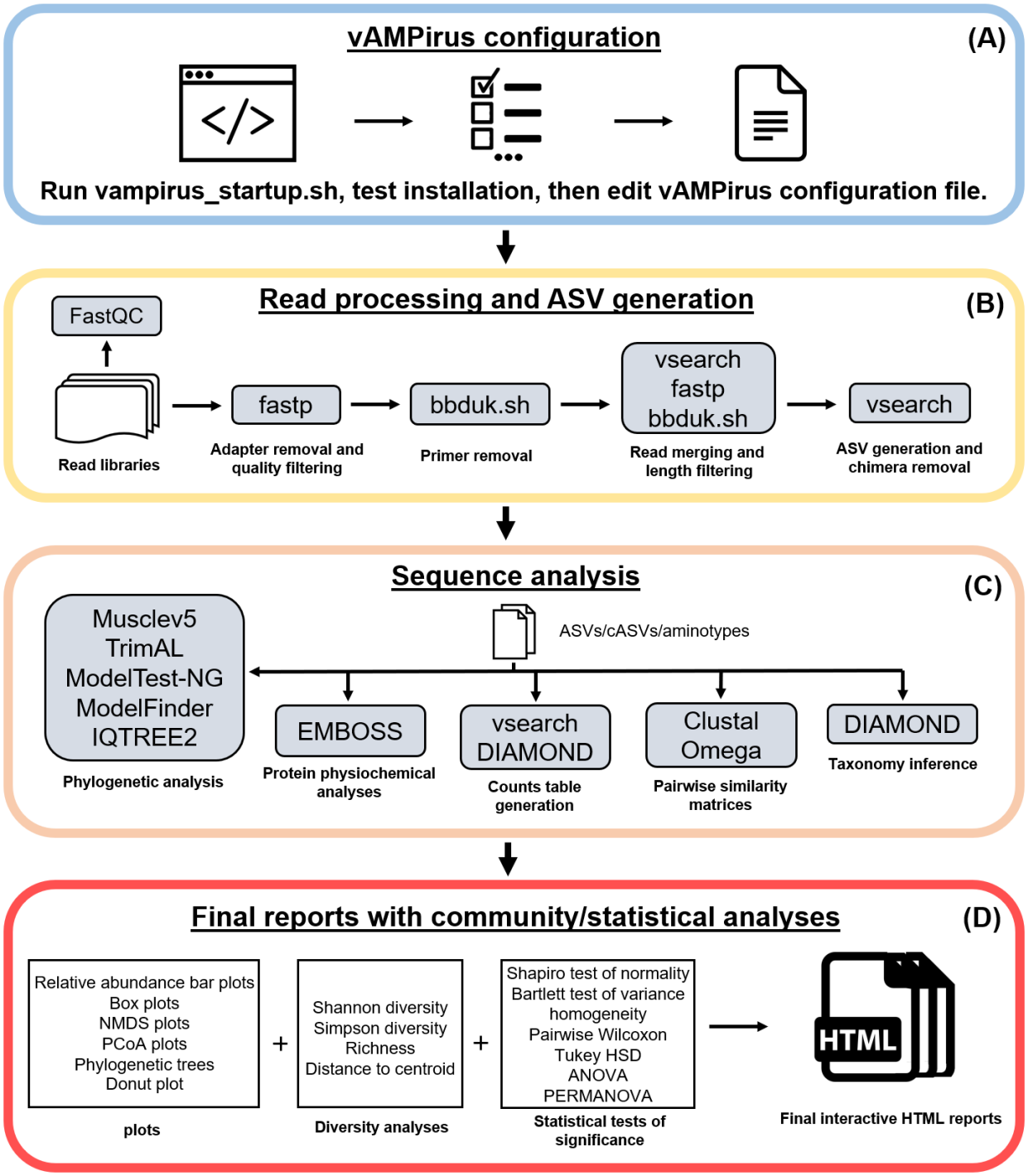


Figure 2. Generalized flowchart of vAMPirus v2.1.0, illustrating its configuration (box A), default analyses and programs used within the read processing (box B) and Analyze (boxes C and D) pipelines. For simplicity, only selected connections between processes are highlighted; processes generating the unique amino acid sequences ('aminotypes') and clustered ASVs ('cASVs'), as well as those specific to the DataCheck pipeline (see Supplemental Figure S2) are omitted. See Supplemental Figures S1, S2, and S3 for a more comprehensive illustration of vAMPirus pipelines and the processes therein.

2.2 Overview of the processes performed within vAMPirus

2.2.1 Read processing and generation of amplicon sequence variants

vAMPirus supports single- and paired-end raw Illumina read libraries as input. By default, read processing and ASV generation processes are performed prior to entering the DataCheck or Analyze pipelines (Figure 2, yellow box; Supplemental Figure S1). The read processing pipeline begins with a check of raw libraries using FastQC (v0.11.9, Andrews 2010), which creates and stores reports for review by the user. As FastQC is running, the program fastp (v0.20.1, Chen et al., 2018) automatically detects and removes adapter contamination, and performs quality/length filtering based on user-set parameters in the configuration file. fastp also performs over-representation analysis and (for paired-end input) base error correction during this step. Next, primers are removed from adapter-less reads using the bbduk.sh program within the BBTools software package (Bushnell 2014), and then another FastQC report is generated and stored. Cleaned reads are then merged using the program VSEARCH (v2.21.1, Rognes et al., 2016) and merged reads (from all libraries) are then concatenated into a single fastq file. For accurate ASV generation, it is imperative that the merged reads be the same length (Edgar, 2016b). To ensure this, merged reads are globally trimmed to a user-specified maximum read length using fastp. Merged reads with the set length are then extracted from the total merged read file using the program bbduk.sh and dereplicated using the program VSEARCH (v2.21.1, Rognes et al., 2016), producing a unique read file containing read representation information. Amplicon sequence variants are then generated from this unique read file with VSEARCH and the UNOISE3 algorithm (Edgar, 2016b; Rognes et al., 2016). Chimeric ASVs are detected and removed using VSEARCH and the UCHIME3 algorithm (Edgar, 2016a). Prior to entering downstream pipelines, vAMPirus provides users the option to filter ASVs with DIAMOND blastx (v2.0.15, Buchfink et al., 2015) to remove non-target sequences or to focus their analyses

on a subset of ASVs/aminotypes. These steps produce a final ASV fasta file that is then used as input for the DataCheck and Analyze pipelines.

2.2.2 Amplicon sequence variants and 'aminotypes'

By default, vAMPirus generates nucleotide-based (ASV) and protein-based (aminotype) results. ASVs support cross-study comparisons and offer a statistically supported view of virus sequence diversity, as biologically inaccurate sequences are removed during denoising (Callahan et al., 2017; Edgar, 2016b). However, ASV results for virus lineages with high mutation rates (e.g., RNA viruses with quasispecies heterogeneity) may still contain high levels of noise that mask biological patterns. It may be beneficial to group ASVs into distinct clusters based on genetic or ecological similarities in such use cases. In vAMPirus, 'aminotypes' (unique amino acid sequences, Grupstra et al. 2022) are generated by translating ASVs with VirtualRibosome (v2.0, Wernersson, 2006) and subsequently dereplicating these translations using the program CD-HIT (v4.8.1, Fu et al., 2012; Li & Godzik, 2006). As direct products of specific ASVs, aminotypes maintain sequence tractability, reproducibility, and comparability, and therefore differ from *de novo* OTUs or cASVs (see Section 2.2.3). The 'aminotyping' approach not only reduces noise; it also removes sequences with internal stop codons (a deleterious mutation) and reveals nonsynonymous mutations that may indicate differences in virus functionality (e.g., infection efficiency, host range; DeFilippis & Villarreal, 2000).

vAMPirus provides two additional (optional) ASV or aminotype "grouping" approaches that are alternatives to *de novo* clustering: Minimum Entropy Decomposition (MED) and phylogeny-based clustering or 'phylogrouping'. MED is a method of sequence clustering that utilizes Shannon entropy (Shannon, 1948) to partition marker gene datasets into 'MED nodes' (Eren et al., 2015). With this approach, users identify sequence positions in a set of ASVs or aminotypes that are information-rich (positions of high variability) or information-poor

(positions of high conservation) and use these positions to assign ASVs/aminotypes to 'MED groups' (sequences with identical bases at specified positions) (Eren et al., 2015). Users can also specify and assign sequences to MED groups based on sequence positions of interest (e.g. positions of a protein sequence known to influence a viral characteristic such as host cell attachment; see Harvey et al., 2021). Phylogrouping is performed with the TreeCluster program (v1.0.3, Balaban et al., 2019). With this approach, ASV or aminotype sequences are assigned to "phylogroups" based on user specified TreeCluster parameters and the phylogenetic tree produced during analysis (see Figure 4-V, VI). All grouping methods can be applied at the same time; coupled with the use of the Nextflow '--resume' feature, adjusting specific parameters and generating new results to review and compare is straightforward and does not require re-running the entire DataCheck or Analyze pipelines.

2.2.3 Optional de novo *sequence clustering*

vAMPirus provides the option to perform *de novo* clustering of ASVs into 'clustered ASVs' or 'cASVs' based on pairwise nucleotide (ncASV) and/or protein (pcASV) sequence similarity using the programs VSEARCH (Rognes et al., 2016) and CD-HIT (Fu et al., 2012; Li & Godzik, 2006), respectively. cASVs differ from traditional *de novo* OTUs because for cASVs, denoising of sequences is done prior to clustering. The *de novo* clustering of ASVs is most useful for more developed virus systems where the degree of sequence divergence between taxonomically or ecologically distinct groups is known. Note that, from a methodological standpoint, representative sequences generated by a cASV approach exhibit the same issues as *de novo* OTUs (e.g., dataset dependence; see Callahan et al., 2017).

2.2.4 vAMPirus DataCheck pipeline and report

The vAMPirus DataCheck pipeline can help investigators determine the optimal parameters for read processing, ASV generation, and other downstream analyses conducted in

the Analyze pipeline. The DataCheck pipeline is particularly beneficial for investigators working on nascent virus systems because it facilitates the informed establishment of gene-, lineage- or system-specific analysis standards. The pipeline produces an HTML report that displays information such as sequencing success per sample, read characteristics (e.g., read length, GC content), and ASV/aminotype sequence properties. The DataCheck pipeline also provides insight into the ASV sequences by clustering them across a range of nucleotide and amino acid similarities and plotting the resultant number of cASVs per similarity value. Briefly, nucleotide-based *de novo* cASVs are produced by clustering ASV sequences using 24 different percent identity values (55%, 65%, 75%, 80-100%) with VSEARCH. To generate *de novo* pcASVs, ASVs are first translated using the program VirtualRibosome (v2.0, Wernersson, 2006), then clustered into *de novo* pcASVs using the same 24 percent identities with the program CD-HIT (v.4.8.1, Fu et al., 2012; Li & Godzik, 2006). For each percent identity value, the number of ncASVs and pcASVs is quantified and visualized as a scatter plot in the DataCheck report. This is a common approach used to determine the clustering percentage (e.g., Gustavsen and Suttle 2021): the percent similarity at which there is no longer a linear rise in the number of cASVs (the inflection point) is selected for sequence clustering. Optionally, users can also apply the program oligotyping (Eren et al., 2015) to calculate Shannon entropy values per sequence position for both ASV and aminotypes, which is then displayed in the report. An example vAMPirus DataCheck report is available at github.com/Aveglia/vAMPirusExamples.

2.2.5 vAMPirus Analyze pipeline and report

The Analyze pipeline includes multiple analyses (e.g., phylogenetics, taxonomy inference); results are summarized in a final HTML report with tables and interactive figures (Supplemental Figure S3). The pipeline also organizes and stores output (e.g., counts tables, similarity matrices, taxonomy files) from these analyses within a user-specified results directory.

Integration into Nextflow allows different analytical approaches (e.g., ASV grouping approach) to be run in parallel. Primary processes and analyses are summarized below; additional processes, such as percent similarity matrix generation and protein physiochemical property analyses are reviewed in the supplemental materials (Section S1).

2.2.5.1 Counts table generation

Nucleotide- and amino acid-based counts tables are generated within the Analyze pipeline. Counts tables for ASVs and ncASVs are produced using the VSEARCH program and the USEARCH algorithm (Edgar, 2010). Optionally, users can replace the use of the USEARCH algorithm with the use of the "--search_exact" feature provided by VSEARCH for exact ASV counts tables (Rognes et al., 2016). The aminotype and pcASV counts tables are generated by aligning translated merged reads to reference amino acid sequences with DIAMOND blastx (Buchfink et al., 2015). Sequence abundance information (in a comma delimited counts table) is then generated with a custom BASH script, which quantifies the number of alignments to each reference aminotype or pcASV from the DIAMOND output file. Users have the option to edit and adjust option parameters for the VSEARCH and DIAMOND counts table generation processes within the configuration file. All count files are stored in the results directory and are processed and visualized within Analyze reports as relative abundance bar plots.

2.2.5.2 Taxonomy inference

Sequence taxonomy is inferred using DIAMOND blastx or blastp via the user-specified protein database and the option parameters specified in the vAMPirus configuration file. The taxonomy process produces several files that are stored within a DIAMOND specific directory: 1. Unmodified DIAMOND output file, 2. fasta file of taxonomy assignments within the sequence headers, and 3. Results summary files (phyloseq taxonomy file, tab-separated summary file, and summary table of the abundance of specific hits). Taxonomy inference results are visualized in

the Analyze report as a donut plot.

2.2.5.3 Phylogenetic analysis

Phylogenetic analyses of ASVs, cASVs and aminotypes are conducted within the Analyze pipeline and all output files are stored in a dedicated directory within the results directory. First, sequences are aligned using the program muscle (v5.1; Edgar, 2021) and then trimmed automatically using the program trimAl (v1.4.1, Capella-Gutiérrez et al., 2009) using a heuristics-based approach. By default, substitution model testing is done with Modeltest-NG (v0.1.7, Darriba et al., 2020). The program IQTREE (v2.2.0.3; Minh et al., 2020) is used to generate a maximum-likelihood tree. The substitution model used to generate the tree can be set by the user, sourced from Modeltest-NG results, or automatically selected with ModelFinder (Kalyaanamoorthy et al., 2017). The tree produced is then used for phylogrouping with TreeCluster and is visualized in the Analyze report. Within the report, the user has the option to color code nodes based on sequence identity, taxonomy hit, MED group, or phylogroup assignment.

2.2.5.4 vAMPirus Analyze reports

The final process within the Analyze pipeline generates HTML reports using R Markdown (v2.3; (Xie et al., 2018). By default, an individual summary report containing community composition/diversity, taxonomy and phylogeny results is generated for ASVs, aminotypes, and cASVs. Users provide a metadata file that includes the sample name and category used to group the samples (i.e., sample treatment, location) for alpha and beta diversity analyses. vAMPirus summary reports are interactive and include: 1. Pre- and post-processing read statistics in tables and plots, 2. Rarefaction curves, 3. Shannon's diversity (*H*), Simpson's diversity (reported as 1-*D*), richness and distance to centroid box plots with statistical tests (Shapiro-Wilk normality test, Bartlett test of variance homogeneity, Kruskal-Wallis rank sum

test, Wilcox test, ANOVA and Tukey HSD, as appropriate), 4. Two- and three-dimensional NMDS plots (if no convergence, then PCoA plots), 5. Relative abundance bar plots, and 6. Taxonomy and phylogenetic results. An example vAMPirus Analyze summary report can be downloaded and reviewed at github.com/Aveglia/vAMPirusExamples.

2.3 vAMPirus Analysis Repository

To encourage and simplify the dissemination of parameters and non-read files needed to reproduce vAMPirus analyses, we created the 'vAMPirus Analysis Repository' (zenodo.org/communities/vampirusrepo/). The vAMPirus Analysis Repository is a Zenodo Community intended as a central location where investigators can deposit vAMPirus configuration files, metadata files, databases used for taxonomy assignment or ASV filtering, and any other files required to reproduce an analysis. Instructions and recommendations for submission are available in the vAMPirus manual (shorturl.at/uCO28). Once uploaded, submissions to the vAMPirus Analysis Repository are given a DOI.

3 Validating the vAMPirus workflow with published double-stranded DNA (dsDNA) virus datasets

We assessed the functionality and performance of vAMPirus' analytical workflow using amplicon sequencing datasets from two previously published dsDNA virus studies (Table 1). Research questions associated with each study are used as examples in Figure 1A (Finke & Suttle 2019; Figure 1A, Q1; Frantzen & Holo 2019; Figure 1A, Q2). For each dataset, we ran a vAMPirus analysis that reproduced the analysis from the associated published paper as closely as possible. For example, if a study generated *de novo* OTUs based on 97% nucleotide identity, the vAMPirus equivalent was ncASVs generated at 97% nucleotide identity with similar data quality control constraints. We then compared the results of the vAMPirus-based analyses to the findings described in each source manuscript. In brief, vAMPirus identified the same biological

patterns as those published by Finke & Suttle (2019, Figure 3) and Frantzen & Holo (2019, Figure 4) from their respective sequence datasets, and detected additional (previously unreported) virus diversity (Table 1). For example, Finke and Suttle (2019) reported increased cyanophage community alpha diversity in samples collected from sites with higher salinity (>27.5 practical salinity units, Figure 3-I, II); this pattern was present in the corresponding vAMPirus results (Figure 3-III, IV, V, VI), which included 86% more cyanophage pcASVs relative to the number of OTUs reported in Finke and Suttle (2019; Table 1). Similarly, the patterns of lactococcal phage OTU richness and relative abundances per sample reported by Franzten and Holo (2019; Figure 4-I) were also present in the vAMPirus results (Table 2; Figure 4-II). vAMPirus reported 43% more lactococcal phage ncASVs, relative to the OTUs reported by Frantzen and Holo (2019; Table 1, Figure 4). In addition, vAMPirus ASV-level analysis (Figure 4-III) revealed high lactococcal phage nucleotide-level diversity (n=531), yet aminotyping results (Figure 4-IV) suggest that the mutations underlying this richness mostly result in synonymous mutations: ASV sequences translated to only 29 aminotypes. Aminotype phylogrouping (see Section 2.2.2) of these data with TreeCluster highlighted a previously hidden overlap of lactococcal phage diversity across samples and dairy plants (Figure 4-VI).

Some variation between results obtained from vAMPirus and previous publications was expected, as the pipelines used in these comparisons were not identical. The only striking difference between the original results (in Finke and Suttle 2019 and Frantzen and Holo 2019) and those produced by vAMPirus is the higher number of pcASVs and ncASVs (respectively) identified via the latter analytic pipeline. Taxonomy results generated with vAMPirus by DIAMOND blastx aligning sequences to the NCBI virus RefSeq database verified that the pcASVs and ncASVs are of cyanophage and lactococcal phage origin, respectively

(Supplemental Figures S4 and S5). The higher diversity identified by vAMPirus may be attributable to differences in reference database used (boutique versus NCBI-curated), handling of singletons, and other factors.

Table 1. Breakdown of test datasets used during vAMPirus development, including the methods and results from the original (published) analysis, as well as results from vAMPirus analysis. vAMPirus results were generated using *de novo* clustering of ASVs into 'clustered ASVs' (cASVs) based on pairwise nucleotide (ncASV) and protein (pcASV) sequence similarity. dsDNA = double-stranded DNA.

Study	Target dsDNA virus group	Target gene	Original Methods	# published OTUs	# cASVs (vAMPirus, this study)
Finke & Suttle 2019	Myoviridae (T4-like cyanophage)	DNA polymerase	OTU clustering at 97% protein identity	606	1,128 97% pcASVs
Frantzen & Holo 2019	Siphoviridae (lactococcal bacteriophage)	portal protein	OTU clustering at 99.5% nucleotide identity	151	216 99.5% ncASVs

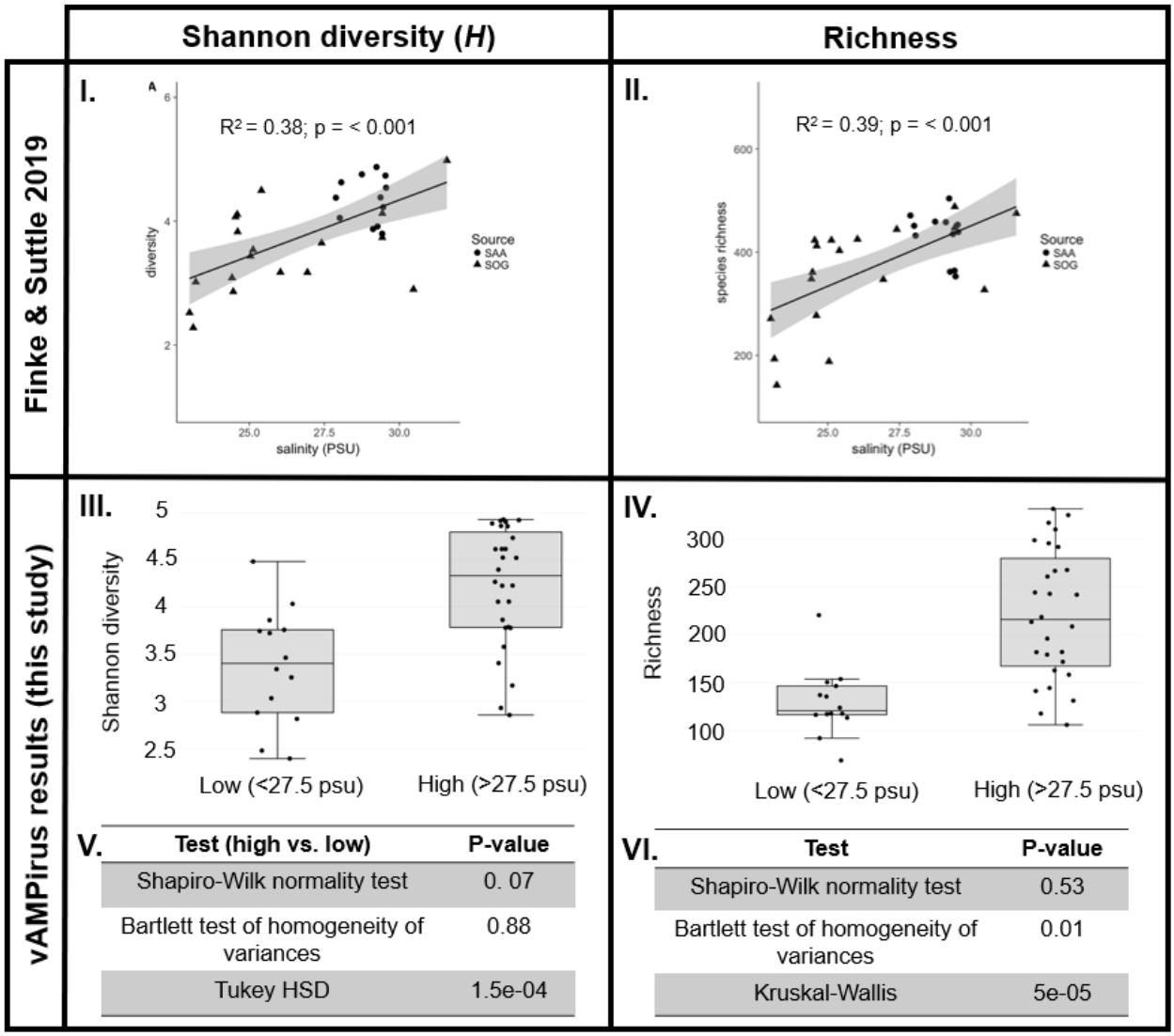


Figure 3. Comparison of results published by Finke & Suttle (2019) to those generated by the vAMPirus workflow for the same virus amplicon sequence dataset. The significant positive relationship between salinity and viral diversity (Shannon diversity (*H*) and richness) reported by Finke and Suttle (2019, panels I and II) was reproduced by vAMPirus (panels III-VI). Panels I and II display original results from Finke and Suttle (2019), based on 97% amino acid similarity OTUs. Panels III, IV, V, and VI include vAMPirus 97% pcASV results. Figures from Finke and Suttle (2019) and vAMPirus analyses were slightly modified for readability. For the vAMPirus analysis, which performs comparisons among categorical sample groups set by the user, all samples were assigned to either 'high' (>27.5) or 'low' (<27.5) salinity (practical salinity units; psu) groups. Figures from Finke and Suttle (2019) reprinted with permission from authors.

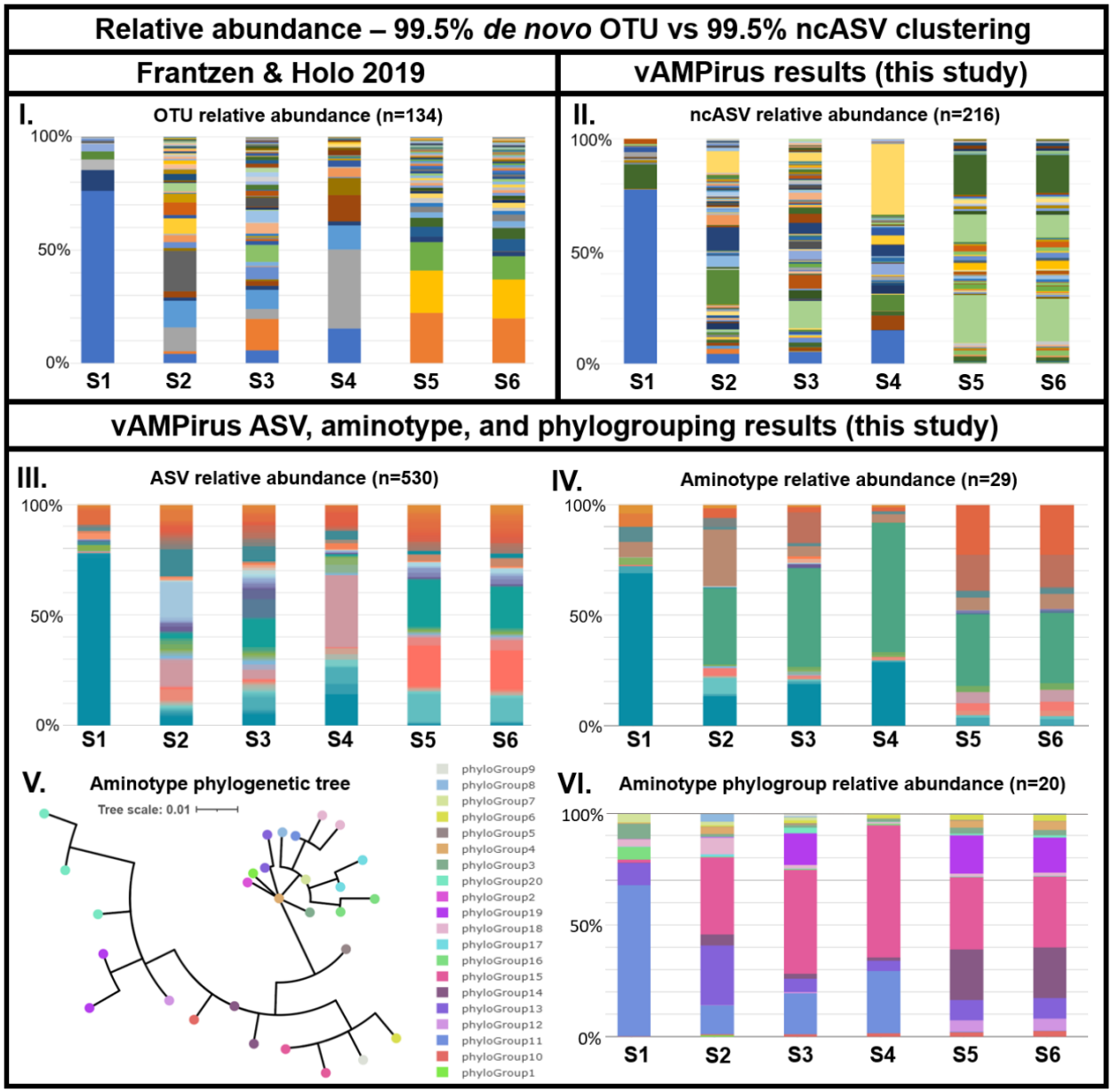


Figure 4. Comparison of results published by Frantzen and Holo (2019) to those generated by vAMPirus for the lactococcal phage amplicon sequence dataset. Patterns of lactococcal phage OTU relative abundance reported by Frantzen and Holo (2019; panel I) were recapitulated in vAMPirus ncASV results (panel II). Panel I displays original relative abundances of *de novo* OTUs based on 99.5% nucleotide similarity from Frantzen and Holo (2019). Panel II displays vAMPirus-generated 99.5% similarity ncASV relative abundance results. Panels III-VI demonstrate additional analyses output by vAMPirus, including ASV relative abundances, aminotype relative abundances, aminotype phylogeny with nodes colored by sequence phylogroup assignment, and aminotype phylogroup relative abundances. X-axis labels for all relative abundance graphs refer to dairy sample identity listed in Table 2. Colors in panels I-IV are plot-specific; similar colors across these panels do *not* indicate same sequences/clusters. However, colors assigned to phylogroups are consistent across panels V and VI. Figure from Frantzen and Holo (2019) reprinted with permission from authors.

Table 2. Comparison of Lactococcus phage sequence richness per sample between Frantzen and Holo (2019) results and vAMPirus (this study) results.

	Dairy Sample	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
	Dairy Plant	**1**	**1**	**1**	**2**	**2**	**2**
Frantzen and Holo 2019	# of OTUs with >100 reads	11	46	55	26	54	59
vAMPirus results	# of ncASVs with >100 reads	25	121	117	47	66	63

4 Applying vAMPirus to study a novel environmental RNA virus dataset

4.1 RNA virus study description

Dinoflagellate-infecting RNA viruses (dinoRNAVs) are positive-sense, single-stranded RNA viruses hypothesized to infect the dinoflagellate symbionts (Family Symbiodiniaceae) that live in the tissues of reef-building stony corals (Correa et al., 2013; Grupstra et al., 2022; Veglia et al., 2022). Although dinoRNAVs can be prevalent in coral colonies (Grupstra et al., 2022; Howe-Kerr et al., 2022; Montalvo-Proaño et al., 2017; Veglia et al., 2022), it is unclear how dinoRNAVs (or Symbiodiniaceae infected by dinoRNAVs) are transmitted among colonies. Recent work has shown that corallivorous (coral-eating) fishes disperse 100s of millions of live Symbiodiniaceae cells across reefscapes in their feces each day (Grupstra et al., 2021). To assess the extent to which corallivorous fish feces disperse dinoRNAVs in their feces (as in Grupstra et al., 2022), we characterized the presence and diversity of dinoRNAVs in various environmental reservoirs using amplicon sequencing of the dinoRNAV major capsid protein (*mcp*) gene. Given that dinoRNAV communities can vary across coral species and colonies (Grupstra et al., 2022; Howe-Kerr, 2022; Montalvo-Proaño et al., 2017) and that corallivorous fish actively 'sample' corals while feeding (Grupstra et al., 2021), we hypothesized that corallivorous fish feces are a reservoir of dinoRNAVs and that fish feces-associated dinoRNAV communities exhibit higher alpha diversity than coral colony-associated dinoRNAV communities. We generated 19 dinoRNAV *mcp* amplicon sequencing libraries from coral colony biopsies (*Acropora hyacinthus*, n=8; *Pocillopora* species complex, n=5), as well as the feces of corallivorous fishes (*Chaetodon*

reticulatus, n=4; *Chaetodon ornatissimus,* n=2). All samples were collected from reefs off the north shore of Moorea, French Polynesia (South Pacific). Methods for sampling and sample processing to generate virus amplicon sequencing libraries are described in Grupstra et al. (2021,2022) and Howe-Kerr et al. (2022). DinoRNAV *mcp* amplicon libraries were processed and analyzed using vAMPirus (see doi.org/10.5281/zenodo.7574173).

4.2 RNA virus study results and discussion

Amplicon sequencing of the dinoRNAV *mcp* gene produced a total of 7.4 million raw reads across 19 samples representing three potential reservoirs of dinoRNAV diversity across the reef. The 7.4 million raw reads were processed and reduced to 2.8 million merged reads at the expected amplicon length of 420 bases. Merged *mcp* amplicons dereplicated into 1.1 million unique sequences from which 481 ASVs and 191 aminotypes were identified. The ASV-level results indicated a potential trend of higher dinoRNAV richness in corallivore feces relative to coral colonies (Kruskal-Wallis H test: p-value = 0.14, Figure 5-I). Aminotype results, however, revealed that dinoRNAV richness is significantly higher in corallivore feces, relative to *Pocillopora* coral colonies (Figure 5-II; Kruskal-Wallis H test: p-value = 0.005; Wilcoxon signed-rank test: *Pocillopora* vs. corallivore, p = 0.01, *Pocillopora* vs. *Acropora*, p = 0.04). We interpret that a biological difference in richness likely does exist between dinoRNAV communities in corallivore feces versus those in at least some species of coral holobionts, and this difference may be more readily detected with aminotype-based analyses (as ASV-based analyses may contain more "noise" due to errors arising during RNA virus replication). This use case illustrates the potential benefits of running nucleotide and protein-based amplicon analyses in tandem when testing hypotheses regarding virus community diversity and dynamics. Furthermore, both ASV and aminotypes differed significantly in composition according to dinoRNAV reservoir (anosim with Bray Curtis distances, R=0.99, p<0.01; Figure 5-III, IV),

although some overlap (14%, 26 of 190 aminotypes) among dinoRNAV communities was observed (Supplemental Figure S6). Overall, this vAMPirus-based analysis of RNA virus amplicon sequencing data further corroborates that dinoRNAV communities differ across reef reservoirs (Grupstra et al. 2022, Montalvo-Proano et al. 2017, Howe-Kerr et al. 2022, Figure 5) and generates a new hypothesis to be tested in future studies: corallivorous fishes are environmental hotspots of dinoRNAV diversity on reefscapes.

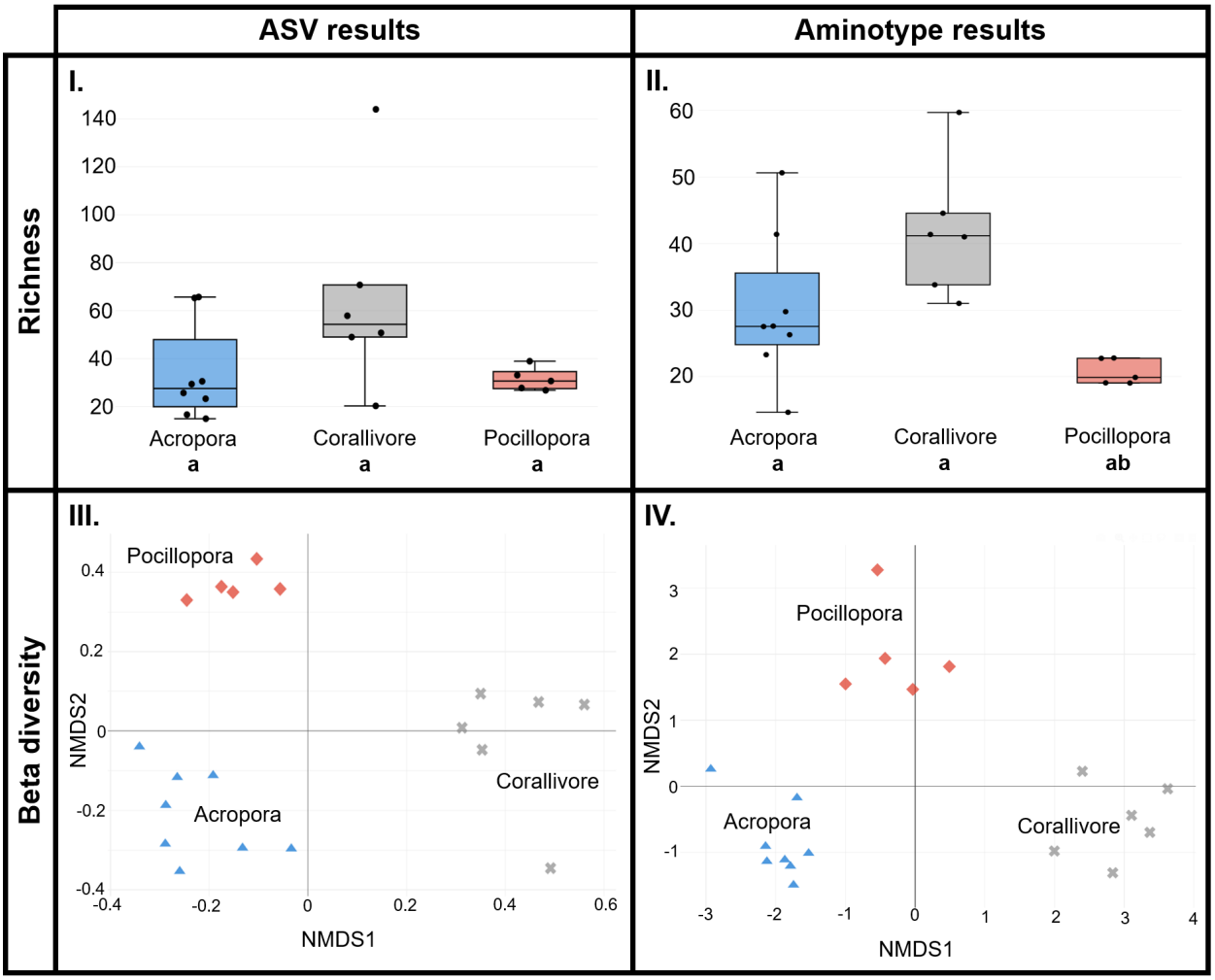


Figure 5. vAMPirus-generated dinoRNAV major capsid protein gene ASV and aminotype alpha (I, II) and beta (III, IV) diversity results from stony coral colonies *(Acropora* sp., *Pocillopora* species complex) and corallivorous (coral-eating) fish feces. Plots include three sample types: 1. *Acropora* biopsies (blue, triangle), 2. Corallivore feces (gray, x), and 3. *Pocillopora* biopsies (red, diamond). Letters beneath x axis labels on richness box plots (I, II) indicate statistically different groups. ASV and aminotype based NMDS plots (III, IV) were generated with Bray Curtis distances (stress values of 0.04 and 0.03, respectively).

5 Discussion

Targeted gene sequencing is increasingly being applied to explore spatiotemporal patterns of viral diversity (Adriaenssens & Cowan, 2014; Finke & Suttle, 2019; Frantzen & Holo, 2019; Grupstra et al., 2022; Gustavsen & Suttle, 2021; Howe-Kerr et al., 2022; Y. Li et al., 2018; Montalvo-Proaño et al., 2017; Prodinger et al., 2020; Short et al., 2010; Tong et al., 2016). The field of virology can now greatly benefit from the development of readily standardizable and reproducible pipelines for analyzing amplicon sequence datasets. Here, we present vAMPirus; a freely available, powerful, and flexible bioinformatics tool that streamlines the processing, analysis, and visualization of virus gene amplicon data. The availability of diverse bioinformatics approaches and tools within the vAMPirus program (*e.g.,* ASV calling, clustering, translation, phylogenetic clustering) empowers the user to adapt and set informed standards for their study system and easily share these standards with colleagues. With a user-friendly design and robust documentation, vAMPirus democratizes comprehensive virus amplicon sequencing analyses, making it a timely and valuable tool for virologists.

To inform virus amplicon data analyses, virologists have primarily relied on pipelines and tutorials geared towards bacterial or microeukaryote amplicon data (e.g., mothur (Schloss, 2020) and QIIME2 (Bolyen et al., 2019)). Although valuable insights have been made using these resources, an accessible virus-focused amplicon analysis pipeline will advance the field by offering via (1) automated pipelines that standardize approaches for viral amplicon analyses (e.g., ASV and aminotype calling); (2) non-cluster-based alternatives to partitioning virus gene sequences (e.g., MED and phylogrouping); and (3) virus-focused taxonomy databases. Virus amplicon analyses have traditionally applied *de novo* clustering of marker gene sequences into *de novo* OTUs based on a percent identity value (i.e., 97% nucleotide identity, Callahan et al.,

2017). However, clustering virus amplicons into biologically accurate *de novo* OTUs is challenging as the optimal clustering percentage is often unknown. vAMPirus provides users with the opportunity to transition from traditional *de novo* OTUs in virus amplicon sequencing analyses to using ASVs and aminotypes. We have illustrated here that ASV and aminotype-based analyses generally recapitulate findings generated via *de novo* OTU-based analyses (Figures 3, 4; Supplemental Table S1), while enabling reproducibility and cross-study comparisons (Callahan et al., 2017). Running analyses of amino acid and nucleotide sequence data in tandem, which is possible in vAMPirus, can aid in resolving virus phylogenies and reveal non-synonymous mutations that indicate virus protein property variability within a community (DeFilippis & Villarreal, 2000). This synergistic approach has been effective in developing dinoRNAVs and their dinoflagellate hosts (family Symbiodiniaceae, endosymbionts of stony corals) as a nascent study system. To characterize dinoRNAVs, studies have used the *mcp* gene, which has a high mutation rate and is hypothesized to be important in host cell attachment (Tomaru et al., 2004). vAMPirus aminotyping uncovered non-synonymous mutations in dinoRNAV *mcp* sequences, which may represent phenotypic differences that correlate with the distribution of host lineages across reefs (Grupstra et al., 2022; Howe-Kerr et al., 2022, this study). Aminotyping also effectively reduced noise from high mutation rates in ASV results, revealing temperature-driven increases in dinoRNAV infection productivity and community diversity across time and space (Grupstra et al. 2022: time only, Howe-Kerr et al., 2022). By making viral protein sequence analyses readily accessible in an amplicon sequence analysis workflow, vAMPirus helps reveal biological patterns in DNA and highly mutable RNA virus lineages by increasing signal-to-noise ratio in results (through collapse of synonymous nucleotide mutations, (Wernersson & Pedersen, 2003).

The increasing application of amplicon sequencing to the study of microbial diversity and dynamics has spurred efforts to improve the proficiency of tools that parse marker gene data. Such tools include the programs TreeCluster (Balaban et al., 2019) and oligotyping (Eren et al., 2015), which were developed as *de novo* clustering alternatives for partitioning genetic sequences into distinct units. In vAMPirus, these programs are utilized to assign ASVs and aminotypes to phylo- or MED groups based on user-set criteria (see Section 2.2.2). Assignment of ASV/aminotype sequences to groups rather than use of cluster representative sequences in analyses (such as, in the case of *de novo* OTUs and cASVs, Callahan et al., 2017) is done by vAMPirus to maintain reproducibility and comparability of results, while still permitting virus sequence classification into phylogenetically or ecologically distinct groups. These grouping approaches are instrumental for investigators because they can expose underlying patterns obscured by high sequence diversity (e.g., lactococcal phage phylogrouping results, Section 3, Figure 4). Phylogeny-based sequence clustering with TreeCluster has been applied to assess the diversity of microorganisms (and barley, Chen et al., 2022) and has been used to resolved virus transmission dynamics (HIV, Balaban et al., 2019; SARS-CoV-2, Plyusnin et al., 2022) and phylogenies (Ni et al., 2023). However, TreeCluster's potential utility for virus amplicon analyses is, for the most part, untapped. The inclusion of TreeCluster in the vAMPirus pipeline also opens the door to epidemiological insights, such as virus genetic linkage, transmission dynamics, and subpopulation mixing, from viral datasets (Balaban et al., 2019; Bezemer et al., 2015; Eshleman et al., 2011; Hué et al., 2014). Similarly, the program oligotyping developed by Eren et al. (2015) has been applied extensively to investigate microorganism diversity from marker gene data (cited 332 times as of February 2, 2023, Web of Science). However, only one published study has applied Minimum Entropy Decomposition sequence clustering with

oligotyping to virus amplicon data (Needham et al., 2017). The MED grouping with oligotyping option provided by vAMPirus is a powerful approach for deciphering virus community diversity because it enables the grouping of sequences based on potential physiologically and/or ecologically relevant similarities. For example, users can identify gene sequence positions with non-synonymous mutations via aminotyping and then specify these positions in MED grouping to partition sequences into units of similar protein phenotypes (i.e., host cell attachment; see Harvey et al., 2021). The option to incorporate cutting-edge bioinformatic approaches, such as phylogrouping and MED grouping, into analyses of virus amplicon data makes vAMPirus a highly useful “raw-reads-to-results" environmental virology workflow. vAMPirus is an easy-to-use, open-source, and flexible tool that streamlines and simplifies the process of analyzing viral amplicon data. vAMPirus is designed to be community-driven; new features and programs (e.g., built-in lineage specific configuration files or databases, new bioinformatic tools) can easily be implemented at the request of investigators or when advances in best practices are made. vAMPirus advances studies of viral community diversity by facilitating informed analyses of amplicon sequence data with its DataCheck and Analyze pipelines in a standardized and reproducible manner.

6 Acknowledgments

The authors would like to thank Jan F. Finke, Curtis A. Suttle, Julia A. Gustavsen, Cyril Frantzen, Hedge Holo, Florian Prodinger and Hiroyuki Ogata (and their respective co-authors) for providing raw virus amplicon sequence data for vAMPirus testing and for permission to reprint original figures. We thank Samantha R. Coy for feedback and insight during early stages of vAMPirus development, and Nikolaos Schizas for access to computational resources during vAMPirus development. This work represents a contribution of the Moorea Coral Reef (MCR)

LTER Site (NSF OCE 16–37396). This research was funded by a U.S. National Science Foundation Grant OCE 22-24354 (and earlier awards) to the Moorea Coral Reef LTER as well as a generous gift from the Gordon and Betty Moore Foundation. Research was completed under permits issued by the Territorial Government of French Polynesia (Délégation à la Recherche) and the Haut-Commissariat de la République en Polynésie Francaise (DTRT) (MCR LTER Protocole d'Accueil 2005–2022; Adrienne Correa Protocole d'Accueil 2013-2019), and we thank the Délégation à la Recherche and DTRT for their continued support. Start-up funds from Rice University, a NSF CAREER Award (OCE-2145472) and an Early-Career Research Fellowship (#2000009651) from the Gulf Research Program of the National Academies of Sciences to AMSC also contributed to this work. Additional funding was provided by Lewis and Clark Grants for Exploration to LHK and CGG, and Wagoner Foreign Study awards and to LHK, CGG, and AJV. The Kirk W. Dotson Endowed Graduate Fellowship in Ecology and Evolutionary Biology also helped support AJV as this work was conducted. RERV acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 764840 (ITN IGNITE).

7 References

Adriaenssens, E. M., & Cowan, D. A. (2014). Using signature genes as tools to assess environmental viral ecology and diversity. *Applied and Environmental Microbiology*, *80*(15), 4470–4480. https://doi.org/10.1128/AEM.00878-14

Aeby, G. S., Ushijima, B., Campbell, J. E., Jones, S., Williams, G. J., Meyer, J. L., Häse, C., & Paul, V. J. (2019). Pathogenesis of a tissue loss disease affecting multiple species of corals along the Florida Reef Tract. *Frontiers in Marine Science*, *6*, 678.

Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology*, *479*, 26–37.

Ambalavanan, L., Iehata, S., Fletcher, R., Stevens, E. H., & Zainathan, S. C. (2021). A Review of Marine Viruses in Coral Ecosystem. *Journal of Marine Science and Engineering*, *9*(7), Article 7. https://doi.org/10.3390/jmse9070711

Balaban, M., Moshiri, N., Mai, U., Jia, X., & Mirarab, S. (2019). TreeCluster: Clustering biological sequences using phylogenetic trees. *PLOS ONE*, *14*(8), e0221068. https://doi.org/10.1371/journal.pone.0221068

Bettarel, Y., Thuy, N. T., Huy, T. Q., Hoang, P. K., & Bouvier, T. (2013). Observation of virus-like particles in thin sections of the bleaching scleractinian coral Acropora cytherea. *Journal of the Marine Biological Association of the United Kingdom*, *93*(4), 909–912. https://doi.org/10.1017/S0025315411002062

Bezemer, D., Cori, A., Ratmann, O., Sighem, A. van, Hermanides, H. S., Dutilh, B. E., Gras, L., Faria, N. R., Hengel, R. van den, Duits, A. J., Reiss, P., Wolf, F. de, Fraser, C., & Cohort, A. observational. (2015). Dispersion of the HIV-1 Epidemic in Men Who Have Sex with Men in the Netherlands: A Combined Mathematical Model and Phylogenetic Analysis. *PLOS Medicine*, *12*(11), e1001898. https://doi.org/10.1371/journal.pmed.1001898

Bigot, T., Temmam, S., Pérot, P., & Eloit, M. (2020). *RVDB-prot, a reference viral protein database and its HMM profiles* (8:530). F1000Research. https://doi.org/10.12688/f1000research.18776.2

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Braga, L. P. P., Spor, A., Kot, W., Breuil, M.-C., Hansen, L. H., Setubal, J. C., & Philippot, L. (2020). Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome*, *8*(1), 52. https://doi.org/10.1186/s40168-020-00822-z

Brandt, M. E., Ennis, R. S., Meiling, S. S., Townsend, J., Cobleigh, K., Glahn, A., Quetel, J., Brandtneris, V., Henderson, L. M., & Smith, T. B. (2021). The Emergence and Initial Impact of Stony Coral Tissue Loss Disease (SCTLD) in the United States Virgin Islands.

Frontiers in Marine Science, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.715329

Breitbart, M., Bonnain, C., Malki, K., & Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nature Microbiology*, *3*(7), 754–766.

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

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), Article 1. https://doi.org/10.1038/nmeth.3176

Buerger, P., Weynberg, K. D., Sato, Y., Willis, B. L., & Oppen, M. van. (n.d.). *Bacteriophages in black band disease Novel T4 bacteriophages associated with black band disease in corals*. https://doi.org/10.1111/1462-2920.14432

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), Article 12. https://doi.org/10.1038/ismej.2017.119

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Cárdenas, A., Ye, J., Ziegler, M., Payet, J. P., McMinds, R., Vega Thurber, R., & Voolstra, C. R. (2020). Coral-Associated Viral Assemblages From the Central Red Sea Align With Host Species and Contribute to Holobiont Genetic Diversity. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.572534

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Y.-Y., Schreiber, M., Bayer, M. M., Dawson, I. K., Hedley, P. E., Lei, L., Akhunova, A., Liu, C., Smith, K. P., Fay, J. C., Muehlbauer, G. J., Steffenson, B. J., Morrell, P. L., Waugh, R., & Russell, J. R. (2022). The evolutionary patterns of barley pericentromeric chromosome regions, as shaped by linkage disequilibrium and domestication. *The Plant Journal*, *111*(6), 1580–1594. https://doi.org/10.1111/tpj.15908

Correa, A. M. S., Ainsworth, T. D., Rosales, S. M., Thurber, R. L. V., Butler, C. R., & Thurber, R. L. V. (2016). Viral outbreak in corals associated with an in situ bleaching event: Atypical herpes-like viruses and a new megavirus infecting Symbiodinium. *Frontiers in Microbiology*, *7*, 127. https://doi.org/10.3389/fmicb.2016.00127

Correa, A. M. S., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., & Weitz, J. S. (2021). Revisiting the rules of life for viruses of microorganisms. *Nature Reviews Microbiology*, *19*(8), 501–513. https://doi.org/10.1038/s41579-021-00530-x

Correa, A. M. S., Welsh, R. M., & Vega Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal*, *7*(1), Article 1. https://doi.org/10.1038/ismej.2012.75

Correa, A. M., Welsh, R. M., & Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA and+ ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal*, *7*(1), 13–27. https://doi.org/10.1038/ismej.2012.81

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Davy, J. E., & Patten, N. L. (2007). Morphological diversity of virus-like particles within the surface microlayer of scleractinian corals. *Aquatic Microbial Ecology*, *47*, 37–44. https://doi.org/10.3354/ame047037

Davy, S. K., Burchett, S. G., Dale, A. L., Davies, P., Davy, J. E., Muncke, C., Hoegh-Guldberg, O., & Wilson, W. H. (2006). Viruses: Agents of coral disease? *Diseases of Aquatic Organisms*, *69*, 101–110. https://doi.org/10.3354/dao069101

DeFilippis, V. R., & VillarreaI, L. P. (2000). An Introduction to the Evolutionary Ecology of Viruses. *Viral Ecology*, 125–208. https://doi.org/10.1016/B978-012362675-2/50005-7

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), Article 4. https://doi.org/10.1038/nbt.3820

Domingo, E., & Perales, C. (2019). Viral quasispecies. *PLOS Genetics*, *15*(10), e1008271. https://doi.org/10.1371/journal.pgen.1008271

Eddy, T. D., Lam, V. W. Y., Reygondeau, G., Cisneros-Montemayor, A. M., Greer, K., Palomares, M. L. D., Bruno, J. F., Ota, Y., & Cheung, W. W. L. (2021). Global decline in capacity of coral reefs to provide ecosystem services. *One Earth*, *4*(9), 1278–1285. https://doi.org/10.1016/j.oneear.2021.08.016

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C. (2016a). *UCHIME2: Improved chimera prediction for amplicon sequencing* (p. 074252). bioRxiv. https://doi.org/10.1101/074252

Edgar, R. C. (2016b). *UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing* (p. 081257). bioRxiv. https://doi.org/10.1101/081257

Edgar, R. C. (2021). *MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping* (p. 2021.06.20.449169). bioRxiv. https://doi.org/10.1101/2021.06.20.449169

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., & Sogin, M. L. (2015). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*, *9*(4), Article 4. https://doi.org/10.1038/ismej.2014.195

Eshleman, S. H., Hudelson, S. E., Redd, A. D., Wang, L., Debes, R., Chen, Y. Q., Martens, C. A., Ricklefs, S. M., Selig, E. J., Porcella, S. F., Munshaw, S., Ray, S. C., Piwowar-Manning, E., McCauley, M., Hosseinipour, M. C., Kumwenda, J., Hakim, J. G.,

Chariyalertsak, S., de Bruyn, G., … Hughes, J. P. (2011). Analysis of Genetic Linkage of HIV From Couples Enrolled in the HIV Prevention Trials Network 052 Trial. *The Journal of Infectious Diseases*, *204*(12), 1918–1926. https://doi.org/10.1093/infdis/jir651

Finke, J. F., & Suttle, C. A. (2019). The Environment and Cyanophage Diversity: Insights From Environmental Sequencing of DNA Polymerase. *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/article/10.3389/fmicb.2019.00167

Frantzen, C. A., & Holo, H. (2019). Unprecedented Diversity of Lactococcal Group 936 Bacteriophages Revealed by Amplicon Sequencing of the Portal Protein Gene. *Viruses*, *11*(5), Article 5. https://doi.org/10.3390/v11050443

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Grupstra, C. G. B., Howe-Kerr, L. I., Veglia, A. J., Bryant, R. L., Coy, S. R., Blackwelder, P. L., & Correa, A. M. S. (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *The ISME Journal*, *16*(5), Article 5. https://doi.org/10.1038/s41396-022-01194-y

Grupstra, C. G. B., Lemoine, N. P., Cook, C., & Correa, A. M. S. (2022). Thank you for biting: Dispersal of beneficial microbiota through "antagonistic" interactions. *Trends in Microbiology*, *30*(10), 930–939. https://doi.org/10.1016/j.tim.2022.03.006

Grupstra, C. G. B., Rabbitt, K. M., Howe-Kerr, L. I., & Correa, A. M. S. (2021). Fish predation on corals promotes the dispersal of coral symbionts. *Animal Microbiome*, *3*(1), 25. https://doi.org/10.1186/s42523-021-00086-4

Grupstra, C. G., Howe-Kerr, L. I., Veglia, A. J., Bryant, R. L., Coy, S. R., Blackwelder, P. L., & Correa, A. (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *The ISME Journal*, 1–12.

Gustavsen, J. A., & Suttle, C. A. (2021). Role of Phylogenetic Structure in the Dynamics of Coastal Viral Assemblages. *Applied and Environmental Microbiology*, *87*(11), e02704-20. https://doi.org/10.1128/AEM.02704-20

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*(7), Article 7. https://doi.org/10.1038/s41579-021-00573-0

Howe-Kerr, L. I. (2022). *Viruses of a key coral symbiont exhibit temperature-driven productivity across a reefscape*. https://doi.org/10.21203/rs.3.rs-1899377/v1

Hué, S., Brown, A. E., Ragonnet-Cronin, M., Lycett, S. J., Dunn, D. T., Fearnhill, E., Dolling, D. I., Pozniak, A., Pillay, D., Delpech, V. C., & Leigh Brown, A. J. (2014). Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS*, *28*(13), 1967. https://doi.org/10.1097/QAD.0000000000000383

Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S. T., Rosenstiel, P., Dutilh, B. E., & Hentschel, U. (2019). A Phage

Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. *Cell Host & Microbe*, *26*(4), 542-550.e5. https://doi.org/10.1016/j.chom.2019.08.019

Jahn, M. T., Lachnit, T., Markert, S. M., Stigloher, C., Pita, L., Ribes, M., Dutilh, B. E., & Hentschel, U. (2021). Lifestyle of sponge symbiont phages by host prediction and correlative microscopy. *The ISME Journal*, *15*(7), Article 7. https://doi.org/10.1038/s41396-021-00900-6

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), Article 6. https://doi.org/10.1038/nmeth.4285

Labadie, T., Batéjat, C., Leclercq, I., & Manuguerra, J.-C. (2020). Historical Discoveries on Viruses in the Environment and Their Impact on Public Health. *Intervirology*, *63*(1–6), 17–32. https://doi.org/10.1159/000511575

Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Jutz, S., Pascelli, C., Botté, E. S., Bell, S. C., Peirce, T. E., Weynberg, K. D., van Oppen, M. J. H., Rattei, T., & Webster, N. S. (2018). Reef invertebrate viromics: Diversity, host specificity and functional capacity. *Environmental Microbiology*, *20*(6), 2125–2141. https://doi.org/10.1111/1462-2920.14110

Leruste, A., Bouvier, T., & Bettarel, Y. (2012). Enumerating Viruses in Coral Mucus. *Applied and Environmental Microbiology*, *78*(17), 6377–6379. https://doi.org/10.1128/AEM.01141-12

Levin, R. A., Voolstra, C. R., Weynberg, K. D., & van Oppen, M. J. H. (2017). Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *The ISME Journal*, *11*(3), 808–812. https://doi.org/10.1038/ismej.2016.154

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, Y., Hingamp, P., Watai, H., Endo, H., Yoshida, T., & Ogata, H. (2018). Degenerate PCR Primers to Reveal the Diversity of Giant Viruses in Coastal Waters. *Viruses*, *10*(9), 496. https://doi.org/10.3390/v10090496

Marhaver, K. L., Edwards, R. A., & Rohwer, F. (2008). Viral communities associated with healthy and bleaching corals. *Environmental Microbiology*, *10*(9), 2277–2286. https://doi.org/10.1111/j.1462-2920.2008.01652.x

Meiling, S. S., Muller, E. M., Lasseigne, D., Rossin, A., Veglia, A. J., MacKnight, N., Dimos, B., Huntley, N., Correa, A. M. S., Smith, T. B., Holstein, D. M., Mydlarz, L. D., Apprill, A., & Brandt, M. E. (2021). Variable Species Responses to Experimental Stony Coral Tissue Loss Disease (SCTLD) Exposure. *Frontiers in Marine Science*, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.670829

Messyasz, A., Rosales, S. M., Mueller, R. S., Sawyer, T., Correa, A. M. S., Thurber, A. R., & Vega Thurber, R. (2020). Coral Bleaching Phenotypes Associated With Differential

Abundances of Nucleocytoplasmic Large DNA Viruses. *Frontiers in Marine Science*, *7*. https://www.frontiersin.org/articles/10.3389/fmars.2020.555474

Metcalf, T. G., Melnick, J. L., & Estes, M. K. (1995). ENVIRONMENTAL VIROLOGY: From Detection of Virus in Sewage and Water by Isolation to Identification by Molecular Biology—A Trip of Over 50 Years. *Annual Review of Microbiology*, *49*(1), 461–487. https://doi.org/10.1146/annurev.mi.49.100195.002333

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, *29*(2), 215–233. https://doi.org/10.1016/S0921-8009(99)00009-9

Montalvo-Proaño, J., Buerger, P., Weynberg, K. D., & van Oppen, M. J. H. (2017). A PCR-Based Assay Targeting the Major Capsid Protein Gene of a Dinorna-Like ssRNA Virus That Infects Coral Photosymbionts. *Frontiers in Microbiology*, *8*. https://www.frontiersin.org/articles/10.3389/fmicb.2017.01665

Needham, D. M., Sachdeva, R., & Fuhrman, J. A. (2017). Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME Journal*, *11*(7), Article 7. https://doi.org/10.1038/ismej.2017.29

Nguyen, M., Wemheuer, B., Laffy, P. W., Webster, N. S., & Thomas, T. (2021). Taxonomic, functional and expression analysis of viral communities associated with marine sponges. *PeerJ*, *9*, e10715. https://doi.org/10.7717/peerj.10715

Nguyen-Kim, H., Bettarel, Y., Bouvier, T., Bouvier, C., Doan-Nhu, H., Nguyen-Ngoc, L., Nguyen-Thanh, T., Tran-Quang, H., & Brune, J. (2015). Coral Mucus Is a Hot Spot for Viral Infections. *Applied and Environmental Microbiology*, *81*(17), 5773–5783. https://doi.org/10.1128/AEM.00542-15

Ni, X.-B., Cui, X.-M., Liu, J.-Y., Ye, R.-Z., Wu, Y.-Q., Jiang, J.-F., Sun, Y., Wang, Q., Shum, M. H.-H., Chang, Q.-C., Zhao, L., Han, X.-H., Ma, K., Shen, S.-J., Zhang, M.-Z., Guo, W.-B., Zhu, J.-G., Zhan, L., Li, L.-J., … Cao, W.-C. (2023). Metavirome of 31 tick species provides a compendium of 1,801 RNA virus genomes. *Nature Microbiology*, *8*(1), Article 1. https://doi.org/10.1038/s41564-022-01275-w

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733-745. https://doi.org/10.1093/nar/gkv1189

Paez-Espino, D., Chen, I.-M. A., Palaniappan, K., Ratner, A., Chu, K., Szeto, E., Pillay, M., Huang, J., Markowitz, V. M., Nielsen, T., Huntemann, M., K. Reddy, T. B., Pavlopoulos, G. A., Sullivan, M. B., Campbell, B. J., Chen, F., McMahon, K., Hallam, S. J., Denef, V.,

… Kyrpides, N. C. (2017). IMG/VR: A database of cultured and uncultured DNA Viruses and retroviruses. *Nucleic Acids Research*, *45*(D1), gkw1030. https://doi.org/10.1093/nar/gkw1030

Pascelli, C., Laffy, P. W., Botté, E., Kupresanin, M., Rattei, T., Lurgi, M., Ravasi, T., & Webster, N. S. (2020). Viral ecogenomics across the Porifera. *Microbiome*, *8*(1), 144. https://doi.org/10.1186/s40168-020-00919-5

Pascelli, C., Laffy, P. W., Kupresanin, M., Ravasi, T., & Webster, N. S. (2018). Morphological characterization of virus-like particles in coral reef sponges. *PeerJ*, *6*, e5625. https://doi.org/10.7717/peerj.5625

Patten, N., Mitchell, J., Middelboe, M., Eyre, B., Seuront, L., Harrison, P., & Glud, R. (2008). Bacterial and viral dynamics during a mass coral spawning period on the Great Barrier Reef. *Aquatic Microbial Ecology*, *50*, 209–220. https://doi.org/10.3354/ame01179

Plyusnin, I., Truong Nguyen, P. T., Sironen, T., Vapalahti, O., Smura, T., & Kant, R. (2022). ClusTRace, a bioinformatic pipeline for analyzing clusters in virus phylogenies. *BMC Bioinformatics*, *23*(1), 196. https://doi.org/10.1186/s12859-022-04709-8

Precht, W. F., Gintert, B. E., Robbart, M. L., Fura, R., & van Woesik, R. (2016). Unprecedented Disease-Related Coral Mortality in Southeastern Florida. *Scientific Reports*, *6*(1), 31374. https://doi.org/10.1038/srep31374

Prodinger, F., Endo, H., Gotoh, Y., Li, Y., Morimoto, D., Omae, K., Tominaga, K., Blanc-Mathieu, R., Takano, Y., Hayashi, T., Nagasaki, K., Yoshida, T., & Ogata, H. (2020). An Optimized Metabarcoding Method for Mimiviridae. *Microorganisms*, *8*(4), Article 4. https://doi.org/10.3390/microorganisms8040506

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Rusanova, A., Fedorchuk, V., Toshchakov, S., Dubiley, S., & Sutormin, D. (2022). An Interplay between Viruses and Bacteria Associated with the White Sea Sponges Revealed by Metagenomics. *Life*, *12*(1), Article 1. https://doi.org/10.3390/life12010025

Schloss, P. D. (2020). Reintroducing mothur: 10 Years Later. *Applied and Environmental Microbiology*, *86*(2). https://doi.org/10.1128/AEM.02343-19

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database: The Journal of Biological Databases and Curation*, *2020*, baaa062. https://doi.org/10.1093/database/baaa062

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Short, S. M., Chen, F., & Wilhelm, S. (2010). The construction and analysis of marker gene libraries. *Manual of Aquatic Viral Ecology*, 82–91.

Soffer, N., Brandt, M. E., Correa, A. M. S., Smith, T. B., & Thurber, R. V. (2014). Potential role of viruses in white plague coral disease. *The ISME Journal*, *8*(2), 271–283. https://doi.org/10.1038/ismej.2013.137

Soffer, N., Zaneveld, J., & Vega Thurber, R. (2015). Phage-bacteria network analysis and its implication for the understanding of coral disease. *Environmental Microbiology*, *17*(4), 1203–1218. https://doi.org/10.1111/1462-2920.12553

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

Sweet, M., & Bythell, J. (2017). The role of viruses in coral health and disease. *Journal of Invertebrate Pathology*, *147*, 136–144. https://doi.org/10.1016/j.jip.2016.12.005

Thurber, R. V., Payet, J. P., Thurber, A. R., & Correa, A. M. S. (2017). Virus–host interactions and their roles in coral reef health and disease. *Nature Reviews Microbiology*, *15*(4), 205–216. https://doi.org/10.1038/nrmicro.2016.176

Tomaru, Y., Katanozaka, N., Nishida, K., Shirai, Y., Tarutani, K., Yamaguchi, M., & Nagasaki, K. (2004). Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga Heterocapsa circularisquama. *Aquatic Microbial Ecology*, *34*(3), 207–218. https://doi.org/10.3354/ame034207

Tong, Y., Liu, B., Liu, H., Zheng, H., Gu, J., Liu, H., Lin, M., Ding, Y., Song, C., & Li, Y. (2016). New universal primers for genotyping and resistance detection of low HBV DNA levels. *Medicine*, *95*(33), e4618. https://doi.org/10.1097/MD.0000000000004618

Uyaguari-Diaz, M. I., Chan, M., Chaban, B. L., Croxen, M. A., Finke, J. F., Hill, J. E., Peabody, M. A., Van Rossum, T., Suttle, C. A., Brinkman, F. S. L., Isaac-Renton, J., Prystajecky, N. A., & Tang, P. (2016). A comprehensive method for amplicon-based and metagenomic characterization of viruses, bacteria, and eukaryotes in freshwater samples. *Microbiome*, *4*(1), 20. https://doi.org/10.1186/s40168-016-0166-1

Veglia, A. J., Beavers, K., Buren, E. W. V., Meiling, S. S., Muller, E. M., Smith, T. B., Holstein, D. M., Apprill, A., Brandt, M. E., Mydlarz, L. D., & Correa, A. M. S. (2022). Alphaflexivirus Genomes in Stony Coral Tissue Loss Disease-Affected, Disease-Exposed, and Disease-Unexposed Coral Colonies in the U.S. Virgin Islands. *Microbiology Resource Announcements*. https://doi.org/10.1128/mra.01199-21

Veglia, A. J., Bistolas, K. S. I., Voolstra, C. R., Hume, B. C. C., Planes, S., Allemand, D., Boissin, E., Wincker, P., Poulain, J., Moulin, C., Bourdin, G., Iwankow, G., Romac, S., Agostini, S., Banaigs, B., Boss, E., Bowler, C., Vargas, C. de, Douville, E., … Thurber, R. L. V. (2022). *Endogenous viral elements reveal associations between a non-retroviral RNA virus and symbiotic dinoflagellate genomes* (p. 2022.04.11.487905). bioRxiv. https://doi.org/10.1101/2022.04.11.487905

Veglia, A. J., Milford, C. R., & Schizas, N. V. (2021). Isolation and genotyping of novel T4 cyanophages associated with diverse coral reef invertebrates. *Coral Reefs*, *40*(2), 485–504. https://doi.org/10.1007/s00338-021-02056-3

Walton, C. J., Hayes, N. K., & Gilliam, D. S. (2018). Impacts of a Regional, Multi-Year, Multi-Species Coral Disease Outbreak in Southeast Florida. *Frontiers in Marine Science*, *5*. https://www.frontiersin.org/articles/10.3389/fmars.2018.00323

Weil, E., & Rogers, C. S. (2011). Coral Reef Diseases in the Atlantic-Caribbean. In *Coral Reefs: An Ecosystem in Transition*. Springer. https://doi.org/10.1007/978-94-007-0114-4

Wernersson, R. (2006). Virtual Ribosome—A comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Research*, *34*(suppl_2), W385–W388. https://doi.org/10.1093/nar/gkl252

Wernersson, R., & Pedersen, A. G. (2003). RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Research*, *31*(13), 3537–3539.

Weynberg, K. D., Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Rattei, T., Webster, N. S., & van Oppen, M. J. H. (2017). Coral-associated viral communities show high levels of diversity and host auxiliary functions. *PeerJ*, *5*, e4054. https://doi.org/10.7717/peerj.4054

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P., & van Oppen, M. J. H. (2015). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. *Scientific Reports*, *5*, 17889. https://doi.org/10.1038/srep17889

Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A., & van Oppen, M. J. H. (2014). Generating viral metagenomes from the coral holobiont. *Frontiers in Microbiology*, *5*. https://www.frontiersin.org/articles/10.3389/fmicb.2014.00206

Wilson, W. H., & Chapman, D. M. (2001). Observation of virus-like particles in thin sections of the plumose anemone, Metridium senile. *Journal of the Marine Biological Association of the United Kingdom*, *81*(5), 879–880. https://doi.org/10.1017/S0025315401004726

Wilson, W. H., Dale, A. L., Davy, J. E., & Davy, S. K. (2005). An enemy within? Observations of virus-like particles in reef corals. *Coral Reefs*, *24*(1), 145–148. https://doi.org/10.1007/s00338-004-0448-0

Wood-Charlson, E. M., Weynberg, K. D., Suttle, C. A., Roux, S., & van Oppen, M. J. H. (2015). Metagenomic characterization of viral communities in corals: Mining biological signal from methodological noise. *Environmental Microbiology*, *17*(10), 3440–3449. https://doi.org/10.1111/1462-2920.12803

Work, T. M., Weatherby, T. M., Landsberg, J. H., Kiryu, Y., Cook, S. M., & Peters, E. C. (2021). Viral-Like Particles Are Associated With Endosymbiont Pathology in Florida Corals Affected by Stony Coral Tissue Loss Disease. *Frontiers in Marine Science*, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.750658

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. https://bookdown.org/yihui/rmarkdown/

Zayed, A. A., Wainaina, J. M., Dominguez-Huerta, G., Pelletier, E., Guo, J., Mohssen, M., Tian, F., Pratama, A. A., Bolduc, B., Zablocki, O., Cronin, D., Solden, L., Delage, E., Alberti, A., Aury, J.-M., Carradec, Q., da Silva, C., Labadie, K., Poulain, J., … Sullivan, M. B. (2022). Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science*, *376*(6589), 156–162. https://doi.org/10.1126/science.abm5847

8 Authors' Contributions

AJV, CBG, LHK conceived of the program with support from AMSC; CBG and LHK contributed R code used in the vAMPirus reports; RERV contributed R code and helped execute vAMPirus incorporation into Nextflow; CBG, LHK and AMSC processed samples and generated the RNA virus dataset; AJV designed the pipelines with input from CBG and LHK; AJV wrote bash and R code used in the program, analyzed data, and wrote the initial draft of the manuscript, with contributions by all authors.

Chapter 3

Investigating the roles of viruses in stony coral tissue loss disease

Veglia, A.J.[1], Beavers, K.[2], Van Buren, E.W.[2], Meiling, S.S.[3], Muller, E.M.[4], Smith, T.B.[3], Holstein, D.M.[5], Apprill, A.[6], Brandt, M.E.[3], Mydlarz, L.D.[2], Correa, A.M.S.[1]

This chapter is in preparation to be submitted to journal *Nature Microbiology*

Abstract

Stony coral tissue loss disease (SCTLD) is a multi-host disease that affects 34 (of 45) Caribbean coral species and has caused unprecedented coral coverage loss in Tropical Atlantic reefs. While the etiology of SCTLD remains unresolved, one identifiable pathological feature of SCTLD infection is the breakdown of the coral-dinoflagellate (Family Symbiodiniaceae) symbiosis. It is hypothesized that the infection of Symbiodiniaceae by viruses, specifically filamentous RNA viruses, initiates this breakdown. To investigate the potential involvement of viruses in SCTLD, we used comparative metatranscriptomics to characterize virus community diversity and expression in corals across SCTLD disease states (apparently healthy, exposed-apparently healthy, and SCTLD diseased). Our findings revealed increased virus sequence abundance in SCTLD-exposed and SCTLD-diseased corals compared to apparently healthy holobionts, indicating heightened virus production in SCTLD-affects tissues. This increase in production was observed in several prokaryotic (e.g., Classes Caudoviricetes and Leviviricetes) and eukaryotic (e.g., Classes Megaviricetes and Duplopiviricetes) virus taxa. Moreover, the absence of sample clustering in ordination plots based on health status suggests that a single viral lineage may not serve as a diagnostic marker for SCTLD. Instead, the composition of holobiont viromes were primarily influenced by holobiont attributes such as coral species and dominant symbiont lineage, with health status playing a secondary role (PERMANOVA: symbiont lineage, $F=38.78$, $p=0.001$; coral species, $F=2.83$, $p=0.016$; health status, $F=1.69$, $p=0.023$). This suggests that the observed increase in virus production is likely evidence of opportunistic infections by resident virus lineages within the sampled coral holobionts. Taken together, the increased productivity of several putative virus orders that infect Symbiodiniaceae (e.g., Algavirales, Durnavirales, Patatavirales) and coral (e.g., Picornavirales) suggests that virus infections in both the symbiont and the coral itself may contribute to the breakdown of symbiosis observed in

SCTLD. However, these findings indicate that the etiology of SCTLD is not driven solely by a single viral group. Overall, this study provides crucial insight into virus community diversity and transcription dynamics associated with SCTLD and proposes several candidate virus groups that may contribute to disease etiology.

Introduction

Global change stressors are increasing the susceptibility of marine organisms, including those on coral reefs, to emerging diseases (Harvell et al., 2002; Jackson et al., 2014). Coral disease epizootics in recent decades have caused significant declines in coral populations and diversity on reefs (Aronson & Precht, 2001; Brandt et al., 2021; Jackson et al., 2014; Weil et al., 2006). Although coral disease outbreaks are not unique to the last several decades, studies show that there has been an increase in their frequency and distribution (Harvell et al., 2007; Sutherland et al., 2004; Tracy et al., 2019). The Caribbean specifically has been dubbed a 'disease hot spot' enduring frequent epizootic events that have contributed to the 80% decline in coral coverage in the region since the 1970's (Jackson et al., 2014; Weil et al., 2006; Weil & Rogers, 2011). Just recently in 2019, a new disease, termed Stony Coral Tissue Loss Disease (SCTLD), was reported on Caribbean reefs threatening already degraded coral reef habitat across the Greater and Lesser Antilles (Brandt et al., 2021).

SCTLD is a multi-host waterborne tissue wasting disease that was reported initially in 2014 on coral reefs in southeast Florida (Aeby et al., 2019; Precht et al., 2016; Walton et al., 2018). Capable of infecting at least 34 (of the 45) Caribbean stony coral species, SCTLD has since spread across the wider Caribbean, with unprecedented detrimental impacts on affected coral populations (Alvarez-Filip et al., 2019; Brandt et al., 2021; Precht et al., 2016; Walton et al., 2018; Williams et al., 2021). Previous work has demonstrated different levels of susceptibility to SCTLD among coral species based on etiological characteristics (e.g., incidence and lesion progression rates) (Aeby et al., 2019; Meiling et al., 2021; Sharp and Maxwell 2018). In addition, key bacterial groups (i.e., Flavobacteriales, Rhodobacteriales) were found to be differentially abundant in SCTLD lesions, although no one group so far has been proven to be

the SCTLD pathogen (Rosales et al. 2020, 2023; Meyer et al. 2019; Becker et al. 2022; Huntley et al. 2022). Furthermore, a pathological characteristic of SCTLD is the breakdown of the symbiosis between the coral and its mutualistic dinoflagellates (in Family Symbiodiniaceae) (Landsberg et al., 2020; Meiling et al. 2021; Beavers et al., *In Press*). Histopathological evidence of a disruption to the coral-dinoflagellate symbiosis in SCTLD-affected tissues includes enlarged symbiosomes (the vacuoles housing the Symbiodiniaceae), physiological deformities of Symbiodiniaceae cells, and apparent exocytosis (expulsion) of Symbiodiniaceae cells from SCTLD-affected tissues (Meiling et al. 2021, Landsberg et al. 2020). Although the etiological agent(s) (biological or environmental (Lesser et al., 2007)) of SCTLD has yet to be determined, mounting evidence suggests that viral infection of Symbiodiniaceae may drive the underlying breakdown of coral-dinoflagellate symbiosis, and thus the development of disease signs.

There are multiple indications that virus infections of Symbiodiniaceae play a role in SCTLD signs, including: i) reduced or halted lesion progression in dinoflagellate-free (or "bleached") coral tissues (S. Meiling et al., 2020), ii) necrosis of Symbiodiniaceae-housing coral gastrodermal cells (Landsberg et al., 2020), iii) microscopic detection of putative filamentous virus-like particles associated with Symbiodiniaceae cells in SCTLD-affected and SCTLD-exposed corals (Work et al., 2021), iv) *in silico* detection of filamentous virus-like genomes in SCTLD-affected/exposed coral metatranscriptomes (Veglia et al., 2022), and v) upregulation of antiviral pathways in SCTLD-affected coral metatranscriptomes (Beavers et al., *In Press*). Comparative omics-based assessments of virus diversity are needed to determine if there are significant differences in the viral community composition and activity across apparently healthy, SCTLD-exposed apparently healthy, and SCTLD-diseased diseased coral colonies.

Virus activity can have negative (parasitic), neutral (commensalistic), or positive

(mutualistic) impacts on coral holobionts (Vega Thurber et al., 2017; Sweet & Bythell, 2017, van Oppen et al. 2009, Correa et al., 2021). Highly abundant in both healthy and diseased coral tissues (as well as in the surrounding water column) (Correa et al., 2013, 2016; Davy & Patten, 2007; Davy et al., 2006; Leruste et al., 2012, Marharver et al., 2008), viruses can act as a key form of top-down control structuring microbial symbiont communities in coral tissues (Thurber et al., 2017). While the exact mechanisms remain unclear, viruses have been hypothesized to play roles in coral diseases like black band disease (T4-like bacteriophages; Beurger et al. 2019), Porites white patch syndrome (Icosahedral virus-like particles, Lawrence et al., 2015), and white plague (Nucleocytoplasmis large DNA viruses; Soffer et al., 2014). In the case of SCTLD, viruses can contribute to disease etiology in several ways. For example, it is possible that a single viral lineage or group is the pathogen, and its invasion of coral tissue and subsequent infection of Symbiodiniaceae triggers dysbiosis (destabilization of holobiont community) and, later, acute tissue wasting. Alternatively, opportunistic virus infections of Symbiodiniaceae could be stimulated if the coral holobiont is stressed by another etiological agent (e.g., environmental change, microbial pathogen). A virus could also be an indirect etiological factor of SCTLD, inducing virulence of its microbial host leading to the development of the SCTLD pathogen. This induced virulence may result from mechanisms such as lysogeny or horizontal gene transfer, as seen with the known coral pathogen *Vibrio coralliilyticus* (Weynberg et al., 2015). Furthermore, collectively, coral virus consortia, through infection of their hosts (causing mortality or modified gene expression), can alter the chemical and physical environment within coral tissues, impacting holobiont gene repertoires and thus potentially, SCTLD susceptibility (Broderick et al. 2014; Roossink 2011; Neil and Cadwell 2018; Nichols and Devenport 2021). Virus-driven variation in coral holobiont-wide gene repertoires/expression could explain

interspecies differences in SCTLD susceptibility (Neil and Cadwell 2018).

To explore the possible involvement of viruses in SCTLD etiology, we utilized comparative metatranscriptomics to analyze the viral diversity in apparently healthy and SCTLD-diseased coral tissues. Our study includes a set of 76 metatranscriptomes from five different coral species, covering various health statuses and susceptibility levels to SCTLD. A subset of which were produced from corals included in a SCTLD transmission experiment conducted in St. Thomas USVI in 2019, shortly after the initial SCTLD report (described in Meiling et al., 2021). The remaining metatranscriptomes were generated from affected and unaffected corals sampled *in situ* from St. Thomas reefs, where SCTLD was endemic or had recently broken out. The objectives of this study were to 1) Characterize virus diversity and assess drivers of virus community expression across coral health states and holobiont attributes (e.g., coral species, dominant symbiont), 2) Assess prevalence and productivity of putative Symbiodiniaceae-infecting virus lineages, and 3) Identify candidate virus lineages that may play important roles in SCTLD etiology using differential expression analysis. This study provides valuable insights into viral community dynamics underlying SCTLD infection and sheds light on the potential role of viruses in the development and progression of this devastating coral disease. Overall, we show virus communities differ among samples holobionts and these differences are likely driven by holobiont attributes and, to a lesser extent, holobiont health state. Our study also reveals higher productivity of Symbiodiniaceae-infecting virus lineages in SCTLD-diseased tissues and identifies multiple eukaryotic virus lineages with increased expression in diseased holobionts.

Materials and Methods

Transmission experiment and field sampling

The April 2019 transmission experiment conducted by Meiling et al. (2021) simultaneously exposed fragments of six (6) coral species (*Colpophyllia natans*, *Montastraea cavernosa*, *Orbicella annularis*, *Porites astreoides*, *Pseudodiploria strigosa,* and *Siderastrea siderea*) to SCTLD-affected colonies of *Diploria labyrinthiformis* (Supplemental Figure 1). Paired fragments from each corresponding genotype were equivalently exposed to apparently healthy colonies of *D. labyrinthiformis* as controls (Figure 1B). Over the course of eight days, we monitored coral fragments for the appearance of expanding lesions in the disease treatment. When lesions were observed, the diseased fragment and its corresponding control fragment from the control treatment were flash frozen in liquid nitrogen and stored frozen until RNA extraction (described below). Meiling et al. (2021) then characterized disease phenotypes via measurements of disease prevalence and incidence, relative susceptibility to lesion development, and lesion progression rates. From this experiment, metatranscriptomes were produced from control- and disease-treated fragments for five of six coral species: 1. *C. natans* (n=4 control, n=5 disease), 2. *P. strigosa* (n=4 control, n=4 disease), 3. *M. cavernosa* (n=6 control, n=7 disease), 4. *O. annularis* (n=6 control, n=6 disease), and 5. *P. astreoides* (n=4 control, n=7 disease) (Table 1). Additional samples of *M. cavernosa* were collected *in situ* for metatranscriptome generation from Buck Island and Black Point, St. Thomas (Table 1). This sample set included diseased tissue collected immediately adjacent to the lesion boundary (edge between live coral tissue and denuded skeleton) (n=8, orange circle on Supplemental Figure 2), apparently healthy tissue collected on the same diseased coral colony (n=8, yellow circle on Supplemental Figure 2), and visually healthy tissue collected from a nearby apparently healthy conspecific (n=6, white circle

on Supplemental Figure 2). All field samples for metatranscriptomics were frozen in liquid nitrogen immediately after sampling and stored frozen in -80 °C until RNA extraction. All collection, preservation, and storage of samples were done following (Vega Thurber et al. 2022).

Table 1. Metatranscriptomes analyzed in this study with sample type and dominant symbiont information. AH = Apparently Healthy; EAH = Exposed-Apparently Healthy; and DD = Disease-Diseased.

Sample source	Coral species	Dominant symbiont (# of libraries)	AH	EAH	DD	Total # of libraries
***ex situ* transmission experiment (Meiling et al. 2021)**	*Colpophyllia natans*	*Durusdinium* (5); *Cladocopium* (4); *Breviolum* (1)	5	0	5	10
	Montastraea cavernosa	*Cladocopium* (13)	6	4	3	13
	Orbicella annularis	*Cladocopium* (11); *Durusdinium* (1)	6	0	6	12
	Pseudodiploria strigosa	*Breviolum* (8)	4	2	2	8
	Porites astreoides	*Symbiodinium* (10); *Breviolum* (1)	4	2	5	11
***in situ* coral holobionts (this study)**	*Montastraea cavernosa*	*Cladocopium* (22)	6	8	8	22
Total	--	*Symbiodinium* (10); *Breviolum* (10); *Cladocopium* (50); *Durusdinium* (6)	31	16	29	76

Sample processing for RNA sequencing

Total RNA was extracted from coral tissue samples using the RNAqueous-4PCR total RNA isolation kit (Invitrogen, Life Technologies AM1914). Tissue was lysed using a refrigerated Qiagen TissueLyser II microcentrifuge with 30 oscillations per second for 30 seconds. Lysing was followed by an elution stage that included two consecutive 30 uL elutions. Next, the Ambion DNase I (RNase-free) kit (Invitrogen, Life Technologies AM2222) was used to remove any contaminant DNA or chromatin in the total RNA. Total RNA was prepared for 150 bp, paired-end sequencing on an Illumina NovaSeq 6000 instrument using polyadenylated (polyA) tail selection with the NEBNext Ultra II RNA library prep kit at Novogene Co., Ltd. (Davis, CA, USA).

RNA sequencing library processing and microbial transcript assembly

For bioinformatic analyses, default parameters for all tools were used unless otherwise stated. To start, raw RNA sequencing read libraries were quality checked and filtered using the program fastp (v0.20.1; Chen et al., 2018). In addition, for all analyses, samples were categorized into three groups: 1) Apparently healthy (AH) – metatranscriptomes from tissues collected from corals that showed no apparent SCTLD symptoms, 2) Exposed-apparently healthy (EAH) – metatranscriptomes from tissue sample from corals exposed to SCTLD but showed no SCTLD symptoms (i.e., experimental samples in disease treatment and visually healthy tissue on an in situ-sampled diseased colony), and 3) Disease-diseased (DD) - metatranscriptomes from tissues collected near a SCTLD lesion on SCTLD-affected corals. To start, the program bbsplit (BBMap v38.90) was used to align cleaned reads to coral and Symbiodiniaceae transcriptomes. This allowed reads to subsequently be processed into three files, which respectively contained: (i) reads aligning to coral transcripts, (ii) reads aligning to Symbiodiniaceae transcripts, and (iii) reads that did not align to coral/Symbiodiniaceae transcripts. Non-coral/non-Symbiodiniaceae reads from all libraries were then pooled and normalized with the program bbnorm.sh within BBMap. Contigs were then assembled from normalized reads with the program TransPi (v1.0.3; Rivera-Vicéns et al., 2022), using a multi-assembler and multi-kmer approach. The multi-assembler/kmer strategy made accessible with TransPi produces a higher quality final reference assembly and enhances the likelihood of detecting virus contigs that would have been missed using a single assembler/kmer. The multi-assembler/kmer approach leverages the unique strengths of each tool and increases the overall robustness of the assembly process. The assemblies were generated using rnaSPADES v3.14.0 (kmer: 75, 85, 91, 107 nucleotides) (Bushmanova et al., 2019), Trans-ABySS v2.0.1 (kmer: 25, 35, 55, 75, 85 nucleotides) (Hölzer

& Marz, 2019), SOAPdenovo-Trans v1.03 (kmer: 25, 35, 55, 75, 85 nucleotides) (Xie et al., 2014), Trinity v2.9.1 (kmer: 35 nucleotides) (Grabherr et al., 2011), and Velvet v1.2.12/Oases v0.2.09 (kmer: 65, 71, 81, 87, 91, 97, 101 nucleotides) (Schulz et al., 2012; Zerbino, 2010). All assemblies were then combined into a single file and redundant, misassembled, and short (<300 bp) contigs were removed with the EvidentialGene tr2aacds pipeline v2019.05.14 (Gilbert, 2019) producing a final, non-redundant assembly file. To quantify contigs, non-normalized, cleaned reads were first aligned to the microbial/virus assembly file using bowtie2 (Langmead & Salzberg, 2012) with the Trinity supplemental script align_and_estimate_abundance.pl. Contig counts matrices were then produced with RSEM (Li & Dewey, 2011) using the abundance_estimates_to_matrix.pl Trinity supplemental script. The non-normalized contigs count matrix was used in differential expression analyses with DESeq2, while the trimmed mean of M-values (TMM) normalized counts table was used in community diversity analyses (analyses described below).

Virus sequence prediction and classification

The identification and evaluation of virus transcripts in the final assembly file involved the application of multiple tools; default parameters were used unless specified otherwise. The program Deep6 (Finke et al., 2023) was employed to differentiate between prokaryotic/eukaryotic and virus transcripts within the assembly file. Sequences predicted to be viral by Deep6 with group scores greater than or equal to 0.7 were then extracted to generate a Deep6-predicted virus-like transcript fasta file. These putative virus sequences were then further scrutinized using four different approaches/tools. The first approach was aligning the predicted virus sequences to the IMR/VR (v4.1; Camargo et al., 2022) protein database with DIAMOND BLASTx (v2.0.15; Buchfink et al., 2015) (with options: --min-score 50 --id 30 -l 60 --ultra-

sensitive) to assess homology to known virus sequences. Next, the potential for sequences to be viral was further explored using the programs VirSorter2 (v2.2.4; Guo et al., 2021) and GeNomad (v1.5.0, Camargo et al., 2023). Finally, the program CheckV (v1.01, Nayfach et al., 2021) was used to assess putative virus sequence quality and completeness. A custom BASH script was developed to evaluate results from all approaches and assign each putative viral sequence a "virus confidence score" and a "classification confidence category". A sequence's virus confidence score (*S)* is calculated using the equation $S = \frac{(g+V+C+D+d)}{5}$ where the sum of the geNomad score (*g*), VirSorter2 score (*V*), CheckV score (*C*), DIAMOND score (*D*), and Deep6 score (*d*) is divided by the total possible score (5) (Table 2). Virus confidence scores closer to 1 suggest higher confidence that that sequence is viral based on evidence produced from the five different programs. Confidence scores were used to ensure that any observed patterns were not artifacts of false positives. One way we did this was by pruning low virus confidence (<0.45) sequences from the counts matrices and re-running the analysis and comparing the results.

Table 2. Scheme used to generate virus confidence scores (*S*) for all Deep6 predicted virus transcripts.

Tool	Result	Score
Deep6 (*d*)	Sequence predicted to be viral; length ≥500 bp	1
	Sequence predicted to be viral; length <500 bp	0.5
DIAMOND BLASTx to IMR/VR v4.1 proteins (*D*)	Sequence aligns to reference virus sequence	1
	Sequence does not align to reference virus sequence	0
geNomad (*g*)	Sequence classified as virus	1
	Sequence classified as plasmid	0.5
	Sequence not classified as virus or plasmid	0
VirSorter2 (*V*)	Sequence classified as virus	1
	Sequence not classified as virus	0
CheckV (*C*)	Sequence classified as complete	1.0
	Sequence classified as high quality	0.75
	Sequence classified as medium quality	0.50
	Sequence classified as low quality	0.25
	Sequence quality not determined	0

Putative filamentous RNA virus genome discovery and analysis

All virus transcripts categorized as "complete" or "high quality" by the program CheckV were examined more closely in terms of confidence score, length, and taxonomic classification. High quality contigs with lengths similar to known filamentous RNA viruses and taxonomic classifications to a filamentous RNA virus order (i.e., Patatavirales, Tymovirales) were extracted from the main assembly file. Open reading frame (ORF) prediction and annotation was done within the program geNomad. Genome-wide percent similarity matrices were generated using the program Clustal-Omega (v1.2.4; Sievers et al., 2011) and visualized as a heat map with *ggplot2* in R (version 4.2.2). Maximum likelihood trees were generated using *RdRp* and polyprotein sequences from discovered Tymovirales and Patatavirales genomes, respectively, as well as reference sequences from NCBI's filamentous RNA viruses. Sequence alignment was performed with muscle (v5; Edgar, 2021) and trimmed with trimAL (v1.4; Capella-Gutiérrez et al., 2009) using the automated heuristics based approach (-automated1). A maximum-likelihood

phylogenetic tree was then generated with IQ-TREE2 (v2.0.3; Minh et al., 2020) with 100 non-parametric bootstraps and the LG+I+G4 and the LG+R3 substitution models (determined by ModelFinder, Kalyaanamoorthy et al., 2017) for the Tymovirales and Patatavirales sequence sets, respectively. The trees were then visualized and branches with <50 bootstrap support values were deleted with iTOL (v5; Letunic and Bork 2021)

Community diversity and differential expression analysis

The sequence classification results, expression counts, and sample metadata were imported into RStudio (v.2022.12.0+353) with R (v.4.2.2) and combined into a single phyloseq object using the phyloseq package (McMurdie & Holmes, 2013). Dominant Symbiodiniaceae genus information per sample was sourced from results reported in Beavers et al. (*In Press*). Relative abundance bar plots were then generated from TMM normalized transcript counts to visualize and compare within-sample virus community composition. Contig-level alpha diversity (Shannon's H' index) was measured for each sample using the *estimate_richness*() (measure = "Shannon") function in the phyloseq package. Testing for statistical significant differences in mean Shannon's H' between sample health state groups (i.e., AH, EAH, and DD) was done using a Kruskal–Wallis rank sum test with the *kruskal.test*() function in R after data normality and homogeneity were assessed with a Shapiro-Wilk normality test and Bartlett test of homogeneity of variances, respectively. To assess beta diversity differences among samples, first a NMDS was generated with the TMM normalized counts and the *ordinate*() function within phyloseq with Bray-Curtis distances. To help further elucidate and verify sample clustering behavior, t-SNE plots based on Bray-Curtis distances were also generated using TMM normalized counts with the R package and function *Rtsne* (with options: perplexity = 21, theta = 0.0, max_iter = 10000, exaggeration_factor = 8). Using the linear NMDS method that best

preserves global (across cluster) structure of data and the non-linear T-SNE method that best preserves local (within cluster) structure of the data in tandem, not only validates observations but provides a more complete understanding of the data (Armstrong et al., 2022). Ellipses were added to NMDS and t-SNE plots using the *stat_ellipse* argument within the ggplot2 command. A PERMANOVA was then performed with the *adonis2*() function within the package *vegan* (v2.6-4) to determine the major predictors of the observed community differences between sample groups. UpSet plots visualizing shared transcripts across groups (e.g., coral species, health states) were then generated using the *get_upset*() function in the *MicrobiotaProcess* package and the *upset*() function within the UpsetR package. A Spearman Correlation analysis was run with the *variable_correlation*() function in the phyloschuyler R package to test for correlation of virus transcript expression and lesion progression rates. Log transformed expression heat maps were generated using the *variable_correlation_heatmap*() function in the phyloschuyler R package. Finally, to attempt to increase the biological signal-to-noise ratio and identify candidate virus pathogens, the package *DESeq2* was then used to identify all differentially expressed virus transcripts between health states, coral species, and dominant symbiont communities with the Wald test, "fittype = local", and independent hypothesis weighting (IHW). Along with pruning of low confidence score virus contigs (described above), the validity of patterns reported in the results were assessed by repeating analyses using only the differentially expressed genes identified by DESeq2.

Results

Virus contig assembly and prediction

RNA sequencing and subsequent processing, including coral/dinoflagellate read filtering and microbial/viral read pooling, resulted in a total of 1.21 x 10^9 reads used for contig assembly.

A total of 922,254 putative microbial/viral non-redundant contigs with lengths >300 nucleotides were assembled from these reads using the TransPi pipeline. Deep6 predicted 332,407 (36.04%) of these transcripts were sourced from cellular organisms belonging to the empires *Prokaryota* (n=41,707) or *Eukaryota* (n=290,700), while 193,114 (20.93%) were identified as being of viral origin. DIAMOND blastx alignment to the IMR/VRv4 protein database revealed 40,320 (4.7%) of the 193,114 virus transcripts showed homology to known viruses. The programs geNomad and VirSorter2 predicted 1,045 and 1,111 of the 193,114 transcripts to be viral, respectively. Finally, the program CheckV provided quality information for 6,125 virus transcripts, of which two (2) transcripts were classified as "complete", 38 were determined to be of high-quality, 106 transcripts were categorized as "medium quality", and 5,979 were labeled as "low quality". All results were then combined to produce a final taxonomy file containing sequence characteristics such as virus confidence scores (*S*) and taxonomic classifications. Virus confidence scores ranged between 0.95 and 0.1 with most sequences (78%) having scores below 0.2 (Supplemental Figure 3). There were 3,529 "high confidence" virus sequences with *S* above 0.45 indicating that evidence of virus origin was sourced from at least three of the five programs used for virus sequence identification.

Virus taxa associated with SCTLD-affected, exposed-apparently healthy, and apparently healthy corals

Taxonomic classifications were obtained from Deep6 for 153,344 contigs, from DIAMOND results for 39,383 contigs, and from geNomad results for 387 contigs (note: VirSorter2 and CheckV do not provide taxonomic classifications; Table 3; Figure 1). Overall, most contigs (n=132,395, 68%) were classified to be of RNA virus origin belonging to Realm Riboviria. The remainder were classified to DNA virus realms (n=60,401, 31%) or to DNA virus

ranks unassigned to a realm (i.e., Class Naldaviricetes, n=230; Family Polydnaviriformidae, n=87; Family Bicaudaviridae, n=1; Figure 1A). Most non-RNA virus contigs were assigned to the double-stranded DNA virus Realms Varidnaviria and Duplodnaviria realms, with 33,360 and 24,641 contigs, respectively. Additionally, the single-stranded DNA virus realm Monodnaviria was represented with 2,400 contigs (Figure 1A).

Out of the 193,114 identified virus contigs, 39,215 (20.3%) were classified to taxonomic ranks beyond the Realm level. The majority (n=36,944; 94%) of contigs with lower rank assignments belonged to the DNA virus Kingdoms Heuggongvirae (n=18,400), which include dsDNA bacteriophages and herpes viruses, and Bamfordviridae (n=18,330), comprised of large dsDNA viruses (Figure 1B). The tailed bacteriophage Class Caudoviricetes within Kingdom Heuggongvirae was the most represented DNA virus Class with 18,150 assigned contigs (Figure 1C). This was followed by the Bamfordviridae Classes Megavircetes and Pokkesviricetes having 16,725 and 717 contigs assigned, respectively (Figure 1C). Notably, there were 1,957 Megavircetes contigs classified as belonging to Algavirales, an order of known algae-infecting dsDNA viruses (Aylward et al., 2021). Sequences of other potential dinoflagellate-infecting large dsDNA viruses including viruses from the Orders Imitervirales (Class Megaviricetes) and Asfuvirales (Class Pokkesviricetes) were also detected. Of the 39,215, just 2,271 (~5.8%) RNA virus contigs (assigned to Realm Riboviria) were classified to taxonomic ranks past Realm, potentially highlighting both the novelty of these sequences as well as the incomplete nature of the reference databases used. Among the 2,271 RNA virus contigs with lower rank classifications were assigned to either the Orthornavirae (n=1,560) or Pararnavirae (n=711) kingdoms of non-retroviral or retroviral RNA viruses, respectively (see Figure 1B). The Orthornavirae Class with the most classified contigs was Duplopiviricetes (n=555), followed by

Pisoniviricetes (n=303), both of which contain eukaryotic viruses potentially capable of infecting coral animals, or associated fungi or Symbiodiniaceae symbionts (Figure 1D). In addition, filamentous RNA virus-like sequences, that potentially infect dinoflagellates, were also detected. Specifically, contigs were most similar to known viruses in the Orders Patatavirales and Tymovirales, belonging to Class Stelpaviricetes and Alsuviricetes, respectively (Figure 1D).

Table 3. Breakdown of the number of contigs/transcripts classified to different taxonomic ranks by Deep6, DIAMOND, or geNomad analyses.

Classification	Number of transcripts
Realm	193,114
Kingdom	39,215
Phylum	38,706
Class	38,545
Order	11,039
Family	7,195
Genus	2,002

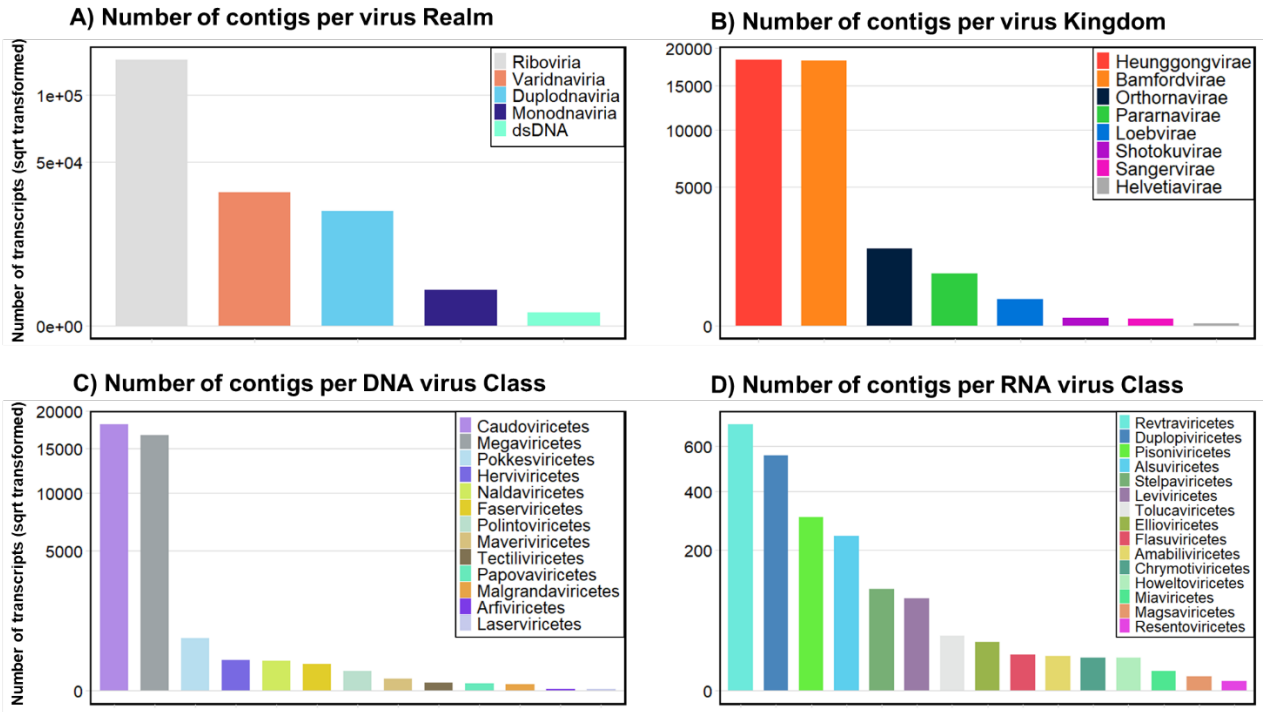


Figure 1. Total number of contigs/transcripts classified to virus Realm (A), Kingdom (B), DNA virus Class (C), and RNA virus Class (D). Contigs/transcripts unassigned to taxonomic ranks lower than Realm (e.g., Kingdom, Class, n=153,899) are not included in 1B-D. Transcripts classified as "dsDNA" at the realm rank (A) represent sequences classified to virus classes unassigned to a realm (e.g., Naldaviricetes). Y-axes were square root transformed using the *scale_y_sqrt*() function within *ggplot2*. Colors are specific to plots. The order of bars in the plot from left to right follows the order of rank names in the legends from top to bottom.

Virus consortia differ among coral holobionts, then disease states

Across health states, the relative proportion of transcripts expressed associated with the four main virus realms remained consistent (Figure 2A). RNA virus contigs were most abundant with little variation across species and health states (Figure 2A). Nucleocytoviricota and Urovircota had the highest proportion of expressed transcripts among the DNA virus phyla (Figure 2B). Although Nucleocytoviricota and Urovircota relative expression remained consistent across species/health, "low expressed" phyla (i.e., Peploviricota, Preplasmiviricota, Hofneiviricota, and Phixviricota) varied among species (Figure 2C). A greater variation in the relative expression of RNA virus phyla was observed, however, Pisuviricota was the most expressed phyla in all species/health states aside from *P. strigosa* AH and EAH sample groups

(Figure 2C). Transcripts from the RNA virus phyla Pisuviricota, Kitrinoviricota, Artverviricota, and Lenarvirivota were observed in nearly all species/health states (Figure 2C). Furthermore, no significant differences in mean contig-level alpha diversity (Shannon's H') was found between sample health groups, as determined by the Kruskal-Wallace rank sum test (chi-squared = 3.08, df = 2, p-value = 0.21).

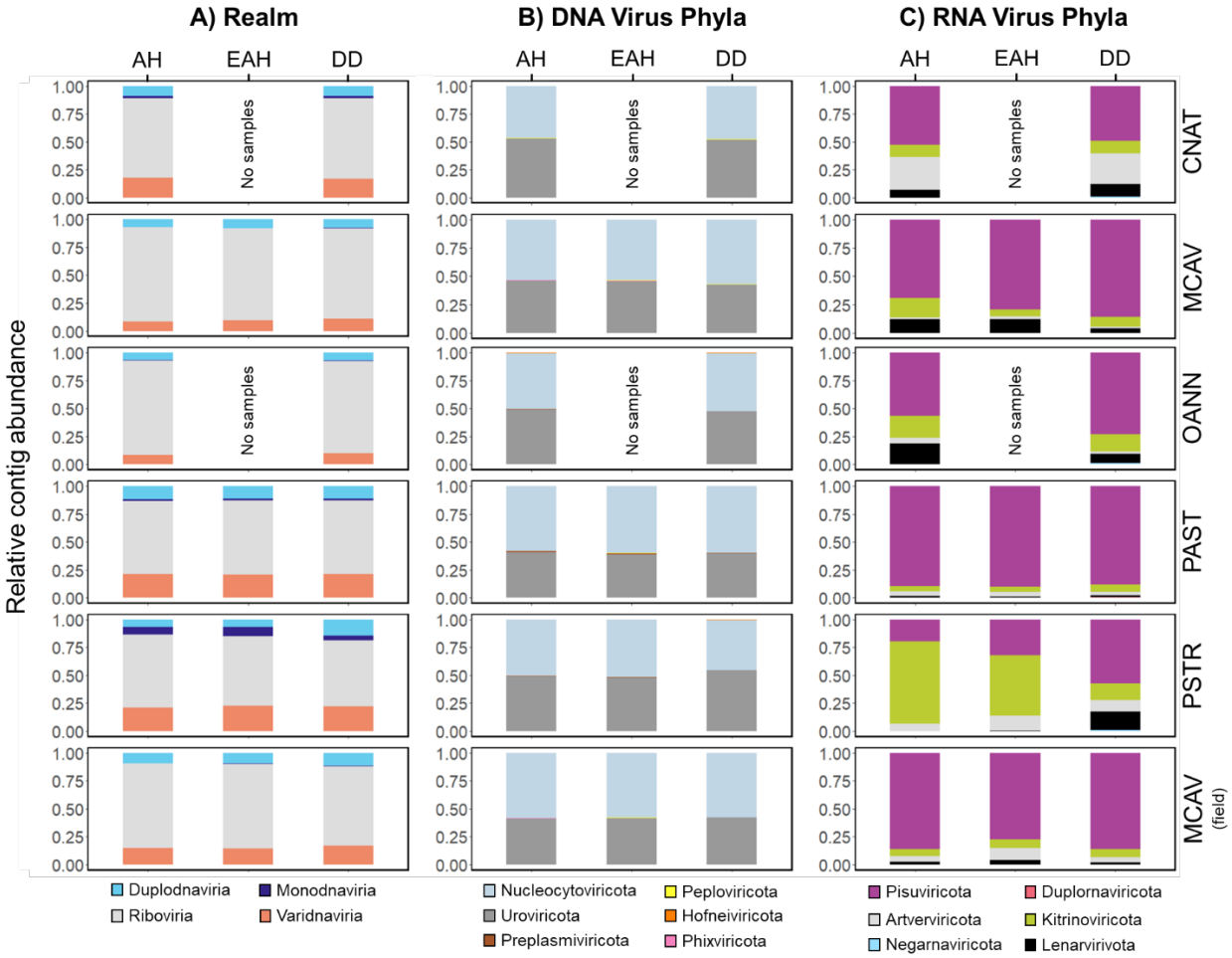


Figure 2. Relative proportion of contigs associated with virus realms (A), DNA virus phyla (B), and RNA virus phyla (C). Relative expression is based on trimmed mean of M-values (TMM) normalized expression counts data and represents the proportion of transcripts expressed across all samples of a given health state (AH, EAH, or DD). AH = Apparently Healthy; EAH = Exposed-Apparently Healthy; and DD = Disease-Diseased. CNAT = *C. natans*; MCAV = *M. cavernosa*; OANN = *O. annularis*; PAST = *P. astreoides*; PSTR = *P. strigosa*.

Ordination plots generated using Brays-Curtis distances illustrated that coral holobiont

attributes (e.g., coral species, dominant Symbiodiniaceae genus) rather than health states seem to be the major drivers of sample clustering (Figure 3, Supplemental Figure 4). Two different approaches were used to increase the biological signal-noise ratio and increase the likelihood of detecting biologically meaningful patterns: 1) Beta diversity analyses based on all high confidence ($S > 0.45$) virus transcripts and 2) Beta diversity analyses based on all differentially expressed transcripts identified by DESeq2. Both removing virus transcript with S below 0.45 and removing transcripts that were not differentially expressed did not influence sample clustering (Supplemental Figure 4C and 4D; Supplemental Figure 5). Moreover, PERMANOVA results indicated that several of the predictors, namely dominant symbiont (F=38.78, p=0.001), coral species (F=1.69, p=0.023), and health status (F=3.27, p=0.004), had significant effects on the virus community transcript composition (Table 4). The interactions between dominant symbiont and coral species (F=2.83, p=0.016), along with health status and dominant symbiont (F=1.69, p=0.023) were also significant (Table 4). In contrast, the interaction between health status and coral species (F=1.34, p=0.187) was not significant.

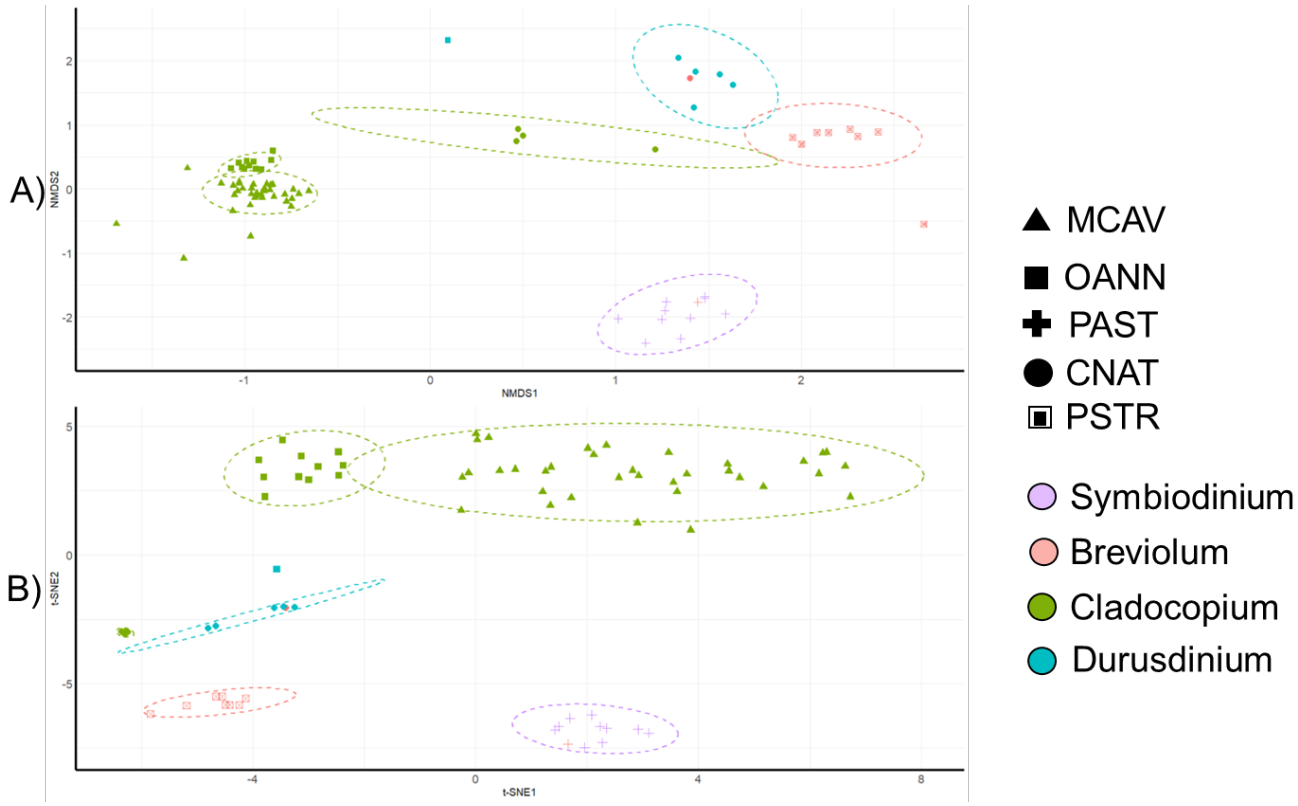


Figure 3. NMDS (A) and t-SNE (B) ordination plots displaying clustering behavior of samples. Shapes represent coral species: *M. cavernosa* (MCAV; triangle), *O. annularis* (OANN, square), *P. astreoides* (PAST, cross), *C. natans* (CNAT, circle), and *P. strigosa* (PSTR, boxed square). Colors represent the dominant symbiont genus: *Symbiodinium* (purple), *Breviolum* (pink), *Cladocopium* (olive), and *Durusdinium* (aquamarine).

Table 4. PERMANOVA results testing for sample attributes that are significant predictors of virus transcript consortia. Df = degrees of freedom. *Indicates a significant p-value.

Predictor	Df	Sum of squares	R2	F	p-value
Health Status	2	0.67	0.03	3.27	0.004*
Dominant symbiont	3	11.91	0.49	38.78	0.001*
Coral Species	4	4.58	0.18	11.18	0.001*
Health Status:Dominant symbiont	5	0.86	0.04	1.69	0.023*
Health Status:Coral Species	2	0.27	0.01	1.34	0.187
Dominant symbiont:Coral Species	1	0.29	0.01	2.83	0.016*
Residual	58	5.94	0.24		
Total	75	24.53	1		

Discovery of novel putative Symbiodiniaceae-infecting filamentous RNA virus genomes

CheckV categorized 40 contigs as "complete" or "high quality"; these contigs potentially represent RNA virus genomes. Of the 40, 13 transcripts were assigned by geNomad to the filamentous RNA virus Orders Tymovirales (n=8) and Patatavirales (n=5). Two of the Tymovirales contigs were previously reported as novel filamentous virus genomes by (Veglia et al., 2022), belong to Family Alphaflexiviridae, and were named Coral Holobiont-associated Filamentous Virus (CHFV) 1 and 2. The other six (previously undetected) putative Tymovirales genomes have similar lengths (between 6,158-6,189 nucleotides) to the CHFV genomes (~6,228 nucleotides) and encode for the same three open reading frames (ORF) (an RNA-dependent RNA polymerase (*RdRp*), a hypothetical protein, and a coat protein) (Table 5, Veglia et al. 2022). In addition, the six newly reported Tymovirales genomes share between ~18-61% genome-wide nucleotide similarity with CHFV1 and 2 and seem to represent two novel "clusters" of coral holobiont-associated Tymoviruses (Figure 4). These newly discovered Tymovirales sequences cluster with high bootstrap support with CHFV and another marine alphaflexivirus (QQG34618.1) in a phylogeny based on translated *RdRp* sequences (Figure 5). Furthermore, three of the five high quality Patatavirales transcripts appear to be full, monopartite potyvirus genomes based on their length (ranging between 10-12 kbp; Table 6). Each putative genome encodes for two open reading frames, a larger one (9,050-10,481 bp) that resembles a polyprotein containing a potyvirus *RdRp* gene (according to geNomad results) and a short ORF (187-365 bp) with ribosomal binding site motifs (Table 6). These newly recovered genomes show high divergence from known patataviruses and likely represent a novel genus within Potyviridae according to the phylogenetic analysis based on translate polyprotein sequences (Figure 6).

Table 5. Summary statistics including genome length and number of open reading frames (ORFs) for nearly complete Tymovirales genomes recovered with comparisons to the previously reported CHFV1 and 2 from Veglia et al. (2022). Pairwise genome similarities to CHFV are based on nucleotide alignments.

Genome ID	Length (bp)	# ORFs	Genome similarity to CHFV1 (%)	Genome similarity to CHFV2 (%)
CHFV1	6,228	3	100	84.57
CHFV2	6,227	3	84.57	100
sctldcontigs1113952	6,189	3	60.90	61.48
sctldcontigs1113954	6,167	3	18.89	18.57
sctldcontigs1351521	6,166	3	61.95	61.47
sctldcontigs1351522	6,158	3	61.92	61.06
sctldcontigs1624265	6,168	3	18.94	18.08
sctldcontigs1638960	6,164	3	60.64	60.85

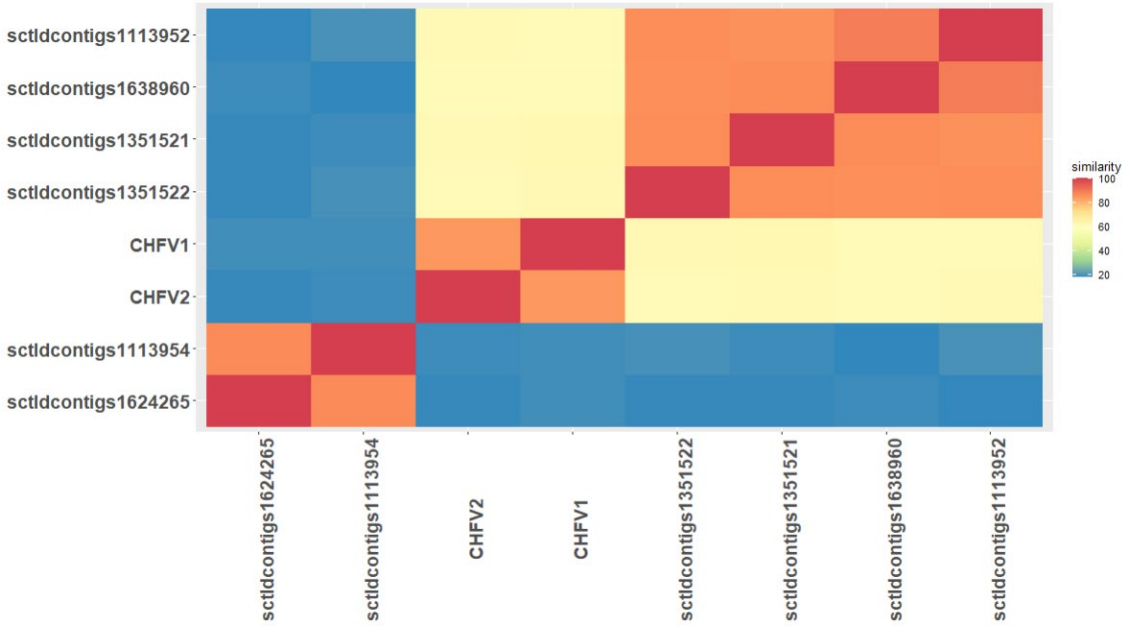


Figure 4. Heatmap visualization of a percent nucleotide identity matrix produced using Clustal-Omega with the newly recovered Tymovirales genomes as well as the CHFV genome previously reported in Veglia et al. (2022). Color gradient represents percent identity, red/orange indicate values >60%, yellow indicates value is less than 60% but greater than 40%, and green/blue indicate values <40%.

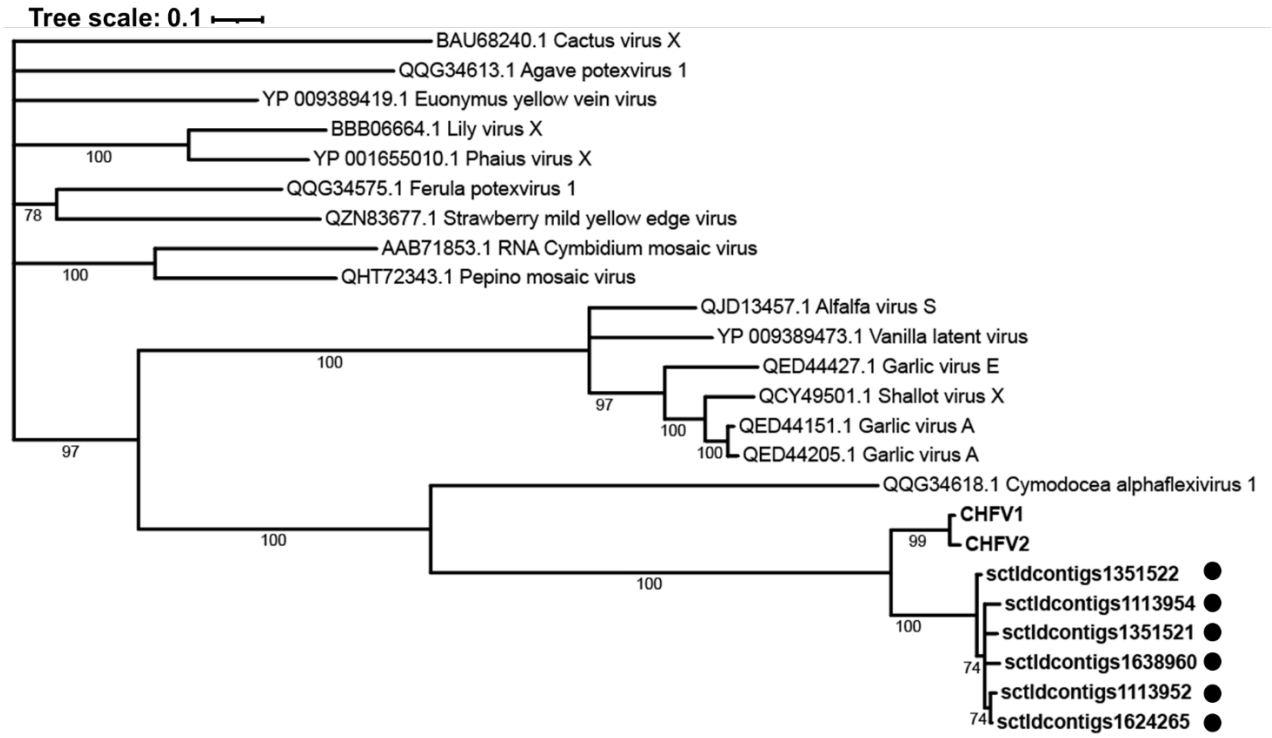


Figure 5. Maximum likelihood phylogeny generated from translated Tymovirales (Family Alphaflexiviridae) RNA-dependent RNA polymerase (*RdRp*) gene sequences sourced from: 1) the CHFV 1 and 2 *RdRp*s reported in Veglia et al. (2022) (bolded), 2) *RdRp* gene sequences from the newly recovered Tymovirales genomes (bolded with black circles), as well as 3) previously described plant-associated alphaflexiviruses. Numbers on branches show bootstrap support values and tree scale represents genetic distance.

Table 6. Description of putative Patatavirales genomes including length information and annotation of identified open reading frames (ORFs).

Genome ID	Genome length	ORF1 length	Putative ORF1 function	ORF2 length	Putative ORF2 function
sctldcontigs1639893	10,025	9,050	Polyprotein (contains an *RdRp* gene)	365	Host ribosome recognition (contains a ribosomal binding site motif)
sctldcontigs1654156	10,529	9,050	Polyprotein (contains an *RdRp* gene)	365	Host ribosome recognition (contains a ribosomal binding site motif)
sctldcontigs1656594	11,284	10,481	Polyprotein (contains an *RdRp* gene)	187	Host ribosome recognition (contains a ribosomal binding site motif)

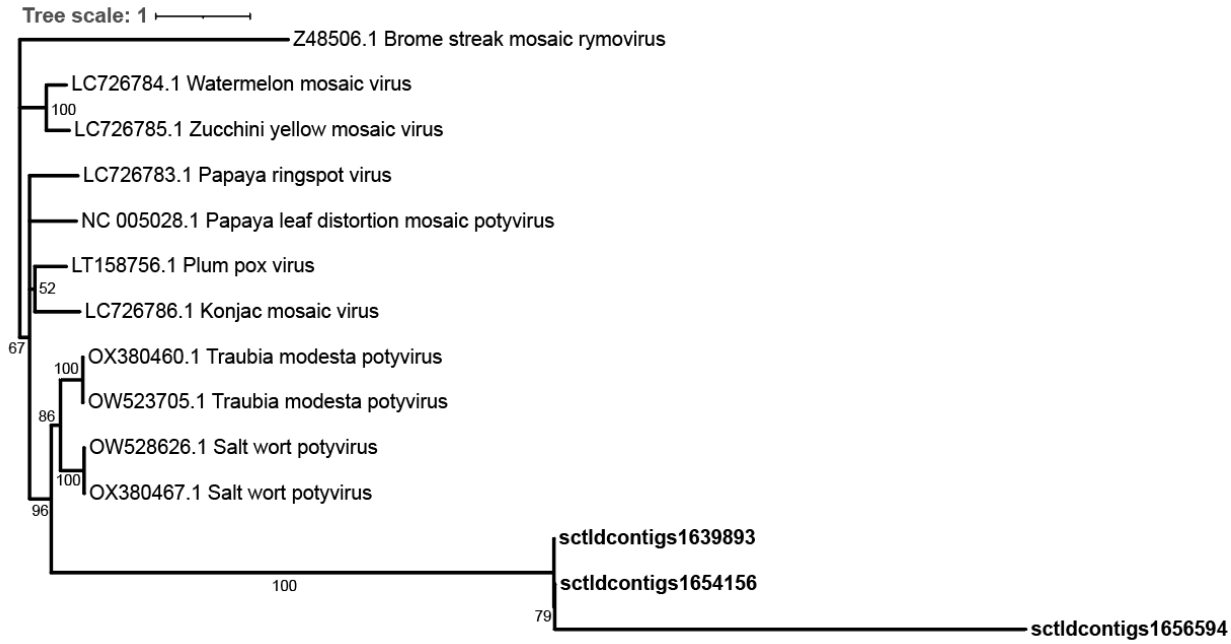


Figure 6. Maximum likelihood phylogeny generated from translated Patatavirales polyprotein gene sequences sourced from: 1) previously described plant-associated Potyviruses, 2) a previously described plant-associated Rymovirus (as outgroup), and 3) the polyprotein sequences from the newly recovered Patatavirales genomes (bolded). Numbers on branches show bootstrap support values and tree scale represents genetic distance.

Differentially expressed virus lineages in diseased coral holobionts

The total number of expressed virus transcripts differed across sample health state groups (Table 7, Figure 7). The DD sample group contained the highest total number of virus contigs/transcripts (n=130,762), followed by the AH group (n=110,372), and the EAH group (n=101,569) (Figure 7). Similarly, the DD and EAH sample groups each had more uniquely associated virus sequences than the HH sample group (Table 7). More sequences classified into hypothesized Symbiodiniaceae-infecting virus orders (including the filamentous virus orders Patatavirales and Tymovirales) were uniquely associated with DD and EAH groups than with AH (Figure 7). Interestingly, even with similar numbers of expressed transcripts per group, the AH and DD groups consistently shared a larger number of transcripts compared to the number shared between EAH and DD (Figure 7). Furthermore, differential expression analysis with DESeq2 revealed 294 virus transcripts differentially expressed ($p<0.1$) between DD and AH

samples, of which, 134 were upregulated in DD samples. Of the 134 upregulated transcripts, the majority were assigned to the realms Varidnaviria (n=64) and Riboviria (n=46). Of these, there were 25 transcripts classified to lower taxonomic ranks (i.e., Order). Most were classified as belonging to likely eukaryote-infecting virus orders, the most represented of which being the dsRNA virus order Durnavirales (Table 8). Durnavirales transcripts were also significantly upregulated in the greatest proportion of DD samples relative to all other identified orders (Figure 8). The majority of remaining upregulated transcript belonged to hypothesized Symbiodiniaceae-infecting virus Orders like Algalvirales, Patatavirales, and Imitervirales, as well as (Table 8, Figure 8). Another notable group containing upregulated transcripts in DD samples is the Order Picornavirales, a potential coral-infecting group of positive-sense, single stranded RNA viruses. Moving on, to identify correlations between virus gene expression and lesion progression rates (ranging from 0.001 to 1.38 cm^2/min; Meiling et al. 2021), a Spearman correlation analysis was performed. The correlation analysis identified a single Imitervirales (Family *Mimiviridae*) transcript as positively correlated with lesion progression rates (rho=0.46, p=0.04).

Table 7. Total and uniquely associated virus contigs with measured expression across sample health groups (AH, EAH, DD).

Sample group	# of libraries	Total number of virus contigs	Virus contigs per library	Total uniquely associated virus contigs	Uniquely associated virus contigs per library
AH	31	110,372	3560.39	16,050	517.7419
EAH	16	101,569	6348.06	29,822	1863.875
DD	29	130,762	4509.03	38,979	1344.103

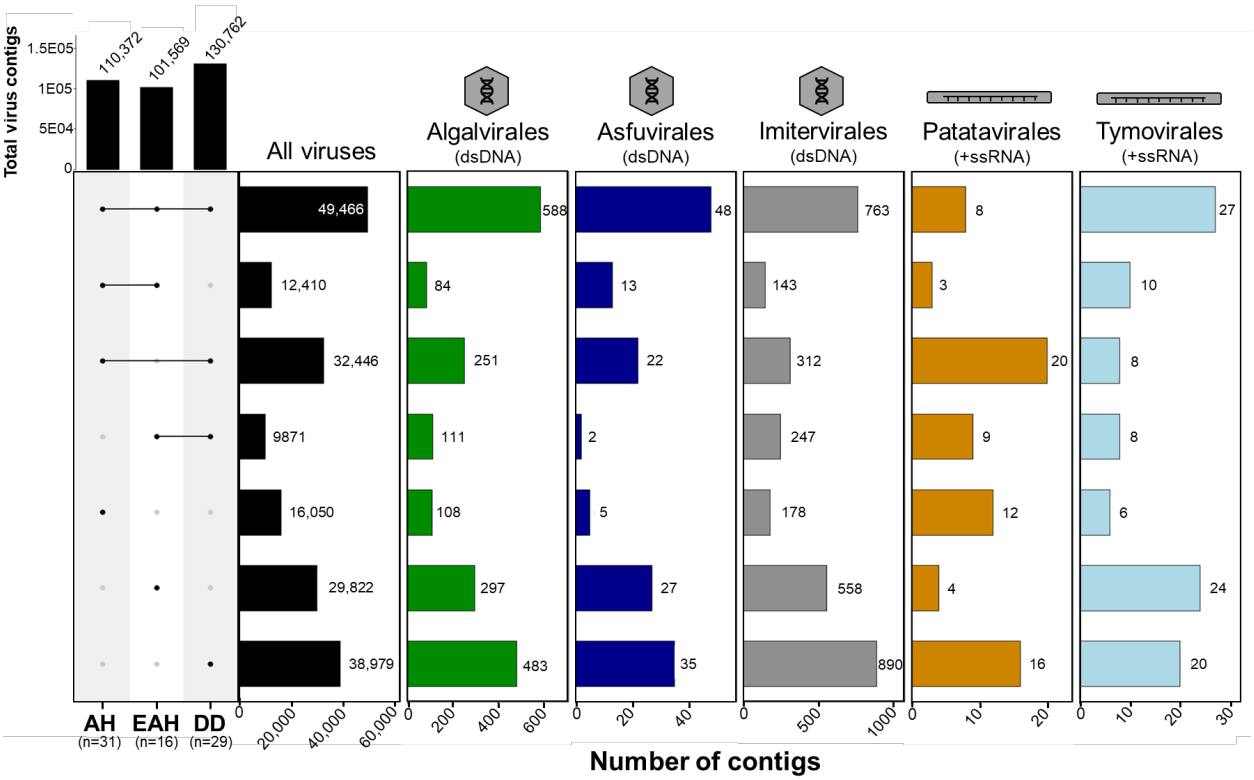


Figure 7. UpSet plot showing shared virus contigs present across sample groups. Intercepts are represented by the black dots with no connection (indicating data specific to that group) or with a connection to other black dots by a line (indicating data represents shared transcripts between groups). AH = Apparently Healthy; EAH = Exposed-Apparently Healthy; and DD = Disease-Diseased.

Table 8. Virus orders associated with upregulated transcripts in Disease-Diseased (DD) samples.

Virus order	Genome type	Hypothesized hosts in coral holobiont	Number of upregulated transcripts
Durnavirales	dsRNA	dinoflagellates or other protists	8
Imitervirales	dsDNA	dinoflagellates or other protists	6
Picornavirales	dsDNA	eukaryotic member of coral holobiont	4
Algavirales	dsDNA	dinoflagellates	3
Pimascovirales	dsDNA	eukaryotic member of coral holobiont	1
Patatavirales	ssRNA	dinoflagellates	1
Norzivirales	ssRNA	prokaryotic member of coral holobiont	1
Chitovirales	dsDNA	eukaryotic member of coral holobiont	1

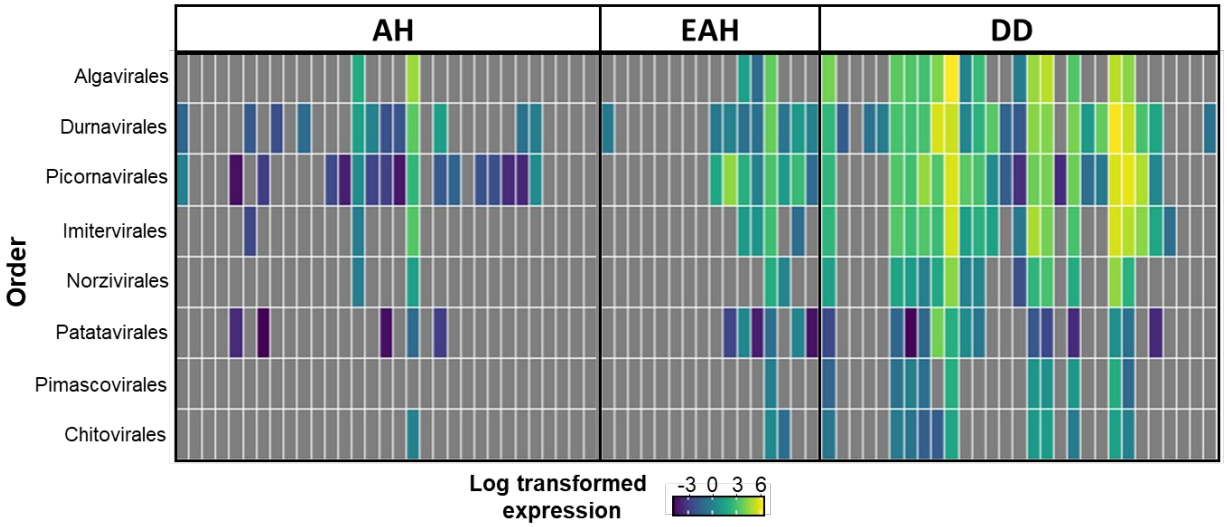


Figure 8. Heatmap displaying log transformed trimmed mean of M-values (TMM) normalized expression values for virus orders with upregulated transcripts in DD samples. Each column represents a sample within the sample group (i.e., AH, EAH, or DD), each box within the columns represents the expression of a given virus genus in that sample. The color of the box represents expression of DESeq2-determined significantly upregulated transcripts in DD samples; higher expression is represented by green or yellow shading, lower expression is represented by blue shading, and grey represents lack of expression. This plot does not include upregulated virus transcripts with no classification to a virus order (n=109). AH = Apparently Healthy; EAH = Exposed-Apparently Healthy; and DD = Disease-Diseased.

Discussion

The frequency of epizootic events affecting biologically and economically productive habitats like coral reefs is increasing (Harvell et al., 2002; Jackson et al., 2014; Vega Thurber et al., 2020; Weil & Rogers, 2011). Effective management and mitigation of coral reef disease outbreaks requires a sufficient understanding of the underlying etiological mechanisms (Vega Thurber et al., 2020). Viruses are ubiquitous members of healthy and diseased stony coral holobionts that are genetically diverse and may impact their hosts and the holobiont in various ways. Currently, we have limited knowledge of how viromes (and specific virus lineages) associated with coral colonies influence coral disease dynamics on reefs. Here we leveraged comparative transcriptomics to elucidate virus community diversity and expression dynamics in the context of the unprecedently lethal stony coral tissue loss disease (SCTLD). We also uncovered novel diversity of RNA viruses associated with coral holobionts, the compartment of

the coral virome we know the least about (Thurber et al., 2017). The results presented here expand our understanding of the potential contributions of viruses to SCTLD etiology and inform the direction of future virus focused SCTLD investigations.

Increased production of diverse viruses in SCTLD-affected holobionts

The breakdown of the coral-Symbiodiniaceae symbiosis is an apparent pathological characteristic of SCTLD infection of a coral holobiont (Landsberg et al., 2020; Meiling et al., 2021; Beavers et al., *In Press*). It has been hypothesized that virus infection of Symbiodiniaceae is the driver of this breakdown in SCTLD-affected tissues, and specifically filamentous RNA viruses were put forth as candidate SCTLD pathogens (Work et al. 2021). In addition, Beavers et al. (*In Press*), who investigated coral and Symbiodiniaceae gene expression from the experiment samples described in this study, reported upregulation of antiviral immunity genes in both the corals and associated Symbiodiniaceae. The results presented here corroborate that observation, indicating that an increase in virus production occurred in SCTLD-affected (DD) and SCTLD-exposed (EAH) tissues, as evidenced by the differences in the number of unique contigs per sample health groups (Figure 7, 8; Table 7). However, the observation of increased sequence abundance for several eukaryotic (and prokaryotic) virus taxa in DD samples, rather than a single lineage as seen with densoviruses in sea star wasting disease (Hewson et al. 2014) and single stranded DNA viruses (SCSDVs) in the putatively viral-associated white plague disease (Soffer et al., 2014), does not support SCTLD as a single-pathogen, viral disease. Even further, the lack of clustering of samples in ordination plots by health state group (i.e., AH, EAH, and DD) likely indicates that there is no one viral lineage that may be diagnostic of SCTLD (Figure 3; Supplemental Figure 5). Instead, altered population dynamics of specific virus groups (e.g., increased relative abundance, changes in alpha diversity) might indicate SCTLD-induced

dysbiosis (example: Norman et al., 2015). Overall, these results suggest that the observed increased virus productivity is likely opportunistic infections by multiple resident lineages of the coral virome.

Symbiont-driven virus community differences across health and diseased holobionts

Beavers et al. (*In Press*) showed that in the samples analyzed here, SCTLD severity was correlated with dominant Symbiodiniaceae identity. Specifically, within *C. natans* holobionts, dominant symbiont was shown to significantly influence host gene expression and lesion acuteness (Beavers et al. *In Press*). Interestingly, our study found that dominant symbiont type was the largest predictor of virus community expression across coral species (Table 4). Virus activity can influence the chemical and physiological environment within animal host tissues and impact (potentially immunity-related) holobiont gene repertoires (Broderick et al., 2014; Neil & Cadwell, 2018; Nichols & Davenport, 2021; Roossinck, 2011). It is thus possible that symbiont-driven virus community differences across corals could result in differences in holobiont susceptibility to dysbiosis via pathogen invasion (Grasis, 2017; Neil & Cadwell, 2018; Roossinck, 2011; Thurber et al., 2017). Such is the case for the enteric virome of humans and mice where the presence of specific virus lineages may help mitigate (Abt et al., 2016; Zhao et al., 2017) or potentiate (Bouziat et al., 2017; Monaco et al., 2016; Norman et al., 2015) diseases and their impact on the animal host (Neil & Cadwell, 2018). In addition, it is even possible that infection by the same virus lineage may have alternative immunopathological outcomes in conspecific hosts (Bouziat et al., 2017). Our results here reveal the upregulation of transcripts assigned to several virus genera (e.g., Algavirales, Durnavirales, Imetervirales) across coral species and dominant symbiont type in the DD sample group (Figure 8). Even further, we detected a positive correlation between the expression of an imetervirus transcript (Family

Mimiviridae) and lesion progression rates within *C. natans* samples. The observed upregulation of transcripts from these resident lineages of the coral virome, interpreted as increased infection activity, suggests that the enhanced productivity of these viral groups could potentially have varying degrees of impacts across different coral species and holobionts. These findings suggest that the altered activity of specific viral lineages, such as Imetervirales, could potentially contribute to the disease severity or susceptibility of coral holobionts. Overall, our results indicate a potential role of resident virus diversity and activity in shaping coral holobiont SCTLD susceptibility and development.

Diverse and prevalent Symbiodiniaceae-infecting RNA virus lineages in healthy and diseased holobionts

Gaining insight into the roles of viruses in coral diseases like SCTLD requires adequate knowledge of the virus diversity commonly associated with coral holobionts and the cellular organisms that inhabit them (e.g., bacteria, Symbiodiniaceae). While coral virology studies over the past three decades have established the core coral DNA virome, the common coral holobiont-associated RNA virus lineages remain enigmatic (Thurber et al. 2017). The results from this study not only shed light on virus community expression dynamics in the context of a deadly disease, but also significantly expand our knowledge of RNA virus diversity associated with healthy and diseased coral holobionts.

One RNA virus group that has garnered considerable interest in the field is the Symbiodiniaceae-infecting filamentous RNA viruses, previously hypothesized to be the pathogenic viral agent of SCTLD (Work et al., 2021). In this study, we provide limited evidence of a single filamentous RNA virus being the sole pathogenic agent of SCTLD. Instead, our results indicate the presence of multiple filamentous RNA virus lineages, including Orders

Tymovirales and Patatavirales, which may be common and diverse residents of both healthy and diseased coral holobionts. However, differential expression analysis revealed the significant upregulation of a Patatavirales transcript in diseased (DD) holobionts, potentially suggesting the contribution of filamentous RNA virus infection of Symbiodiniaceae to disease sign development (i.e., breakdown of coral-Symbiodiniaceae symbiosis). Overall, we recovered a total of eight novel Tymovirales genomes (including those presented in Veglia et al., 2022) and three novel Patatavirales genomes, representing the first of their kind associated with coral tissues. These results, combined with frequent transmission electron microscopy (TEM) observations of filamentous virus-like particles in healthy and diseased coral tissues (Work et al., 2021; Howe-Kerr et al., In Prep), suggest that these viruses are more prevalent in coral holobionts than previously thought.

Similarly, the Order Durnavirales, consisting of dsRNA viruses capable of infecting diverse eukaryotes such as plants, protists, and fungi (Charon et al., 2021; Zhu et al., 2022), remains an underrecognized group with limited reports in coral holobionts. As such, information regarding durnavirus diversity and ecology within coral tissues is currently scarce. However, recent findings revealed the abundance of Durnavirales in marine unicellular algae metatranscriptomes (Charon et al., 2021), suggesting that the observed Durnavirales in this study might be infecting Symbiodiniaceae. Interestingly, Durnavirales was observed to have several highly expressed contigs across coral species and dominant symbiont type (Table 8, Figure 8). Upregulation of durnavirus transcripts was seen in almost all DD samples implying a potentially significant contribution of durnavirus infections to the breakdown of the coral-Symbiodiniaceae symbiosis of SCTLD infection (Figure 8; Landsberg et al., 2020; Meiling et al., 2021; Beavers et al., *In Press*). The prevalence and high expression levels of dsRNA viruses belonging to the

order Durnavirales across various coral species and dominant symbiont types, particularly in SCTLD-diseased corals, highlight the urgent need for further investigation. Durnaviruses have the potential to be a significant contributor to SCTLD disease sign development and can greatly impact the health and resilience of coral holobionts.

Further research is needed to better understand the global distribution of filamentous RNA viruses (Orders Tymovirales and Patatavirales) and dsRNA viruses (Order Durnavirales) in coral holobionts and their impact on the associated Symbiodiniaceae (Howe-Kerr et al., In Prep). This deeper understanding will contribute to a more confident determination of the roles played by these virus lineages in the health of the coral holobiont, especially in the context of diseases like SCTLD.

Evidence for virus-driven alteration to coral antiviral mechanism

Analyzing virus expression dynamics in samples with characterized cellular (host) gene expression can provide crucial context to observations shedding light on the interplay between host and viral gene expression. Beavers et al. (*In Press*) previously revealed alterations in expression of coral antiviral mechanism-related genes in SCTLD-affected samples included in this study. Notably, Beavers and colleagues (*In Press*) identified the upregulation of *Abce1* gene across coral species. The Abce1 gene serves as a negative regulator of RNase L, an enzyme involved in single stranded RNA degradation and shown to reduce virus propagation (Drappier & Michiels, 2015; Zhou et al., 1997). Interestingly, human +ssRNA viruses have been shown to inhibit and evade RNase L antiviral mechanisms through the upregulation of host *Abce1* (Drappier & Michiels, 2015). One such virus, the encephalomyocarditis-virus (Martinand et al., 1998), is a +ssRNA virus that belongs to Picornavirales, an virus order observed to have upregulated transcripts in 21 of 29 DD samples (Figure 8). Picornavirales sequences were also

detected in AH and EAH samples, suggesting that picornaviruses may establish persistent infections in coral hosts, as observed in other animals (Colbère-Garapin et al., 2002). It is possible that coral-associated picornaviruses increase activity as a response to stress (in this case SCTLD infection) potentially driving some observed changes in coral gene expression such as upregulation of *Abce1* (Figure 8). If the upregulation of *Abce1* is driven by coral-infecting picornaviruses, this viral group may play a significant role in influencing coral antivirus capabilities during stress (or SCTLD infection). This, in turn, could facilitate opportunistic infections by various resident viral lineages of coral holobionts, as observed in our results (Figure 7, 8).

Future directions for deciphering the role of viruses in SCTLD

In conclusion, this study has provided crucial insights into virus diversity and activity in coral holobionts across different SCTLD health states. The upregulation of antivirus immunity genes in both the corals and Symbiodiniaceae, as reported by Beavers et al. (*In Press*), aligns with the observed increase in productivity by putative coral-infecting virus lineages (e.g., Picornavirales) and Symbiodiniaceae-infecting virus lineages (e.g., Algalvirales, Imitervirales, Patatavirales). Collectively, these findings support the hypothesis of viral involvement in SCTLD etiology, particularly through increased opportunistic infections by several resident virus groups affecting both the coral and its symbiotic Symbiodiniaceae (e.g., Durnavirales, Picornavirales). However, more investigation is needed to better understand these virus groups in terms of their diversity, prevalence, and impact on coral holobionts both in and outside the context of SCTLD. Future investigations should look to: 1) isolate and culture members of these virus groups to facilitate comprehensive characterization (e.g., whole genome sequencing, determination of virus particle structure) and lab-based experimentation to determine role in

coral holobionts, 2) develop targeted PCR primer sets to leverage alternative sequencing techniques like amplicon sequencing to decipher community diversity/dynamics of these virus groups on SCTLD-affect and SCTLD-unaffected reefs (see Grupstra et al., 2022; Howe-Kerr, 2022), and 3) use developed primer sets for more quantitative molecular approaches like quantitative PCR or droplet digital PCR to assess true abundance of these virus groups in healthy and diseased coral tissues.

Acknowledgements

This work was supported by National Science Foundation awards 2109622 to M.E.B., 1928753 to M.E.B. and T.B.S., 1928609 to A.M.S.C., 1928817 to E.M.M., 19228771 to L.D.M., 1927277 to D.M.H., and 1928761 to A.A., as well as by VI EPSCoR (NSF numbers 0814417 and 1946412). This work could not have been completed without the hours of lab and fieldwork dedicated by Adam Glahn, Danielle Lasseigne, Amanda Long, Bradley Arrington, Daniel Mele, Kathryn Cobleigh, Bradford Dimos, Nicholas Macknight, Naomi Huntley, and Alexandra Gutting. We also thank Nikoloas Schizas and the Marine Genomic Biodiversity Lab of the Department of Marine Sciences at the University of Puerto Rico—Mayaguez for providing us with access to computational resources used to conduct this study. Field collections were authorized by the Department of Planning and Natural Resources Coastal Zone Management permit #DFW19057U.

References

Abt, M. C., Buffie, C. G., Sušac, B., Becattini, S., Carter, R. A., Leiner, I., Keith, J. W., Artis, D., Osborne, L. C., & Pamer, E. G. (2016). TLR-7 activation enhances IL-22–mediated colonization resistance against vancomycin-resistant enterococcus. *Science Translational Medicine*, *8*(327), 327ra25-327ra25. https://doi.org/10.1126/scitranslmed.aad6663

Aeby, G. S., Ushijima, B., Campbell, J. E., Jones, S., Williams, G. J., Meyer, J. L., Häse, C., & Paul, V. J. (2019). Pathogenesis of a tissue loss disease affecting multiple species of corals along the Florida Reef Tract. *Frontiers in Marine Science*, *6*, 678.

Alvarez-Filip, L., Estrada-Saldívar, N., Pérez-Cervantes, E., Molina-Hernández, A., & González-Barrios, F. J. (2019). A rapid spread of the stony coral tissue loss disease outbreak in the Mexican Caribbean. *PeerJ*, *7*, e8069. https://doi.org/10.7717/peerj.8069

Armstrong, G., Rahman, G., Martino, C., McDonald, D., Gonzalez, A., Mishne, G., & Knight, R. (2022). Applications and Comparison of Dimensionality Reduction Methods for Microbiome Data. *Frontiers in Bioinformatics*, *2*, 821861. https://doi.org/10.3389/fbinf.2022.821861

Aronson, R. B., & Precht, W. F. (2001). White-band disease and the changing face of Caribbean coral reefs. *Hydrobiologia*, *460*, 25–38. https://doi.org/10.1007/978-94-017-3284-0_2

Aylward, F. O., Moniruzzaman, M., Ha, A. D., & Koonin, E. V. (2021). A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLOS Biology*, *19*(10), e3001430. https://doi.org/10.1371/journal.pbio.3001430

Beavers, K. M., Van Buren, E., Rossin, A., Emery, M., Veglia, A. J., Karrick, C. E., MacKnight, N. J., Dimos, B. A., Meiling, S. S., Smith, T. B., Apprill, A., Muller, E. M., Holstein, D. M., Correa, A. M. S., Brandt, M. E., Mydlarz, L. D. (2023). Stony Coral Tissue Loss Disease Induces Transcriptional Signatures of in situ Degradation of Dysfunctional Symbiodiniaceae. *Nature Communications. In press*

Buerger, P., Weynberg, K. D., Sato, Y., Willis, B. L., & Oppen, M. van. (2019). Bacteriophages in black band disease Novel T4 bacteriophages associated with black band disease in corals. *Environmental Microbiology.* https://doi.org/10.1111/1462-2920.14432

Bouziat, R., Hinterleitner, R., Brown, J. J., Stencel-Baerenwald, J. E., Ikizler, M., Mayassi, T., Meisel, M., Kim, S. M., Discepolo, V., Pruijssers, A. J., Ernest, J. D., Iskarpatyoti, J. A., Costes, L. M. M., Lawrence, I., Palanski, B. A., Varma, M., Zurenski, M. A., Khomandiak, S., McAllister, N., … Jabri, B. (2017). Reovirus infection triggers inflammatory responses to dietary antigens and development of celiac disease. *Science*, *356*(6333), 44–50. https://doi.org/10.1126/science.aah5298

Brandt, M. E., Ennis, R. S., Meiling, S. S., Townsend, J., Cobleigh, K., Glahn, A., Quetel, J., Brandtneris, V., Henderson, L. M., & Smith, T. B. (2021). The Emergence and Initial Impact of Stony Coral Tissue Loss Disease (SCTLD) in the United States Virgin Islands. *Frontiers in Marine Science*, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.715329

Broderick, N. A., Buchon, N., & Lemaitre, B. (2014). Microbiota-induced changes in Drosophila melanogaster host gene expression and gut morphology. *MBio*, *5*(3), e01117-14. https://doi.org/10.1128/mBio.01117-14

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), Article 1. https://doi.org/10.1038/nmeth.3176

Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, *8*(9), giz100. https://doi.org/10.1093/gigascience/giz100

Camargo, A. P., Nayfach, S., Chen, I.-M. A., Palaniappan, K., Ratner, A., Chu, K., Ritter, S. J., Reddy, T. B. K., Mukherjee, S., Schulz, F., Call, L., Neches, R. Y., Woyke, T., Ivanova, N. N., Eloe-Fadrosh, E. A., Kyrpides, N. C., & Roux, S. (2022). IMG/VR v4: An expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Research*, gkac1037. https://doi.org/10.1093/nar/gkac1037

Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., & Kyrpides, N. C. (2023). *You can move, but you can't hide: Identification of mobile genetic elements with geNomad* (p. 2023.03.05.531206). bioRxiv. https://doi.org/10.1101/2023.03.05.531206

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Charon, J., Murray, S., & Holmes, E. C. (2021). Revealing RNA virus diversity and evolution in unicellular algae transcriptomes. *Virus Evolution*, *7*(2), veab070. https://doi.org/10.1093/ve/veab070

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Colbère-Garapin, F., Pelletier, I., & Ouzilou, L. (2002). Persistent Infections by Picornaviruses. In *Molecular Biology of Picornavirus* (pp. 437–448). John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555817916.ch35

Correa, A. M. S., Ainsworth, T. D., Rosales, S. M., Thurber, R. L. V., Butler, C. R., & Thurber, R. L. V. (2016). Viral outbreak in corals associated with an in situ bleaching event: Atypical herpes-like viruses and a new megavirus infecting Symbiodinium. *Frontiers in Microbiology*, *7*, 127. https://doi.org/10.3389/fmicb.2016.00127

Correa, A. M. S., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., & Weitz, J. S. (2021). Revisiting the rules of life for viruses of microorganisms. *Nature Reviews Microbiology*, *19*(8), 501–513. https://doi.org/10.1038/s41579-021-00530-x

Correa, A. M. S., Welsh, R. M., & Vega Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal*, *7*(1), Article 1. https://doi.org/10.1038/ismej.2012.75

Davy, J. E., & Patten, N. L. (2007). Morphological diversity of virus-like particles within the surface microlayer of scleractinian corals. *Aquatic Microbial Ecology*, *47*, 37–44. https://doi.org/10.3354/ame047037

Drappier, M., & Michiels, T. (2015). Inhibition of the OAS/RNase L pathway by viruses. *Current Opinion in Virology*, *15*, 19–26. https://doi.org/10.1016/j.coviro.2015.07.002

Edgar, R. C. (2021). *MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping* (p. 2021.06.20.449169). bioRxiv. https://doi.org/10.1101/2021.06.20.449169

Finke, J. F., Kellog, C., & Suttle, C. A. (2023). *Deep6: Classification of Metatranscriptomic Sequences into Cellular Empires and Viral Realms Using Deep Learning Models |*

Microbiology Resource Announcements. https://journals.asm.org/doi/10.1128/mra.01079-22

Gilbert, D. G. (2019). *Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes?* (p. 829184). bioRxiv. https://doi.org/10.1101/829184

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Grasis, J. A. (2017). The Intra-Dependence of Viruses and the Holobiont. *Frontiers in Immunology*, *8*. https://www.frontiersin.org/articles/10.3389/fimmu.2017.01501

Grupstra, C. G., Howe-Kerr, L. I., Veglia, A. J., Bryant, R. L., Coy, S. R., Blackwelder, P. L., & Correa, A. (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *The ISME Journal*, 1–12.

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

Harvell, C. D., Mitchell, C. E., Ward, J. R., Altizer, S., Dobson, A. P., Ostfeld, R. S., & Samuel, M. D. (2002). Climate Warming and Disease Risks for Terrestrial and Marine Biota. *Science*, *296*(5576), 2158–2162. https://doi.org/10.1126/science.1063699

Harvell, D., Jordán-Dahlgren, E., Merkel, S., Rosenberg, E., Raymundo, L., Smith, G., Weil, E., & Willis, B. (2007). Coral Disease, Environmental Drivers, and the Balance Between Coral and Microbial Associates. *Oceanography*, *20*(1), 172–195. https://doi.org/10.5670/oceanog.2007.91

Hewson, I., Button, J. B., Gudenkauf, B. M., Miner, B., Newton, A. L., Gaydos, J. K., Wynne, J., Groves, C. L., Hendler, G., Murray, M., Fradkin, S., Breitbart, M., Fahsbender, E., Lafferty, K. D., Kilpatrick, A. M., Miner, C. M., Raimondi, P., Lahner, L., Friedman, C. S., … Harvell, C. D. (2014). Densovirus associated with sea-star wasting disease and mass mortality. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(48), 17278–17283. https://doi.org/10.1073/pnas.1416625111

Hölzer, M., & Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience*, *8*(5), giz039. https://doi.org/10.1093/gigascience/giz039

Howe-Kerr, L. I. (2022). Viruses of a key coral symbiont exhibit temperature-driven productivity across a reefscape. https://doi.org/10.21203/rs.3.rs-1899377/v1

Howe-Kerr, L. I., Knochel, A. M., Meyer, M. M., Sims, J. A., Karrick, C. E., Grupstra, C. G. B., Veglia, A. J., Thurber, A. R., Vega Thurber, R. L., Correa, A. M. S. (*In Prep*). Filamentous virus-like particles are present in coral dinoflagellates across genera and ocean basins.

Jackson, J., Donovan, M., Cramer, K., & Lam, V. Y. Y. (2014). *Status and trends of Caribbean coral reefs: 1970-2012*. https://www.semanticscholar.org/paper/Status-and-trends-of-Caribbean-coral-reefs-%3A-Jackson-Donovan/6a7d1f5fb6784b5444191bf1cc39f79fd97d454b

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), Article 6. https://doi.org/10.1038/nmeth.4285

Landsberg, J. H., Kiryu, Y., Peters, E. C., Wilson, P. W., Perry, N., Waters, Y., Maxwell, K. E., Huebner, L. K., & Work, T. M. (2020). Stony Coral Tissue Loss Disease in Florida Is Associated With Disruption of Host–Zooxanthellae Physiology. *Frontiers in Marine Science*, *7*. https://www.frontiersin.org/articles/10.3389/fmars.2020.576013

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

Lawrence, S. A., Floge, S. A., Davy, J. E., Davy, S. K., & Wilson, W. H. (2017). Exploratory analysis of Symbiodinium transcriptomes reveals potential latent infection by large dsDNA viruses. *Environmental Microbiology*, *19*(10), 3909–3919. https://doi.org/10.1111/1462-2920.13782

Lesser, M. P., Bythell, J. C., Gates, R. D., Johnstone, R. W., & Hoegh-Guldberg, O. (2007). Are infectious diseases really killing corals? Alternative interpretations of the experimental and ecological data. *Journal of Experimental Marine Biology and Ecology*, *346*(1), 36–44. https://doi.org/10.1016/j.jembe.2007.02.015

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Marhaver, K. L., Edwards, R. A., & Rohwer, F. (2008). Viral communities associated with healthy and bleaching corals. *Environmental Microbiology*, *10*(9), 2277–2286. https://doi.org/10.1111/j.1462-2920.2008.01652.x

Martinand, C., Salehzada, T., Silhol, M., Lebleu, B., & Bisbal, C. (1998). RNase L inhibitor (RLI) antisense constructions block partially the down regulation of the 2-5A/RNase L pathway in encephalomyocarditis-virus-(EMCV)-infected cells. *European Journal of Biochemistry*, *254*(2), 248–255. https://doi.org/10.1046/j.1432-1327.1998.2540248.x

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Meiling, S., Muller, E. M., Smith, T. B., & Brandt, M. E. (2020). 3D Photogrammetry Reveals Dynamics of Stony Coral Tissue Loss Disease (SCTLD) Lesion Progression Across a Thermal Stress Event. *Frontiers in Marine Science*, *7*. https://www.frontiersin.org/articles/10.3389/fmars.2020.597643

Meiling, S. S., Muller, E. M., Lasseigne, D., Rossin, A., Veglia, A. J., MacKnight, N., Dimos, B., Huntley, N., Correa, A. M. S., Smith, T. B., Holstein, D. M., Mydlarz, L. D., Apprill, A., & Brandt, M. E. (2021). Variable Species Responses to Experimental Stony Coral Tissue Loss Disease (SCTLD) Exposure. *Frontiers in Marine Science*, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.670829

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Monaco, C. L., Gootenberg, D. B., Zhao, G., Handley, S. A., Ghebremichael, M. S., Lim, E. S., Lankowski, A., Baldridge, M. T., Wilen, C. B., Flagg, M., Norman, J. M., Keller, B. C., Luévano, J. M., Wang, D., Boum, Y., Martin, J. N., Hunt, P. W., Bangsberg, D. R., Siedner, M. J., … Virgin, H. W. (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host & Microbe*, *19*(3), 311–322. https://doi.org/10.1016/j.chom.2016.02.011

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

Neil, J. A., & Cadwell, K. (2018). The Intestinal Virome and Immunity. *The Journal of Immunology*, *201*(6), 1615–1624. https://doi.org/10.4049/jimmunol.1800631

Nichols, R. G., & Davenport, E. R. (2021). The relationship between the gut microbiome and host gene expression: A review. *Human Genetics*, *140*(5), 747–760. https://doi.org/10.1007/s00439-020-02237-0

Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., Kambal, A., Monaco, C. L., Zhao, G., Fleshner, P., Stappenbeck, T. S., McGovern, D. P. B., Keshavarzian, A., Mutlu, E. A., Sauk, J., Gevers, D., Xavier, R. J., Wang, D., Parkes, M., & Virgin, H. W. (2015). Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell*, *160*(3), 447–460. https://doi.org/10.1016/j.cell.2015.01.002

Precht, W. F., Gintert, B. E., Robbart, M. L., Fura, R., & van Woesik, R. (2016). Unprecedented Disease-Related Coral Mortality in Southeastern Florida. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep31374

Rivera-Vicéns, R. E., Garcia-Escudero, C. A., Conci, N., Eitel, M., & Wörheide, G. (n.d.). TransPi—A comprehensive TRanscriptome ANalysiS PIpeline for de novo transcriptome assembly. *Molecular Ecology Resources*, *n/a*(n/a). https://doi.org/10.1111/1755-0998.13593

Roossinck, M. J. (2011). The good viruses: Viral mutualistic symbioses. *Nature Reviews Microbiology*, *9*(2), 99–108. https://doi.org/10.1038/nrmicro2491

Rosario, K., Van Bogaert, N., López-Figueroa, N. B., Paliogiannis, H., Kerr, M., & Breitbart, M. (2022). Freshwater macrophytes harbor viruses representing all five major phyla of the RNA viral kingdom Orthornavirae. *PeerJ*, *10*, e13875. https://doi.org/10.7717/peerj.13875

Schulz, M. H., Zerbino, D. R., Vingron, M., & Birney, E. (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, *28*(8), 1086–1092. https://doi.org/10.1093/bioinformatics/bts094

Sharp, W., & Maxwell, K. (2018). Investigating the ongoing coral disease outbreak in the Florida Keys: collecting corals to diagnose the etiological agent (s) and establishing sentinel sites to monitor transmission rates and the spatial progression of the disease. *Florida Fish and Wildlife Conservation Commission*.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Soffer, N., Brandt, M. E., Correa, A. M. S., Smith, T. B., & Thurber, R. V. (2014). Potential role of viruses in white plague coral disease. *The ISME Journal*, *8*(2), 271–283. https://doi.org/10.1038/ismej.2013.137

Sutherland, K. P., Porter, J. W., & Torres, C. (2004). Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. *Marine Ecology Progress Series*, *266*, 273–302. https://doi.org/10.3354/meps266273

Sweet, M. & Bythell, J. (2017). The role of viruses in coral health and disease. *Journal of Invertebrate Pathology*, 147, 136-144. https://doi.org/10.1016/j.jip.2016.12.005

Thurber, R. V., Payet, J. P., Thurber, A. R., & Correa, A. M. S. (2017). Virus–host interactions and their roles in coral reef health and disease. *Nature Reviews Microbiology*, *15*(4), 205–216. https://doi.org/10.1038/nrmicro.2016.176

Thurber, R. V., Schmeltzer, E. R., Grottoli, A. G., Woesik, R. van, Toonen, R. J., Warner, M., Dobson, K. L., McLachlan, R. H., Barott, K., Barshis, D. J., Baumann, J., Chapron, L., Combosch, D. J., Correa, A. M., DeCarlo, T. M., Hagedorn, M., Hédouin, L., Hoadley, K., Felis, T., … Sawall, Y. (2022). Unified methods in collecting, preserving, and archiving coral bleaching and restoration specimens to increase sample utility and interdisciplinary collaboration. *PeerJ*, *10*, e14176. https://doi.org/10.7717/peerj.14176

Tracy, A. M., Pielmeier, M. L., Yoshioka, R. M., Heron, S. F., & Harvell, C. D. (2019). Increases and decreases in marine disease reports in an era of global change. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1912), 20191718. https://doi.org/10.1098/rspb.2019.1718

van Oppen, M. J., Leong, J. A., & Gates, R. D. (2009). Coral-virus interactions: a double-edged sword?. *Symbiosis*, 47, 1-8. https://doi.org/10.1007/BF03179964

Vega Thurber, R., Mydlarz, L. D., Brandt, M., Harvell, D., Weil, E., Raymundo, L., Willis, B. L., Langevin, S., Tracy, A. M., Littman, R., Kemp, K. M., Dawkins, P., Prager, K. C., Garren, M., & Lamb, J. (2020). Deciphering Coral Disease Dynamics: Integrating Host, Microbiome, and the Changing Environment. *Frontiers in Ecology and Evolution*, *8*. https://www.frontiersin.org/articles/10.3389/fevo.2020.575927

Veglia, A. J., Beavers, K., Buren, E. W. V., Meiling, S. S., Muller, E. M., Smith, T. B., Holstein, D. M., Apprill, A., Brandt, M. E., Mydlarz, L. D., & Correa, A. M. S. (2022). Alphaflexivirus Genomes in Stony Coral Tissue Loss Disease-Affected, Disease-Exposed, and Disease-Unexposed Coral Colonies in the U.S. Virgin Islands. *Microbiology Resource Announcements*. https://doi.org/10.1128/mra.01199-21

Walton, C. J., Hayes, N. K., & Gilliam, D. S. (2018). Impacts of a Regional, Multi-Year, Multi-Species Coral Disease Outbreak in Southeast Florida. *Frontiers in Marine Science*, *5*. https://www.frontiersin.org/articles/10.3389/fmars.2018.00323

Weil, E., & Rogers, C. S. (2011). Coral Reef Diseases in the Atlantic-Caribbean. In *Coral Reefs: An Ecosystem in Transition*. Springer. https://doi.org/10.1007/978-94-007-0114-4

Weil, E., Smith, G., & Gil-Agudelo, D. L. (2006). Status and progress in coral reef disease research. *Diseases of Aquatic Organisms*, *69*(1), 1–7. https://doi.org/10.3354/dao069001

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P., & van Oppen, M. J. H. (2015). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. *Scientific Reports*, *5*, 17889. https://doi.org/10.1038/srep17889

Williams, S. M., García-Sais, J., & Sabater-Clavell, J. (2021). Prevalence of Stony Coral Tissue Loss Disease at El Seco, a Mesophotic Reef System off Vieques Island, Puerto Rico.

Frontiers in Marine Science, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.668669

Work, T. M., Weatherby, T. M., Landsberg, J. H., Kiryu, Y., Cook, S. M., & Peters, E. C. (2021). Viral-Like Particles Are Associated With Endosymbiont Pathology in Florida Corals Affected by Stony Coral Tissue Loss Disease. *Frontiers in Marine Science*, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.750658

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G. K.-S., & Wang, J. (2014). SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*(12), 1660–1666. https://doi.org/10.1093/bioinformatics/btu077

Zerbino, D. R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *CHAPTER*, Unit-11.5. https://doi.org/10.1002/0471250953.bi1105s31

Zhao, G., Vatanen, T., Droit, L., Park, A., Kostic, A. D., Poon, T. W., Vlamakis, H., Siljander, H., Härkönen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Ilonen, J., Wang, D., Knip, M., Xavier, R. J., & Virgin, H. W. (2017). Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proceedings of the National Academy of Sciences*, *114*(30), E6166–E6175. https://doi.org/10.1073/pnas.1706359114

Zhou, A., Paranjape, J., Brown, T. L., Nie, H., Naik, S., Dong, B., Chang, A., Trapp, B., Fairchild, R., Colmenares, C., & Silverman, R. H. (1997). Interferon action and apoptosis are defective in mice devoid of 2′,5′-oligoadenylate-dependent RNase L. *The EMBO Journal*, *16*(21), 6355–6363. https://doi.org/10.1093/emboj/16.21.6355

CONCLUDING REMARKS

Background and summary

Stony corals and their symbiotic microbial communities, consisting of diverse bacteria, archaea, fungi, protists, and viruses, coexist and evolve as a single biological unit known as the "holobiont" (Bordenstein & Theis, 2015). Alterations to the community of symbiotic microbes, or interactions among them, can influence emergent coral holobiont phenotypes and fitness (Bordenstein & Theis, 2015; Howe-Kerr et al., 2020; Hussa & Goodrich-Blair, 2013; Pita et al., 2016). These alterations in the holobiont can help corals acclimate to novel environments/resources, however, they can also increase susceptibility to stress and/or disease driven dysbiosis (Pita et al., 2016). Viruses are diverse and abundant within coral tissues and their activity is hypothesized to influence symbiont community structure/evolution, nutrient cycling, metabolic interactions, and holobiont immunity/stress resilience (Ambalavanan et al., 2021; Planes et al., 2019; Thurber et al., 2017). However, despite the existing hypotheses, there is no empirical evidence linking the coral virome to specific functions within the coral holobiont (Thurber et al., 2017, Planes et al., 2019). Furthermore, although we have some understanding of the virus lineages that are frequently found in coral holobionts, the specific target hosts for the majority of these virus remain unidentified or unverified (Ambalavanan et al., 2021; Thurber et al., 2017). With the current state of coral reefs, there is thus an urgent need to intensify research efforts and foster collaborative initiatives to address the knowledge gaps in coral reef virology. Understanding the role of viruses in potentially mitigating or exacerbating stressors (e.g., bleaching, disease outbreaks) is crucial for the conservation of these habitats. By improving our comprehension of the interactions between viruses, coral reefs, and their inhabiting holobionts, we can effectively develop informed conservation and management strategies to protect these

already threatened ecosystems.

Virus diversity and ecology within marine holobionts

High throughput sequencing techniques like shotgun metagenomics and amplicon sequencing have been critical in expanding our comprehension of global virus diversity (Camargo et al., 2022; Roux et al., 2015). Despite our newly expanded knowledge of virus diversity across diverse biomes, virologists at times are still grappling with fundamental questions such as "what is the target host of a virus?" and "what impact does this virus have on its host/ecosystem it inhabits?". Coral reef virology presents a particular challenge due to the complicated nature of the coral holobiont, which involves multiple potential eukaryotic and prokaryotic hosts for detected symbiotic viruses (Thompson et al., 2015; Thurber et al., 2017). This, coupled with the complexity of virus genomes and the lack of coral-specific virus gene databases, has hindered our ability to fully detect and interpret the dynamics of known coral-associated viruses such as dinoRNAV. Fortunately for investigators, virus infection can sometimes result in the incorporation of virus genes into their host genomes (Aiewsakun & Katzourakis, 2015). These virus-derived genes (or endogenous viral elements, EVEs) within host genomes can provide hints into virus evolutionary history and ecology in multipartite systems like coral holobionts (Aiewsakun & Katzourakis, 2015). In my first chapter, I illustrate the effectiveness of using EVEs to identify the target host of dinoRNAVs as Symbiodiniaceae (the mutualistic dinoflagellates) within coral tissues (Chapter 1, Veglia et al. 2023). With the increasing number of coral, dinoflagellate, and bacterial genomes becoming available each year, there is an opportunity for coral reef virologists to leverage EVEs to develop host predictions for more commonly coral-associated virus lineages (e.g., Imetervirales, Picornavirales, Durnavirales identified in Chapter 3). Future coral virology research should focus investigations on

characterizing the diversity, ecology, and prevalence of known coral-associated virus groups to determine those which may have substantial influences on coral health.

Amplicon sequencing is another promising, yet underutilized, tool available to coral reef virologists; this approach can provide insights into the prevalence, diversity, and population dynamics of a given virus lineage on reefs. In the last two decades, four studies have used amplicon sequencing to study coral-associated viruses (T4 bacteriophages, Buerger et al., 2019; dinoRNAV, Grupstra et al., 2022; Howe-Kerr et al., 2022; Montalvo-Proaño et al., 2017). With vAMPirus (Chapter 2), investigators can now design and share lineage-specific analyses, facilitating the characterization of viruses in underexplored environments like coral reef holobionts. vAMPirus not only enhances accessibility to comprehensive virus amplicon analyses but also promotes the comparability and expandability of findings across various studies. As demonstrated by Grupstra et al. (2022), Howe-Kerr et al. (2022), this allows investigators to efficiently compare their results with those of other studies, build on previous findings, and generate new insights into a given virus system. As a result, we now know that dinoRNAV is prevalent across coral species (Grupstra et al., 2022; Howe-Kerr et al. 2022; Montalvo-Proaño et al., 2017; Chapter 2, Veglia et al., 2023), exhibits temperature-driven community changes and productivity increases (Grupstra et al. 2022, Howe-Kerr et al., 2022), and are potentially dispersed across reefs through corallivore feces (Chapter 2, Veglia et al., 2023). The application of high throughput sequencing methods, coupled with the *in silico* detection of EVEs has thus been crucial in developing the dinoRNAV system, which can be considered a model of what the field can do for a plethora of other coral-associated virus lineages in the future.

Viruses and coral disease traits

Viruses can play fundamental roles in either mitigating or potentiating diseases in metazoan holobionts (Neil & Cadwell, 2018; Roossinck, 2011). In corals, previous studies have reported increased abundances of specific virus groups in holobionts experiencing stress and/or disease states, suggesting their potential contribution (Buerger et al., 2019; Correa et al., 2013, 2016; Messyasz et al., 2020; Soffer et al., 2014; Thurber et al., 2017). However, the effects of the individual lineages, as well as the virome as a whole, on coral disease traits, such as susceptibility and severity, have remained largely unknown.

In Chapter 3 of this dissertation, I explore the potential for viral mechanisms underlying the etiology of a multi-species coral disease (stony coral tissue loss disease (SCTLD)) using comparative metatranscriptomics. The coral samples analyzed in this chapter were sourced from an *ex situ* transmission experiment (described in Meiling et al., 2021) as part of a multidisciplinary collaboration aimed at characterizing various aspects of SCTLD etiology (e.g., relative risk of infection, lesion progression rates, microbiome dynamics). As a result, Beavers et al. (*In Press*) assessed and previously reported the coral and Symbiodiniaceae gene expression of the same samples analyzed in Chapter 3. This information provided a unique opportunity to assess not only the viral dynamics in response to SCTLD but also the interplay between coral/Symbiodiniaceae gene expression and viral gene expression before and during infection. In Chapter 3, limited evidence was found for a single virus pathogen of SCTLD. Instead, the results did suggest that core virome members (e.g., Imetervirals, Algavirales, Patatavirales, and Picornavirales) may contribute to SCTLD etiology collectively via opportunistic infections. The increase in virus opportunistic infections in SCTLD-affected coral holobionts coincides with the reported upregulation of antivirus immunity genes by the coral and Symbiodiniaceae (Beavers et

al., *In Press*). One such upregulated gene reported for all coral species by Beavers et al., (*In Press*) was coral *Abce1*, the negative regulator of RNase L – an enzyme known to impede viral infection via degradation of viral RNA (Drappier & Michiels, 2015). Viruses belonging to the order Picornavirales are known to induce the upregulation of *Abce1* in their hosts to evade the RNase L antiviral mechanism (Drappier & Michiels, 2015; Martinand et al., 1998). Interestingly, in Chapter 3, picornavirus transcripts were significantly upregulated in 21 of 29 SCTLD-affected coral holobionts. This could suggest that increased picornavirus activity in SCTLD-affected holobionts may drive this observed upregulation of *Abce1*, reducing RNase L activity, weakening the coral's antivirus capability, and facilitating opportunistic virus infections. For the first time, these results provide empirical evidence linking a specific virus lineage to a mechanism that contributes to changes in coral gene expression, which in turn may influence coral susceptibility to disease.

The results presented in Chapter 3 underscore the value of conducting multi-disciplinary collaborative investigations to better understand the functional roles of viruses in coral health and disease. While further research is needed to confirm the precise contribution of virus opportunistic infections to SCTLD, these efforts have yielded findings of great interest to the field, revealing promising future virus research directions within and beyond the context of SCTLD. Moving forward, it is important for coral reef virus investigations to prioritize characterizing the diversity and dynamics of individual virus lineages that show potential for significant roles in the health of the coral holobiont (e.g., dinoRNAV, Picornavirales). One approach to achieve this is by isolating and culturing these viruses (see Veglia et al., 2021), which would enable comprehensive phenotypic and genomic characterization. Additionally, developing lineage specific PCR primer sets would allow the use amplicon sequencing or

quantitative PCR techniques to gain insight into prevalence and diversity of these virus lineages on the reef (see Grupstra et al., 2022; Howe-Kerr et al., 2023; Montalvo-Proaño et al., 2017). Furthermore, it is crucial for the coral reef virus community to collaborate and collectively consolidate efforts to develop coral reef virus-specific laboratory methods (as demonstrated in Veglia et al., 2021) and bioinformatic resources (e.g., reference sequence databases, lineage-specific vAMPirus analyses). Advancing our knowledge of individual coral-associated virus lineages in tandem with community-driven methods/resource development, will greatly enhance our ability to decipher and interpret virus diversity, prevalence, and gene expression data in future studies. This, in turn, will significantly broaden our understanding of the roles of viruses in coral health and diseases such as SCTLD, and ultimately support effective conservation and management strategies for coral reef ecosystems.

CHAPTER 1 APPENDIX

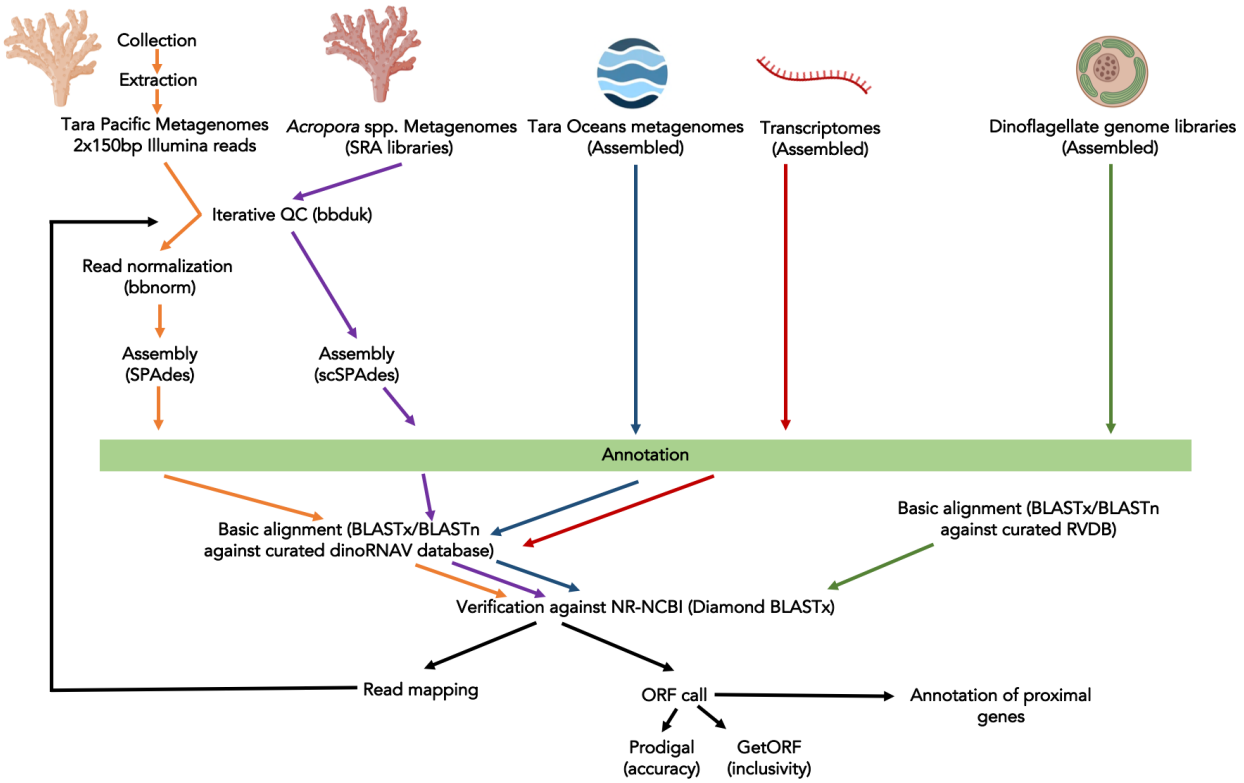
Chapter 1 Supplementary Tables and Figures

This section includes all Supplementary Figures and Table captions, all supplementary tables can be downloaded and reviewed here: www.biorxiv.org/content/10.1101/2022.04.11.487905v1.supplementary-material.

Supplemental Table 1A. Metadata for metagenomic libraries of coral and water queried for dinoRNAV EVEs indicating sequencing depth, quality control, and assembly statistics. Table also includes metagenome species (with taxonomic ID) and location (site and colony) per library. Also find a Google Earth map of dinoRNAV EVE-positive sites here: https://bit.ly/3ri6m1V

Table parameters: Sample - unique library name; Location – island and other geographic indicator, Site_Colony - unique coral colony indicator; Sample_material – type of biological material collected; TaxID – NCBI taxonomic indicator of biological material collected; #raw.PE.reads/bases – number of paired end reads or bases sequenced; #QC.PE.reads/bases – number of quality controlled reads or bases; #contigs – number of contigs assembled from quality controlled reads; N50 – average N50 of contigs; positive_dinoRNAV – number of dinoRNAV EVE sequences identified in library

Supplemental Table 1B. Accession numbers, taxonomy, and submitter data (when available) for publicly available genomes, chromosomes, and metagenomes queried for dinoRNAV EVEs. Table includes datasets for coral genomes, dinoflagellate genomes, dinoflagellate chromosomes, *Acropora* spp. metagenomes, pelagic water metagenomes, and Symbiodiniaceae transcriptomes.



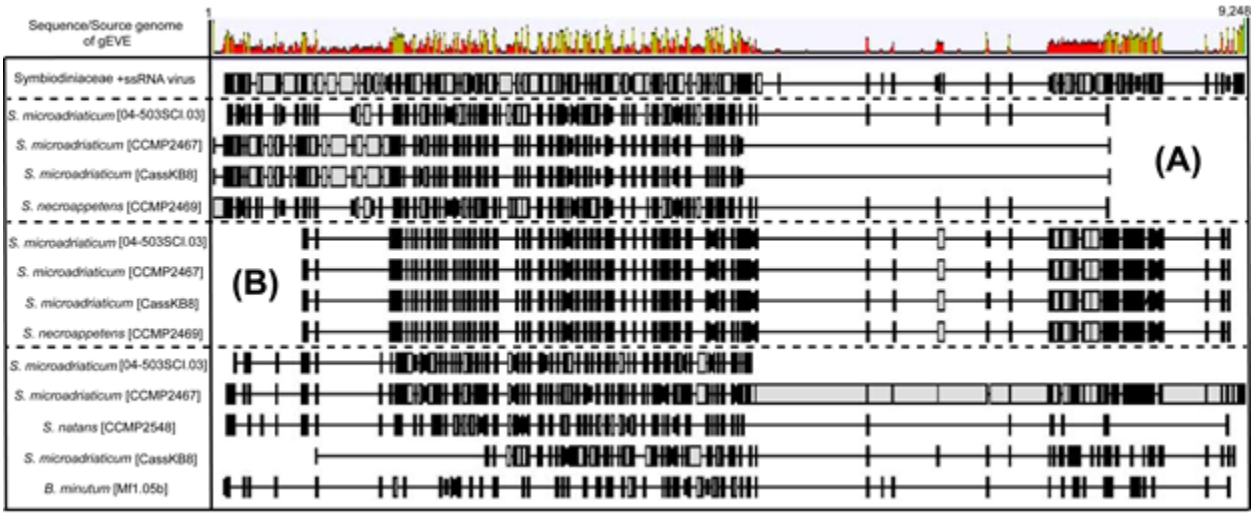
Supplemental Fig 1. Acquisition and bioinformatic analysis pipeline utilized for dinoRNAV EVE identification and verification. Five sources of data (newly assembled metagenomes from Tara Pacific, newly assembled metagenomes from publicly available *Acropora* holobionts, pre-assembled metagenomes from Tara Ocean databases, pre-assembled transcriptomes from Baumgarten et al, 2020, and publicly available genomes from dinoflagellates) were quality controlled/assembled per a variety of software in the bbtools suite and SPAdes prior to alignment-based annotation against several databases (NR-NCBI, UniProt, RVDB, or manually curated databases of dinoRNAV genomes. Details not available in "Methods" can be found below:

Collection and sequencing of Tara Pacific coral metagenomes (see full methods in doi:10.5281/zenodo.4068293): Briefly, coral nubbins were collected from replicate colonies of a fire coral (genus: *Millepora*, n=60), and two stony corals (genera: *Porites*, n=108; and *Pocillopora*, n=101), as well as from immediately surrounding water. DNA was extracted via Quick-DNA/RNA Kit with supplemental enzymatic lysis (Zymo Research, Irvine, CA, USA). mechanically fragmented (300bp; Covaris E210, Covaris, Inc., USA), end-repaired and polyadenylated (NEBNext DNA Modules; New England Biolabs, MA, USA) before being ligated to adapters (NextFlex DNA barcodes; Bio Scientific Co) for clean-up and amplification. DNA was subjected to 2x150bp paired-end Illumina sequencing (Genoscope; Évry, France). Metagenomic libraries were trimmed for length, quality, and adapters (bbduk v.38.06), normalized (bbnorm v. 38.06; target depth 40x coverage) and assembled *de novo* (SPAdes v.3.12).Publicly available metagenomes were quality controlled as detailed above, non-normalized, and assembled via SPAdes utilizing the single cell modifier (v.3.12).Collection and sequencing of Tara Pacific coral amplicon libraries (see full methods in doi:10.5281/zenodo.4061797 and Belser et al, *in prep*): DNA was extracted using the Quick-DNA/RNA kit with supplemental enzymatic lysis (Zymo Research, Irvine, CA, USA), PCR amplified in triplicate using SYM-VAR-5.8S2 / SYM-VAR-REV (Hume et al. 2019), and cleaned (AMPure XP; Beckman Coulter, Brea, CA) prior to pooling and library construction (NEBNext DNA Modules; New England Biolabs, MA, USA). Amplicons were sequenced on the Illumina platform. "Defining intragenomic variants" were utilized to differentiate taxonomic

profiles within the Symportal analytical framework post-minimum entropy decomposition (Hume et al, 2019).

Transcriptomes: To assess (1) if dinoRNAV EVEs were present in poly(A)-selected dinoflagellate transcriptomes, and (2) if these transcripts resembled endogenized dinoRNAVs in proximal gene composition, presence of a characteristic pre-mRNA spliced leader (SL) sequence (SL; as in Levin et al, 2017), or were differentially expressed under variable environmental conditions. Publicly accessible transcriptomes from the genus *Symbiodinium* (n=11 smRNA, n=11 mRNA; Supplemental Table ST.1B) were queried for dinoRNAV-like sequences. Reads were trimmed and quality controlled per the same description as the original study (cutadapt v. 3.1; Baumgarten et al, 2020). Transcripts were annotated for dinoRNAVs in congruent manner as metagenomes (described in Methods), reads were mapped via bbmap (v.38.84), and RNA secondary structure was predicted via mfold (v.3.5).

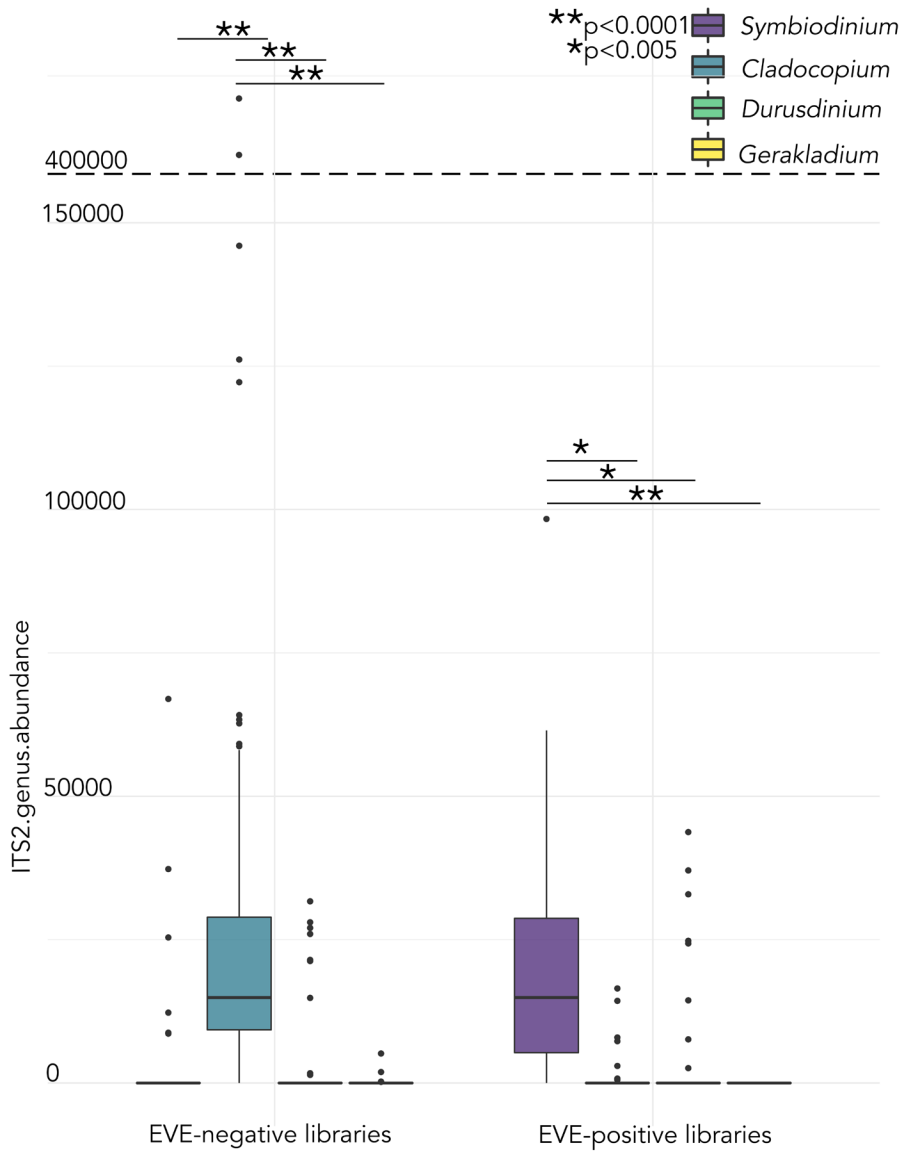
Supplemental Table 2. Reference database of accession numbers (including HcRNAV, +ssRNAV genomes referenced in Levin et al, 2017, and other dinornavirus-like sequences) utilized to query metagenomes and genomes for endogenous dinoRNA. Includes both nucleotide (nt) and amino-acid coding sequences (CDS/aa) accession numbers. 'Source' indicates if sequence was derived from an environmental source other than a dinoflagellate (e.g. *genomic* – from a viral genome, *envm* - from a source such as sediment, as submitted by Nakayama & Hamaguchi et al, submitted to NCBI 2020).



Supplementary Figure 2. Alignment of putative endogenous genomic dinoRNAV EVEs to the Symbiodiniaceae-infecting +ssRNA virus genome assembled by Levin et al (2017) from a *Symbiodinium* spp. transcriptome. Repeated "whole" genome EVEs (A & B) showing similarity across *Symbiodinium* genome are marked with dashed boxes. Putative whole genome EVEs were extracted from their respective scaffolds and aligned to the dinoRNAV reference genomes using MAFFT (v7.464, Katoh and Standley 2013) and visualized in Geneious Prime (v.2021.0.1, restricted).

Supplemental Table 3. Quantification of hydrocoral metagenome assemblies with and without dinoRNAVs. Hydrocoral metagenomes were sequenced at equivalent depth as scleractinian corals and had comparable level of annotation, precluding methodological bias towards dinoRNAV prevalence among these libraries.

Supplementary Table 4. Most common genes immediately proximal to putative dinoRNAV EVEs identified in metagenomes, including accession numbers and name. dinoRNAV EVE open reading frames (ORFs) appeared to be immediately adjacent to ORFs identified as dinoflagellate genes.

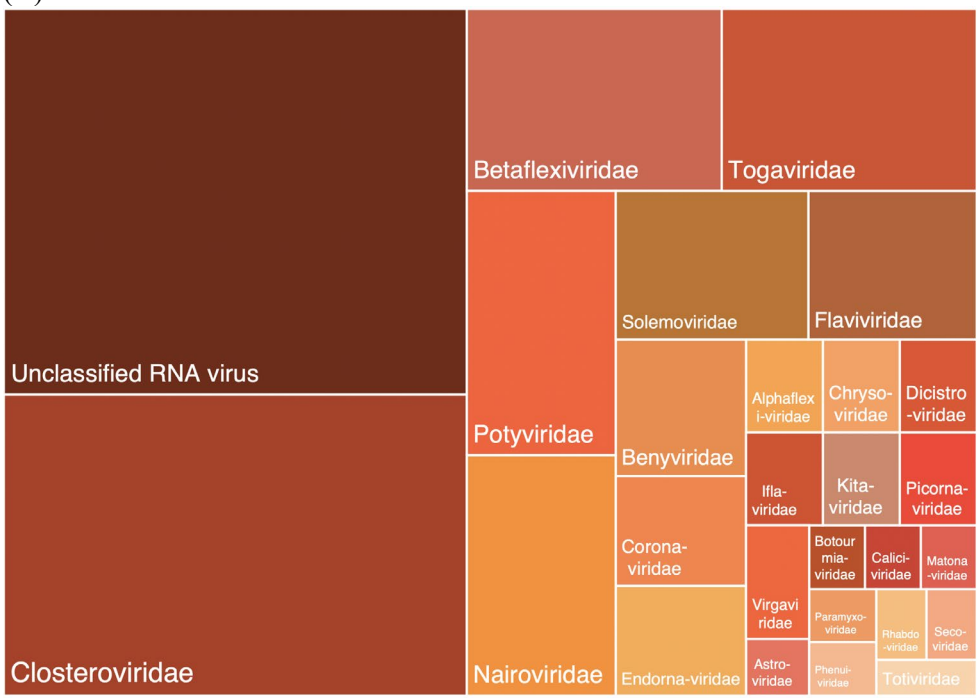


Supplementary Figure 3. DinoRNAV EVE detection - regardless of host species - largely correlated with Symbiodinium dominated communities. $F_{2,1044}$=25.8, p<0.0001, Type III ANOVA with Satterthwaite's. *Posthoc*: pairwise Tukey with Kenward-Roger approximation for linear mixed models.

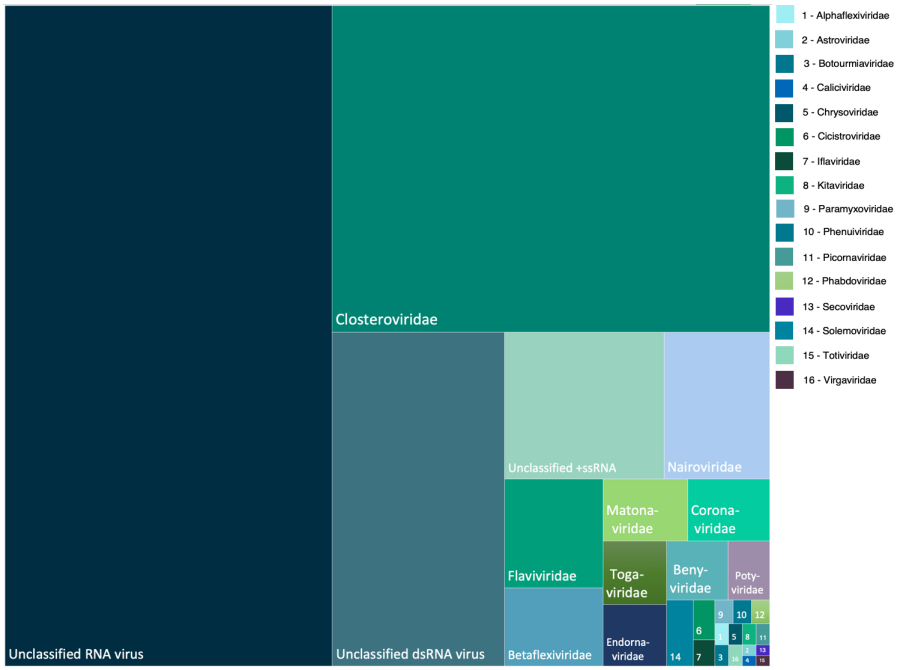
Supplemental Table 5. Completeness of dinoflagellate genomes queried for dinoRNAV EVEs measured in BUSCO (Benchmarking Universal Single-Copy Orthologs) scores. While a linear model suggested that there was no relationship between dinoRNAV detection and genome completeness, greater genome completeness could lend more opportunities for dinoRNAV EVE detection

Supplementary Figure 4. Example secondary structural elements of *Symbiodinium* sp. (A1) transcript Locus_83084 (accn: GAKY01194223.1) including hairpin folds. *Symbiodinium* genome contains all core RNAi protein machinery, including Argonaute and Dicer and GAKY01194223.1 folds into several hairpins (ΔG = -142.5kcal/mol; mfold v. 3.5)

(A)



(B)



Supplemental Figure 5. Composition of RNA virus families detected in Tara Pacific metagenomes where DinoRNAV EVE was also detected per identification and characterization of *RdRp* gene via Diamond BLASTx. (A) Map describes the composition of *unique* contigs affiliated with each RNA viral family (per *RdRp* detection) across 42 dinoRNAV EVE-positive metagenomes (representing n=26 families and 50 RNA viral contigs containing *RdRps* with unknown viral families; total of n=187 *unique* contigs); (B) Map describes the *total* number of contigs identified within each RNA viral family across 42 dinoRNAV EVE-positive metagenomes (per *RdRp* detection; n=3,5069 *total* contigs identified; average 834.97 per library).

Supplementary Table 6. Annotation of genes proximal to putative dinoRNAV EVEs on genome scaffolds. Genome scaffolds annotated with maker3 with repeatmasking and augustus gene prediction using the Chlorella gene model. Protein annotations were inferred with blastp against a hybrid database containing the protein version of the Reference Viral database and the uniprot_spot database (blastp -query $x -db $db -evalue 1e-7 -max_hsps 1 -max_target_seqs 1 -outfmt 6 -out "$name"_output.blastp -num_threads 23) and protein families were inferred with Interproscan w/ HMMER and the PFAM database (interproscan.sh -appl pfam -dp -f TSV -goterms -iprlookup -pa -t p -i $x -o "$name"_output.iprscan -cpu 23).

CHAPTER 2 APPENDIX

Chapter 2 Supplementary Methods Tables and Figures

S1. Supplemental Methods

S1.1 Pairwise similarity matrices

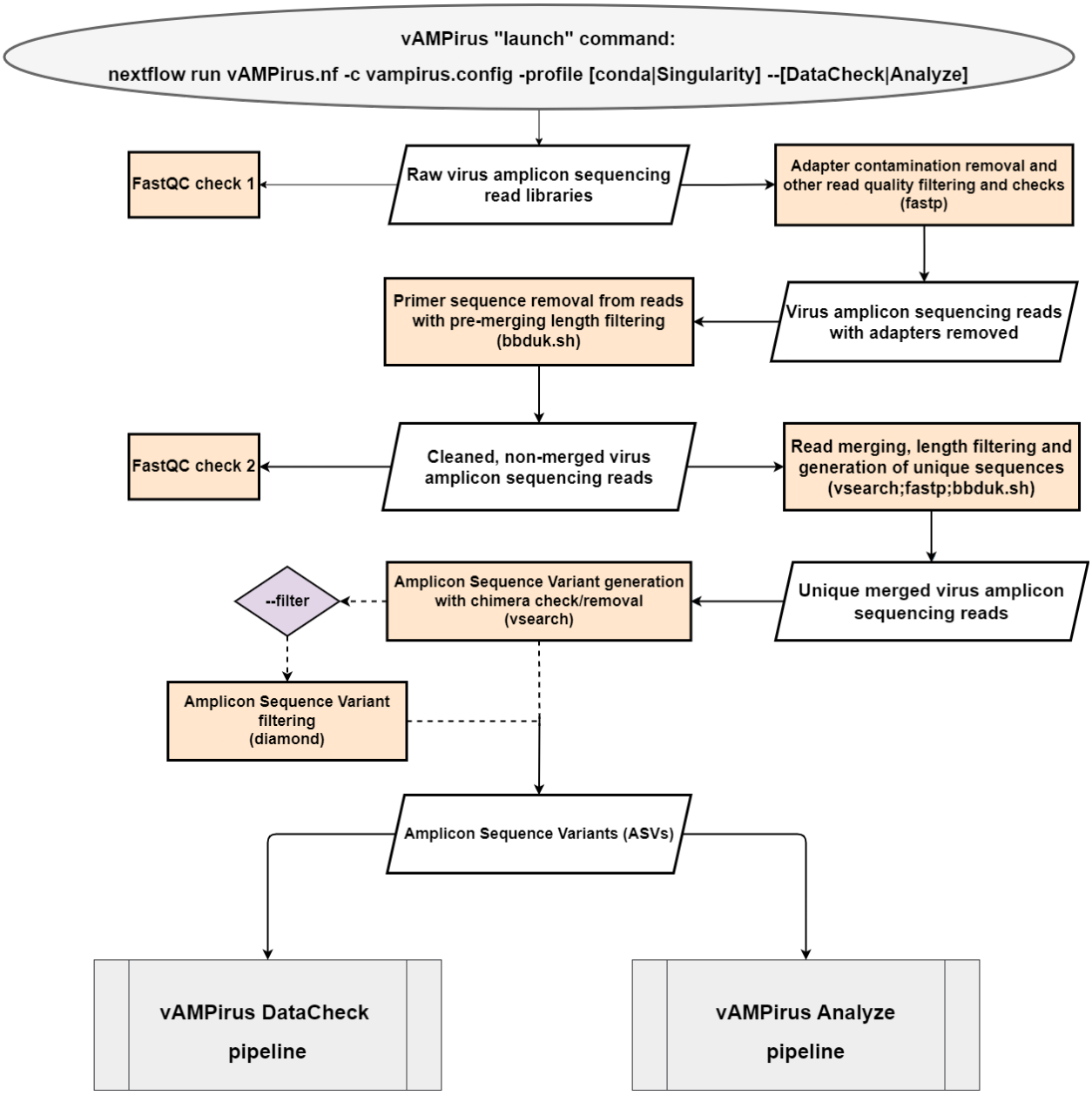
Pairwise percent similarity and pairwise distance matrices are generated for sets of ASVs, cASVs, and aminotypes with the program Clustal Omega (Sievers et al., 2011). These pairwise similarity matrices are stored within the results directory for the user to review as comma-separated variables (csv) files and are visualized as interactive heat maps within Analyze HTML reports.

S1.2 Analysis of protein physiochemical properties

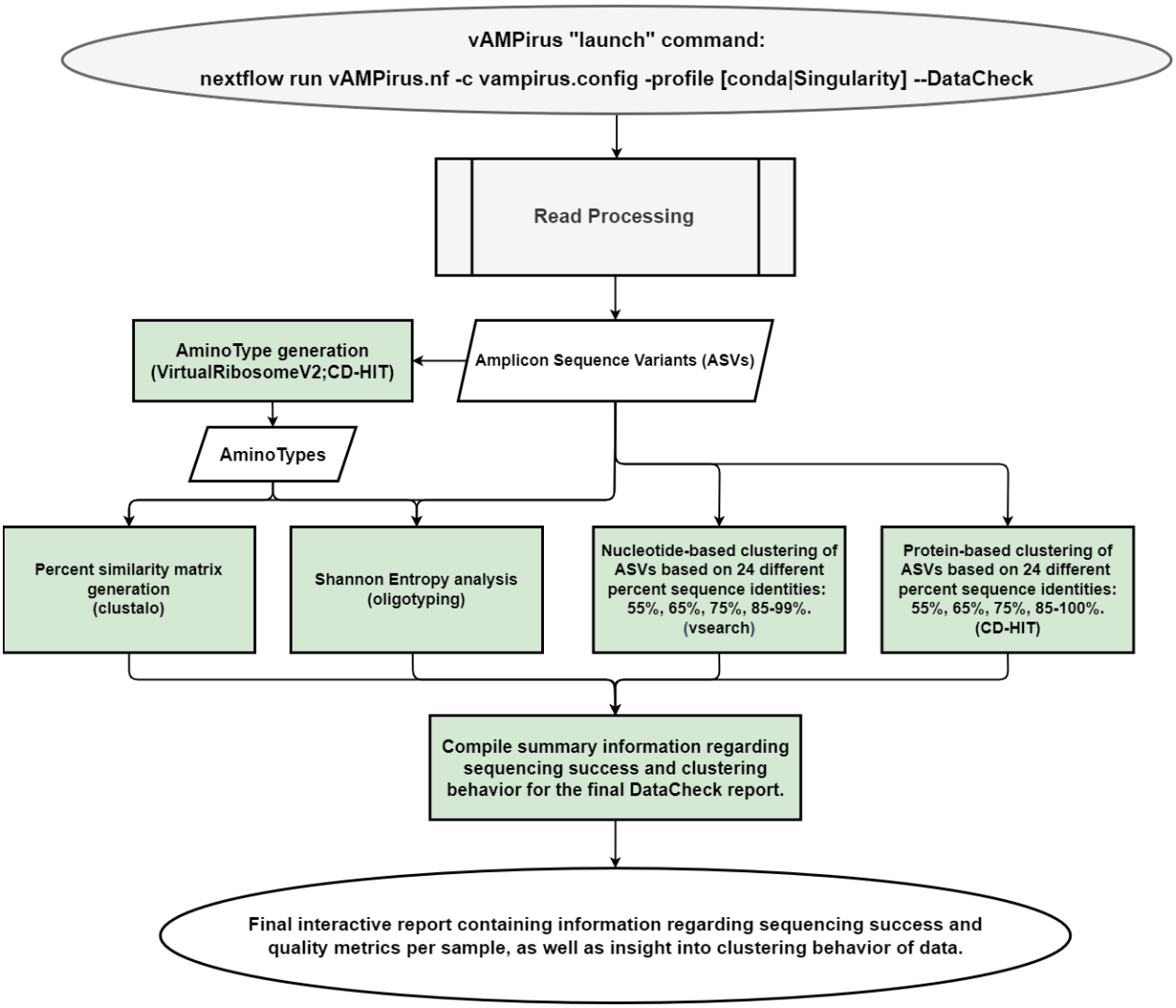
To provide users with the ability to determine the biological implications of amino acid sequence differences, the Analyze pipeline includes an analysis of protein physiochemical properties using scripts within EMBOSS (Rice et al., 2000) for all amino acid sequences (i.e., pcASVs and aminotypes). These scripts include: 1. pepstats – reads protein sequences and provides an output file containing various statics on protein properties (e.g., molecular weight, average residue weight, charge, molar extinction coefficient, and more), 2. iep – estimates the protein isoelectric point from the amino acid sequence and plots an ionization curve with respect to pH, 3. hmoment – provides hydrophobic moments information for each protein sequence, and 4. garnier, pepinfo, and pepwheel – Generates two-dimensional structure information from protein sequences. All outputs from these analyses are stored and organized within a dedicated EMBOSS directory.

S1.3 vAMPirus report

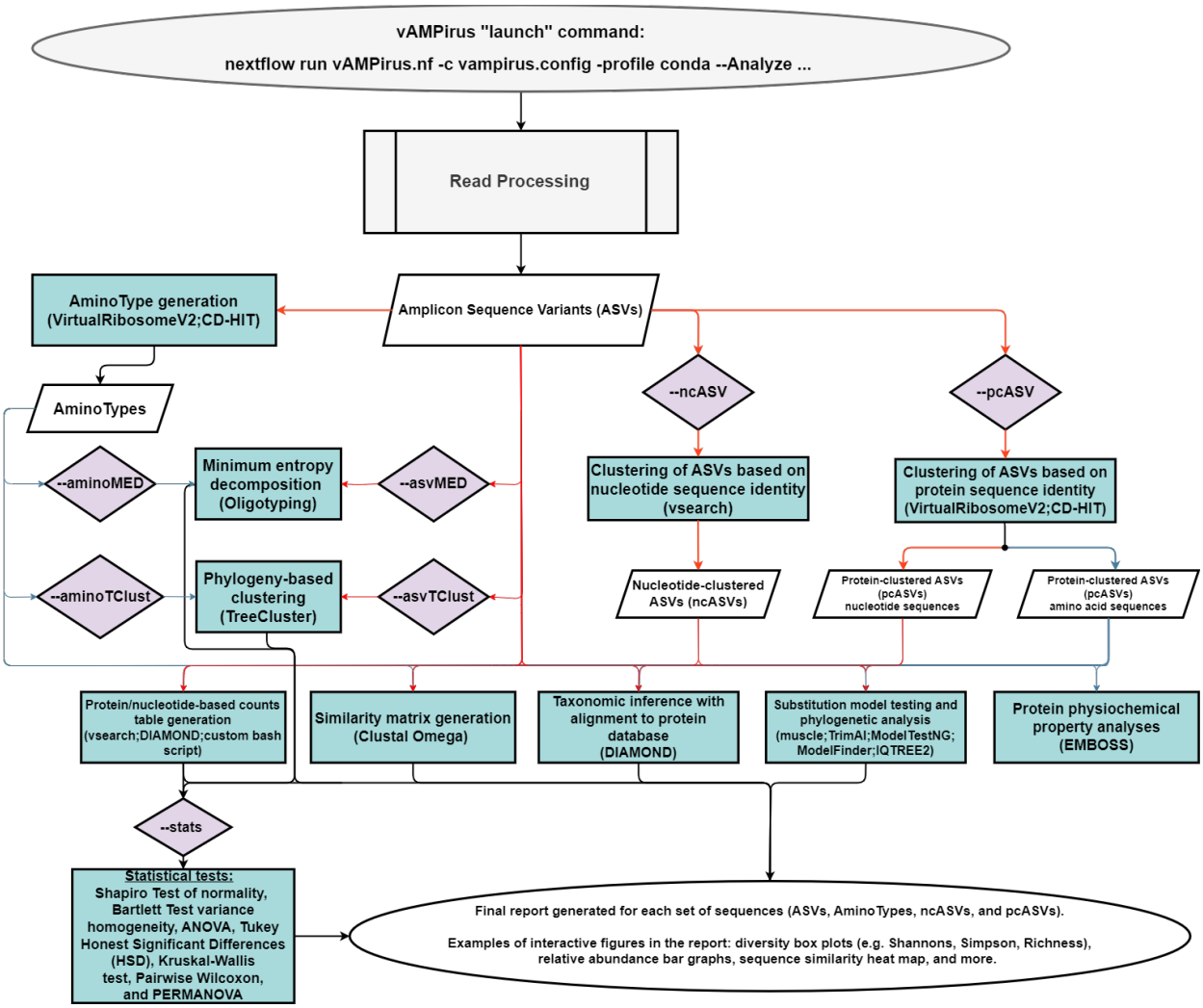
To start, the counts tables output by vAMPirus, and the user-generated metadata file are merged. The total number of reads per sample is calculated and visualized in a scatterplot. Then, samples with low reads are filtered to a user-configurable threshold set in the configuration file; the default is 1000. Rarefaction curves are generated with rarecurve() in *vegan* (v. 2.6.4) and data are then rarefied with rrarefy (*vegan* v. 2.6.4) to the number of reads present in the sample with the fewest reads. Shannon and Simpson diversity indices are calculated based on these rarefied libraries using the diversity() command in *vegan* (v. 2.6.4). Richness is calculated by repeatedly subsampling libraries (sample size = 0.95*reads in the library with the fewest reads) using the rarefy function in *vegan*. To test for compositional differences between sample types, adonis2() is used based on Bray-Curtis distances calculated from untransformed rarefied data (vegdist in *vegan* v. 2.6.4); betadisper() is used to calculate distance to centroid values. Finally, metaMDS is used to calculate NMDS scores on bray-curtis distances using K of 2 (2D NMDS) and 3 (3D NMDS) with autotransform set to TRUE. If NMDS fails to converge, 2D and 3D Principal component analysis (PCOA) plots are generated using pcoa(). All plots within the report are interactive and generated with plotly (v. 4.10.0).



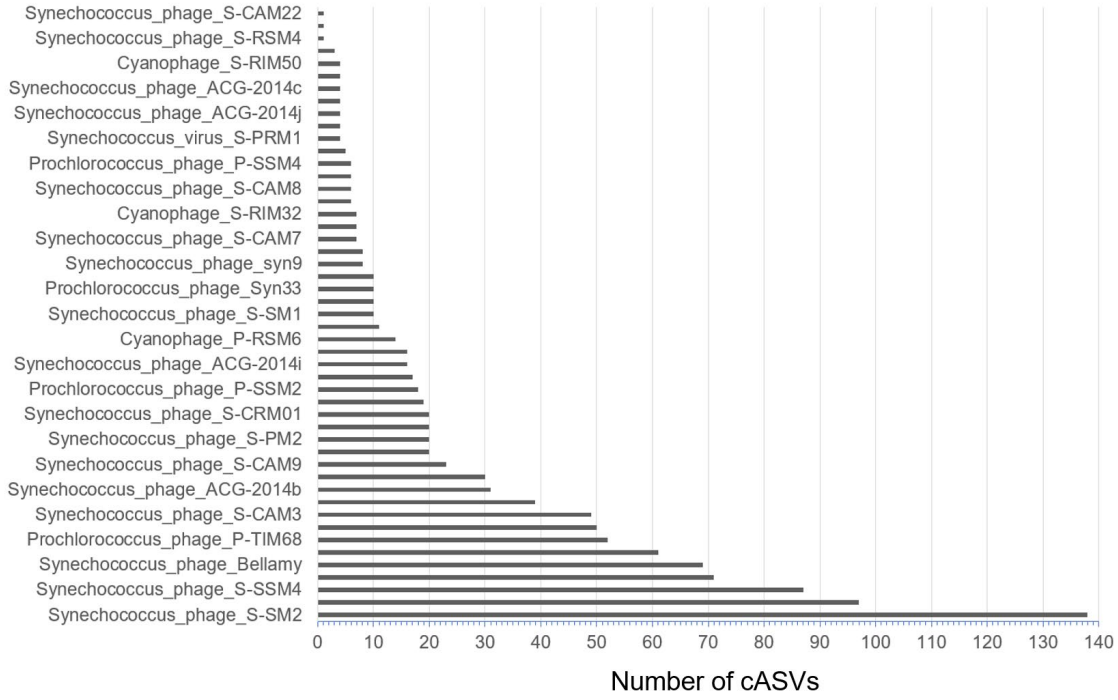
Supplemental Figure 1. Schematic of vAMPirus v2.1.0 read processing pipeline including programs used in processes. An example vAMPirus startup command to initiate read processing prior to the DataCheck or Analyze pipeline is displayed in the gray oval. Orange boxes are actions performed with the input raw virus amplicon read libraries; the main output files are represented by white parallelograms. The final set of ASV sequences produced are then input to the DataCheck and/or Analyze pipelines.



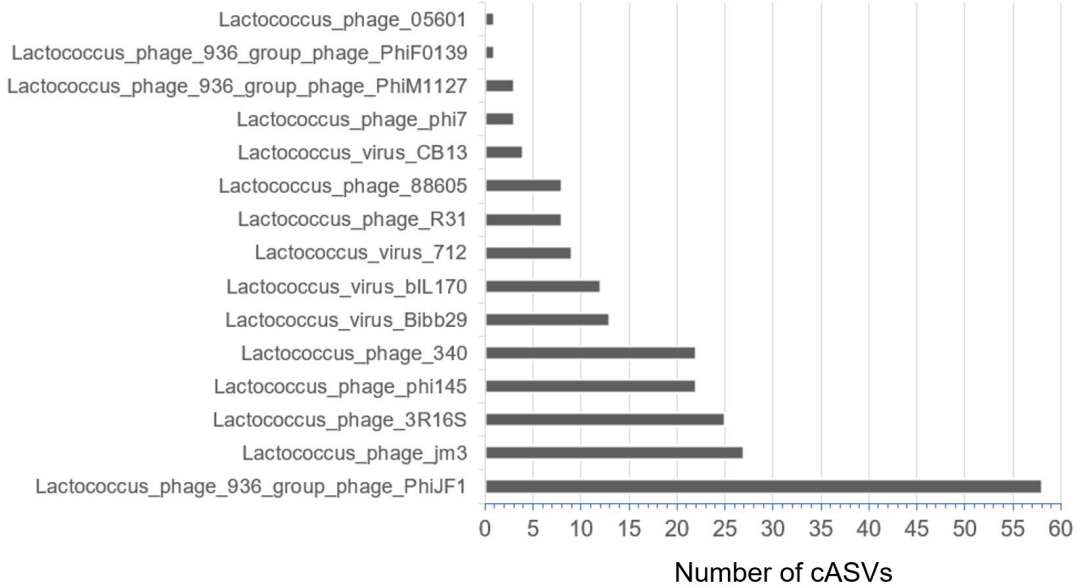
Supplemental Figure 2. Schematic of vAMPirus v2.1.0 DataCheck pipeline including programs used in processes. An example vAMPirus startup command to run the DataCheck pipeline is displayed in the gray oval. Green boxes are actions performed in this pipeline with the input ASV sequences (white parallelograms); the results of this pipeline are compiled and displayed in a final html report (white oval).



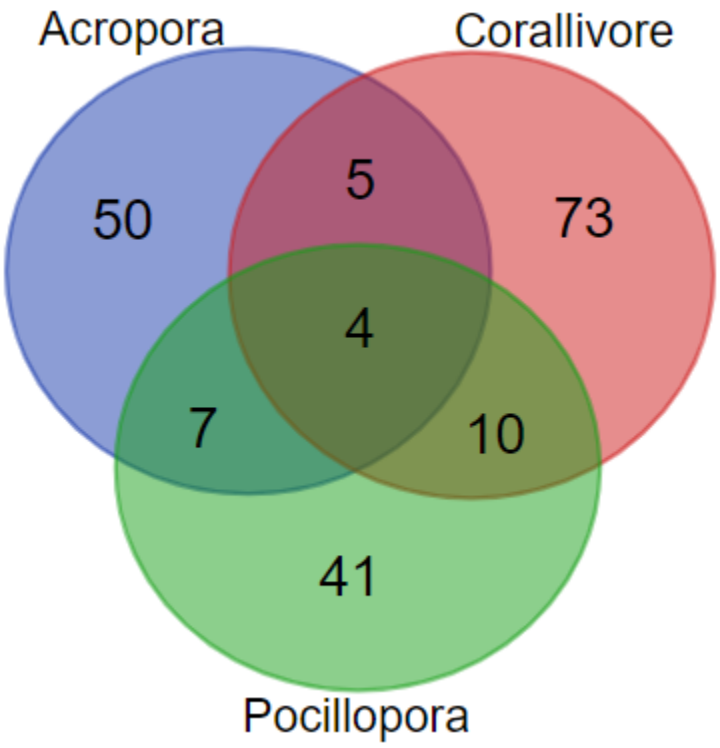
Supplemental Figure 3. Visualization of vAMPirus v2.1.0 Analyze workflow including programs used in processes. Steps within read processing have been collapsed. Red lines represent path for nucleotide-based sequence analyses and blue lines represent path for amino acid-based sequence analyses. Shapes represent: gray oval– input vAMPirus launch command; white parallelogram – sequence fasta files produced; green box – process ran within the pipeline; purple diamonds – sequence grouping options; and white oval– report output file.



Supplemental Figure 4. Taxonomy results for cyanophage pcASVs produced from Finke & Suttle, (2019) dataset using DIAMOND blastp and the NCBI virus RefSeq database within the vAMPirus Analyze pipeline (Section 3, this study).



Supplemental Figure 5. Taxonomy results for Lactococcus phage ncASV produced from Frantzen & Holo (2019) dataset using DIAMOND blastx and the NCBI virus RefSeq database within the vAMPirus Analyze pipeline (Section 3, this study).



Supplemental Figure 6. Venn diagram of dinoRNAV major capsid protein gene aminotypes indicates that 14% (26) of 190 unique aminotypes are shared among stony corals *(Acropora* sp., *Pocillopora* sp.) and/or corallivorous (coral-eating) fish feces. The figure was generated using the aminotype counts table results produced by vAMPirus Analyze pipeline. Venn diagram was generated using the webtool hosted by UGent Bioinformatics & Evolutionary Genomics: https://bioinformatics.psb.ugent.be/webtools/Venn/.

Supplemental Table 1. Results of statistical tests showing significant differences in cyanophage ASV and aminotype alpha diversity in samples collected in low (<27.5 psu) and high (>27.5) salinity sites. Results were generated from Finke & Suttle (2019) cyanophage g43 amplicon dataset by the vAMPirus Analyze pipeline (section 3, this study).

Statistical test	ASV results	aminotype results
	Shannon diversity	
Shapiro-Wilk normality test	W = 0.97, p-value = 0.26	W = 0.97, p-value = 0.43
Bartlett test of homogeneity of variances	Bartlett's K-squared = 0.1, df = 1, p-value = 0.75	Bartlett's K-squared = 1.95e-05, df = 1, p-value = 0.99
Tukey HSD (low vs high)	p=9.22e-05	p=8.37e-05
	Richness	
Shapiro-Wilk normality test	W = 0.98, p-value = 0.5	W = 0.97, p-value = 0.37
Bartlett test of homogeneity of variances	Bartlett's K-squared = 7.62, df = 1, p-value = 0.005	Bartlett's K-squared = 7.02, df = 1, p-value = 0.008
Kruskal-Wallis rank sum test (low vs high)	Kruskal-Wallis chi-squared = 17.99, df = 1, p-value = 2.21e-05	Kruskal-Wallis chi-squared = 17.10, df = 1, p-value = 3.54e-05

References Cited

Finke, J. F., & Suttle, C. A. (2019). The Environment and Cyanophage Diversity: Insights From Environmental Sequencing of DNA Polymerase. *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/article/10.3389/fmicb.2019.00167

Frantzen, C. A., & Holo, H. (2019). Unprecedented Diversity of Lactococcal Group 936 Bacteriophages Revealed by Amplicon Sequencing of the Portal Protein Gene. *Viruses*, *11*(5), Article 5. https://doi.org/10.3390/v11050443

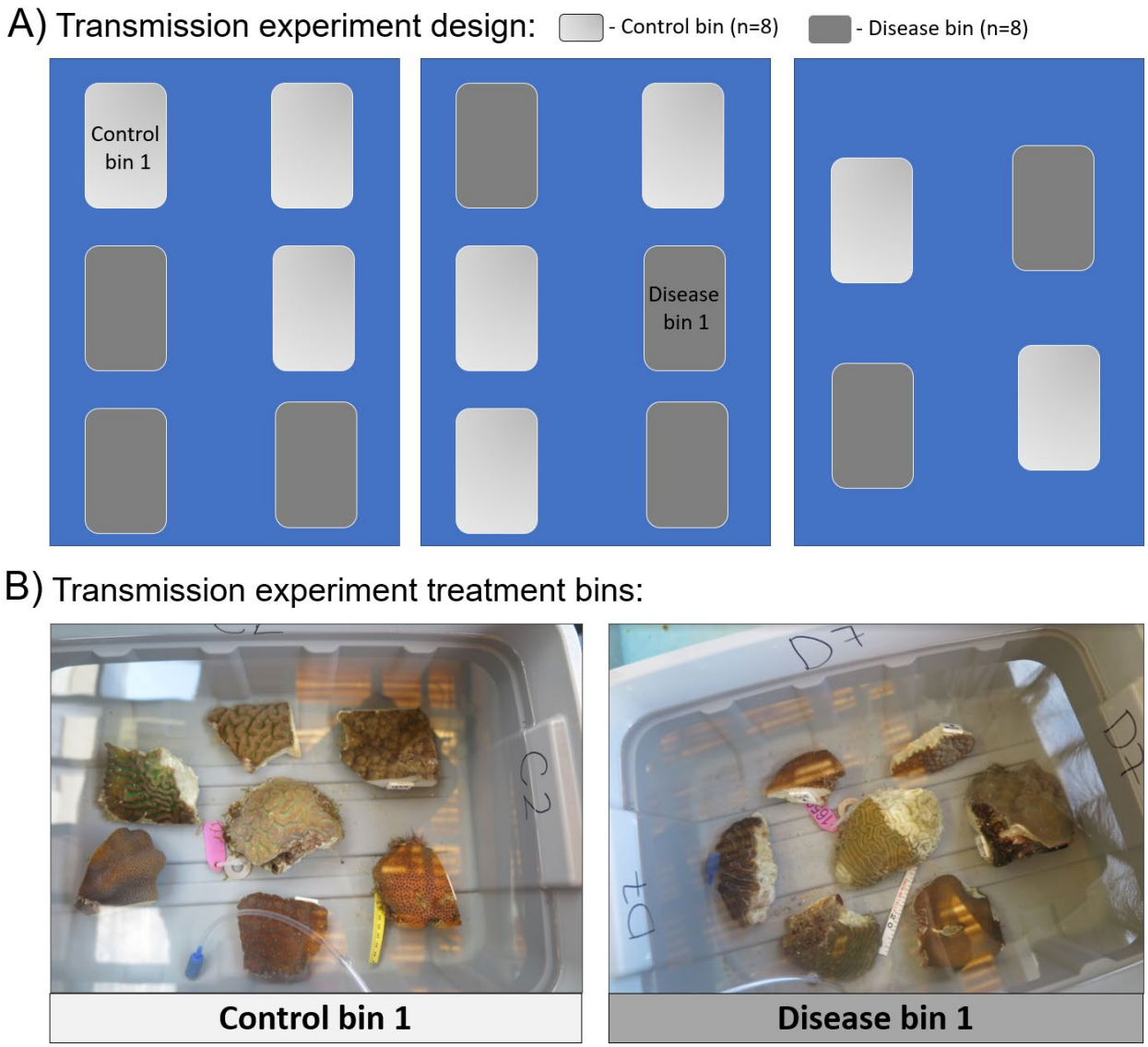
Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics: TIG*, *16*(6), 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539. https://doi.org/10.1038/msb.2011.75

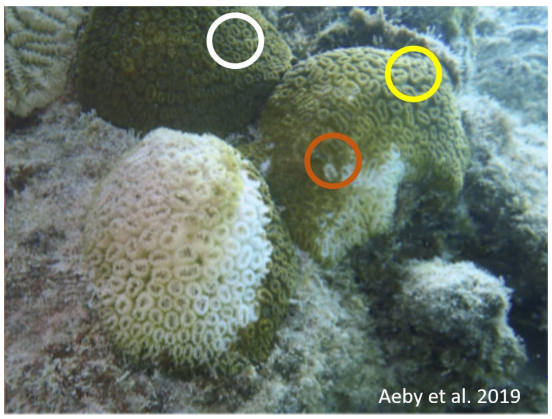
CHAPTER 3 APPENDIX

Chapter 3 Supplementary Figures

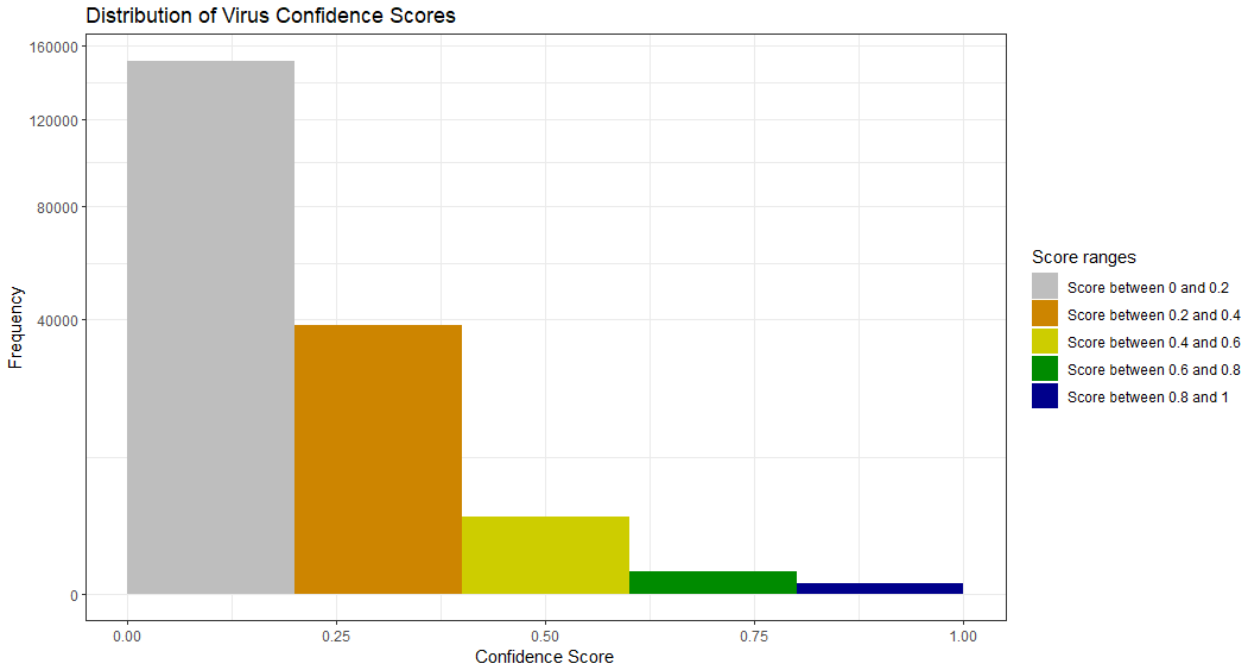
S1. Supplemental Figures



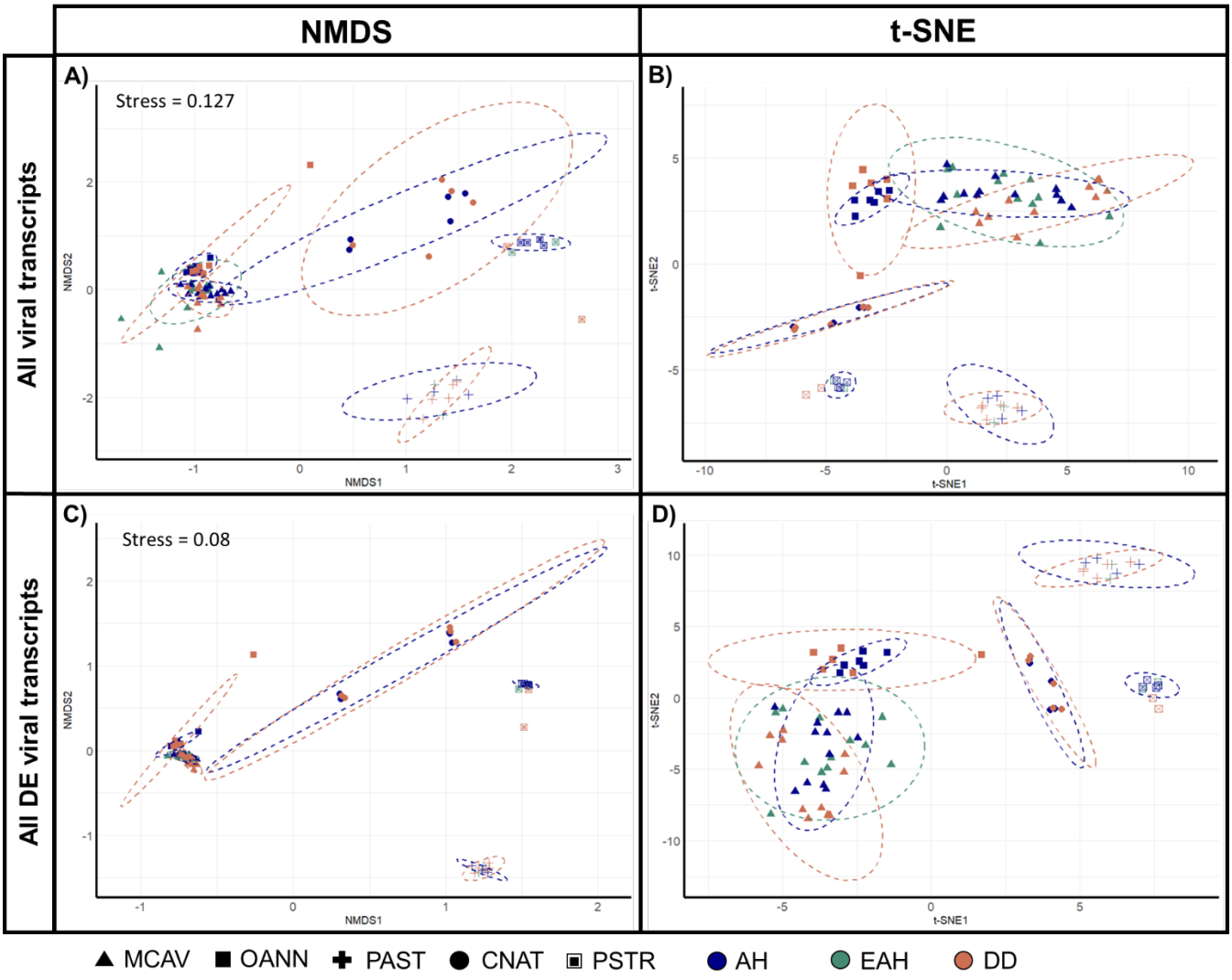
Supplemental Figure 1. Overview of SCTLD transmission experiment conducted in April 2019 and described by Meiling et al. (2021). A) General experimental design for transmission experiment; blue boxes represent water tables with circulating and temperature control water where the treatment bins (control and disease) were held during the experiment. Placement of bins in water tables were random and changed daily throughout the experiment. B) Photos of control and disease treatment bins from the transmission experiment, stars indicates that fragments are sourced from the same coral holobiont.



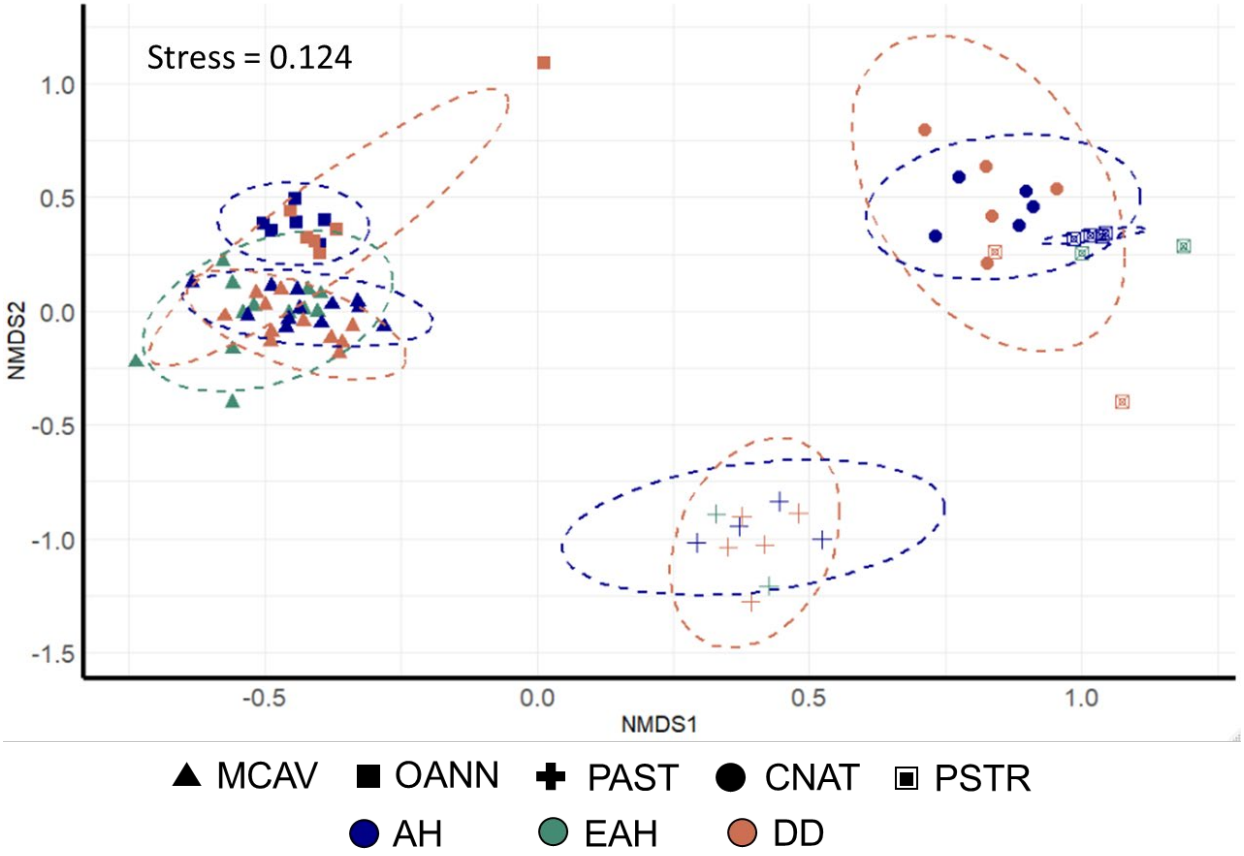
Supplemental Figure 2. Example of in situ sampling design: i. Orange circle – disease tissue sample collected immediately adjacent to the lesion boundary (the edge between live coral tissue and denuded skeleton), ii. Yellow circle – apparently healthy tissue sample from diseased coral colony, and iii. White circle – healthy tissue sample from nearby conspecific.



Supplemental Figure 3. Histogram illustrating distribution of virus confidence scores assigned to Deep6 predicted transcripts.



Supplemental Figure 4. Ordination plots based on Brays-Curtis distances displaying sample clustering behavior. Shapes represent coral species: *M. cavernosa* (MCAV; triangle), *O. annularis* (OANN, square), *P. astreoides* (PAST, cross), *C. natans* (CNAT, circle), and *P. strigosa* (PSTR, boxed square). Colors represent health status of sampled tissue: Apparently Healthy (AH, blue), Exposed-Apparently Healthy (EAH, green), Diseased (DD, orange). Ordination plots A and B are NMDS and t-SNE plots, respectively, based on all virus transcripts. Plots C and D are NMDS and t-SNE, respectively, based on all DESeq2-determined differentially expressed transcripts.



Supplemental Figure 5. NMDS ordination plot illustrating clustering behavior of samples with low confidence (*S*<0.45) virus transcripts excluded from the analysis. Shapes represent coral species: *M. cavernosa* (MCAV; triangle), *O. annularis* (OANN, square), *P. astreoides* (PAST, cross), *C. natans* (CNAT, circle), and *P. strigosa* (PSTR, boxed square). Colors represent coral health: Apparently Health (AH, blue), Expose-Apparently Healthy (EAH, green), Diseased (DD, orange).

INTRODUCTION AND CONCLUDING REMARKS REFERENCES

Aeby, G. S., Ushijima, B., Campbell, J. E., Jones, S., Williams, G. J., Meyer, J. L., Häse, C., & Paul, V. J. (2019). Pathogenesis of a tissue loss disease affecting multiple species of corals along the Florida Reef Tract. *Frontiers in Marine Science*, *6*, 678.

Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology*, *479*, 26–37.

Ambalavanan, L., Iehata, S., Fletcher, R., Stevens, E. H., & Zainathan, S. C. (2021). A Review of Marine Viruses in Coral Ecosystem. *Journal of Marine Science and Engineering*, *9*(7), Article 7. https://doi.org/10.3390/jmse9070711

Bao, X., & Roossinck, M. J. (2013). A life history view of mutualistic viral symbioses: Quantity or quality for cooperation? *Current Opinion in Microbiology*, *16*(4), 514–518. https://doi.org/10.1016/j.mib.2013.05.007

Beavers, K. M., Van Buren, E., Rossin, A., Emery, M., Veglia, A. J., Karrick, C. E., MacKnight, N. J., Dimos, B. A., Meiling, S. S., Smith, T. B., Apprill, A., Muller, E. M., Holstein, D. M., Correa, A. M. S., Brandt, M. E., Mydlarz, L. D. (2023). Stony Coral Tissue Loss Disease Induces Transcriptional Signatures of in situ Degradation of Dysfunctional Symbiodiniaceae. *Nature Communications. In press*

Bettarel, Y., Thuy, N. T., Huy, T. Q., Hoang, P. K., & Bouvier, T. (2013). Observation of virus-like particles in thin sections of the bleaching scleractinian coral Acropora cytherea. *Journal of the Marine Biological Association of the United Kingdom*, *93*(4), 909–912. https://doi.org/10.1017/S0025315411002062

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bordenstein, S. R., & Theis, K. R. (2015). Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biology*, *13*(8), e1002226. https://doi.org/10.1371/journal.pbio.1002226

Brandt, M. E., Ennis, R. S., Meiling, S. S., Townsend, J., Cobleigh, K., Glahn, A., Quetel, J., Brandtneris, V., Henderson, L. M., & Smith, T. B. (2021). The Emergence and Initial Impact of Stony Coral Tissue Loss Disease (SCTLD) in the United States Virgin Islands. *Frontiers in Marine Science*, *8*. https://www.frontiersin.org/articles/10.3389/fmars.2021.715329

Brown, B. E. (1997). Coral bleaching: Causes and consequences. *Coral Reefs*, *16*, 129–138. https://doi.org/10.1007/s003380050249

Buerger, P., Weynberg, K. D., Sato, Y., Willis, B. L., & Oppen, M. van. (2019). *Bacteriophages in black band disease Novel T4 bacteriophages associated with black band disease in corals*. https://doi.org/10.1111/1462-2920.14432

Camargo, A. P., Nayfach, S., Chen, I.-M. A., Palaniappan, K., Ratner, A., Chu, K., Ritter, S. J., Reddy, T. B. K., Mukherjee, S., Schulz, F., Call, L., Neches, R. Y., Woyke, T., Ivanova, N. N., Eloe-Fadrosh, E. A., Kyrpides, N. C., & Roux, S. (2022). IMG/VR v4: An expanded database of uncultivated virus genomes within a framework of extensive

functional, taxonomic, and ecological metadata. *Nucleic Acids Research*, gkac1037. https://doi.org/10.1093/nar/gkac1037

Cárdenas, A., Ye, J., Ziegler, M., Payet, J. P., McMinds, R., Vega Thurber, R., & Voolstra, C. R. (2020). Coral-Associated Viral Assemblages From the Central Red Sea Align With Host Species and Contribute to Holobiont Genetic Diversity. *Frontiers in Microbiology*, *11*. https://www.frontiersin.org/articles/10.3389/fmicb.2020.572534

Correa, A. M. S., Ainsworth, T. D., Rosales, S. M., Thurber, R. L. V., Butler, C. R., & Thurber, R. L. V. (2016). Viral outbreak in corals associated with an in situ bleaching event: Atypical herpes-like viruses and a new megavirus infecting Symbiodinium. *Frontiers in Microbiology*, *7*, 127. https://doi.org/10.3389/fmicb.2016.00127

Correa, A. M. S., Welsh, R. M., & Vega Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal*, *7*(1), Article 1. https://doi.org/10.1038/ismej.2012.75

Correa, A. M., Welsh, R. M., & Thurber, R. L. (2013). Unique nucleocytoplasmic dsDNA and+ ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal*, *7*(1), 13–27. https://doi.org/10.1038/ismej.2012.81

Davy, J. E., & Patten, N. L. (2007). Morphological diversity of virus-like particles within the surface microlayer of scleractinian corals. *Aquatic Microbial Ecology*, *47*, 37–44. https://doi.org/10.3354/ame047037

Davy, S. K., Burchett, S. G., Dale, A. L., Davies, P., Davy, J. E., Muncke, C., Hoegh-Guldberg, O., & Wilson, W. H. (2006). Viruses: Agents of coral disease? *Diseases of Aquatic Organisms*, *69*, 101–110. https://doi.org/10.3354/dao069101

Drappier, M., & Michiels, T. (2015). Inhibition of the OAS/RNase L pathway by viruses. *Current Opinion in Virology*, *15*, 19–26. https://doi.org/10.1016/j.coviro.2015.07.002

Eddy, T. D., Lam, V. W. Y., Reygondeau, G., Cisneros-Montemayor, A. M., Greer, K., Palomares, M. L. D., Bruno, J. F., Ota, Y., & Cheung, W. W. L. (2021). Global decline in capacity of coral reefs to provide ecosystem services. *One Earth*, *4*(9), 1278–1285. https://doi.org/10.1016/j.oneear.2021.08.016

González-Pech, R. A., Stephens, T. G., Chen, Y., Mohamed, A. R., Cheng, Y., Shah, S., Dougan, K. E., Fortuin, M. D. A., Lagorce, R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2021). Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus Symbiodinium. *BMC Biology*, *19*(1), 73. https://doi.org/10.1186/s12915-021-00994-6

Grasis, J. A. (2017). The Intra-Dependence of Viruses and the Holobiont. *Frontiers in Immunology*, *8*. https://www.frontiersin.org/articles/10.3389/fimmu.2017.01501

Grupstra, C. G. B., Howe-Kerr, L. I., Veglia, A. J., Bryant, R. L., Coy, S. R., Blackwelder, P. L., & Correa, A. M. S. (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *The ISME Journal*, *16*(5), Article 5. https://doi.org/10.1038/s41396-022-01194-y

Grupstra, C. G., Howe-Kerr, L. I., Veglia, A. J., Bryant, R. L., Coy, S. R., Blackwelder, P. L., & Correa, A. (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *The ISME Journal*, 1–12.

Hayes, R. L., & Bush, P. G. (1990). Microscopic observations of recovery in the reef-building scleractinian coral, Montastrea annularis, after bleaching on a Cayman reef. *Coral Reefs*, *8*(4), 203–209. https://doi.org/10.1007/BF00265012

Howe-Kerr, L. I. (2022). *Viruses of a key coral symbiont exhibit temperature-driven productivity across a reefscape*. https://doi.org/10.21203/rs.3.rs-1899377/v1

Howe-Kerr, L. I., Bachelot, B., Wright, R. M., Kenkel, C. D., Bay, L. K., & Correa, A. M. S. (2020). Symbiont community diversity is more variable in corals that respond poorly to stress. *Global Change Biology*, *26*(4), 2220–2234. https://doi.org/10.1111/gcb.14999

Howe-Kerr, L. I., Grupstra, C. G. B., Rabbitt, K. M., Conetta, D., Coy, S. R., Klinges, J. G., Maher, R. L., McConnell, K. M., Meiling, S. S., Messyasz, A., Schmeltzer, E. R., Seabrook, S., Sims, J. A., Veglia, A. J., Thurber, A. R., Thurber, R. L. V., & Correa, A. M. S. (2023). Viruses of a key coral symbiont exhibit temperature-driven productivity across a reefscape. *ISME Communications*, *3*(1), 27. https://doi.org/10.1038/s43705-023-00227-7

Hussa, E. A., & Goodrich-Blair, H. (2013). It Takes a Village: Ecological and Fitness Impacts of Multipartite Mutualism. *Annual Review of Microbiology*, *67*(1), 161–178. https://doi.org/10.1146/annurev-micro-092412-155723

Jackson, J., Donovan, M., Cramer, K., & Lam, V. Y. Y. (2014). *Status and trends of Caribbean coral reefs: 1970-2012*. https://www.semanticscholar.org/paper/Status-and-trends-of-Caribbean-coral-reefs-%3A-Jackson-Donovan/6a7d1f5fb6784b5444191bf1cc39f79fd97d454b

Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S. T., Rosenstiel, P., Dutilh, B. E., & Hentschel, U. (2019). A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. *Cell Host & Microbe*, *26*(4), 542-550.e5. https://doi.org/10.1016/j.chom.2019.08.019

Jahn, M. T., Lachnit, T., Markert, S. M., Stigloher, C., Pita, L., Ribes, M., Dutilh, B. E., & Hentschel, U. (2021). Lifestyle of sponge symbiont phages by host prediction and correlative microscopy. *The ISME Journal*, *15*(7), Article 7. https://doi.org/10.1038/s41396-021-00900-6

Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Jutz, S., Pascelli, C., Botté, E. S., Bell, S. C., Peirce, T. E., Weynberg, K. D., van Oppen, M. J. H., Rattei, T., & Webster, N. S. (2018). Reef invertebrate viromics: Diversity, host specificity and functional capacity. *Environmental Microbiology*, *20*(6), 2125–2141. https://doi.org/10.1111/1462-2920.14110

LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Current Biology*, *28*(16), 2570-2580.e6. https://doi.org/10.1016/j.cub.2018.07.008

Leruste, A., Bouvier, T., & Bettarel, Y. (2012). Enumerating Viruses in Coral Mucus. *Applied and Environmental Microbiology*, *78*(17), 6377–6379. https://doi.org/10.1128/AEM.01141-12

Levin, R. A., Voolstra, C. R., Weynberg, K. D., & van Oppen, M. J. H. (2017). Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *The ISME Journal*, *11*(3), 808–812. https://doi.org/10.1038/ismej.2016.154

Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., Cooke, I., Aranda, M., Bourne, D. G., Forêt, S., Miller, D. J., van Oppen, M. J. H., Voolstra, C. R., Ragan, M. A., & Chan, C. X. (2018). Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Communications Biology*, *1*(1), Article 1. https://doi.org/10.1038/s42003-018-0098-3

Marhaver, K. L., Edwards, R. A., & Rohwer, F. (2008). Viral communities associated with healthy and bleaching corals. *Environmental Microbiology*, *10*(9), 2277–2286. https://doi.org/10.1111/j.1462-2920.2008.01652.x

Martinand, C., Salehzada, T., Silhol, M., Lebleu, B., & Bisbal, C. (1998). RNase L inhibitor (RLI) antisense constructions block partially the down regulation of the 2-5A/RNase L pathway in encephalomyocarditis-virus-(EMCV)-infected cells. *European Journal of Biochemistry*, *254*(2), 248–255. https://doi.org/10.1046/j.1432-1327.1998.2540248.x

Martinez, S., Grover, R., & Ferrier-Pagès, C. (2022). Symbiont starvation affects the stability of the coral–Symbiodiniaceae symbiosis. *Frontiers in Marine Science*, *9*. https://www.frontiersin.org/articles/10.3389/fmars.2022.979563

Messyasz, A., Rosales, S. M., Mueller, R. S., Sawyer, T., Correa, A. M. S., Thurber, A. R., & Vega Thurber, R. (2020). Coral Bleaching Phenotypes Associated With Differential Abundances of Nucleocytoplasmic Large DNA Viruses. *Frontiers in Marine Science*, *7*. https://www.frontiersin.org/articles/10.3389/fmars.2020.555474

Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, *29*(2), 215–233. https://doi.org/10.1016/S0921-8009(99)00009-9

Montalvo-Proaño, J., Buerger, P., Weynberg, K. D., & van Oppen, M. J. H. (2017). A PCR-Based Assay Targeting the Major Capsid Protein Gene of a Dinorna-Like ssRNA Virus That Infects Coral Photosymbionts. *Frontiers in Microbiology*, *8*. https://www.frontiersin.org/articles/10.3389/fmicb.2017.01665

Neil, J. A., & Cadwell, K. (2018). The Intestinal Virome and Immunity. *The Journal of Immunology*, *201*(6), 1615–1624. https://doi.org/10.4049/jimmunol.1800631

Nguyen, M., Wemheuer, B., Laffy, P. W., Webster, N. S., & Thomas, T. (2021). Taxonomic, functional and expression analysis of viral communities associated with marine sponges. *PeerJ*, *9*, e10715. https://doi.org/10.7717/peerj.10715

Nguyen-Kim, H., Bettarel, Y., Bouvier, T., Bouvier, C., Doan-Nhu, H., Nguyen-Ngoc, L., Nguyen-Thanh, T., Tran-Quang, H., & Brune, J. (2015). Coral Mucus Is a Hot Spot for Viral Infections. *Applied and Environmental Microbiology*, *81*(17), 5773–5783. https://doi.org/10.1128/AEM.00542-15

Pascelli, C., Laffy, P. W., Botté, E., Kupresanin, M., Rattei, T., Lurgi, M., Ravasi, T., & Webster, N. S. (2020). Viral ecogenomics across the Porifera. *Microbiome*, *8*(1), 144. https://doi.org/10.1186/s40168-020-00919-5

Pascelli, C., Laffy, P. W., Kupresanin, M., Ravasi, T., & Webster, N. S. (2018). Morphological characterization of virus-like particles in coral reef sponges. *PeerJ*, *6*, e5625. https://doi.org/10.7717/peerj.5625

Patten, N., Mitchell, J., Middelboe, M., Eyre, B., Seuront, L., Harrison, P., & Glud, R. (2008). Bacterial and viral dynamics during a mass coral spawning period on the Great Barrier Reef. *Aquatic Microbial Ecology*, *50*, 209–220. https://doi.org/10.3354/ame01179

Pita, L., Fraune, S., & Hentschel, U. (2016). Emerging Sponge Models of Animal-Microbe Symbioses. *Frontiers in Microbiology*, *7*. https://www.frontiersin.org/articles/10.3389/fmicb.2016.02102

Planes, S., Allemand, D., Agostini, S., Banaigs, B., Boissin, E., Boss, E., Bourdin, G., Bowler, C., Douville, E., Flores, J. M., Forcioli, D., Furla, P., Galand, P. E., Ghiglione, J.-F., Gilson, E., Lombard, F., Moulin, C., Pesant, S., Poulain, J., … Zoccola, D. (2019). The Tara Pacific expedition—A pan-ecosystemic approach of the "-omics" complexity of

coral reef holobionts across the Pacific Ocean. *PLoS Biology*, *17*(9), e3000483. https://doi.org/10.1371/journal.pbio.3000483

Precht, W. F., Gintert, B. E., Robbart, M. L., Fura, R., & van Woesik, R. (2016). Unprecedented Disease-Related Coral Mortality in Southeastern Florida. *Scientific Reports*, *6*(1), 31374. https://doi.org/10.1038/srep31374

Quistad, S. D., Grasis, J. A., Barr, J. J., & Rohwer, F. L. (2017). Viruses and the origin of microbiome selection and immunity. *The ISME Journal*, *11*(4), Article 4. https://doi.org/10.1038/ismej.2016.182

Roossinck, M. J. (2011). The good viruses: Viral mutualistic symbioses. *Nature Reviews Microbiology*, *9*(2), 99–108. https://doi.org/10.1038/nrmicro2491

Roossinck, M. J. (2015). Plants, viruses and the environment: Ecology and mutualism. *Virology*, *479–480*, 271–277. https://doi.org/10.1016/j.virol.2015.03.041

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

Rusanova, A., Fedorchuk, V., Toshchakov, S., Dubiley, S., & Sutormin, D. (2022). An Interplay between Viruses and Bacteria Associated with the White Sea Sponges Revealed by Metagenomics. *Life*, *12*(1), Article 1. https://doi.org/10.3390/life12010025

Schloss, P. D. (2020). Reintroducing mothur: 10 Years Later. *Applied and Environmental Microbiology*, *86*(2). https://doi.org/10.1128/AEM.02343-19

Soffer, N., Brandt, M. E., Correa, A. M. S., Smith, T. B., & Thurber, R. V. (2014). Potential role of viruses in white plague coral disease. *The ISME Journal*, *8*(2), 271–283. https://doi.org/10.1038/ismej.2013.137

Soffer, N., Zaneveld, J., & Vega Thurber, R. (2015). Phage-bacteria network analysis and its implication for the understanding of coral disease. *Environmental Microbiology*, *17*(4), 1203–1218. https://doi.org/10.1111/1462-2920.12553

Stephens, T. G., González-Pech, R. A., Cheng, Y., Mohamed, A. R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2020). Genomes of the dinoflagellate Polarella glacialis encode tandemly repeated single-exon genes with adaptive functions. *BMC Biology*, *18*(1), 56. https://doi.org/10.1186/s12915-020-00782-8

Sweet, M., & Bythell, J. (2017). The role of viruses in coral health and disease. *Journal of Invertebrate Pathology*, *147*, 136–144. https://doi.org/10.1016/j.jip.2016.12.005

Thompson, J. R., Rivera, H. E., Closek, C. J., & Medina, M. (2015). Microbes in the coral holobiont: Partners through evolution, development, and ecological interactions. *Frontiers in Cellular and Infection Microbiology*, *4*. https://www.frontiersin.org/articles/10.3389/fcimb.2014.00176

Thurber, R. V., Payet, J. P., Thurber, A. R., & Correa, A. M. S. (2017). Virus–host interactions and their roles in coral reef health and disease. *Nature Reviews Microbiology*, *15*(4), 205–216. https://doi.org/10.1038/nrmicro.2016.176

van Woesik, R., & Kratochwill, C. (2022). A global coral-bleaching database, 1980–2020. *Scientific Data*, *9*(1), Article 1. https://doi.org/10.1038/s41597-022-01121-y

Veglia, A. J., Beavers, K., Buren, E. W. V., Meiling, S. S., Muller, E. M., Smith, T. B., Holstein, D. M., Apprill, A., Brandt, M. E., Mydlarz, L. D., & Correa, A. M. S. (2022). Alphaflexivirus Genomes in Stony Coral Tissue Loss Disease-Affected, Disease-Exposed, and Disease-Unexposed Coral Colonies in the U.S. Virgin Islands. *Microbiology Resource Announcements*. https://doi.org/10.1128/mra.01199-21

Veglia, A. J., Milford, C. R., & Schizas, N. V. (2021). Isolation and genotyping of novel T4 cyanophages associated with diverse coral reef invertebrates. *Coral Reefs*, *40*(2), 485–504. https://doi.org/10.1007/s00338-021-02056-3

Veglia, A., Rivera-Vicens, R. E., Grupstra, C., Howe-Kerr, L., & Correa, A. (2023). *vAMPirus: A versatile amplicon processing and analysis program for studying viruses* [Preprint]. Preprints. https://doi.org/10.22541/au.167592965.50010700/v1

Walton, C. J., Hayes, N. K., & Gilliam, D. S. (2018). Impacts of a Regional, Multi-Year, Multi-Species Coral Disease Outbreak in Southeast Florida. *Frontiers in Marine Science*, *5*. https://www.frontiersin.org/articles/10.3389/fmars.2018.00323

Weil, E., & Rogers, C. S. (2011). Coral Reef Diseases in the Atlantic-Caribbean. In *Coral Reefs: An Ecosystem in Transition*. Springer. https://doi.org/10.1007/978-94-007-0114-4

Weynberg, K. D., Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Rattei, T., Webster, N. S., & van Oppen, M. J. H. (2017). Coral-associated viral communities show high levels of diversity and host auxiliary functions. *PeerJ*, *5*, e4054. https://doi.org/10.7717/peerj.4054

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P., & van Oppen, M. J. H. (2015). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. *Scientific Reports*, *5*, 17889. https://doi.org/10.1038/srep17889

Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A., & van Oppen, M. J. H. (2014). Generating viral metagenomes from the coral holobiont. *Frontiers in Microbiology*, *5*. https://www.frontiersin.org/articles/10.3389/fmicb.2014.00206

Wilson, W. H., & Chapman, D. M. (2001). Observation of virus-like particles in thin sections of the plumose anemone, Metridium senile. *Journal of the Marine Biological Association of the United Kingdom*, *81*(5), 879–880. https://doi.org/10.1017/S0025315401004726

Wilson, W. H., Dale, A. L., Davy, J. E., & Davy, S. K. (2005). An enemy within? Observations of virus-like particles in reef corals. *Coral Reefs*, *24*(1), 145–148. https://doi.org/10.1007/s00338-004-0448-0

Wood-Charlson, E. M., Weynberg, K. D., Suttle, C. A., Roux, S., & van Oppen, M. J. H. (2015). Metagenomic characterization of viral communities in corals: Mining biological signal from methodological noise. *Environmental Microbiology*, *17*(10), 3440–3449. https://doi.org/10.1111/1462-2920.12803