ABSTRACT

Title of Thesis: IDENTIFICATION AND CHARACTERIZATION OF VIRUSES ASSOCIATED WITH THE ATLANTIC BLUE CRAB, *CALLINECTES SAPIDUS,* ACROSS ITS UNITED STATES GEOGRAPHIC RANGE

Jennifer Herrera, Master of Science, 2025

Thesis Directed By: Associate Research Professor, Eric Schott, University of Maryland Center for Environmental Science

Across two hemispheres, including in the United States, the Atlantic blue crab, *Callinectes sapidus*, supports a multi-million dollar fishery and plays critical roles in benthic food webs. Recent blue crab population declines in the Mid-Atlantic highlight the need to better understand potential threats to the fishery, including infectious diseases. Viruses can cause disease outbreaks and mortality yet remain relatively understudied compared to other microbial pathogens of marine shellfish. Movement of harvested blue crabs between states may pose a threat of unknown magnitude to the health of wild crabs in the receiving location.

To gain a preliminary understanding of the viral communities associated with the blue crab, we performed a virus-enrichment protocol on specimens collected from six U.S. states, from New York to Texas along the U.S. Atlantic and Gulf Coasts. Enrichment was followed by RNA

extraction, cDNA synthesis, and high-throughput sequencing. Bioinformatic tools were used to identify and characterize putative virus genomes in gill collections.

We discovered a total of 97 putative viral sequences, of which 29 were high-quality genomes with a k-mer coverage >10. Positive-sense RNA viruses were especially prominent, with provisional Picornavirales members constituting 19 of the 29 genomes. Beyond the Picornavirales, seven other positive-sense, two negative-sense, and one double-stranded RNA virus were identified. Finally, this survey yielded potential insights into viral geographic distributions.

This research expands our current knowledge of viruses associated with the blue crab across multiple U.S. states. Similar to several prior studies of crustacean viruses, this study identified a large number of entities that putatively belong to viruses of the Picornavirales. Future work will investigate whether the observation of regionally-restricted virus genomes is confirmed when additional crabs are screened using specific PCR assays. The knowledge generated here will enable investigation of the pathogenicity of novel blue crab viruses, which is especially relevant in the face of climate change and the interstate transport of marine resources. Ultimately, such research can help monitor and mitigate the spread of shellfish pathogens.

IDENTIFICATION AND CHARACTERIZATION OF VIRUSES ASSOCIATED WITH THE ATLANTIC BLUE CRAB, *CALLINECTES SAPIDUS,* ACROSS ITS UNITED STATES GEOGRAPHIC RANGE

by

Jennifer Herrera

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2025

Advisory Committee:
Professor Eric Schott, Chair
Dr. Tsvetan Bachvaroff, Co-chair
Dr. Allision Tracy

Acknowledgements

I would like to express my gratitude to everyone who has supported me throughout this journey.

First, I extend my thanks to my primary advisor, Dr. Eric Schott, for welcoming me into his lab and providing me with opportunities to grow as a scientist. I am also grateful to my co-advisor, Dr. Tsetso Bachvaroff, for his unwavering support, patience, and for always welcoming my many questions about bioinformatics with kindness and clarity. A heartfelt thank you to Dr. Allison Tracy for enthusiastically joining my committee, offering valuable feedback, and encouraging me at just the right time.

I am grateful to the NOAA-Living Marine Resources Cooperative Science Center (LMRCSC) for funding my graduate work and supporting my summer research in Connecticut. Thank you to Dr. Shannon Meseck for being a generous NOAA mentor, and to everyone at the Milford Lab for making my time there both productive and memorable. I also thank the LMRCSC for introducing me to a fun, supportive group of peers—your presence truly shaped this experience.

Thank you to my labmates for helping with sample collection and for their steady support. I also appreciate the continued encouragement from colleagues and friends from the Marine Animal Disease Laboratory at Stony Brook University, who never failed to stay in touch. To my roommate Sasha and my friends, thank you for making life in Baltimore so memorable.

To everyone who listened to my frustrations about failed experiments or long days of data analysis, I couldn't have done it without you.

Finally, I’m forever grateful to my parents, Julian and Adriana, for supporting my decision to study diseases in shellfish instead of people, and for always encouraging my curiosity. To my sister Daniella and my dog Brownie, thank you for being a constant source of love and support.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

1 The blue crab, *Callinectes sapidus*

1.1 Ecological importance

The Atlantic blue crab, *Callinectes sapidus*, is a benthic coastal crustacean renowned for its ecological, economic, and cultural importance. This species ranges from Nova Scotia, Canada, through the Gulf of Mexico and as far as South America (e.g., Brazil and northern Argentina) (Mancinelli et al., 2021). As integral members of estuarine food webs, blue crabs function as both predators and prey. By feeding on bivalves and infauna, they can influence soft-bottom community composition (Lipcius et al., 2007; Seitz et al., 2005) and juvenile crabs serve as a key dietary component for fish such as striped bass, channel bass, and croaker (Orth et al., 1999). Additionally, blue crabs are highly cannibalistic, with larger individuals preying on smaller conspecifics, a behavior that influences population dynamics and recruitment success (Hines & Ruiz, 1995; Moksnes et al., 1997). By linking lower trophic levels and higher predators, *C. sapidus* helps maintain the balance and productivity of coastal ecosystems (Orth et al., 1999).

1.2 Cultural and social importance

Beyond its ecological roles, the blue crab carries significant social and cultural value, especially in regions like the Chesapeake Bay. It is considered an iconic species—indeed the state crustacean of Maryland—and has long been embedded in local traditions and cuisine (Chesapeake Bay Foundation, n.d.; NOAA Fisheries, 2025). For generations, coastal communities have engaged in crabbing as a way of life, and seasonal crab feasts are hallmark traditions in these areas (Chesapeake Bay Foundation, n.d.). This longstanding cultural association reinforces the blue crab’s status as a prominent symbol of regional identity and heritage.

1.3 Economic importance

Economically, the Atlantic blue crab supports one of the most valuable fisheries along the Atlantic and Gulf coasts, with annual U.S. commercial landings generating approximately $200 million in revenue (NMFS Commercial Landings, 2022). The Mid-Atlantic accounts for slightly less than half of this total ($80–$84 million per year in recent years), making it the largest regional contributor to the national fishery. Substantial blue crab fisheries also exist in the Southeast and Gulf of Mexico, with Louisiana consistently ranking as the top state for harvest revenue nationwide.

1.4 Interstate transport and regulatory gaps

Louisiana’s status as the leading state for blue crab landings is partly attributable to the interstate transport of live crabs to the Mid-Atlantic, where they supply a robust seafood industry, and help meet regional market demands during periods of local scarcity (Perry & VanderKooy, 2015). This seasonal movement of product is especially pronounced during the winter months when harvest restrictions in the temperate waters of the Mid- and North Atlantic are implemented to protect hibernating crabs (Cameron’s Seafood, 2023; Maryland Department of Natural Resources, 2025).

Despite the economic and logistical importance of this trade, regulations governing the interstate transport of blue crabs are highly variable and fragmented across jurisdictions. Commercial handlers typically prioritize maintaining crab viability during transit—such as by controlling temperature and humidity in transport containers—to preserve product quality and reduce mortality-related losses (Louisiana Direct Seafood, n.d.). At the federal level, the U.S. Food and Drug Administration (FDA) requires that food products be designated as adulterated—and consequently prohibited from sale—if they contain "in whole or in part any filthy, putrid, or decomposed substance" (Saxowsky, n.d.; U.S. Congress, n.d.; U.S. Food and Drug Administration, 2023). This regulatory standard applies to crabs that perish during transport.

Notably, while catch reports and trip tickets are used to document the number of live crabs sold from the point of harvest to market, there are no federal or state-level regulatory requirements for reporting "market-bound mortalities." Consequently, moribund/dead crabs are generally excluded from official landing data (Chesapeake Bay Stock Assessment Committee, 2021; Virginia Marine Resources Commission, 2024; Walters et al., 2025). A recent study of Florida's blue crab fishery highlighted that market-bound mortalities are not accounted for in fishery statistics, leading to the systematic underestimation of harvest-associated losses (Walters et al., 2025). While some businesses may internally log such mortalities for inventory control or supplier accountability, these practices are not standardized or legally mandated. In rare cases involving large-scale mortality events, businesses may notify public health authorities; however, this mostly remains discretionary and informal.

In terms of blue crab disposal, regulatory oversight does exist but is inconsistently enforced. Under the U.S. Environmental Protection Agency's Marine Protection, Research, and Sanctuaries Act (MPRSA), the unpermitted disposal of fishery waste into harbors and protected or enclosed coastal waters is prohibited (U.S. Environmental Protection Agency, 2024). Subsequently, most localities and environmental groups encourage the disposal of moribund/dead crabs through approved waste management systems such as landfilling, composting, or rendering. Nevertheless, anecdotal accounts from fishers suggest that potentially non-compliant disposal practices like overboard dumping continue in some regions, revealing gaps in regulatory enforcement.

1.5 Blue crab population trends

The Mid-Atlantic blue crab population size has consistently been below average since 2019, and 2022 marked the lowest number of blue crabs ever estimated to be in the Chesapeake Bay, with 2025 ranking a close second (Chesapeake Bay Stock Assessment Committee, 2025; **Figure 1.1**). These recent declines have introduced ecological and economic uncertainty to the communities that benefit from the fishery. In each of the past five years, the Chesapeake Bay Stock Assessment Advisory Report mentions the need to better understand the role of diseases in blue crab mortality (Chesapeake Bay Stock Assessment Committee, 2021, 2022, 2023, 2024, 2025).

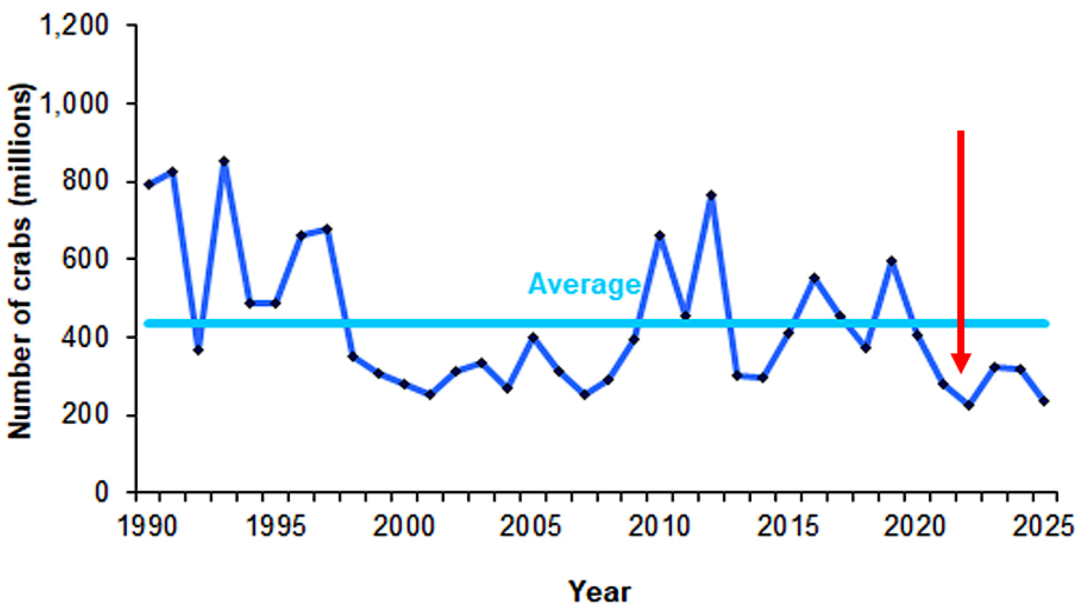


Figure 1.1. Total number of blue crabs estimated to be in the Chesapeake Bay. Red arrow points to 2022, the year with the lowest number of blue crabs ever estimated to be in the bay by the Winter Dredge Survey (Chesapeake Bay Stock Assessment Committee, 2025).

2 Viruses

Viruses are the most numerous biological entities in the ocean (Bergh et al., 1989; Fuhrman, 1999). Despite their small size, research over several decades has unraveled their critical and extensive roles at the foundation of marine ecosystems. As obligate parasites, viruses hijack their hosts' molecular machinery to replicate and propagate, sometimes influencing host immunity, physiology, and overall health. Since the turn of the century, virus research has revolutionized our understanding of marine food webs by revealing the key roles of these non-living entities in nutrient cycling within microbial and phytoplankton communities (Flynn et al., 2022; Fuhrman, 1999; Proctor, 1990; Rhodes et al., 2008; Suttle, 2007; Suttle et al., 1990; Wommack & Colwell, 2000). More recently, molecular methods commonly applied to microbial and unicellular marine organisms have begun uncovering the diversity of viruses in marine metazoans.

2.1 Viral diversity in the blue crab

Given the broad prevalence of viruses in marine ecosystems, it is not surprising that blue crabs harbor a variety of viruses. Early investigations of blue crab diseases in the 1970s and 1980s uncovered several distinct viruses infecting this species (Messick & Sindermann, 1992). At the forefront of this research was Dr. Phyllis Johnson, who described various viral infections in blue crabs through histopathology and electron microscopy (Johnson 1977, 1983, 1984). Her work documented viruses that are likely benign, including two baculoviruses (baculoviruses "A" and "B"). Furthermore, Johnson described viruses with potential synergistic interactions (i.e., viruses with enhanced function/pathology in the presence of another viral infection), such as paramyxo-like enveloped helical virus (EHV) and rhabdo-like virus A, as well as pathogens that are lethal to crabs via injection or natural infection, including a reo-like virus (RLV), bi-facies virus (BFV), and the picorna-like Chesapeake Bay virus (CBV) (Messick & Sindermann, 1992).

Based in Oxford, Maryland, Johnson initially characterized these viruses in blue crabs from Mid-Atlantic waters. To date, baculovirus "B", BFV, and CBV have not been reported outside this region. Other viruses have broader known distributions—for example, baculovirus "A" has been detected in the Long Island Sound (Connecticut), and EHV has been found in Florida (Shields & Overstreet, 2003). Rhabdo-like virus A is believed to be widespread along both Atlantic and Gulf coasts, while RLV has been definitively shown to be distributed throughout most of the Americas (Shields & Overstreet, 2003).

More blue crab virus discoveries have been made in the 21st century, including the identification of two novel toti-like viruses infecting the species in the northern range of the United States (Zhao et al., 2022). Overall, the identification of double-stranded DNA (i.e., baculo-like, toti-like, BFV) and RNA (i.e., reo-like) viruses, as well as positive-sense (i.e., picorna-like) and negative-sense (i.e., rhabdo-like, paramyxo-like) RNA viruses, highlight the marked differences and wide diversity in the viruses that infect the blue crab.

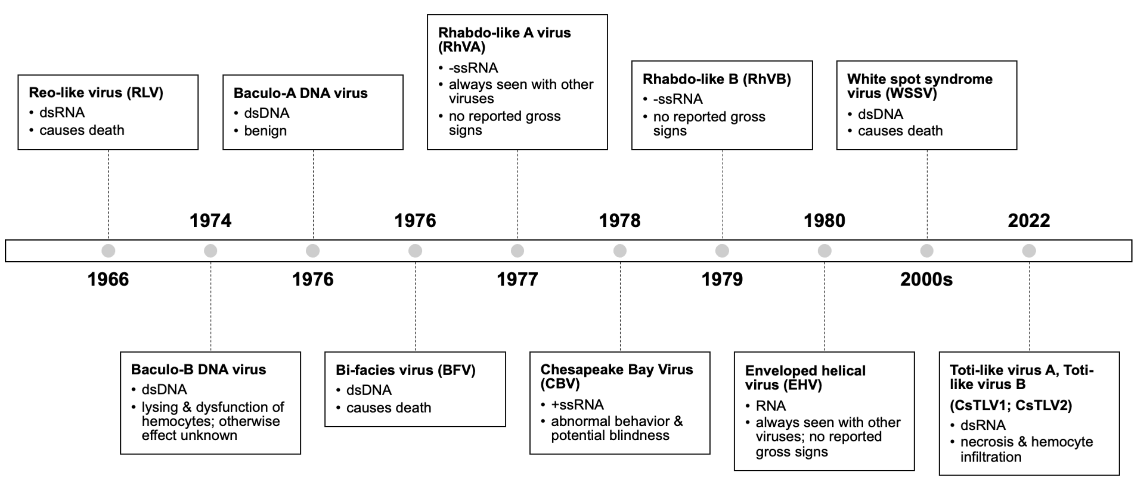


Figure 1.2. Timeline of blue crab virus discovery.

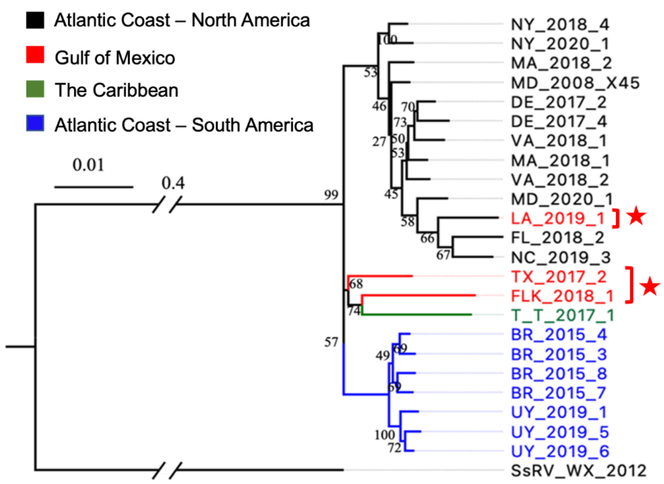
2.2 *Callinectes sapidus* reovirus 1

The blue crab reo-like virus (RLV), first described to cause tremors and paralysis in blue crabs of the Chesapeake Bay (Johnson, 1977), has emerged as a focal point in blue crab virology. In 2010, molecular characterization confirmed it as a distinct reovirus, subsequently named *Callinectes sapidus* reovirus 1 (CsRV1) (Bowers et al., 2010). This virus localizes in crab hemocytes and hypodermal tissue, leading to a systemic disease (Bowers et al., 2010; Johnson 1977, 1984). Infected crabs often become lethargic and may develop paralysis of limbs, ultimately resulting in death at high viral loads (Bowers et al., 2010, 2011).

Anthropogenic factors influence the geographic distribution and abundance of CsRV1. The virus has been associated with mortalities at soft-shell shedding facilities, where pre-molt crabs are held in shallow troughs until they complete ecdysis (Bowers et al., 2010, 2011; Johnson 1977; Spitznagel et al. 2019). Furthermore, CsRV1 infections in crabs are more abundant in waterways near these facilities rather than in waterways further away from shedding activity (Flowers et al., 2018). These studies gave preliminary insights into the ability of fishery and aquaculture practices to influence CsRV1 disease dynamics.

Since its rediscovery in 2010, studies examining the cross-hemispheric genetic diversity of CsRV1 have revealed distinct phylogeographic patterns and provided insights into the virus's population structure. As shown in **Figure 1.3**, findings from Zhao et al. (2023) support that CsRV1 strains generally cluster with others from proximate regions. However, the researchers identified one notable exception: a Louisiana strain of the virus is more similar to strains from the Atlantic Coast than to other strains from the Gulf of Mexico. These findings are inconsistent with gene flow due to natural migration along the coast, and suggest that the interstate transport of blue crabs could be influencing the spread of a CsRV1 subtype.

(A) **(B)**



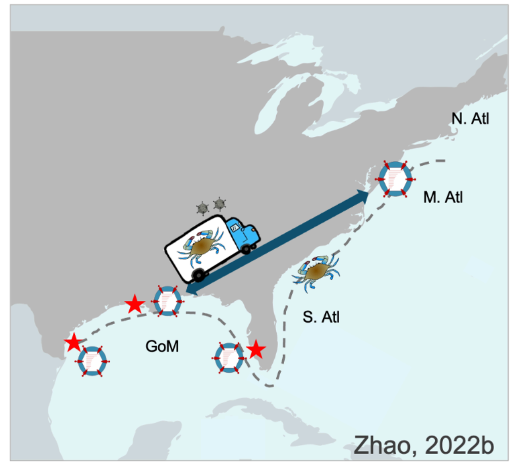


Figure 1.3. Phylogenetic tree of CsRV1 genotypes across the U.S. and entire native range of the blue crab, showing evidence of unexpected gene flow between regions in the Gulf of Mexico and the Mid-Atlantic. (A) Maximum Likelihood phylogenetic tree of concatenated genome sequences of CsRV1 across the Atlantic Coast (black), the Gulf of Mexico (red), the Caribbean (green), and the South Atlantic Coast (blue). The Gulf of Mexico CsRV1 sequences, shown in red, show that the Louisiana (LA) strain groups with strains of the Atlantic Coast rather than other strains in the Gulf of Mexico. **(B)** Map showing the native range of the blue crab in the U.S. including the North Atlantic (N. Atl), the Mid-Atlantic (M. Atlantic), and the South Atlantic (S. Atl). Red stars correspond to the origins of the Gulf of Mexico genotypes. The dashed grey line represents the expected natural crab migration along the U.S. coasts and the teal arrow represents a hypothetical pathway for crab movement via trucks for the seafood industry.

These genetic connectivity findings, with the knowledge of large-scale interstate transport, highlight that anthropogenic movement of marine species can facilitate the spread of their viruses, bypassing natural geographic barriers.

3 Virus discovery and its evolution

Understanding the ‘virosphere’ has been tightly linked to the tools and approaches available to discover viruses. Over the past century, the field of virology has evolved from relying on simple observation of diseased organisms to using sophisticated molecular techniques that can reveal an enormous diversity of viruses. The history of virus discovery in marine crustaceans reflects this evolution.

3.1 Microscopy and histopathology (mid-20th century)

For much of the 20th century, viruses could only be detected by observing their effects on host tissues or by visualizing virus particles under a microscope. In the 1960s-1980s, transmission electron microscopy (TEM) and light microscopy of stained tissue sections were pioneering methods to find new viruses. Dr. Phyllis Johnson and other scientists subsequently used these techniques to publish some of the earliest descriptions of blue crab viruses (Messick & Sindermann, 1992). However, only viruses present at relatively high abundances or causing prominent tissue abnormalities were noted. Moreover, microscopy alone could not determine much about a virus's genetic identity or its broader distribution in the host population.

3.2 Molecular detection (PCR and RT-PCR, 1990s–2010s)

The advent of molecular biology techniques significantly advanced virus discovery and diagnostics. Polymerase chain reaction (PCR) enabled researchers to amplify specific nucleic acid sequences even when virus particles could not be directly observed by microscopy. In crustacean pathology, this advancement facilitated the confirmation of known viruses and the detection of pathogens through genetic signatures. For instance, following the genomic characterization of CsRV1, researchers developed RT-PCR assays to screen wild crab populations (Flowers et al., 2016). PCR-based surveys conducted in the 2010s confirmed that CsRV1 is present in hard crabs along nearly the entire U.S. Atlantic and Gulf Coasts (Bowers et al., 2011; Flowers et al., 2016; Rogers et al., 2014). Similarly, PCR has become a standard diagnostic tool for globally significant crustacean viruses, such as white spot syndrome virus (WSSV) in farmed shrimp (Nunan & Lightner, 2011). These molecular approaches dramatically increased diagnostic sensitivity, facilitating extensive prevalence studies and enabling the detection of low-level infections previously undetectable by microscopy. However, PCR methods require prior sequence knowledge and therefore excel at monitoring known pathogens but are less effective for discovering entirely novel viruses.

3.3 Viromics and metagenomics (high-throughput sequencing, 2010s–present)

The most recent leap in virus discovery comes from high-throughput sequencing and metagenomics. Virus metagenomics involves sequencing all genetic material (i.e., shotgun sequencing) and using bioinformatics to identify viral genomes. Unlike molecular techniques, this shotgun approach does not require prior sequence knowledge.

Over the past decade, viromics has revealed an extraordinary breadth of previously unknown viruses across a wide range of hosts. One of the most influential studies characterized RNA virus diversity across more than 220 species of invertebrates—including insects, crustaceans, nematodes, and mollusks—and identified 1,445 novel RNA viruses (Shi et al., 2016). Subsequent studies have reinforced these findings, identifying hundreds of novel viruses in *C. hongkongensis* (Jiang et al., 2023) and marine environments (Liang et al., 2019; Wolf et al., 2020).

The strength of viromics lies in its ability to detect both highly abundant and rare viruses, offering a more complete view of viral community structure. However, this approach also presents new challenges: it often identifies virus sequences detached from any information about their infectious particles or pathology. In other words, sequencing can tell us a virus genome is present, but additional work (e.g. electron microscopy, cell culture, or animal infection studies) is needed to confirm the virus is actively infecting the host and to understand its effects. Despite these challenges, the trend is clear—metagenomics has rapidly advanced the field of marine virology, moving us from a sparse list of known crustacean viruses to a much richer inventory of "virus-like" entities that invites further biological and ecological investigation.

Through these evolving methodologies, our understanding of viruses in blue crabs has steadily progressed. Each technological advance has addressed distinct gaps in knowledge: microscopy provided the first visual evidence of viral infections in blue crabs; PCR-based assays demonstrated that some of these infections are prevalent in wild and aquaculture populations; and now, metagenomics is revealing the hidden diversity of the crab virome and the broader marine virosphere. The ongoing challenge lies in translating this genomic data into biological insight about virus ecology, host specificity, and pathogenicity.

4 Dissertation goals and objectives

As highlighted throughout this chapter, the Atlantic blue crab (*C. sapidus*) is a marine species of considerable ecological, economic, and sociocultural importance. Like many marine invertebrates, it is subject to a variety of environmental and anthropogenic stressors (e.g., elevated water temperatures, interstate transport for the seafood industry, etc.) that may increase its susceptibility to infectious agents, including viruses.

The primary objective of this dissertation is to provide a snapshot and characterization of RNA viruses associated with commercially harvested *C. sapidus* across six sampled U.S. states—Florida, Louisiana, Maryland, Texas, North Carolina, and New York—spanning a substantial portion of the species' range along the Atlantic and Gulf coasts. This work aims to establish a baseline understanding of viruses with potential ecological implications for blue crab populations.

Such information will aid future research efforts focused on developing diagnostic tools for the detection of specific viruses in blue crab populations, as well as biological assays to investigate the pathogenicity of these viruses and their effects on host immunity and physiology. In turn, these insights may inform evidence-based management practices, including strategies for monitoring interstate crab transport, protocols for disease surveillance and reporting, and responses to emerging viral threats.

More broadly, this research contributes to the growing body of knowledge on marine invertebrate virology, highlighting the role of viruses as influential components of host population dynamics. The continued characterization of crustacean viromes advances our understanding of viral ecology in the marine environment, with implications for ecosystem health and the sustainability of fisheries worldwide.

Chapter Two: Identification and Characterization of Novel Viruses Associated with the Atlantic Blue Crab, *Callinectes sapidus,* Across Its United States Geographic Range

1 Introduction

In the United States, the Atlantic blue crab, *Callinectes sapidus*, supports a commercial fishery of around $200 million annually and symbolizes a way of life for many coastal communities (Chesapeake Bay Foundation, n.d.; NMFS Commercial Landings, 2022). These vibrant animals, whose scientific name translates to "beautiful savory swimmer" in Latin, play key ecological roles as predators and prey in benthic food webs of estuarine habitats. For example, blue crab predation on bivalves and infauna can significantly alter soft-bottom community structures (Lipcius et al., 2007; Seitz et al., 2005), and juvenile crabs are an important food source for fish such as striped bass, channel bass, and croaker (Orth et al., 1999).

In the Mid-Atlantic, commercial fishing generates approximately $84 million in annual blue crab landings, making the region the largest contributor to the U.S. fishery (NMFS Commercial Landings, 2022). However, the Southeast and the Gulf of Mexico also have substantial blue crab fisheries. Louisiana is the top state for blue crab harvest revenue nationwide, partially due to interstate sale of crabs to the Mid-Atlantic region for the seafood industry. This transport peaks during the winter when regulations prevent the harvest of blue crabs hibernating in the temperate waters of the Mid- and North Atlantic (Cameron's Seafood, 2023; Maryland Department of Natural Resources, 2025).

Mid-Atlantic blue crab populations steadily declined from 2019 to 2022, and 2022 marked the lowest number of blue crabs ever recorded in the Chesapeake Bay (Chesapeake Bay Stock Assessment Committee, 2024). Since then, their populations within the region have consistently been estimated to be below average, with 2025 ranking a close second in terms of the lowest crab numbers. These recent declines have introduced ecological and economic uncertainty to the communities that benefit from the fishery. In each of the past five years, the Chesapeake Bay Stock Assessment Advisory Report mentions the need to better understand the role of diseases in blue crab mortality (Chesapeake Bay Stock Assessment Committee, 2021, 2022, 2023, 2024, 2025).

Viruses are the most numerous biological entities in the ocean (Bergh et al., 1989; Fuhrman, 1999). Despite their small size, research over several decades has unraveled their critical and extensive roles at the foundation of marine ecosystems. As obligate parasites, viruses hijack their hosts' molecular machinery to replicate and propagate, sometimes influencing host immunity, physiology, and overall health. Since the turn of the century, virus research has revolutionized our understanding of marine food webs by revealing the key roles of these non-living entities in nutrient cycling within microbial and phytoplankton communities (Flynn et al., 2022; Fuhrman, 1999; Proctor, 1990; Rhodes et al., 2008; Suttle, 2007; Suttle et al., 1990; Wommack & Colwell, 2000). More recently, molecular methods commonly applied to microbial and unicellular marine organisms have begun uncovering the diversity of viruses in marine metazoans.

The spatial and temporal distribution of marine microbial viruses may be restricted by the viruses' sensitivity to specific environmental conditions, and by the tolerance limits of their hosts (Mojica & Brussaard, 2014). Notably, while many viruses withstand high temperatures, such conditions may alter their structures and be unfavorable for viral proliferation. The distributions of metazoan viruses also respond to environmental conditions. In the case of the fatal blue crab pathogen, *Callinectes sapidus* reovirus 1 (CsRV1), prevalence is higher in temperate regions compared to tropical or subtropical areas (Zhao et al., 2020), yet in culture, CsRV1-related disease progression is accelerated by higher temperatures (Chung et al., 2015). These seemingly inconsistent observations underscore that the geographic distribution of viruses will be influenced by the life history and range of their hosts, as well as the physiological limits of the viruses themselves. Therefore, is it likely that while some crab viruses may be widely distributed and abundant, others may have a limited range.

This pattern is particularly evident in the case of two blue crab toti-like viruses, *Callinectes sapidus* toti-like virus 1 and *Callinectes sapidus* toti-like virus 2. These viruses were detected exclusively in northeastern samples from Massachusetts, Rhode Island, and New York, but were entirely absent from lower-latitude samples from Delaware, Maryland, and North Carolina, as well as from Gulf Coast samples from Louisiana and Texas (Zhao et al., 2022). This stark geographic disparity raises important questions about the ecological and environmental factors shaping viral distributions. One possibility is that these viruses have simply not yet been introduced to lower-latitude regions due to barriers in natural host migration or viral dispersal. Alternatively, abiotic factors such as temperature, salinity, or other oceanographic conditions may impose physiological constraints that limit their ability to establish and persist in these environments.

Anthropogenic behaviors can also directly affect the spread of viruses and, more generally, putative pathogens. In the case of Ostreid herpesvirus-1 (OsHV-1), a virus devastating populations of the Pacific oyster, *Crassostrea gigas*, research strongly suggests that the pathogen’s introduction to the U.S. Pacific Coast was via the ballast water in ships transporting oysters from France for human consumption (Paul-Pont et al., 2013). Furthermore, research has shown that CsRV1 is more abundant in waterways near crab soft-shedding flow-through aquaculture facilities rather than the waterways further away from human activity (Flowers et al., 2018). Previous research has also shown that a Louisiana strain of CsRV1 is more similar to strains from the Mid-Atlantic than to other strains from the Gulf of Mexico (Zhao et al., 2023). These findings are inconsistent with gene flow due to natural migration along the coast, and suggest that the interstate transport of blue crabs could be influencing the spread of a CsRV1 subtype.

Despite the significance of the fishery, the pathogens that infect blue crabs and potentially pose risks to their populations remain relatively understudied. Therefore, this study used a virus-like particle (VLP) enrichment protocol and high-throughput sequencing techniques to identify and characterize putative RNA virus genomes associated with the blue crab along the host's U.S. geographic range.

2 Methods

2.1 Crab sampling

Live crabs were obtained from commercial sources that could verify their origin. Crabs from Louisiana and Maryland were obtained from Blake's Crab House in Baltimore, MD; New York crabs from Two Cousins Fish Market in Freeport, Long Island, NY; Florida crabs from Hull's Seafood in Ormond Beach, FL; North Carolina crabs from E. Goodwin & Sons Seafood in Jessup, MD; finally, Texas crabs from Paul's Seafood in Corpus Christi, TX. Crabs were chilled and transported on ice packs to the Institute of Marine and Environmental Technology (IMET) in

Baltimore, Maryland. A total of 30 crabs from each state were dissected, for a total of 2.5 g of gill tissue per state.

2.2 Virus-like particle enrichment

The steps followed were a modified version of the protocol by Jiang and Wei (2018): https://www.protocols.io/view/isolation-of-viral-like-particles-vlp-from-tissues-dm6gp1xjgzpn/v2?step=12. The 2.5 g of gill tissues collected per state were combined into three subsamples of equivalent weight to simplify sample processing. Each subsample was homogenized on ice in 5 volumes of SB buffer (0.2 M NaCl, 50 mM Tris–HCl, 5 mM CaCl2, 5 mM MgCl2; pH 7.5) at 3000 rpm using a Brinkman/Polytron PT3000.

After homogenization, each sample underwent three freeze-thaw cycles using dry ice and 40 °C baths, followed by a final buffer addition to reach 12 volumes of SB buffer. The samples were centrifuged at 3,500 xg for 15 min at 4 °C. Subsequently, the supernatants were saved, and centrifuged again under the same conditions for 5 min. At this point, all supernatants were passed through a 0.45 um filter, resulting in a pooled filtrate. The virus filtrate was split evenly between six tubes and underlayed with ~28% (w/w) sucrose cushions. Finally, the tubes were ultracentrifuged at 30,000 rpm (154,426 xg) for two hours.

Following ultracentrifugation, pellets were resuspended in 382 uL of water, 45 uL 10x DNAse I buffer, and 22.5 uL DNAse I (2U/uL). The samples were mixed, incubated for an hour with shaking, and stored overnight at 4 °C.

2.3 Nucleic acid extraction

Extractions were performed using the Purelink™ Viral RNA/DNA Mini Kit (without carrier RNA). A Qubit™ RNA HS Assay Kit (Thermofisher) was used to quantify RNA concentrations, and the Bioanalyzer 2100 Eukaryote Total RNA Nano Kit was used to examine RNA quality (Agilent Technologies, CA, USA).

2.4 DNA wipeout, library construction, and sequencing

Samples were submitted to CD Genomics (https://www.cd-genomics.com/) for DNA wipeout, library construction, and sequencing. After a wipeout using DNAse I to remove DNA contamination, RNA samples were subjected to random shotgun library preparation using the NEBNext® UltraTM II Directional RNA Library Prep Kit for Illumina®. Briefly, RNA was sheared into ~200 bp fragments, and random hexamer primers were added, which hybridized to the RNA for reverse transcription. First and second cDNA strands were synthesized, and the resulting double-stranded RNA molecules were end-repaired to create blunt phosphorylated ends. A single adenine overhang was then incorporated into the 3' end to allow for the ligation of a hairpin loop-shaped NEBNext® adapter. The second strands were removed with the NEBNext USER® Enzyme, and a size selection for 150 bp fragments was performed. Finally, the six libraries were PCR amplified using Q5 High-Fidelity DNA Polymerase master mix and sequenced on the NovaSeq X platform.

2.5 Bioinformatic pipeline

2.5.1 Read filtering and contig assembly

Read overhangs and adapter sequences were trimmed, and any reads that mapped to the blue crab reference genome (Bachvaroff et al., 2021) were removed using BBDuk v39.01 from the Joint Genome Institute (*bbduk.sh rcomp=f hdist=1 minlength=800 ktrimtips=70 &* for adapter sequence trimming and *bbduk.sh ftr=149 k=40 rieb=f &* for overhang trimming and host removal). The cleaned reads from all states were pooled into one meta-assembly using rnaviralSPAdes v3.15.5 (Antipov et al., 2020).

2.5.2 Contig filtering and taxonomic assignment

Contigs less than 3,000 bp were removed, followed by the use of VIBRANT v1.2.1 (Kieft et al., 2020), DeepVirFinder v1.0 (Ren et al., 2020), viralVerify v1.1 (Antipov et al., 2020), and HMMER v3.3.2 employing the NeoRdRp database v1.1 (Sakaguchi et al., 2022) to identify viral-like contigs. Subsequently, these putative virus contigs were subjected to further analysis.

After cross-comparing sequences with blastn to confirm that each represented a unique entity within the dataset (<95% nucleotide similarity), the contigs were compared to NCBI's nr database, and the top blastx hit for each sequence was determined based on the highest bit-score followed by the lowest e-value. Ultimately, contigs with high similarity to non-viral sequences were filtered from the analysis.

To gather additional information on the top blastx hits of the putative virus contigs, TaxonKit v0.19.0 (Shen & Ren, 2021) was used to extract their taxonomic lineages from the NCBI Taxonomy database (*taxonkit lineage*). For top blastx hits that lacked order and/or family information in NCBI Taxonomy, provisional taxonomic classifications were assigned by referencing the literature (Shi et al., 2016; van der Loos et al., 2023). Specifically, publications describing these genomes were reviewed to determine their proposed phylogenetic placement. The NCBI-derived lineages and literature-based provisional classifications are hereafter referred to as the **related taxonomic classifications** (i.e., **related order** and **family**) of the contigs. Notably, the related taxonomic classifications were revised if subsequent analyses (e.g., phylogenetics) provided strong evidence supporting the assignment of the genome to a different taxonomic group.

2.5.3 Relative abundance of contigs

Here, two terms are used to describe relative abundance estimates of sequences. The SPAdes assembler calculates a k-mer coverage for each contig it assembles, which can be found in the sequence's unique ID (i.e., NODE_1_length_21095_cov_**102.649795**). Given our meta-assembly approach to this study, any reference of k-mer coverage refers to the coverage of a sequence across all locations sampled.

In the second method, transcripts per million (TPM) were calculated using RSEM (Li & Dewey, 2011) based on read mapping with bowtie2 v2.3.4.1 (*rsem-calculate-expression --bowtie2 --paired-end*) to measure the normalized relative abundance of each contig per state.

2.5.4 Genome completeness

The genome completeness of contigs was assessed using CheckV v1.5 (*checkv end_to_end*) (Nayfach et al., 2020). Briefly, CheckV compares input protein sequences to the CheckV genome database using AAI (average amino acid identity), and completeness is computed as the ratio between the contig length and the length of matched reference genomes. CheckV assigns low, medium, and high-quality to contigs with <50%, 50-90%, and >90% genome completeness, respectively. If a contig lacks a high- or medium- confidence estimate based on AAI, a more sensitive but less accurate HMM-based approach is used.

2.5.5 Identification of putative open reading frames (ORFs) and functional domains

To identify likely open reading frames (ORFs) in the predicted viral contigs, the GeneMark.hmm model v3.25 (https://exon.gatech.edu/heuristic_gmhmmp.cgi) was used with the settings "genetic code 1" and "Heuristic parameters", as in Besemer and Borodovsky (1999). Putative functional domains within these ORFs were then identified by using HMMscan v3.1b2 to compare the ORF sequences to the Pfam database (Pfam-A.hmm). To ensure the identification of high-confidence regions, e-value thresholds for all domains were set to <1e-3, while independent e-value thresholds for polymerases and capsid proteins were set to <1e-2 and <1, respectively. In the case of the identification of domains that were fully nested within larger domains, the larger domain was reported. In the case of putative domains that partially overlapped but had Pfam hits to domains with the same biological function, the two domains were combined and the larger region was reported. Data visualization was performed in RStudio using the 'ggplot2' and 'gggenes' packages.

2.5.6 Phylogenetic analysis

As an additional line of evidence to support the provisional assignment of genomes into two of the most well-characterized RNA virus taxonomies—the Mononegavirales and Picornavirales—phylogenetic analysis of high-quality genomes with a k-mer coverage >10 was performed. The phylogenetic trees were constructed using protein sequences characteristic of the specific viral orders. For the Mononegavirales tree, RdRp/L protein sequences from the NCBI

Protein Database were aligned using the COBALT Multiple Alignment tool (Papadopoulos and Agarwala, 2007). The phylogenetic tree was built using the IQTREE web server with 'Protein' set as the sequence type, Branch support analysis set to 'Ultrafast,' and the number of Bootstrap alignments and maximum iterations set to 1,000 (http://iqtree.cibiv.univie.ac.at/). The tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Similarly, the Picornavirales tree was constructed using non-structural regions of known picornaviruses and picorna-like viruses identified in metagenomic studies (van der Loos et al., 2023, Shi et al., 2016, Guo et al., 2023). The protein sequences were aligned using COBALT, high dissimilarity regions were trimmed using Mesquite v3.81, and the phylogenetic tree was built using IQTREE and visualized using FigTree.

2.5.7 Assessing the novelty of high-quality putative virus genomes

To most accurately report the novelty of the high-quality putative virus genomes, further analyses were performed. These genomes were considered novel if there was <90% amino acid similarity across all of their ORFs to proteins of viruses in the NCBI nr database.

2.5.8 Assessing the relatedness of high-quality putative virus genomes

Putative virus genomes were compared using blastp to assess whether any represented variants of the same species. Genomes were considered to belong to distinct species if there was <90% amino acid similarity across all of their ORFs.

3 Results

This study employed RNA sequencing to identify and characterize viruses associated with commercially harvested blue crabs collected from six states along the U.S. Atlantic Coast. From

180 crabs, a total of 387,029,192 reads were sequenced, with each state accounting for more than 40 million reads **(Supplementary Table 2.1)**. The raw sequence reads are publicly available and can be found here: bioproject/PRJNA1259537. Mapping the read data to the blue crab genome demonstrated that 49.7% to 91.5% of reads from each state mapped to the blue crab host. Following the removal of host reads, the remaining 93,892,582 reads assembled into 64,192 SPAdes contigs of which 4,956 were over 500 bp with an N50 of 1,017 bp.

3.1 Identification and validation of putative virus genomes

Following assembly, contigs were filtered to prioritize sequences most likely representing complete or near-complete viral genomes. Filtering criteria consisted of a CheckV validation step and a 3,000 bp threshold. This resulted in 156 contigs ranging from 3,014 to 21,095 bp. Of these, 97 were identified as putative virus genomes through viral identification programs, domain-based annotation, and comparison with existing metagenomic datasets. Detailed information about these 97 contigs (e.g., top blastx hits, percent identity values, checkV qualities, and taxonomies) can be found in **Supplementary Table 2.2**. Notably, no RNA transcripts of DNA viruses were identified.

The classification of a contig as "virus-like" varied considerably across the four bioinformatic programs employed, with each contig being detected as such by one to three programs. Additional lines of evidence supported that these 97 contigs belonged to viruses, based on their amino acid similarity to proteins of known viruses in the BLAST nr database, and, in many cases, the presence of putative virus protein domains (e.g., RNA-dependent RNA polymerases and viral methyltransferases) within their predicted open reading frames.

3.2 Overview of putative virus genomes

3.2.1 Viral taxonomic assignment

The 97 contigs exhibited 21.1% to 99.8% amino acid similarity to (1) proteins of viruses recognized by the International Committee on Taxonomy of Viruses (ICTV) or (2) proteins derived from viral genomes identified in existing datasets. Notably, 22 contigs had top blastx hits to proteins of the Picornavirales. In contrast, more than half (N = 58) showed blastx similarity to viral sequences lacking order classifications in NCBI Taxonomy. However, a review of the literature showed that 27 of these 58 contigs clustered phylogenetically within the Picornavirales. Therefore, in total, 49 of the 97 putative viral genomes were assigned to the Picornavirales related order. Of the remaining 48 contigs, 29 were assigned to 10 other related orders, and eight additional contigs were provisionally assigned to 'order-like' categories due to their resemblance to virus families spanning two distinct orders (Shi et al., 2016). With the related orders of 86 of the putative virus genomes accounted for, only 11 contigs remained unclassified at the order level. Classifications are illustrated in **Figure 2.1**, where each circle represents an assembled genome, circle size corresponds to nucleotide length, and facet labels indicate related order classifications. The x-axis represents k-mer coverage, while the y-axis reflects percent identity to the top blastx hit. As previously described in **Section 2.5.2** of this chapter, related orders derive primarily from NCBI lineages and literature-based provisional classifications of the top blastx hits; therefore, they do not constitute formal taxonomic assignments of the contigs but rather indicate phylogenetic relatedness based on sequence identity.

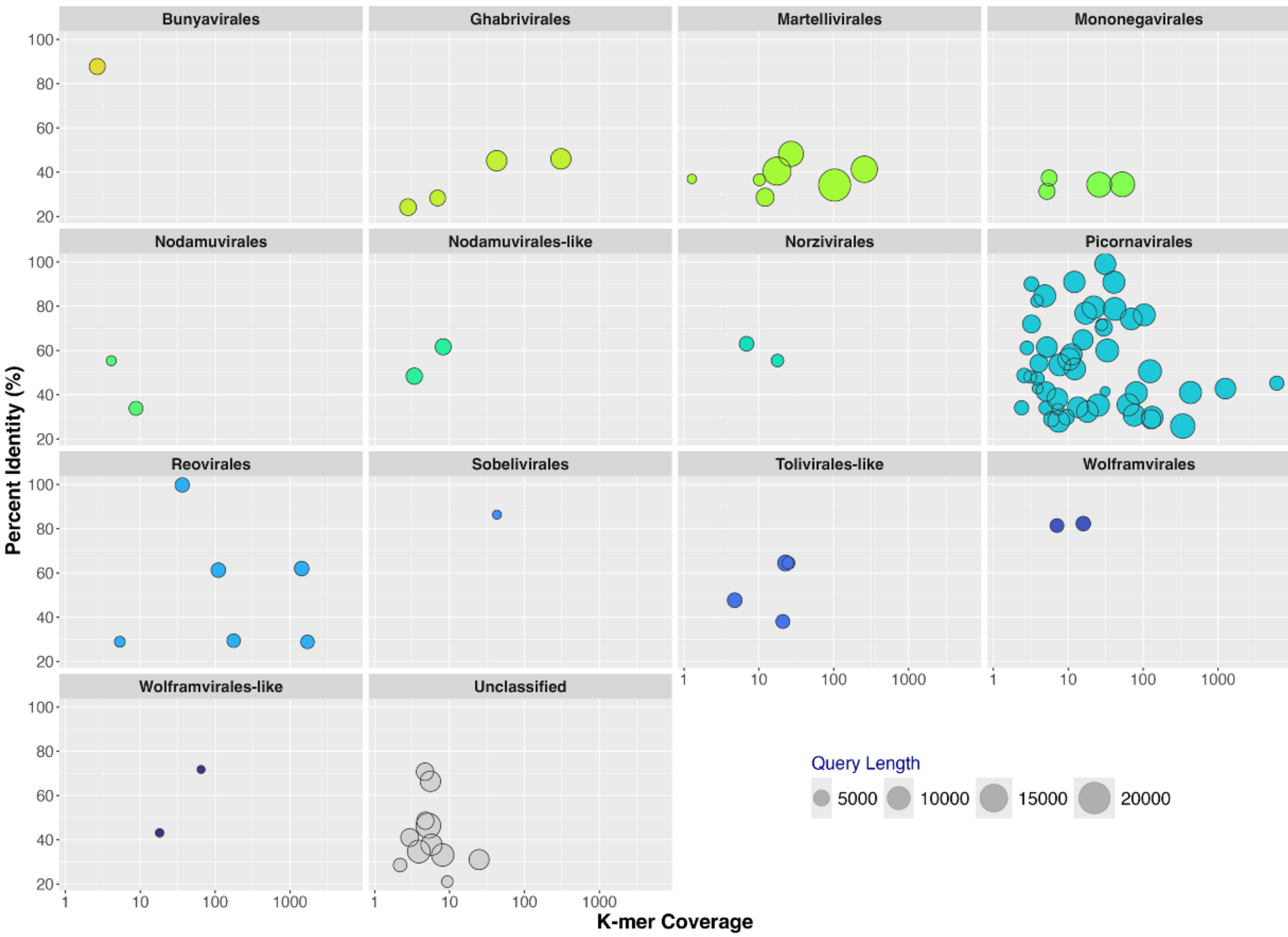


Figure 2.1. Related order classifications of 97 contigs. Each circle represents an assembled genome, with circle size corresponding to nucleotide length and facet labels indicating related order classifications. The x-axis represents k-mer coverage, while the y-axis reflects percent identity to the top blastx hit.

3.2.2 Completeness of putative virus genomes

Of the 97 putative virus contigs, 43 were classified as high-quality (>90% complete), 25 as medium-quality (50-90% complete), 18 as low-quality (<50% complete), and 11 as undetermined based on comparison to reference genomes in the CheckV database **(Supplementary Table 2.2)**.

3.2.3 High-quality putative virus genomes with a k-mer coverage >10

To emphasize the most abundant of the 43 high-quality putative virus genomes, our analysis hereinafter focused on those with a k-mer coverage >10. This resulted in 28 genomes that were provisionally named ***Callinectes*-associated viruses #1-28 (Figure 2.2)**, as well as the identification of a known virus, ***Ulva* picorna-like virus 1** (van der Loos et al., 2023). The completeness estimates of these genomes ranged from 93.11% to 100% and SPAdes k-mer coverages from 10.2 to 429.1. The taxonomic assignment of the 29 top blastx hits supported assigning related order classifications for 19 genomes to the Picornavirales, for two genomes each to the Mononegavirales and Tolivirales-like, and for one genome each to the Martellivirales, Norzivirales, Sobelivirales, Wolframvirales, Wolframvirales-like, and Ghabrivirales. Of these, all are putative positive-sense single-stranded RNA (+ssRNA) viruses, with the exception of two negative-sense single-stranded RNA (-ssRNA) genomes grouped into the Mononegavirales and one double-stranded RNA (dsRNA) genome grouped into the Ghabrivirales.

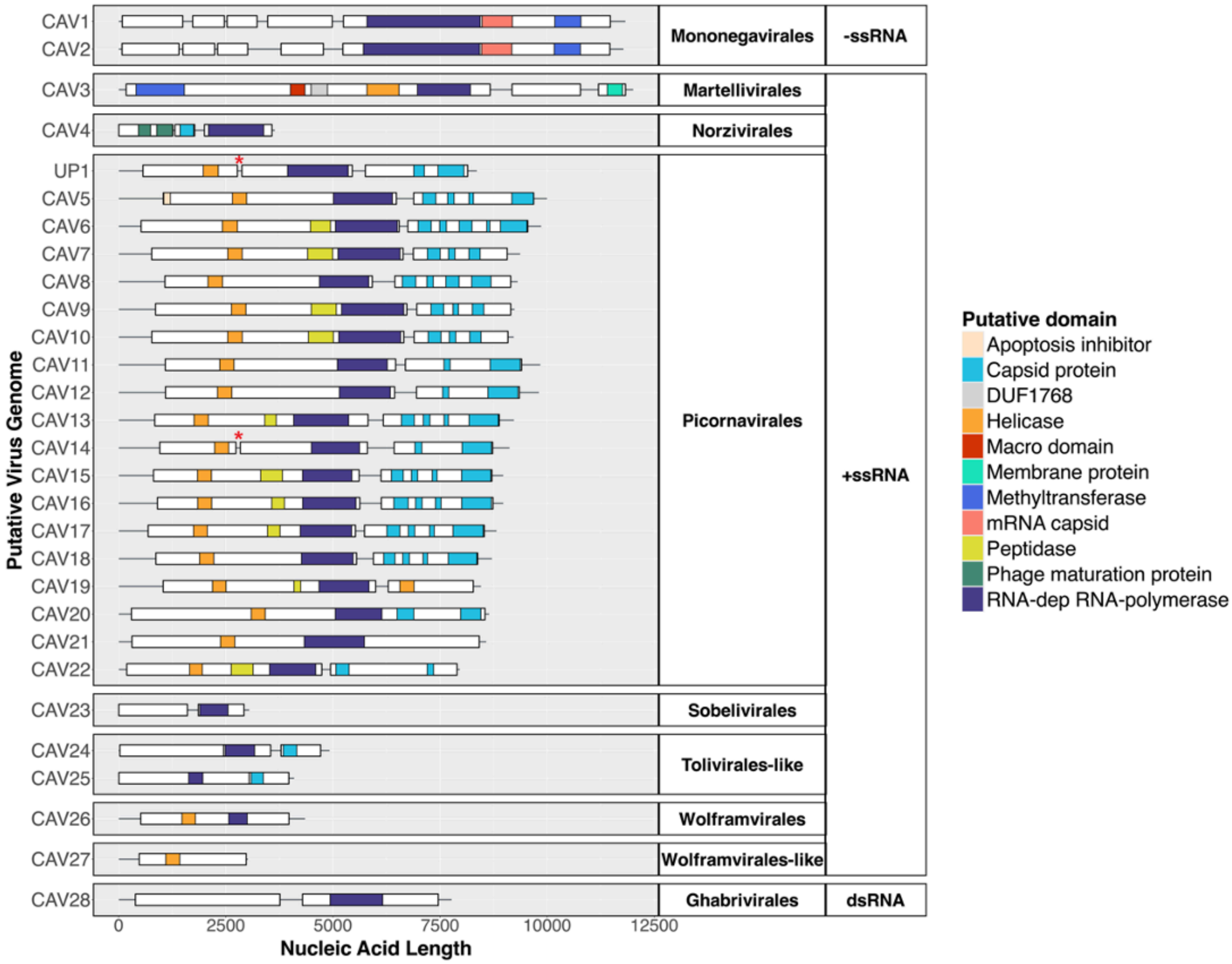


Figure 2.2. Genomic architecture of high-quality putative virus genomes (k-mer coverage >10) identified from blue crab gill RNA. A total of 28 previously uncharacterized virus genomes were provisionally designated ***Callinectes*-associated viruses #1-28 (CAV1-28)**, alongside the detection of a known virus, ***Ulva* picorna-like virus 1 (UP1)**. Genomes were categorized into negative-sense, positive-sense, or double-stranded RNA virus groups based on similarity to sequences in the NCBI nr database. They were further organized into their related order classifications. Each horizontal grey line represents a putative genome, with superimposed white rectangles denoting predicted ORFs. Colored segments within ORFs indicate regions with similarity to Pfam domains, as annotated in the legend. Gaps in sequence data disturbing ORF prediction in **CAV14** and **UP1** are annotated with red asterisks.

3.3 High-quality putative -ssRNA virus genomes

3.3.1 Two putative viruses of the Mononegavirales

Two novel high-quality putative -ssRNA viruses—hereinafter referred to as ***Callinectes*-associated virus 1 (CAV1)** and ***Callinectes*-associated virus 2 (CAV2)**—were discovered. Five open reading frames (ORFs) were identified within each genome, with the ORF closest to the 5' end (ORF1) displaying limited similarity (55–65% query coverage; 30.5–33.1% identity) to the nucleoprotein (N protein) of *Dicrocoelium* rhabdo-like virus 2 (WFD53217). Similarly, the ORF nearest the 3' end (ORF5) of both genomes displayed limited similarity (98–99% query coverage; 34.5% identity) to the Large protein (L protein) of the same virus (WFD53221). Genome schematics of **CAV1** and **CAV2** are illustrated in **Figure 2.3** and the top blastx hits for the ORFs are shown in **Supplementary Tables 2.3a and b**.

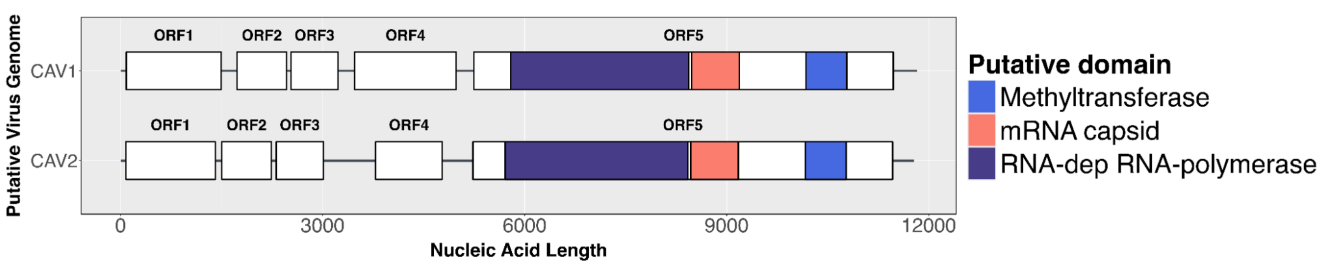


Figure 2.3. Genomic architecture of two putative Mononegavirales genomes, CAV1 and CAV2. Amino acid identities between the analogous ORFs of **CAV1** and **CAV2** are as follows: ORF1 were 59.5% identical; ORF2 53.8%; ORF3 62.7%; ORF4 68.7%; ORF5 78.6%. Based on comparison of the ORFs to that of known Mononegavirales, the sequenced contigs are believed to be the anti-genomes of negative-sense single-stranded RNA viruses.

Comparisons with the Pfam database identified putative domains in ORF5 of both viruses associated with mRNA capping (PF14318.5), RNA-dependent RNA polymerase activity (PF00946.18), and viral methyltransferase function (PF14314.5)—features characteristic of the Mononegavirales L protein. Ultimately, blastx similarity to viruses of the Mononegavirales and the presence of putative Mononegavirales domains, paired with a phylogenetic analysis of both ORF5 sequences, supports the classification of **CAV1** and **CAV2** as related viruses within the Mononegavirales and firmly within the Rhabdoviridae **(100% bootstrap support; Figure 2.4)**. Furthermore, a comparison of the ORFs of **CAV1** and **CAV2** supports the assignment of these genomes as two distinct viruses, as even the ORF with the highest similarity between the two viruses (ORF5) showed limited amino acid identity (78.61%).

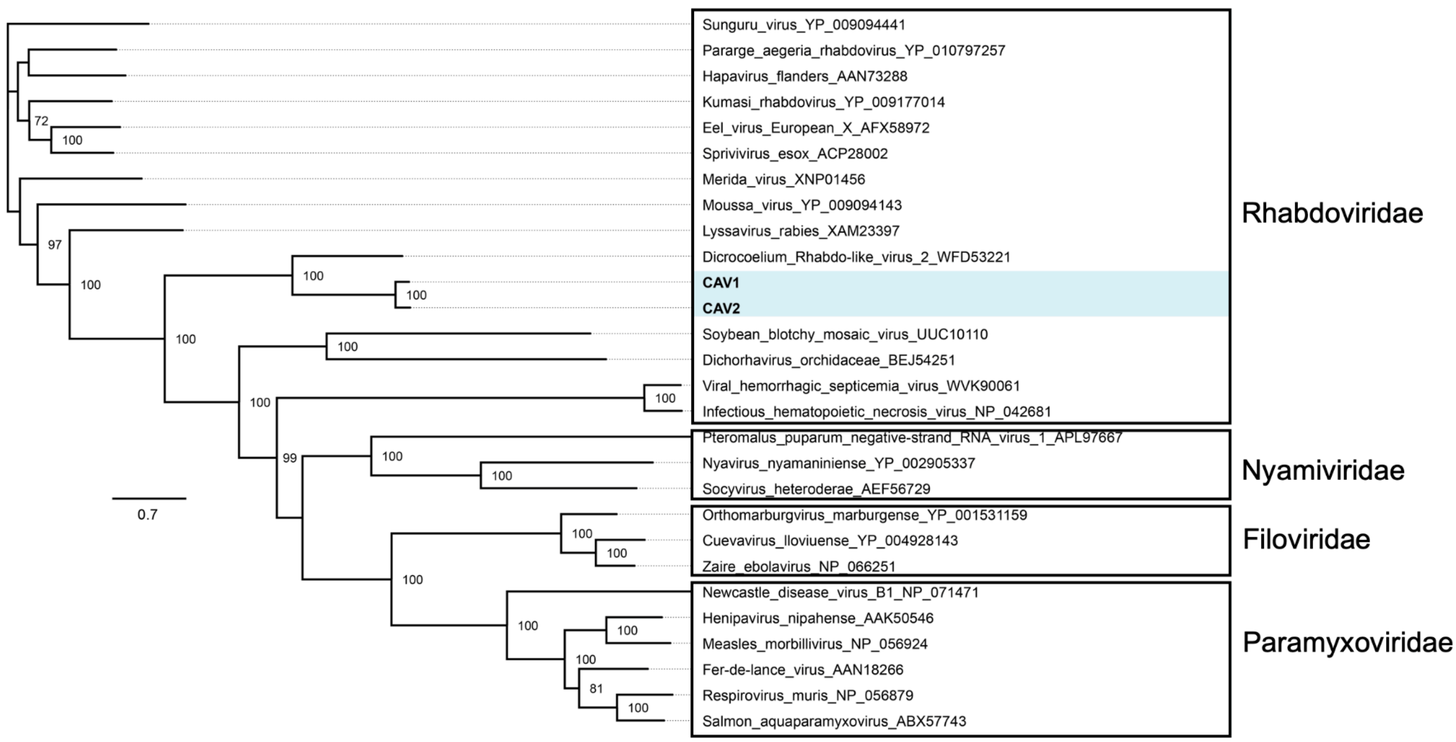


Figure 2.4. Phylogenetic placement of CAV1 and CAV2 within the Mononegavirales and Rhabdoviridae. Alignment of L protein sequences of the Mononegavirales, including the two putative Mononegavirales genomes identified in this study (highlighted in light blue). The tree supports the assignment of the two related yet distinct novel genomes into the Mononegavirales and Rhabdoviridae (100% bootstrap support).

3.4 High-quality putative +ssRNA virus genomes

3.4.1 Putative viruses of the Picornavirales

A majority (19/25) of the high-quality putative +ssRNA virus genomes identified in this study had top blastx hits to (1) proteins of viruses belonging to the Picornavirales or (2) proteins of viruses that clustered phylogenetically within the Picornavirales based on the literature review discussed in **Section 2.5.2** of this chapter. The 19 contigs ranged from 7,960 to 9,997 bp and varied considerably in their similarity (33.4% to 99.1%) to viral proteins in publicly available databases **(Supplementary Table 2.2)**.

Fifteen of the 19 contigs shared blastx similarity to viral proteins from the Dicistroviridae and Marnaviridae. Four remained as ‘Unclassified’ Picornavirales because (1) their top blastx hits were to sequences for which no family information was available or could be derived from the literature, and (2) blastp and phylogenetic analyses did not provide a clear classification into a specific family.

Consistent with the ICTV categorization of the Picornavirales, these putative viruses had monopartite or bipartite genomes, all encoding putative non-structural proteins such as RNA helicases and RNA-dependent RNA polymerases, and most also encoding RNA peptidases (ICTV 9th Report 2011). In addition, all genomes, with the exception of two, encoded putative structural (capsid) proteins. Ultimately, six genomes were classified into the Dicistroviridae, nine into the Marnaviridae, and four were left 'Unclassified' to a family **(Figure 2.5).**

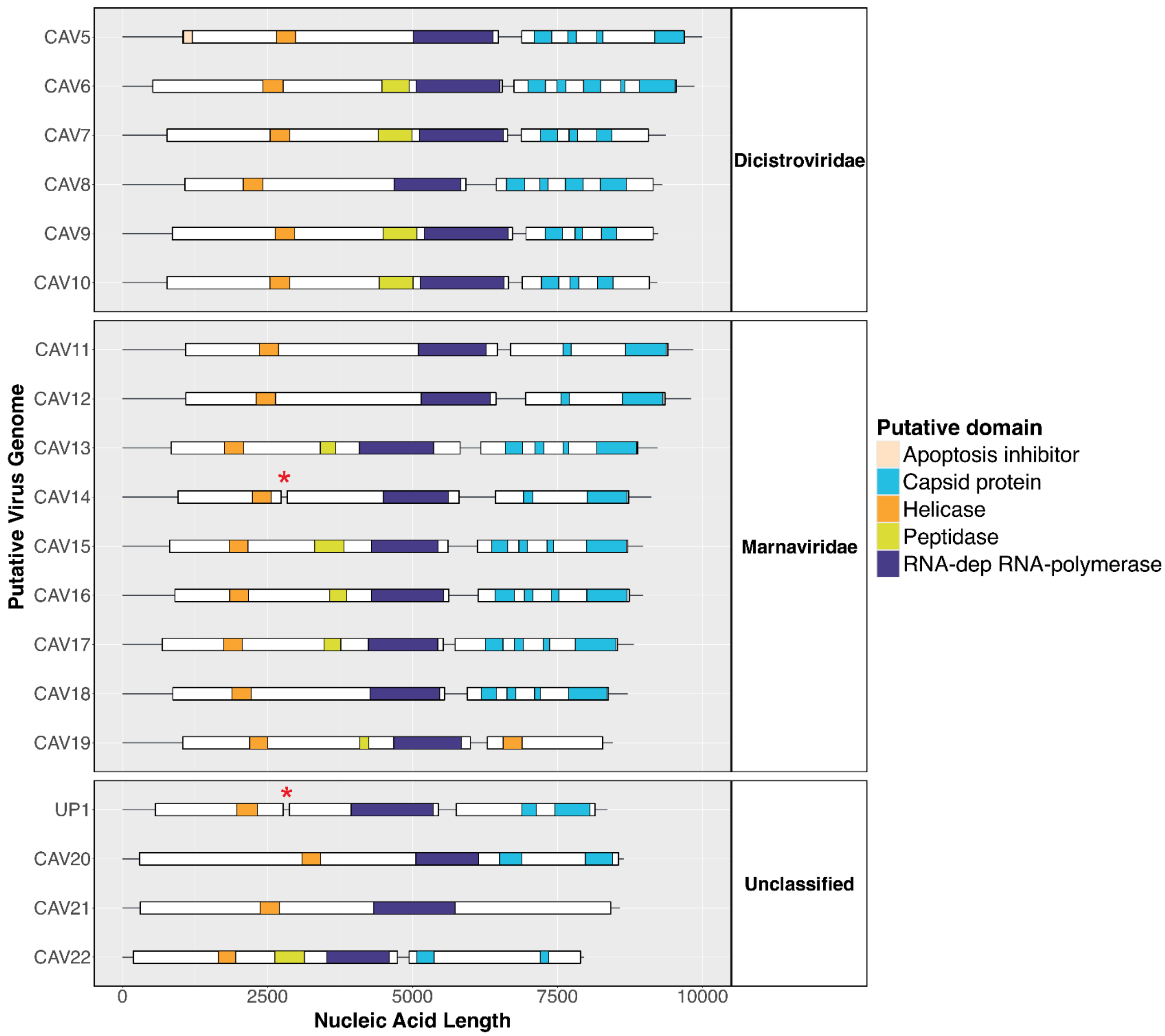


Figure 2.5. Genomic architecture of high-quality putative Picornavirales. A total of 19 high-quality contigs (k-mer coverage >10) were provisionally assigned to the Picornavirales related order. One was virtually identical to **Ulva picorna-like virus 1 (UP1)**. Additionally, **CAV21** shared 91% amino acid identity with the hypothetical protein of Beihai picorna-like virus 107 (YP_009333563). The remaining 17 were novel putative virus genomes. Genomes were further categorized into related families. As in **Figures 2.2** and **2.3**, each horizontal grey line represents a putative genome, with superimposed white rectangles denoting predicted ORFs. Colored segments within ORFs indicate regions with similarity to Pfam domains, as annotated in the legend. Gaps in sequence data interrupting ORF predictions in **CAV14** and **UP1** are annotated with red asterisks.

3.4.1.1 Putative viruses of the Dicistroviridae

Within the Picornavirales related order, **CAV5–CAV10** displayed similarity to the Dicistroviridae. These genomes each contained two ORFs, with the ORFs nearest the 5' ends encoding putative non-structural proteins (PF00910.21; PF00548.19; PF12381.7; PF00680.19) and those nearest the 3' ends encoding three to five capsid proteins (PF11492.7; PF00073.19; PF00915.19; PF08762.9).

Of these sequences, **CAV9** shared limited similarity (38% query coverage; 41.0% identity) with a non-structural protein (WBY66453) derived from *Eriocheir sinensis* dicistrovirus 2—an ICTV-recognized member of the Dicistroviridae. The five remaining genomes displayed amino acid similarity to proteins of viruses discovered through metagenomics that, although unclassified at the order or family level, cluster phylogenetically within the Dicistroviridae (Shi et al., 2016), including Wenling picorna-like virus 3, Beihai picorna-like virus 70, Wenzhou shrimp virus 5, and Beihai picorna-like virus 82.

To further support the assignment of **CAV5-CAV10** within the Dicistroviridae, phylogenetic analysis of the non-structural region was conducted using recognized and likely members of this order, as well as of the Marnaviridae and the Picornaviridae **(Figure 2.6)**. This analysis supported the placement of **CAV5–CAV10** within the Picornavirales and the Dicistroviridae (95% bootstrap support). Furthermore, to assess the novelty of these genomes, the predicted proteins of **CAV5-CAV10** were compared to those of publicly available viruses in the NCBI nr database. The amino acid similarities between the proteins of **CAV5-CAV10** and their top blastp hits were consistently below 66.14%, suggesting the identification of novel viruses (refer to **Section 2.5.7** of this chapter).

To determine whether **CAV7**, **CAV9**, and **CAV10** represented variants of the same virus, pairwise comparisons of ORF1 and ORF2 were conducted. ORF1 and ORF2 of **CAV7** and **CAV9** shared 83.0% and 87.9% identity, respectively; ORF1 and ORF2 of **CAV7** and **CAV10** shared 87.0% and 92.2% identity, respectively; ORF1 and ORF2 of **CAV9** and **CAV10** shared 85.5% and 88.1% identity, respectively. These results suggest that **CAV7**, **CAV9**, and **CAV10** represent related yet distinct viral species (refer to **Section 2.5.8** of this chapter).

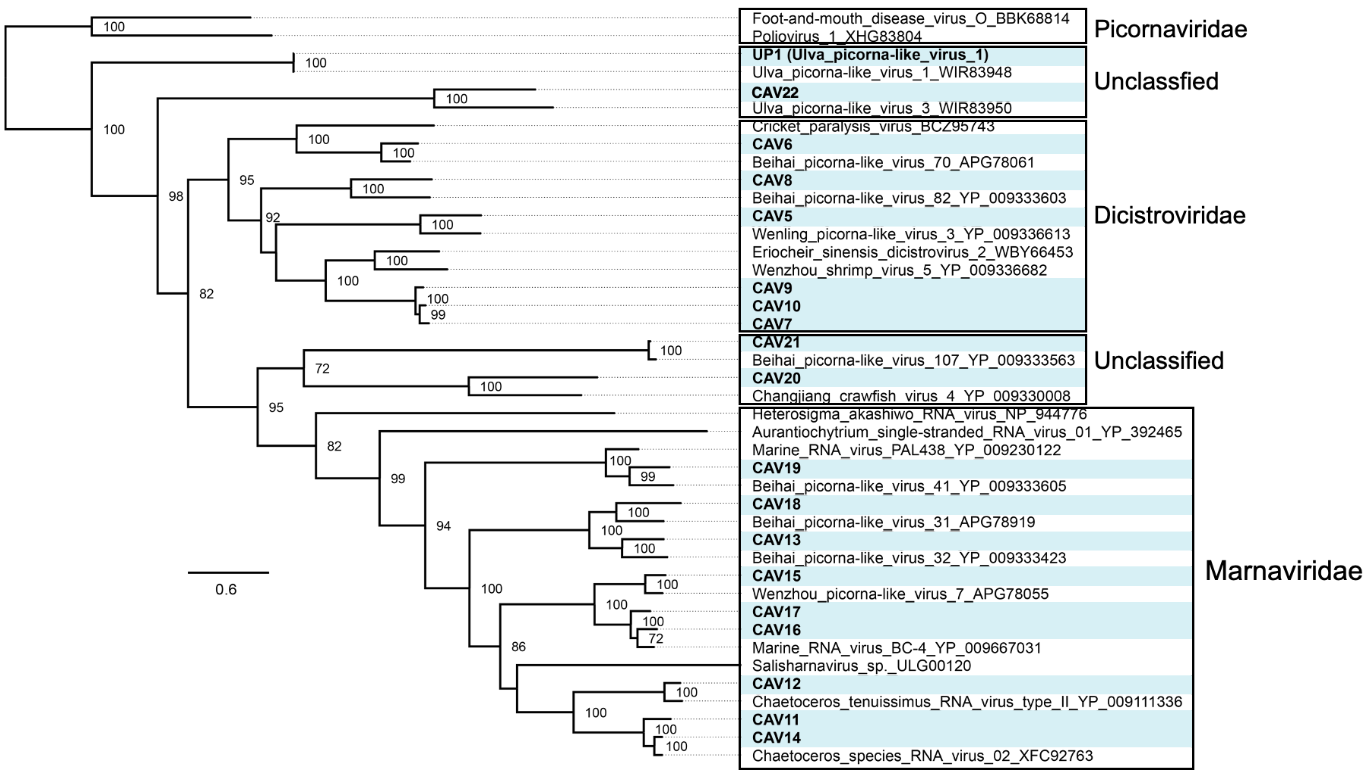


Figure 2.6. Phylogenetic placement of putative Picornavirales genomes within their related families. Alignment of the putative non-structural proteins of the Picornavirales identified in this study (highlighted in light blue) with those of picornaviruses and picorna-like viruses. Bootstrap values of 72-100 support the assignment of the novel genomes into the Picornavirales and within their respective related families.

3.4.1.2 Putative viruses of the Marnaviridae

Putative viruses of the Marnaviridae consisted of two ORFs, with the ORFs nearest the 5' ends encoding putative non-structural proteins (PF00910.21; PF00548.19; PF12381.7; PF00680.19) and those nearest the 3' ends encoding three to four capsid proteins (PF11492.7; PF08762.9; PF00073.19). The exception was **CAV19**, where the 3’ ORF encodes a putative RNA helicase.

Five genomes—hereinafter referred to as **CAVs #11-12, 14,** and **16-17**—had top blastx hits to proteins of viruses officially classified within the Marnaviridae, including to a non-structural polyprotein in Marine RNA virus BC-4 (YP_009667031), a predicted association protein in *Chaetoceros tenuissimus* RNA virus type II (BAP99820), and a putative replication related protein in *Chaetoceros* species RNA virus 02 (YP_010084314). Placement of these five genomes within the Marnaviridae was supported by phylogenetic analysis of the non-structural region (**Figure 2.6; 82% bootstrap support)**.

A literature review revealed that the top blastx hits of four additional genomes—hereinafter referred to as **CAVs #13, 15, 18,** and **19**—grouped into an operational “Aquatic picorna-like cluster” within the Picornavirales (Shi et al. 2016). Furthermore, phylogenetic analysis showed clustering of the predicted non-structural proteins with protein sequences from known viruses of the Marnaviridae, including *Heterosigma akashiwo* RNA virus (NP_944776), *Aurantiochytrium* single-stranded RNA virus (YP392465), and *Chaetoceros* species RNA virus 02, supporting the placement of **CAVs #13, 15, 18** and **19** within the Marnaviridae **(Figure 2.6; 82% bootstrap support)**.

Finally, pairwise amino acid comparison between the ORFs of **CAV11-CAV19,** as well as to the proteins of known viruses, suggested that all the putative Marnaviridae genomes identified in this study represent novel and distinct viruses.

3.4.1.3 Putative 'Unclassified' Picornavirales

Several putative Picornavirales genomes identified in this study—hereinafter referred to as **CAVs #20-22** and **UP1**—could not be provisionally assigned to a family and were thus categorized as 'Unclassified' **(Figure 2.6)**. **CAV20** and **CAV21** had top blastx hits to hypothetical proteins in Changjiang crawfish virus 4 (YP_009330008) and Beihai picorna-like virus 107 (YP_009333563), viruses discovered through metagenomics that clustered outside of the Dicistroviridae and the "Aquatic Picorna-like Cluster," without clear assignment to a specific family (Shi et al., 2016). Furthermore, phylogenetic analysis of **CAV20** and **CAV21** did not yield a clear family classification, but did suggest possible clustering with viruses of the Marnaviridae (95% bootstrap analysis). For these reasons, **CAV20** and **CAV21** remained 'Unclassified.' Notably, these were also the only picorna-like genomes described in this study with one predicted ORF and **CAV21** had high similarity (94% query coverage; 91% identity) to Beihai picorna-like virus 107 (YP_009333563), representing a potential variant.

The **UP1** and **CAV22** genomes exhibited top blastx similarity to the polyproteins of *Ulva* picorna-like virus 1 (WIR83948) and *Ulva* picorna-like virus 3 (WIR83950), respectively (van der Loos et al., 2023). All ORFs of **UP1** were nearly identical to regions of the polyprotein of *Ulva* picorna-like virus 1 (100% query coverage; 99.1-99.9% identity), confirming the presence of a known virus within the data. On the other hand, the limited amino acid similarity between the putative proteins of **CAV22** and its top ORF hits supports its classification as a novel virus genome. Phylogenetic analysis of **UP1** and **CAV22** did not clearly support their placement within the Dicistroviridae **(Figure 2.6)**. As a result, they were listed as 'Unclassified' within the Picornavirales related order.

3.4.2 Other positive single-stranded RNA virus genomes

Of the 26 putative +ssRNA virus genomes identified in this study, there were seven that were not assigned to the Picornavirales related order. Instead, one genome each was assigned to Martellivirales, Norzivirales, Sobelivirales, Wolframvirales, and Wolframvirales-like related order classifications. Two additional genomes were categorized within the Tolivirales-like related order **(Figure 2)**.

The genome with similarity to the Martellivirales—hereinafter referred to as **CAV3**—was 12,007 bp in length and encoded three ORFs. The most 5’ ORF had Pfam hits to regions encoding a putative viral methyltransferase (PF01660.16), a Macro domain (PF01661.20), a domain of unknown function (PF08719.10), a viral RNA helicase (PF01443.17), and an RNA-dependent RNA polymerase (PF00978.20). ORF2 had no Pfam hits that met our filtering criteria while ORF3 (nearest the 3’ end) encoded a putative membrane protein (PF16504.4).

The genome in the Norzivirales related order—hereinafter referred to as **CAV4**—was much smaller than most genomes described in this study. This 3,642 bp genome contained three ORFs, with Pfam hits encoding phage maturation proteins (PF03863.12), a Levivirus coat (capsid) protein (PF01819.16), and an RNA-dependent RNA polymerase (PF03431.12). **CAV4**’s top blastx hit was to the RdRp of a ssRNA bacteriophage (YP_010769536) that has been formally classified within the Norzivirales known as SRR6960802 (43% query coverage; 55.4% identity).

The genome with similarity to the Sobelivirales—hereinafter referred to as **CAV23**—was 3,044 bp in length and consisted of two ORFs. While the first ORF did not have any Pfam hits that met our filtering criteria, ORF2 had a hit encoding a putative RNA-dependent RNA polymerase (PF00680.19).

CAV26 was assigned to the Wolframvirales related order due to its blastx similarity to a putative RNA-dependent RNA polymerase of *Narnaviridae* sp., a member of the Wolframvirales (UJQ92779). On the other hand, **CAV27** was provisionally assigned to the Wolframvirales-like related order, due to the resemblance of its top blastx hit to both the Narnaviridae and Leviviridae, but closer phylogenetic relationship to the Leviviridae within the Wolframvirales (YP_009337422). **CAV26** was 4,348 bp in length and consisted of one ORF with Pfam hits to domains encoding a putative helicase (PF00910.21) and RdRp (PF05919.10). **CAV27** also only encodes one ORF but is much smaller at 3,014 bp. This genome encodes only a putative RNA helicase domain (PF00910.21), making it the only one among the 29 high-quality genomes with a k-mer coverage >10 that lacked a detectable RdRp domain **(Figure 2.2)**.

Finally, the **CAV25** and **CAV24** genomes were 4,093 and 4,919 bp, respectively. **CAV25** consisted of two ORFs, encoding a putative RdRp (PF00680.19) and a putative capsid protein (PF00729.17). While the top blastx similarity of the full **CAV25** genome was to a Picornavirales protein (QJI53534), ORF1 resembled a partial hypothetical protein of XiangYun tombus-noda-like virus 7 (UUG74216) while ORF2 resembled hypothetical protein 3 in Wenling tombus-like virus 4 (YP_009337116). Therefore **CAV25** was classified into the Tolivirales-like related order rather than the Picornavirales.

The **CAV24** genome consisted of three ORFs, with ORF1 having no Pfam hits to putative domains, ORF2 encoding an RdRp (PF00998.22), and ORF3 encoding a capsid protein (PF00729.17). The top blastx hit of this putative virus genome was to a hypothetical protein of Wenzhou tombus-like virus 18 (YP_009342270).

3.5 A high-quality putative dsRNA virus genome

One putative dsRNA virus—referred to as **CAV28**—was identified in this study. The 7,769 bp genome consisted of two ORFs, with the 3' ORF encoding a putative RNA-dependent RNA polymerase (PF00910.21). Given the limited amino acid similarity of its ORFs to hypothetical protein 1 (APG76059.1; 23% query coverage; 26.52% identity) and hypothetical protein 2 (APG76060.1; 78% query coverage; 47.47% identity) of Wenling toti-like virus 1, this novel virus was provisionally assigned to the Ghabrivirales related order.

3.6 State-by-state abundances of high-quality putative virus genomes

To assess the geographic distribution of the 29 high-quality putative virus genomes (k-mer coverage >10), transcripts per million (TPM) values—normalized metrics of relative abundance—were calculated using RSEM. TPM values for each genome were designated as TPM_FL, TPM_LA, TPM_MD, TPM_TX, TPM_NC, and TPM_NY, corresponding to samples from Florida, Louisiana, Maryland, Texas, North Carolina, and New York, respectively.

As illustrated in **Figure 2.7**, the TPM values (i.e., TPM_FL, TPM_LA, etc.) for the 29 genomes ranged from 0 to 414.64, with a cumulative total (TPM_total) of 1,055.42 across all states. The most abundant genome was **CAV9**, which had a TPM_FL of 414.64 and was largely absent in other states **(Figure 2.7)**. The minimum, maximum, and total TPM values of the 29 genomes by state are summarized in **Table 2.1**.

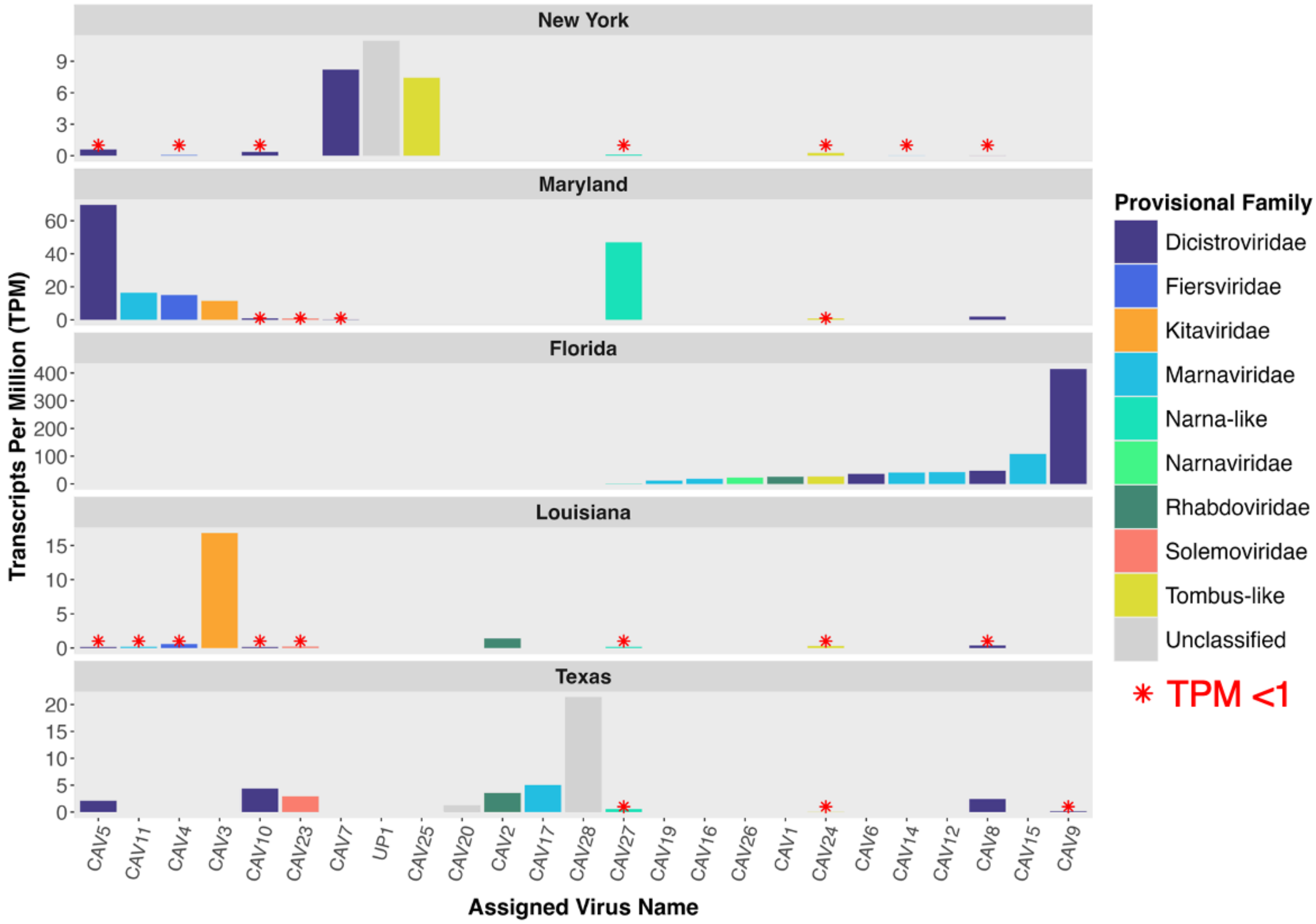


Figure 2.7. State-by-state abundances of high-quality putative virus genomes. The x-axis displays the assigned virus acronyms, and the y-axis represents relative abundance in transcripts per million (TPM). Provisional virus family assignments are represented by the color of the bars. None of the 29 high-quality genomes were detected in North Carolina; therefore, data from that state are not shown. TPM values <1 are represented by red asterisks. Genomes with TPMs <1 across all of the states (**CAV13**, **CAV18**, **CAV21**, and **CAV22**) are excluded from the figure.

Table 2.1. Summary of TPM values of 29 high-quality putative virus genomes by state.

State	Minimum TPM	Maximum TPM	Total TPM
Florida	0	414.64	795.7
Louisiana	0	16.82	20.22
Maryland	0	69.61	164.11
Texas	0	21.38	47.27
New York	0	10.95	28.12
North Carolina	0	0	0*
Cumulative Total (TPM_total)	—	—-	**1,055.42**

*No viral copies of the 29 genomes were detected in North Carolina.

3.7 Partial putative virus genomes

Up to this point, the analysis has focused on the complete or near-complete genomes assembled in this study. However, the sequence data include 16 distinct partial genomes with blastx similarity to the polyprotein of one virus in the database, *Macrobrachium rosenbergii* virus 7, a virus identified in a virome study of the giant freshwater prawn (Zhou et al., 2022). Many of the 16 contigs (also referred to as NODES in **Figure 2.8**), are highly abundant, but were below 3,000 bp in length so did not meet our filtering criteria for detailed analysis. Nevertheless, the sheer abundance and geographic range of these contigs deserves mention. For brevity, we refer to these as MR7-like contigs.

A state-by-state breakdown of relative abundances (TPMs) shows that the highest coverage of MR7-like contigs was in the Florida crab sample. Notably, when considering all viral-like contigs ≥1000 bp, 89% of the TPM in Florida was accounted for by contigs with best blastx hits to the polyprotein of MR7. Furthermore, almost all of the sequences that make up the 89% can be traced back to four contigs (NODES 343, 225, 191, and 245), each representing an abundant (aggregate TPM of 35,186 to 41,091), overlapping, yet partial, putative virus genome. These same four contigs were also similarly abundant in the Texas sequence data, less so in Maryland and Louisiana, and nearly undetected in North Carolina.

NODES 343, 225, 191, and 245 had limited blastx similarity to MR7: 1) From the available contigs lengths of 1,907 to 2,647 nt, the lengths of alignments to the MR7 genome were 142 to 879 nt, and 2) the identity to the MR7 polyprotein sequence was from 43.9% to 61.3%. The limited alignment to known viruses precluded additional analysis as conducted with the 29 near-complete genomes.

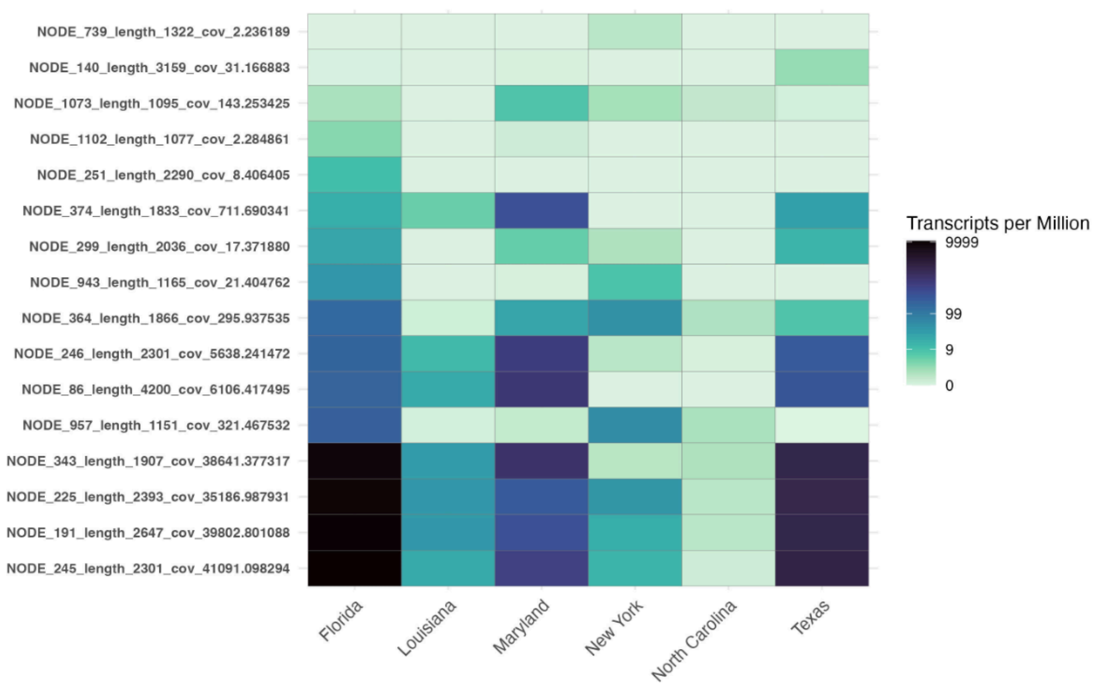


Figure 2.8. State-by-state abundances of partial putative virus genomes with blastx similarity to *Macrobrachium rosenbergii* virus 7. The heat map shows contig relative abundance in transcripts per million (TPM). SPAdes descriptors on the left include the NODE, length, and k-mer coverage (estimated coverage across all locations) of each contig.

4 Discussion

4.1 Viral community composition and taxonomic diversity

4.1.1 Overview

This study examined the RNA virus communities associated with commercially harvested Atlantic blue crabs (*C. sapidus*) sourced from Maryland or transported to Maryland from five additional states along the U.S. Atlantic and Gulf Coasts. In total, we identified 97 viral-like contigs (≥3,000 bp) from blue crab gill tissues. Positive-sense single-stranded RNA (+ssRNA) segments were especially prominent, with provisionally assigned Picornavirales contigs constituting a large fraction of all 97 **(Figure 2.1; Supplementary Table 2.2)**. Smaller numbers of negative-sense (e.g., Mononegavirales-related) and double-stranded RNA contigs (e.g., Ghabrivirales-related) were also identified. Similar trends were observed when the analysis focused on the 29 high-quality (i.e., complete or near-complete) genomes with a k-mer coverage >10. Positive-sense single-stranded RNA genomes (mainly Picornavirales-related) dominated the data while -ssRNA and dsRNA genomes trailed behind **(Figure 2.2)**. Overall, this first viromics survey in *C. sapidus* revealed a notable diversity of RNA virus orders represented in blue crabs across its U.S. geographic range.

To maximize the taxonomic and genomic diversity of viruses reported in this study, we: 1) collected crabs from different regions, and 2) utilized blast, Pfam, and literature searches to assign provisional taxonomies, elucidate genome structures, and identify putative functional domains.

4.1.2 Dominance of +ssRNA viruses

The dominance of blue crab-associated +ssRNA viruses is consistent with prior metagenomic surveys of marine invertebrates, which have similarly reported a prevalence of picorna-like viruses in crustaceans (Dong et al., 2024; Shi et al., 2016). For example, a recent large-scale survey of 13 economically important crustacean species across 24 locations in China uncovered 90 distinct RNA viruses, over 80% of which had +ssRNA genomes, including dozens of novel picorna-like viruses (Dong et al., 2024). Likewise, analyses of diverse invertebrate transcriptomes have revealed that small RNA viruses (often related to the Picornavirales) dominate the invertebrate virosphere (Shi et al., 2016). We speculate that the characteristics of Picornavirales genomes—small and monopartite or bipartite, and capable of serving directly as messenger RNA—facilitate rapid virus evolution and adaptation, contributing to their abundance and remarkable diversity. We did not detect any putative bipartite Picornavirales genomes. Overall, the vast diversity of +ssRNA viruses described here align with the broader patterns observed in the crustacean literature, expanding the known diversity of this viral group even further.

4.1.3 Viral novelty, taxonomy, host association, and challenges

Like other recent metagenomic studies in marine invertebrates nearly every virus contig we identified was from a novel virus, and each was highly divergent of any currently known virus (Dong et al., 2024; Rosani et al., 2019). The exceptions were one genome **(UP1)** that was confidently identified as a described virus of a macroalga (van der Loos et al., 2023), and a second **(CAV21)** that was identified as a potential variant of Beihai picorna-like virus 107 (Shi et al., 2016). The remaining 27 genomes generally were at less than 50% amino acid identity to any described virus and met the novelty criteria defined in **Section 2.5.7** of this chapter (i.e., <90% amino acid identity across ORFs), highlighting the considerable gaps that remain in our understanding of the blue crab virome and the broader marine virosphere.

The putative virus genomes we identified had taxonomic affiliations with previously-described virus genomes associated with a wide range of host types. The Dicistroviridae within the Picornavirales are recognized for their associations with arthropods (ICTV, 2023); however, most of the provisional viral taxonomies assigned in this study infect organisms outside of the invertebrate clade. The Mononegavirales are known to infect a wide array of hosts, including humans, vertebrate and invertebrate animals, and plants (ICTV, 2011); the Kitaviridae within the Martellivirales is mainly comprised of plant-infecting viruses (ICTV, 2023), while the Marnaviridae commonly infect marine protists (ICTV, 2021); the Norzivirales and Sobelivirales are generally known to infect prokaryotes and eukaryotes, respectively (ICTV, 2021; Sõmera et al., 2021). Finally, the Tolivirales are recognized for their infections of insects and plants (ICTV, 2012), whereas the known Ghabrivirales and Wolframvirales are primarily associated with fungal hosts (SIB Swiss Institute of Bioinformatics, n.d.; Ghabrivirales 2025).

The identification of viruses phylogenetically related to those infecting non-crustacean hosts is not surprising. Part of this diversity may be attributed to the nature of our sampling and processing methods. Gill tissue is exposed to the surrounding marine environment, increasing the likelihood of capturing viruses that infect not only blue crabs but also organisms present in seawater. Additionally, because our protocol did not include steps to rigorously eliminate potential environmental viruses (e.g., rinsing or removing fluid surrounding the gills), it is likely that some non-crab-infecting viruses were co-extracted and sequenced. Our detection of *Ulva* picorna-like virus 1, a virus previously identified in a marine alga (van der Loos et al., 2023), exemplifies the risk of sequencing "contaminating genomes" in high-throughput metagenomic surveys.

Furthermore, while phylogenetic inference can sometimes provide valuable clues (e.g., crustacean viruses clustering with other invertebrate viruses), metagenomic data alone are insufficient to resolve host-virus relationships. For example, Zhao et al. (2022) discovered two viruses affiliated with the Totiviridae—a family initially known for infecting fungi—in RNA extracted from blue crab muscle. Transmission electron microscopy (TEM) confirmed the presence of virions within cells of crabs identified by PCR to carry the toti-like virus genomes. Without the TEM evidence, prior studies and phylogenetic information alone would have suggested these two viruses were present in fungal hosts.

Large advances in the number of metagenomic studies over the past two decades can partly explain why invertebrate studies are increasingly identifying viruses resembling groups initially associated with non-invertebrate hosts. Metagenomics, will continue to uncover untold numbers of novel viral genomes with limited amino acid similarity to sequences in existing databases, revealing genomes that straddle the boundaries of defined taxa. For example, the genome of a novel virus whose closest relatives belong to the Totiviridae may be sufficiently distinct to infect an entirely different set of hosts. Indeed, metagenomics has expanded the recognized host range of toti-like viruses to include arthropods, crustaceans, fish, and plants (Abreu et al., 2015; Chen et al., 2015; Haugland et al., 2011; Koyama et al., 2015, 2016; Poulos et al., 2006; Wu et al., 2010) and of kita-like viruses to include arthropods and crustaceans (Guo et al., 2023).

As virus diversity is revealed through ongoing metaviromic studies, it may be expected that each taxonomic order of viruses will be seen in association with a wide variety of hosts. Although many of the viruses identified here resemble those known to infect unicellular eukaryotes or other non-crustacean organisms, definitive host assignment cannot be made solely on the basis of metagenomic evidence. The possibility that these viruses infect blue crabs remains plausible and warrants further investigation. This uncertainty highlights a broader and persistent challenge in virus metagenomics: the difficulty of accurately establishing both taxonomic placement and host-virus relationships without experimental validation (Shi et al., 2016; Wolf et al., 2020). To address this challenge, the International Committee on Taxonomy of Viruses (ICTV) has increasingly moved toward recognizing new virus groups discovered through metagenomics, even in the absence of virus isolation (Koonin et al., 2020).

4.1.4 Concluding thoughts on viral taxonomy and host association

Our findings contribute to this evolving landscape, emphasizing that in the age of metagenomics, both viral taxonomy and host association are dynamic, provisional constructs—subject to ongoing revision as additional genomic, ecological, and experimental data emerge. Importantly, this study also establishes *C. sapidus* as a promising model for advancing crustacean virology. By contributing valuable sequences to reference databases and providing baseline information on the diversity, novelty, and taxonomic complexity of RNA viruses associated with blue crabs, our work lays the foundation for future metagenomic surveys and experimental investigations. Building upon this foundation, subsequent studies can further explore the geographic distribution, relative abundance, and ecological impacts of blue crab-associated viruses, topics addressed in the following section.

4.2 Geographic patterns in viral community composition (high-quality and k-mer coverage >10)

4.2.1 An abundance of virus diversity and representation in Florida

Florida stands out as the location with a high diversity of putative viruses and the greatest representation and abundance of viral copies among all sampled states. A total of 12 virus genomes were detected, 11 of which had TPM_FL values >10, ranging from 11.84 to 414.64. In fact, seven of the 11 genomes (**CAV15**, **CAV12**, **CAV6**, **CAV1**, **CAV26**, **CAV16**, **CAV19**) were exclusive to Florida, and two (**CAV9** and **CAV14**) were detected in only one other state (albeit at low TPMs). In contrast, the two remaining virus genomes (**CAV8** and **CAV24**) were present in every state, with the exception of North Carolina, the only state for which no virus genomes met our filtering criteria. Notably, **CAV8** and **CAV24** also displayed potential regional preferences in abundance, with the TPM_FL values being substantially higher than those of any other state samples.

Our results create an apparent dichotomy of viruses that are abundant in just one sample/location versus others that are present across most of the sampled states. As discussed further in the following sections, such findings can have important implications for virus-crab coevolution, as well as the biosecurity of commercial crab harvest and interstate transport.

4.2.2 Virus diversity in Maryland

Maryland was the state with the second overall viral abundance. A total of 13 virus genomes were identified, of which five had TPM_MD values >10, ranging from 11.42 to 69.61. Of the 13 genomes, three (**CAV27**, **CAV8**, and **CAV24**) were detected in each of the state samples except

North Carolina. Of these broadly distributed genomes, **CAV27** had a TPM_MD >10, while **CAV24** and **CAV8** were predominantly detected in Florida and had low TPMs in Maryland.

One pattern that emerged in this study was the mutually exclusive nature of the viruses in the Maryland and Florida samples. Of the 22 viruses detected in either Maryland or Florida, 10 were detected in Maryland but not in Florida, nine in Florida but not Maryland, and only three were present in both locations. More research is needed to determine whether these viruses are indeed mutually exclusive between Maryland and Florida blue crab populations and subsequently at risk of being transported from one state to another.

4.2.3 Other states with highly abundant virus genomes

The Louisiana, Texas, and New York samples each contained one putative virus genome that met the contig filtering criteria and was relatively abundant. **CAV28** was uniquely identified in Texas and **UP1** in New York, with TPMs of 21.38 and 10.95, respectively. **CAV3** was only found in two states—Louisiana and Maryland (TPM_LA = 16.82; TPM_MD = 11.42). The detection of this genome in two distinct regions with active blue crab fisheries and nowhere else potentially reflects viral spread through interstate transport; however, more data is needed to support this claim. Overall, the identification of viruses that were solely sequenced in one or two locations suggests their potential regional specificity, and can have important implications for blue crab disease ecology, as well as the biosecurity of commercial crab harvest.

The discovery of virus-like contigs that were highly abundant, yet apparently incomplete, and that had limited similarity to the top blastx hit of the polyprotein of MR7 presents a puzzle: Why would a putative virus contig with such high representation in multiple states fail to assemble into a complete genome? It may illustrate two different possibilities (which may both be true): 1)

There is far more virus diversity to be mined from deep sequencing of *C. sapidus*; 2) These contigs may derive from the existence of RNA entities that are virus-like, but not from complete, encapsidated viruses. For example, these could be transcripts from a defective virus or virus-like sequences, reminiscent of endogenous virus elements (EVEs) in the crab genome itself (Brait et al., 2024).

4.2.4 Our results in the context of other blue crab studies

A distinct pattern emerged from this study, wherein some of the novel virus genomes exhibited apparent geographic restriction, while others were broadly distributed across multiple sampling locations. This dichotomy parallels observations from previous blue crab virus research. For example, Zhao et al. (2022) identified two novel toti-like viruses that were prevalent in *C. sapidus* specimens from northern locations (Massachusetts, Rhode Island, and New York), yet absent from more southern sites, from Delaware, to Texas. Such findings suggest that viral distribution may be shaped either by: 1) ecological or physiological constraints that limit a virus's range, or 2) a lack of introduction into certain regions via natural host migration, viral dispersal, or anthropogenic behaviors.

Although many of the viruses identified in our study did not exhibit sharply delineated biogeographic patterns, several showed evidence of regional specificity. For instance, seven genomes with TPM values >10 were uniquely detected in samples from Florida, while one genome was found exclusively in Texas and another in New York. Similar to the findings from Zhao et al., (2022), our results suggest that these viruses are either restricted by ecological or physiological limits, or haven't been physically introduced to other locations.

In addition to regionally constrained viruses, we also identified several genomes that were detected across nearly all sampled states. This pattern is reminiscent of the distribution of *C. sapidus* reovirus 1 (CsRV1), a pathogenic virus known to infect blue crabs throughout their range in the United States. Specifically, three putative viruses in our dataset—**CAV8**, **CAV24**, and **CAV27**—were present in all states except North Carolina. However, **CAV8** and **CAV24** were found at elevated levels only in Florida (TPM_FL >10), while **CAV27** was abundant only in Maryland (TPM_MD >10). These observations suggest that while some viruses may have a broad geographic presence, their relative abundance may vary substantially across locations, potentially due to differences in environmental conditions, host susceptibility, or seasonal dynamics. For example, **CAV8** and **CAV24** may be more active in subtropical climates, whereas **CAV27** may resemble CsRV1 in its preferential circulation in temperate environments (Zhao et al., 2020).

Notably, although CsRV1 contigs were detected in our dataset, they were excluded from downstream analyses due to the incomplete recovery of the virus's 12-segment genome. Given CsRV1's well-documented prevalence and wide distribution along the U.S. coast, this exclusion underscores the broader constraints of the present study, as discussed below.

4.2.5 Constraints and considerations in interpreting viral distributions

While the preceding observations highlight potential geographic trends in viral diversity and distribution, it is essential to address the unexpected findings and limitations that constrain the interpretation of these results. Despite the detection of viruses with apparent regional specificity and others found across multiple states, the overall lack of a ubiquitous, core community among sampling sites was striking. Even the most abundant virus genomes were inconsistently distributed, often occurring at high levels in only one location and being nearly absent from neighboring states. This pattern was evident in Louisiana and Texas, where each state's most abundant virus—**CAV3** in Louisiana and **CAV28** in Texas—was not detected in the other. In fact, despite their geographic proximity, Texas and Louisiana exhibited markedly different viral community compositions. They shared only seven of the 25 viruses detected in either state, all of which were present at low abundances. Finally, another unexpected finding was the lack of a geographic pattern among states with the highest cumulative viral abundances, as the sampling sites for the top-ranking locations—Florida and Maryland—were situated in entirely separate regions along the U.S. Atlantic Coast.

It is tempting to wonder whether the observed differences in viral composition may reflect temporal fluctuations or stochastic variation rather than true biogeographic structuring. The intent of this study was to be a first look at the diversity of RNA viruses in the commercial product stream, which precludes us from making any inferences from the patterns of occurrence or abundance. To more definitively distinguish between transient and persistent viral patterns, future investigations will require sampling across multiple timepoints and environmental contexts, ideally spanning seasons and years.

Methodological constraints also limited our ability to comprehensively characterize the diversity of blue crab-associated viruses. The use of an intricate virus-like particle (VLP) enrichment protocol, although intended to enhance viral yield, likely contributed to genome degradation, which might be reflected in our limited recovery of complete virus genomes. Of the 97 viral-like contigs identified, fewer than half met high-quality criteria (>90% genome completeness), despite the prior application of a 3,000 bp threshold intended to filter out degraded sequences.

In addition to laboratory constraints, bioinformatic choices, though intentional and carefully considered, also shaped our results. Specifically, we implemented a 3,000 bp threshold and CheckV filter following assembly to exclude partial or low-quality contigs. While these measures streamlined downstream analyses and improved confidence in genome completeness, they likely excluded valid sequences from segmented viruses, which often comprise multiple shorter genome fragments. Prioritizing the efficiency and feasibility of the bioinformatic workflow came at the expense of potentially underrepresenting the diversity and distribution of segmented viruses in the blue crab virome.

Together, these constraints emphasize that the virus diversity reported here represents only a first glimpse of what is sure to be a much bigger and more diverse community. Nevertheless we recovered numerous novel, high-quality genomes—some exhibiting potential regional structuring. Future studies will benefit from deeper sequencing, refined sample preparation protocols, and the integration of metagenomic binning approaches. Such improvements will be critical for fully elucidating the composition and structure of the blue crab virome.

4.3 Advantages of molecular approaches

The use of high-throughput sequencing in this study enabled the rapid discovery of a diverse array of RNA viruses associated with the Atlantic blue crab, *C. sapidus*. Our findings expand the catalog of traceable virus genomes found in tissue samples of this economically and culturally important crustacean, representing a crucial step forward in our ability to monitor and understand blue crab virus dynamics in wild and commercial settings.

One of the principal advantages of this molecular approach is its utility for long-term and large-scale virus surveillance. By generating high-quality genome sequences, this study establishes a foundation for the development of diagnostic multiplex PCR assays, which can be used to track these viruses across time and space with greater efficiency and specificity. The ability to detect multiple viruses simultaneously through multiplexed platforms enhances throughput and reduces costs, making such tools especially valuable for routine health monitoring.

Moreover, these molecular tools offer the opportunity for early detection and proactive management of emerging viral pathogens. Because viromics does not rely on prior sequence knowledge, it is uniquely suited to uncover novel or divergent viruses that may go undetected using traditional methods. Once identified, these putative pathogens can be integrated into targeted surveillance programs, allowing resource managers to screen for viruses before they manifest as clinical diseases or cause population-level impacts. In this way, metagenomic discovery pipelines can serve not only as exploratory tools but also as practical assets for biosecurity and ecosystem health monitoring.

The molecular approach adopted in this study offers significant advantages for both basic research and applied virology. It accelerates virus discovery, facilitates diagnostic development, and supports surveillance frameworks aimed at mitigating the impacts of viral pathogens on blue crab populations and the fisheries they support.

4.4 Implications for blue crab health and fisheries management

The discovery of numerous blue crab-associated viruses raises important questions about their potential impact on the blue crab fishery. Because they are newly discovered, the pathogenicity of these viruses to blue crabs is unknown. They were found in presumably healthy, marketable fishery harvest. However, it is plausible that some of the viruses characterized in this study could be influencing crab health in sublethal ways (e.g. reducing growth, fecundity, or stress tolerance), even if they do not cause detectable mortalities. Future research may show that individuals that harbor these viruses may appear healthy in the wild, but exhibit symptoms under stressful conditions. This appears to be the case for CsRV1, which is associated with the majority of blue crab mortality in soft crab aquaculture (Bowers et al., 2010, 2011; Spitznagel et al., 2019).

It will be important to investigate whether the viruses that appear to be limited to one state are indeed geographically restricted, or whether this observation is a consequence of the sampling and methodological limitations of this initial study. Marine resources globally are known to be subject to losses from introduction of new infectious agents (Stentiford et al., 2012). The global spread of shrimp aquaculture viruses is well known (Jones, 2012), and pathogenic viruses of oysters have upended the culture of *Crassostrea* on three continents (Fuhrmann et al., 2022). These examples are the reason that interstate and international movement of shellfish is typically highly regulated. However, there are currently few regulations on the transport of blue crabs between states in the United States, and many states (Maryland included) do not track the sources of crabs entering the marketplace. Zhao et al. (2023) observed that Maryland and Louisiana genotypes of the well-studied blue crab virus, CsRV1, were more similar to each other than to CsRV1 genotypes in intervening states. These findings were consistent with a direct transfer of this genotype—likely through the disposal or repurposing (e.g., bait) of moribund and dead animals—between two states with intense commercial crab transport.

Finally, it is worth noting that our identification of dicistro-like viruses is particularly relevant from a fisheries management perspective. Members of the Dicistroviridae are known pathogens of crustaceans, including Taura syndrome virus in penaeid shrimp and mud crab dicistrovirus in *Scylla paramamosain*, both of which have caused significant economic losses (Dhar et al., 2004;

Guo et al., 2013, 2014; Kibenge, 2024). By expanding the list of viruses with potential impacts on blue crab health, this study informs future pathogenicity experiments, helping researchers decide which viruses they want to prioritize in their investigations. Furthermore, by providing the genome sequences for these putative pathogens, we enable the development of molecular diagnostic tools for their proactive tracking and monitoring. Significant changes in viral prevalence or geographic distribution over time could then be investigated as part of ongoing management and biosecurity efforts.

5 Conclusion

To our knowledge, this study presents the first metagenomic survey of viruses associated with the Atlantic blue crab (*C. sapidus*) across a significant portion of its United States range. By applying a virus-like particle enrichment protocol and high-throughput RNA sequencing to gill tissue from specimens collected in six states, we identified 97 putative viral contigs, 43 of which were classified as high-quality genomes. Among these, 29 exhibited a k-mer coverage >10, indicating substantial representation within metagenomic libraries. Positive-sense RNA viruses were especially prominent, with provisional Picornavirales members constituting 19 of the 29 genomes. Beyond the Picornavirales, seven other positive-sense, two negative-sense, and one double-stranded RNA virus were identified. These findings place the blue crab virome in context with those of other marine invertebrates, revealing a considerable amount of viral novelty and diversity.

While preliminary in scope, this study provides a foundational framework for future research of an economically and culturally important species by (1) establishing a baseline catalog of virus genomes associated with the blue crab, (2) revealing an apparent dichotomy of viruses that are widespread versus those that are regionally specific, and (3) addressing the advantages and challenges associated with metaviromic surveys. The next steps researchers should take include the development of molecular diagnostic tools followed by experimental infection trials to assess whether the viruses described here replicate in *C. sapidus* and induce pathology. These same tools can then be used to monitor those viruses identified as novel blue crab pathogens and subsequently inform fisheries management practices.

Crustacean virology research will also benefit from broader metagenomic surveys conducted not only across geographic regions but also seasons and years—helping us to distinguish transient from persistent viruses. Furthermore, this study is part of an evolution of methods in which virus identification is rapidly becoming more accessible. Future metagenomic surveys should incorporate optimized laboratory and sequencing protocols that capture the genomes of DNA viruses, enhance the number of raw reads, and limit viral community bias, which has become more feasible through advancements in sequencing capabilities and a decrease in costs. The enhancements described here will substantially expand our understanding of the blue crab virome, taking us from having the genome sequences and provisional taxonomies of viruses to being able to elucidate their roles in blue crab health and population dynamics.

Finally, with climate-driven warming of coastal waters potentially compromising blue crab immunity, the emergence and spread of viral pathogens could become an increasingly important concern (Danovaro et al., 2011). Warmer temperatures have already been shown to accelerate the progression of several crustacean diseases and may similarly enhance viral replication and transmission. Thus, maintaining vigilance through pathogen monitoring will be important for sustaining blue crab populations and, by extension, the fisheries they support.

The present study lays the groundwork by cataloging the RNA viral baseline; it is our hope that this will spur further research into the role of viruses in blue crab biology and promote a more virus-informed framework in marine resource management.

6 Supplementary material

Supplementary Table 2.1

Number of reads per state.

Reads	State						
	Louisiana SRX28690409	Maryland SRX28690410	North Carolina SRX28690412	Florida SRX28690408	Texas SRX28690413	New York SRX28690411	**Total**
Initial	51,695,864	53,267,244	50,946,184	78,569,842	107,526,748	45,023,310	387,029,192
Initial - Host	47,312,723 (91.52%)	44,078,530 (82.75%)	45,574,547 (89.46%)	71,871,723 (91.47%)	53,454,744 (49.7%)	33,178,933 (73.69%)	295,471,200
Host Removed	46,900,396 (90.72%)	43,801,784 (82.23%)	45,006,748 (88.34%)	71,445,488 (90.93%)	53,032,300 (49.32%)	32,949,894 (73.18%)	293,136,610
Final	**4,795,468 (9.28%)**	**9,465,460 (17.77%)**	**5,939,436 (11.66%)**	**7,124,354 (9.07%)**	**54,494,448 (50.68%)**	**12,073,416 (26.82%)**	**93,892,582**
Contigs							**64,192**

Supplementary Table 2.2

All 97 contigs >3000 bp identified in the study.

Assigned Name	Contig Length	K-mer Coverage	Top Blastx Accession	Related Order	Related Family	Top Blastx Hit	Percent Identity	Query Coverage	CheckV Quality
CAV9	9238	429.14	WBY66453.1	Picornavirales	Dicistroviridae	MAG: nonstructural protein [Eriocheir sinensis dicistrovirus 2]	41	38	High
CAV28	7769	306.68	APG76060.1	Ghabrivirales	Unclassified	hypothetical protein 2, partial [Wenling toti-like virus 1]	46	33	High
CAV5	9997	123.49	YP_009336613.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Wenling picorna-like virus 3]	50.6	47	High
CAV15	8974	103.88	YP_009336714.1	Picornavirales	Unclassified	hypothetical protein 1 [Wenzhou picorna-like virus 7]	76.1	53	High
CAV8	9309	81.03	YP_009333603.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Beihai picorna-like virus 82]	40.8	53	High
CAV17	8816	69.35	YP_009667031.1	Picornavirales	Marnaviridae	non-structural polyprotein [Marine RNA virus BC-4]	74.3	56	High
CAV27	3014	64.9	YP_009337422.1	Wolframvirales-like	Narna-like	RNA-dependent RNA polymerase [Wenzhou shrimp virus 10]	71.8	83	High
CAV10	9219	63.13	YP_009336682.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Wenzhou shrimp virus 5]	35.5	54	High
CAV2	11780	52.72	WFD53221.1	Mononegavirales	Rhabdoviridae	MAG: L protein [Dicrocoelium Rhabdo-like virus 2]	34.5	52	High
CAV23	3044	42.93	WBY66479.1	Sobelivirales	Solemoviridae	MAG: putative peptidase [Eriocheir sinensis sobemo-like virus 2]	86.4	46	High
CAV12	9805	41.76	BAP99820.1	Picornavirales	Marnaviridae	predicted replication-associated protein [Chaetoceros tenuissimus RNA virus type II]	78.7	55	High
CAV14	9121	40.76	YP_010084314.1	Picornavirales	Marnaviridae	putative replication related protein [Chaetoceros species RNA virus 02]	90.9	52	High
CAV6	9859	33.28	YP_009330121.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Beihai picorna-like virus 70]	60	57	High

UP1	8357	31.25	WIR83948.1	Picornavirales	Unclassified	MAG: polyprotein [Ulva picorna-like virus 1]	99.1	57	High
CAV3	12007	26.84	WBY66456.1	Martellivirales	Kitaviridae	MAG: polyprotein [Eriocheir sinensis kita-like virus]	48.3	70	High
CAV1	11825	25.92	WFD53221.1	Mononegavirales	Rhabdoviridae	MAG: L protein [Dicrocoelium Rhabdo-like virus 2]	34.4	52	High
CAV7	9370	25.14	YP_009336682.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Wenzhou shrimp virus 5]	35.3	53	High
CAV24	4919	22.68	YP_009342270.1	Tolivirales-like	Tombus-like	hypothetical protein 2 [Wenzhou tombus-like virus 18]	64.5	36	High
CAV11	9839	21.84	YP_010084314.1	Picornavirales	Marnaviridae	putative replication related protein [Chaetoceros species RNA virus 02]	79.5	52	High
CAV25	4093	20.9	QJI53534.1	Tolivirales-like	Tombus-like	MAG: hypothetical protein 1 [Picornavirales sp.]	38.1	53	High
CAV20	8645	18.02	YP_009330008.1	Picornavirales	Unclassified	hypothetical protein [Changjiang crawfish virus 4]	32.4	73	High
CAV4	3642	17.7	YP_010769536.1	Norzivirales	Fiersviridae	RNA-directed RNA polymerase [ssRNA phage SRR6960802_1]	55.4	43	High
CAV16	8974	17.1	YP_009667031.1	Picornavirales	Marnaviridae	non-structural polyprotein [Marine RNA virus BC-4]	76.8	54	High
CAV26	4348	15.89	UJQ92779.1	Wolframvirales	Narnaviridae	MAG: putative RNA-dependent RNA polymerase [Narnaviridae sp.]	82.3	79	High
CAV22	7960	13.46	WIR83950.1	Picornavirales	Unclassified	MAG: polyprotein [Ulva picorna-like virus 3]	34.3	51	High
CAV18	8710	12.24	APG78919.1	Picornavirales	Unclassified	hypothetical protein 1 [Beihai picorna-like virus 31]	51.6	48	High
CAV21	8578	12.05	YP_009333563.1	Picornavirales	Unclassified	hypothetical protein [Beihai picorna-like virus 107]	91	94	High
CAV19	8454	11.05	YP_009333605.1	Picornavirales	Unclassified	hypothetical protein 1 [Beihai picorna-like virus 41]	58.3	58	High
CAV13	9224	10.21	YP_009333423.1	Picornavirales	Unclassified	hypothetical protein [Beihai picorna-like virus 32]	56.1	51	High

NODE89	4102	8.77	QUS52665.1	Nodamuvirales	Nodaviridae	RNA-dependent RNA polymerase [Mute swan feces associated noda-like virus 4]	33.8	59	High
NODE63	5068	8.22	YP_009337425.1	Nodamuvirales-like	Noda-like	hypothetical protein 1 [Wenzhou bivalvia virus 3]	61.7	66	High
NODE15	9427	8.09	WDS50619.1	Unclassified	Unclassified	MAG: nonstructural protein, partial [Riboviria sp.]	33.2	36	High
NODE26	8935	7.69	ULG00077.1	Picornavirales	Marnaviridae	MAG: hypothetical protein [Salisharnavirus sp.]	53.5	56	High
NODE27	8908	7.47	UNY42005.1	Picornavirales	Unclassified	MAG: polyprotein [Picornavirales sp.]	28	78	High
NODE93	4045	7.06	UJQ92717.1	Wolframvirales	Narnaviridae	MAG: putative RNA-dependent RNA polymerase, partial [Narnaviridae sp.]	81.4	62	High
NODE78	4382	6.84	APG77015.1	Norzivirales	Fiersviridae	hypothetical protein [Beihai levi-like virus 32]	63	43	High
NODE33	8470	5.74	QNM80717.1	Unclassified	Unclassified	replicative protein [Haldis virus]	37.8	49	High
NODE41	7828	5.55	QBP32757.1	Unclassified	Unclassified	MAG: replicative protein [Cragig virus 7]	66.4	58	High
NODE7	11746	5.2	YP_009337358.1	Unclassified	Hepe-Virga-like	replicase [Wenling hepe-like virus 2]	46.4	49	High
NODE37	8037	5.18	YP_009333585.1	Picornavirales	Unclassified	hypothetical protein [Beihai picorna-like virus 122]	61.5	86	High
NODE21	9168	4.85	UNY42064.1	Picornavirales	Unclassified	MAG: polyprotein 1 [Picornavirales sp.]	84.7	54	High
NODE61	5183	3.38	APG76212.1	Nodamuvirales-like	Noda-like	hypothetical protein 1 [Beihai noda-like virus 22]	48.4	62	High
NODE101	3970	2.18	WKV34149.1	Unclassified	Unclassified	MAG: RNA-dependent RNA polymerase, partial [Riboviria sp.]	28.6	55	High
NODE38	7955	1257.51	YP_009333603.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Beihai picorna-like virus 82]	42.8	57	Medium
NODE29	8859	132.42	QZZ63368.1	Picornavirales	Unclassified	hypothetical protein [Leuven Picorna-like virus 5]	29.8	28	Medium

NODE46	6891	128.38	UWY63979.1	Picornavirales	Solinviviridae	polyprotein [Penaeus vannamei solinvivirus]	28.9	65	Medium
NODE23	8985	75.97	QJI53584.1	Picornavirales	Unclassified	MAG: hypothetical protein, partial [Picornavirales sp.]	30.8	25	Medium
NODE40	7841	42.6	APG76060.1	Ghabrivirales	Unclassified	hypothetical protein 2, partial [Wenling toti-like virus 1]	45.2	31	Medium
NODE58	5645	29.81	YP_009336862.1	Picornavirales	Unclassified	hypothetical protein 1 [Wenzhou picorna-like virus 40]	70.4	78	Medium
NODE108	3718	24.87	YP_009342270.1	Tolivirales-like	Tombus-like	hypothetical protein 2 [Wenzhou tombus-like virus 18]	64.5	48	Medium
NODE44	7539	24.78	QKN89018.1	Unclassified	Unclassified	MAG: hypothetical protein, partial [Riboviria sp.]	31.1	29	Medium
NODE154	3020	18.14	YP_009337422.1	Wolframvirales-like	Narna-like	RNA-dependent RNA polymerase [Wenzhou shrimp virus 10]	43.2	82	Medium
NODE43	7649	15.69	YP_009047193.1	Picornavirales	Marnaviridae	predicted replication-associated protein [Asterionellopsis glacialis RNA virus]	64.8	62	Medium
NODE70	4843	9.45	YP_009336787.1	Picornavirales	Unclassified	hypothetical protein [Wenzhou picorna-like virus 46]	29.9	67	Medium
NODE120	3400	9.32	YP_009337663.1	Unclassified	Unclassified	RNA-dependent RNA polymerase [Hubei zhaovirus-like virus 2]	21.1	35	Medium
NODE36	8288	7.17	YP_009333602.1	Picornavirales	Unclassified	hypothetical protein [Beihai sesarmid crab virus 2]	38.2	77	Medium
NODE69	4907	6.93	APG76060.1	Ghabrivirales	Unclassified	hypothetical protein 2, partial [Wenling toti-like virus 1]	28.3	28	Medium
NODE45	7002	5.03	YP_009333481.1	Picornavirales	Unclassified	hypothetical protein 1 [Beihai picorna-like virus 39]	41.5	57	Medium
NODE74	4425	4.75	APG76440.1	Tolivirales-like	Tombus-like	hypothetical protein 3 [Sanxia water strider virus 14]	47.7	29	Medium
NODE57	5665	4.75	AYN75548.1	Unclassified	Unclassified	replicative protein [Halhan virus 1]	48.6	51	Medium
NODE56	5781	4.7	QBP32757.1	Unclassified	Unclassified	MAG: replicative protein [Cragig virus 7]	70.8	71	Medium

NODE142	3135	4.1	QYF49899.1	Nodamuvirales	Nodaviridae	MAG: RNA-dependent RNA polymerase, partial [Nansha Islands sediment noda-like virus 1]	55.4	82	Medium
NODE53	5989	4.06	YP_009336646.1	Picornavirales	Unclassified	hypothetical protein 1 [Wenzhou picorna-like virus 2]	54.2	56	Medium
NODE134	3271	3.9	ASG92545.1	Picornavirales	Unclassified	putative non-structural protein [Picornavirales Q_sR_OV_023]	42.9	63	Medium
NODE51	6088	3.23	YP_009333605.1	Picornavirales	Unclassified	hypothetical protein 1 [Beihai picorna-like virus 41]	72	57	Medium
NODE110	3706	3.11	QJI53534.1	Picornavirales	Unclassified	MAG: hypothetical protein 1 [Picornavirales sp.]	48.1	55	Medium
NODE52	6086	2.94	YP_009553171.1	Unclassified	Unclassified	replicative protein [Cragig virus 1]	41	65	Medium
NODE59	5359	2.79	YP_009336788.1	Ghabrivirales	Unclassified	hypothetical protein [Hubei toti-like virus 9]	24.2	20	Medium
NODE86	4200	6106.42	UUV42150.1	Picornavirales	Dicistroviridae	MAG: polyprotein [Macrobrachium rosenbergii virus 7]	45.2	87	Low
NODE3	13414	255.91	APG77570.1	Martellivirales	Endornaviridae	hypothetical protein [Behai endorna-like virus 1]	41.3	11	Low
NODE1	21022	102.65	WJN66647.1	Martellivirales	Endornaviridae	polyprotein [Grapevine endornavirus 2]	34.2	12	Low
NODE82	4278	36.39	YP_009246448.1	Reovirales	Unclassified	VP1 [Callinectes sapidus reovirus 1]	99.8	99	Low
NODE117	3476	27.91	YP_009336863.1	Picornavirales	Unclassified	hypothetical protein 2 [Wenzhou picorna-like virus 40]	71.7	91	Low
NODE115	3540	10.17	UXE05527.1	Martellivirales	Endornaviridae	MAG: RNA-dependent polymerase, partial [Althepos endorna-like virus 1]	36.6	34	Low
NODE119	3455	7.28	APG78048.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Wenzhou shrimp virus 6]	33.4	50	Low
NODE65	4974	5.56	WFD53221.1	Mononegavirales	Rhabdoviridae	MAG: L protein [Dicrocoelium Rhabdo-like virus 2]	37.4	95	Low
NODE64	5100	5.19	WFD53217.1	Mononegavirales	Rhabdoviridae	MAG: nucleoprotein [Dicrocoelium Rhabdo-like virus 2]	31.4	15	Low

NODE100	3981	5	YP_009333524.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Beihai picorna-like virus 90]	34	44	Low
NODE103	3848	3.89	YP_009333584.1	Picornavirales	Unclassified	hypothetical protein 2 [Beihai picorna-like virus 19]	47.4	58	Low
NODE12	9814	3.88	YP_009337201.1	Unclassified	Hepe-Virga-like	RdRp [Wenling crustacean virus 1]	34.7	39	Low
NODE112	3588	3.81	YP_009336706.1	Picornavirales	Dicistroviridae	hypothetical protein [Wenzhou picorna-like virus 27]	82.4	77	Low
NODE80	4346	3.21	YP_009330023.1	Picornavirales	Dicistroviridae	hypothetical protein [Beihai picorna-like virus 73]	90	98	Low
NODE96	4023	2.8	QIN96625.1	Picornavirales	Marnaviridae	MAG: hypothetical protein 1 [Marnaviridae sp.]	61.2	40	Low
NODE66	5040	2.66	UZT75550.1	Bunyavirales	Unclassified	putative RdRP [Whenzhou Shrimp Virus 2]	87.7	99	Low
NODE81	4327	2.57	YP_009333481.1	Picornavirales	Unclassified	hypothetical protein 1 [Beihai picorna-like virus 39]	48.7	95	Low
NODE83	4232	2.38	YP_009336700.1	Picornavirales	Dicistroviridae	hypothetical protein 1 [Wenling picorna-like virus 5]	34.1	78	Low
NODE90	3971	1713.91	APG79086.1	Reovirales	Reoviridae	RdRp [Beihai reo-like virus 1]	28.9	74	Not determined
NODE77	4348	1431.13	APG79214.1	Reovirales	Reoviridae	RdRp [Wenling reo-like virus 1]	62	95	Not determined
NODE8	11463	338.34	YP_009333320.1	Picornavirales	Unclassified	hypothetical protein [Beihai blue swimmer crab virus 1]	25.8	37	Not determined
NODE91	3971	176.72	APG79086.1	Reovirales	Reoviridae	RdRp [Beihai reo-like virus 1]	29.4	74	Not determined
NODE75	4340	110.5	APG79214.1	Reovirales	Reoviridae	RdRp [Wenling reo-like virus 1]	61.3	96	Not determined
NODE140	3116	31.17	UUV42150.1	Picornavirales	Dicistroviridae	MAG: polyprotein [Macrobrachium rosenbergii virus 7]	41.5	8	Not determined
NODE2	15448	17.46	QYF50063.1	Martellivirales	Endornaviridae	MAG: polyprotein, partial [Henan sediment alphaendornavirus 1]	40.5	6	Not determined

NODE50	6129	12.09	BCL84886.1	Martellivirales	Endornaviridae	polyprotein [Phytophthora endornavirus 2]	28.7	14	Not determined
NODE71	4733	5.94	QJI52022.1	Picornavirales	Dicistroviridae	MAG: hypothetical protein 1 [Dicistroviridae sp.]	29	14	Not determined
NODE136	3255	5.34	APG79086.1	Reovirales	Reoviridae	RdRp [Beihai reo-like virus 1]	28.9	90	Not determined
NODE148	3067	1.27	QGA70918.1	Martellivirales	Endornaviridae	RNA-dependent RNA polymerase, partial [Forneby virus]	37	14	Not determined

Supplementary Table 2.3a: *Callinectes*-associated virus 1

The top blastx hits of the putative open reading frames of ***Callinectes-associated virus 1***.

Assigned Name	Query	Percent Identity	Query Length (AA)	Query Coverage	Subject Title
CAV1	ORF1	33.1	470	55	MAG: nucleoprotein [Dicrocoelium Rhabdo-like virus 2]
	ORF2	No Hits	246	No Hits	No Hits
	ORF3	No Hits	233	No Hits	No Hits
	ORF4	22.6	502	94	TPA_asm: hypothetical protein [Sphaeridiorhabdovirus 1]
	ORF5	34.5	2076	99	MAG: L protein [Dicrocoelium Rhabdo-like virus 2]

Supplementary Table 2.3b: *Callinectes*-associated virus 2

The top blastx hits of the putative open reading frames of ***Callinectes-associated virus 2***.

Assigned Name	Query	Percent Identity	Query Length (AA)	Query Coverage	Subject Title
CAV2	ORF 1	30.5	442	65	MAG: nucleoprotein [Dicrocoelium Rhabdo-like virus 2]
	ORF2	No Hits	246	No Hits	No Hits
	ORF3	No Hits	233	No Hits	No Hits
	ORF4	No Hits	329	No Hits	No Hits
	ORF5	34.5	2077	98	MAG: L protein [Dicrocoelium Rhabdo-like virus 2]

Chapter Three: Findings and Future Directions

1 Introduction

As expressed throughout this thesis, the Atlantic blue crab, *Callinectes sapidus*, is a benthic coastal crustacean of major ecological, economic, and cultural significance. Ecologically, blue crabs play pivotal roles in estuarine food webs as both predators and prey. They influence soft-bottom community structure through predation (Lipcius et al., 2007; Seitz et al., 2005), serve as forage for fish species (Orth et al., 1999), and shape recruitment and population dynamics through cannibalism (Hines & Ruiz, 1995; Moksnes et al., 1997). In addition to these ecological functions, blue crabs hold considerable social and cultural value, particularly in the Chesapeake Bay region, where the crustacean is celebrated in regional cuisine and local traditions (Chesapeake Bay Foundation, n.d.; NOAA Fisheries, 2025). Economically, the species supports one of the most valuable commercial fisheries along the U.S. Atlantic and Gulf Coasts, with the Mid-Atlantic being the largest regional contributor and a net importer of blue crabs. Substantial blue crab fisheries also exist in the Southeast and Gulf of Mexico, with Louisiana consistently ranking as the top state for harvest revenue nationwide (NMFS Commercial Landings).

Of the millions of blue crabs imported into Maryland from other states, it is reasonable to estimate that over 5% die during transit (Schott, personal observation). Anecdotal accounts suggest that it is not uncommon for dead crabs to be discarded directly into Chesapeake Bay waters. While Maryland lacks a formal system for tracking the geographic origins of its imported crabs, Louisiana is likely the largest source of live crab imports into the state. Notably, a prior study found that genotypes of *C. sapidus* reovirus 1 (CsRV1) in Maryland and Louisiana were more closely related to each other than to those from intervening regions (Zhao et al., 2023). This finding reminds us of the possibility that not only viral genotypes, but also entirely distinct viruses, may be unintentionally transported between major fishery regions via interstate commerce.

Unfortunately, recent blue crab population declines have introduced ecological and economic uncertainty to the communities that depend on this fishery. In each of the past five years, the Chesapeake Bay Stock Assessment Advisory Report mentions the need to better understand the role of diseases—including those caused by viruses—in blue crab mortality (Chesapeake Bay Stock Assessment Committee, 2021, 2022, 2023, 2024, 2025).

Given the essential roles viruses play in marine ecosystems—including their potential to impair host immunity, physiological performance, and overall health—this dissertation aimed to establish a baseline understanding of the RNA viruses associated with *C. sapidus*, a valued marine invertebrate. The work described here increases the number of viruses associated with the blue crab by more than 10-fold over what has been reported in the past 50 years. The discovery of these viruses through molecular methods means that the development of highly specific tools to monitor them is all but one step away, yet leaves the need for more biological characterization to assess the impacts of each virus on crab health. Ultimately, an important goal of this research is to inform fisheries practices in ways that mitigate the spread of pathogens through interstate commerce.

2 Findings

To our knowledge, this study presents the first metagenomic survey of viruses associated with the Atlantic blue crab across a significant portion of its United States range. By applying a virus-like particle enrichment protocol and high-throughput RNA sequencing to gill tissue from specimens collected in six states, we identified 97 putative viral contigs, 43 of which were classified as high-quality genomes. Among these, 29 exhibited a k-mer coverage >10, indicating substantial representation within metagenomic libraries. Positive-sense RNA viruses were especially prominent, with provisional Picornavirales members constituting 19 of the 29 genomes. Beyond the Picornavirales, seven other positive-sense, two negative-sense, and one double-stranded RNA virus were identified. These findings place the blue crab virome in context with those of other marine invertebrates, revealing a considerable amount of viral novelty and diversity.

Ultimately, this work (1) established a baseline catalog of blue crab-associated virus genomes, (2) revealed an apparent dichotomy of viruses that are widespread versus those that are regionally specific, and (3) illustrated the advantages and challenges associated with metagenomic surveys.

3 Future directions

The catalog of genomes identified here can be used in numerous ways. First, researchers can use these sequences to develop molecular diagnostic tools (i.e., PCR assays) for viruses of interest (e.g., dicistro-like viruses in the Florida sample). In particular, the design and optimization of multiplex assays that target *Callinectes*-associated viruses will allow for hundreds of specimens to be screened at once. As a targeted approach that requires specific sequence information, PCR will not identify novel viruses but will streamline the detection and quantification of the viruses identified within this metagenomic study. Fortunately, quantitative PCR technology is readily accessible to us, with IMET's BASLab capable of running multiplex assays for up to five viral targets at once. The development of these molecular diagnostic tools, followed by large-scale and widespread geographic screening of blue crabs across multiple seasons and years, will dramatically enhance our understanding of the prevalence and abundance patterns of the viruses identified in **Chapter Two**.

While PCR-based diagnostics are invaluable for detecting and quantifying known viruses, they provide no information on the pathogenicity of the viruses detected. For example, high prevalence or abundance of a virus within a population, even if persistent across seasons, do not necessarily correlate with pathogenic effects on the host (Holt et al., 2019; Johnson, 1984). Nonetheless, large-scale PCR surveys, when combined with environmental and observational data, can offer valuable clues about the ecological roles of the viruses under investigation. For instance, a virus that exhibits high abundance during summer months and is frequently associated with moribund blue crabs may warrant further examination. Therefore, a critical next step is to conduct experimental infection trials to assess whether the viruses identified in this study can replicate within *C. sapidus* and induce pathological effects. Such studies are essential for determining causality and understanding viral impacts on blue crab immunity, physiology, and overall health.

As our understanding of the viruses we've described in **Chapter Two** and their effects on blue crabs deepens, these scientific insights can be leveraged to inform and enhance fisheries management decisions. For instance, if experimental infection studies reveal that a virus is highly pathogenic to blue crabs, and PCR-based surveillance confirms its presence in southern U.S. states but absence in northern states, targeted biosecurity measures can be recommended to prevent the introduction of this pathogen into naive waterways. Such measures may include restrictions on the movement of live crabs and enhanced monitoring of aquaculture facilities.

Beyond the development of molecular diagnostic tools for virus detection and the performance of infection studies to examine pathogenicity, blue crab virology research will also benefit from broader metagenomic surveys. As previously noted, the virus diversity we report likely represents only an initial glimpse of a much larger and more complex community; therefore, future studies should address our limitations through deeper sequencing, improved sample preparation protocols, and the incorporation of metagenomic binning approaches. Such improvements will be critical for fully characterizing the composition and structure of the blue crab virome. In the following sections, we address the common limitations and evolving methodologies in viromics, contextualizing them alongside the approaches applied in our study.

4 Addressing the common limitations and evolving methodologies in viromics

Viruses remain underexplored relative to other microbial groups, and several challenges unique to viromics contribute to this disparity. One of the foundational challenges is the absence of a universally conserved genomic region across viruses—unlike the 16S gene in bacteria or ITS regions in fungi—rendering amplicon-based approaches unavailable for viral diversity studies. Consequently, viromics relies heavily on untargeted shotgun DNA and RNA sequencing, a strategy that, while comprehensive, lacks widely-adopted laboratory and bioinformatic protocols and can be less accessible to labs new to this research area. Sample preparation and bioinformatic analysis are discussed further below.

4.1 Sample processing

Tissue processing for virus metagenomics has been accomplished using a wide range of methods, many of which aim to enrich for virus-like particles (VLPs) prior to nucleic acid extraction. In our RNA virus study, we leveraged viruses' small size and unique densities to enrich viral particles before RNA extraction. Drawing on a comprehensive review of the literature, we implemented a series of steps that included tissue homogenization and freeze-thaw cycles of to lyse host cells (Guo et al., 2023; Zhou et al., 2022), differential centrifugation and ultracentrifugation for particle separation (Conceição-Neto et al., 2015), filtration through a 0.45 µm membrane to exclude larger debris, sucrose gradient purification to further concentrate viral particles (Hurwitz et al., 2012), and DNAse treatment to eliminate free nucleic acids. Ultimately, our protocol closely followed that of Jiang and Wei (2018), with only minor modifications.

Despite our efforts to optimize virus enrichment through a multi-step protocol, several methodological constraints likely limited our ability to recover a greater diversity of complete virus genomes. While the VLP enrichment process was designed to increase viral yield by reducing host contamination, its complexity may have inadvertently contributed to degradation of fragile RNA genomes. In addition, we opted to omit ribosomal RNA (rRNA) depletion during library preparation, reasoning that the extensive VLP enrichment would sufficiently reduce host-derived RNA. However, mapping the sequencing reads to the *C. sapidus* genome revealed that 49.7% to 91.5% of reads still aligned to host sequences. This indicates that host RNA removal was incomplete, and a substantial host signal persisted through to sequencing. As a result, the effective sequencing depth available for viral genome recovery was reduced, ultimately decreasing the coverage, completeness, and diversity of the viral contigs recovered.

In light of the limitations associated with complex VLP enrichment protocols, advances in high-throughput sequencing have shifted virus metagenomics toward simpler, less intrusive approaches. Rather than relying on extensive mechanical manipulation to enrich for virus-like particles, many researchers now favor sequencing total RNA directly from samples. This strategy is particularly advantageous for characterizing RNA virus communities, as it minimizes physical stress on viral particles and reduces the risk of genome degradation. By incorporating selection steps during library preparation—such as poly(A) selection or ribosomal RNA (rRNA) depletion—researchers can enrich for virus genomes or transcripts while substantially reducing the signal of host-derived rRNA. Coupled with deeper sequencing, which has become increasingly feasible due to lower costs and greater efficiency of sequencing platforms, this approach enhances the detection of viruses and supports more comprehensive virome profiling.

Importantly, the choice of RNA selection method during library preparation should be informed by the specific goals of the study. For example, rRNA depletion is better suited for comprehensive virome surveys, as it preserves both coding and non-coding components of RNA genomes. In contrast, poly(A) enrichment targets only polyadenylated transcripts and is therefore biased toward detecting the transcripts of active viruses. This approach may be particularly useful for identifying viruses with high transcriptional activity that could play significant roles in host immunity, physiology, or disease. Ideally, parallel library preparations using both selection strategies would offer a more holistic view—capturing not only the viral genomes associated with *C. sapidus*, but also the transcripts that reflect their functional activity within the host.

Thanks to the continuing declines in sequencing costs, a similar trend toward reduced sample manipulation has been observed in marine virome studies targeting DNA viruses, where some researchers now sequence total DNA without VLP enrichment (Messyasz et al., 2020; Herrera, personal communication). However, DNA virome studies present distinct challenges, as there is no equivalent of rRNA depletion for selectively removing host or non-viral DNA during library preparation.

Regardless of genome type, the overarching objective in sample processing for computational virology research remains consistent: to preserve viral integrity, minimize methodological bias, and maximize the recovery of high-quality, complete virus genomes or transcripts.

4.2 Bioinformatic analysis

The variability observed in sample preparation for viromics also extends to bioinformatic analyses, where a multitude of virus identification programs exist. Although many experts converge on similar analytical frameworks, these practices are more clearly conveyed through informal discussions than documented in the literature, making it difficult for newcomers to identify and adopt the most effective approaches. In contrast, while ambiguities and challenges associated with amplicon sequencing analyses also remain, such analyses are significantly more standardized, with numerous user-friendly online resources and bioinformatic pipelines that typically rely on a core set of tools to streamline the process. Future efforts should prioritize the accessibility and transparency of virus bioinformatics resources including workflows and documentation to encourage broader participation in viromics.

The effectiveness of reference-based bioinformatic tools is also dependent on the completeness of global databases. The large number of novel virus genomes identified in **Chapter Two** reinforces that scientific understanding of the marine virome is still in an early, discovery-driven phase. In this study, the provisional taxonomic assignment of viruses were particularly challenging when their top blastx hits had missing order and family-level annotations, often necessitating extensive literature reviews and de novo phylogenetic analyses. These difficulties highlight that viral taxonomic assignment based on pairwise identity to reference sequences remains inherently constrained. Ultimately, the accuracy and efficiency of referenced-based taxonomic assignment will only improve with increasing amounts of publicly available virome data, continuous characterization of novel taxa, and revisioning of existing taxonomic frameworks.

5 Conclusion

Despite the challenges outlined in this chapter, the integration of high-throughput sequencing into virus discovery continues to offer transformative and once-unimaginable potential. As sequencing technologies become increasingly powerful and cost-effective, researchers are able to capture novel viruses—both rare and abundant—at unprecedented depth. This expansion of viromics will continue to deepen our understanding of marine viral diversity, drive the discovery of novel taxa, and prompt the revision of existing taxonomic frameworks. Looking ahead, there is little doubt that metagenomic studies will remain a cornerstone of biological research, reshaping our understanding of biodiversity for generations to come.

Within the context of *C. sapidus*, our study establishes a baseline of RNA viruses associated with the species, contributing valuable sequences to global viral reference databases. Simultaneously, this virus catalog informs the development of molecular diagnostic tools and the design of infection studies aimed at assessing viral impacts on host physiology and health. Finally, the methodological refinements and strategic recommendations presented in this chapter offer a framework for improving viromics, with relevance not only to blue crabs but to other ecologically and economically important marine species.

Ultimately, our efforts support a critical shift in the field—from exploratory virus discovery and provisional taxonomic assignment toward functional investigations of viral ecology, host-pathogen interactions, and population-level outcomes. Insights from such studies will be essential for informing sustainable fisheries management and guiding evidence-based policies that protect marine resources in an era of anthropogenic change.

Bibliography

Abreu, E. F., Daltro, C. B., Nogueira, E. O., Andrade, E. C., & Aragão, F. J. (2015). Sequence and genome organization of papaya meleira virus infecting papaya in Brazil. *Archives of Virology*, *160*(12), 3143–3147. https://doi.org/10.1007/s00705-015-2605-x

Antipov, D., Raiko, M., Lapidus, A., & Pevzner, P. A. (2020). Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics*, *36*(14), 4126-4129. https://doi.org/10.1093/bioinformatics/btaa490

Bachvaroff, T. R., McDonald, R. C., Plough, L. V., & Chung, J. S. (2021). Chromosome-level genome assembly of the blue crab, *Callinectes sapidus. G3*, *11*(9), jkab212. https://doi.org/10.1093/g3journal/jkab212

Bergh, Ø., Børsheim, K. Y., Bratbak, G., & Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature*, *340*(6233), 467-468. https://doi.org/10.1038/340467a0

Besemer, J., & Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Research*, *27*(19), 3911-3920. https://doi.org/10.1093/nar/27.19.3911

Bowers, H. A., Carrion, L., Hanif, A., Messick, G. A., Zmora, O., & Schott, E. J. (2011). A reo-like virus associated with mortalities of blue crab *Callinectes sapidus*: Development of tools to improve soft-shell crab aquaculture. *Journal of Shellfish Research*, *30*(2), 487.

Bowers, H. A., Messick, G. A., Hanif, A., Jagus, R., Carrion, L., Zmora, O., & Schott, E. J. (2010). Physicochemical properties of double-stranded RNA used to discover a reo-like virus from blue crab Callinectes sapidus. *Diseases of Aquatic Organisms*, *93*(1), 17-29. https://doi.org/10.3354/dao02280

Brait, N., Hackl, T., Morel, C., Exbrayat, A., Gutierrez, S. and Lequime, S., (2024). A tale of caution: How endogenous viral elements affect virus discovery in transcriptomic data. *Virus Evolution,* 10(1), p.vead088. https://doi.org/10.1093/ve/vead088

Bushnell, B. (2014). BBMap: *A fast, accurate, splice-aware aligner.* Joint Genome Institute. https://sourceforge.net/projects/bbmap/

Cameron's Seafood. (2023, July 5). *When is the Maryland Blue Crab season?* https://www.cameronsseafood.com/blogs/faq/maryland-blue-crab-season

Chen, S., Huang, Q., Wu, L., and Qian, Y. (2015). Identification and characterization of a maize-associated *mastrevirus* in China by deep sequencing small RNA populations. *Virology Journal.* 12, 156–159. https://doi.org/10.1186/s12985-015-0384-3

Chesapeake Bay Foundation. (n.d.). *Blue crabs*. Retrieved April 29, 2025, from https://www.cbf.org/about-the-bay/chesapeake-wildlife/blue-crabs/index.html

Chesapeake Bay Stock Assessment Committee. (2021, December). *Blue crab harvest reporting document*. Chesapeake Bay Program. https://d18lev1ok5leia.cloudfront.net/chesapeakebay/documents/2021_cbsac_harvest_reporting_doc1.pdf

Chesapeake Bay Stock Assessment Committee. (2021, June 22). *2021 Chesapeake Bay Blue Crab Advisory Report*. Chesapeake Bay Program.

https://www.chesapeakebay.net/files/documents/2021_Blue-Crab-Advisory-Report_Final_06-22-21.pdf

Chesapeake Bay Stock Assessment Committee. (2022, July 6). *2022 Chesapeake Bay Blue Crab Advisory Report*. Chesapeake Bay Program. https://d38c6ppuviqmfp.cloudfront.net/documents/2022_Blue_Crab_Advisory_Report_Final.pdf

Chesapeake Bay Stock Assessment Committee. (2023, June 27). *2023 Chesapeake Bay Blue Crab Advisory Report*. Chesapeake Bay Program. https://d18lev1ok5leia.cloudfront.net/chesapeakebay/documents/2023_Blue-Crab-Advisory-Report_FINAL.pdf

Chesapeake Bay Stock Assessment Committee. (2024, June 27). *2024 Chesapeake Bay Blue Crab Advisory Report*. Chesapeake Bay Program. https://www.chesapeakebay.net/files/documents/2024_Blue-Crab-Advisory-Report_Final.pdf

Chesapeake Bay Stock Assessment Committee. (2025, July 2). *2025 Chesapeake Bay Blue Crab Advisory Report*. Chesapeake Bay Program. https://www.chesapeakebay.net/files/2025_Blue-Crab-Advisory-Report_FINAL-DRAFT.pdf

Chung, J. S., Pitula, J. S., Schott, E., Alvarez, J. V., Maurer, L., & Lycett, K. A. (2015). Elevated water temperature increases the levels of reo-like virus and selected innate immunity genes in hemocytes and hepatopancreas of adult female blue crab, *Callinectes sapidus*. *Fish & Shellfish Immunology*, *47*(1), 511-520. https://doi.org/10.1016/j.fsi.2015.09.027

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C. K., Lavigne, R., Maes, P., Van Ranst, M., Heylen, E., & Matthijnssens, J. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Scientific Reports*, *5*(1), 16532. https://doi.org/10.1038/srep16532

Danovaro, R., Corinaldesi, C., Dell'anno, A., Fuhrman, J. A., Middelburg, J. J., Noble, R. T., & Suttle, C. A. (2011). Marine viruses and global climate change. *FEMS Microbiology Reviews*, *35*(6), 993–1034. https://doi.org/10.1111/j.1574-6976.2010.00258.x

Dhar, A. K., Cowley, J. A., Hasson, K. W., & Walker, P. J. (2004). Genomic organization, biology, and diagnosis of Taura syndrome virus and yellowhead virus of penaeid shrimp. *Advances in Virus Research*, *63*, 353–421. https://doi.org/10.1016/S0065-3527(04)63006-5

Dong, X., Meng, F., Zhou, C., Li, J., Hu, T., Wang, Y., Wang, G., Luo, J., Li, X., Liu, S., Huang, J., & Shi, W. (2024). Enormous diversity of RNA viruses in economic crustaceans. *mSystems*, *9*(10), e0101624. https://doi.org/10.1128/msystems.01016-24

Flowers, E. M., Johnson, A. F., Aguilar, R., & Schott, E. J. (2018). Prevalence of the pathogenic crustacean virus Callinectes sapidus reovirus 1 near flow-through blue crab aquaculture in Chesapeake Bay, USA. *Diseases of Aquatic Organisms*, *129*(2), 135-144. https://doi.org/10.3354/dao03232

Flowers, E. M., Simmonds, K., Messick, G. A., Sullivan, L., & Schott, E. J. (2016). PCR-based prevalence of a fatal reovirus of the blue crab, Callinectes sapidus (Rathbun) along the northern Atlantic coast of the USA. *Journal of Fish Diseases*, *39*(6), 705–714. https://doi.org/10.1111/jfd.12403

Flynn, K. J., Mitra, A., Wilson, W. H., Kimmance, S. A., Clark, D. R., Pelusi, A., & Polimene, L. (2022). 'Boom‐and‐busted'dynamics of phytoplankton–virus interactions explain the paradox of the plankton. *New Phytologist*, *234*(3), 990-1002. https://doi.org/10.1111/nph.18042

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, *399*(6736), 541-548. https://doi.org/10.1038/21119

Fuhrmann, M., Pathirana, E., de Kantzow, M., & Hick, P. (2022). Ostreid herpesvirus disease. In *Aquaculture Pathophysiology* (pp. 473-488). Academic Press. https://doi.org/10.1016/B978-0-323-95434-1.00063-2

Ghabrivirales**.** (2025, March 17). *In* Wikipedia. https://en.wikipedia.org/w/index.php?title=Ghabrivirales&oldid=1280964195

Guo, Z. X., He, J. G., Xu, H. D., & Weng, S. P. (2013). Pathogenicity and complete genome sequence analysis of the mud crab dicistrovirus-1. *Virus Research*, 171(1), 8-14. https://doi.org/10.1016/j.virusres.2012.10.002

Guo, G., Wang, M., Zhou, D., He, X., Han, P., Chen, G., Zeng, J., Liu, Z., Wu, Y., Weng, S., & He, J. (2023). Virome Analysis Provides an Insight into the Viral Community of Chinese

Mitten Crab *Eriocheir sinensis*. *Microbiology Spectrum*, *11*(4), e0143923. https://doi.org/10.1128/spectrum.01439-23

Guo, Z., Zhang, D., Ma, H., Su, Y., Feng, J., & Xu, L. (2014). Rapid detection of mud crab dicistrovirus-1 using loop-mediated isothermal amplification. *Journal of Virological Methods*, 208, 171-176. https://doi.org/10.1016/j.jviromet.2014.08.014

Haugland, Ø., Mikalsen, A. B., Nilsen, P., Lindmo, K., Thu, B. J., Eliassen, T. M., et al. (2011). Cardiomyopathy syndrome of Atlantic salmon (*Salmo salar* L.) is caused by a double-stranded RNA virus of the *Totiviridae* family. *Journal of Virology.* 85, 5275–5286. https://doi.org/10.1128/JVI.02154-10

Hines, A. H., & Ruiz, G. M. (1995). Temporal variation in juvenile blue crab mortality: nearshore shallows and cannibalism in Chesapeake Bay. *Bulletin of Marine Science*, *57*(3), 884-901. https://repository.si.edu/handle/10088/17787

Holt, C. C., Stone, M., Bass, D., Bateman, K. S., van Aerle, R., Daniels, C. L., van der Giezen, M., Ross, S. H., Hooper, C., & Stentiford, G. D. (2019). The first clawed lobster virus Homarus gammarus nudivirus (HgNV n. sp.) expands the diversity of the Nudiviridae. *Scientific Reports*, *9*(1), 10086. https://doi.org/10.1038/s41598-019-46008-y

Hurwitz, B. L., Deng, L., Poulos, B. T., & Sullivan, M. B. (2013). Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology*, *15*(5), 1428-1440. https://doi.org/10.1111/j.1462-2920.2012.02836.x

International Committee on Taxonomy of Viruses (ICTV). (2023). *Dicistroviridae*. In *ICTV Virus Taxonomy: 2023 Release*. https://ictv.global/report/chapter/dicistroviridae/dicistroviridae

International Committee on Taxonomy of Viruses (ICTV). (2023). *Kitaviridae*. In *ICTV Virus Taxonomy: 2023 Release*. https://ictv.global/report/chapter/kitaviridae/kitaviridae

International Committee on Taxonomy of Viruses (ICTV). (2011). *Order Mononegavirales*. In A. M. Q. King, M. J. Adams, E. B. Carstens, & E. J. Lefkowitz (Eds.), *Virus taxonomy: Ninth report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press. https://ictv.global/report_9th/RNAneg/Mononegavirales

International Committee on Taxonomy of Viruses (ICTV). (2024). *Norzivirales*. Retrieved May 5, 2025, from https://ictv.global/taxonomy/taxondetails?taxnode_id=202407201&taxon_name=Norzivirales

Jiang, J. Z., Fang, Y. F., Wei, H. Y., Zhu, P., Liu, M., Yuan, W. G., Yang, L. L., Guo, Y. X., Jin, T., Shi, M., Yao, T., Lu, J., Ye, L. T., Shi, S. K., Wang, M., Duan, M., & Zhang, D. C. (2023). A remarkably diverse and well-organized virus community in a filter-feeding oyster. *Microbiome*, *11*(1), 2. https://doi.org/10.1186/s40168-022-01431-8

Jiang, Jing-Zhe (Ginger), & Wei, H. Y. (2018). Isolation of Viral Like Particles (VLP) from Tissues of Molluscs V.2.Version 2. protocols.io. https://dx.doi.org/10.17504/protocols.io.m4yc8xw

Johnson, P. T. (1977). A viral disease of the blue crab, *Callinectes sapidus*: histopathology and differential diagnosis. *Journal of Invertebrate Pathology*, *29*(2), 201-209. https://doi.org/10.1016/0022-2011(77)90194-X

Johnson, P. T. (1983). Diseases caused by viruses, rickettsiae, bacteria, and fungi. *The biology of Crustacea*, 1-78. https://doi.org/10.1016/B978-0-12-106406-8.50012-1

Johnson, P. T. (1984). Viral diseases of marine invertebrates. *Helgoländer Meeresuntersuchungen*, *37*, 65-98. https://doi.org/10.1007/bf01989296

Jones B. (2012). Transboundary movement of shrimp viruses in crustaceans and their products: a special risk?. *Journal of Invertebrate Pathology*, *110*(2), 196–200. https://doi.org/10.1016/j.jip.2012.01.012

Kibenge, F. S. (2024). Dicistroviruses of crustaceans. In Aquaculture Virology (pp. 617-620). Academic Press. https://doi.org/10.1016/b978-0-323-91169-6.00011-x

Kieft, K., Zhou, Z., & Anantharaman, K. (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*, *8*(1), 90. https://doi.org/10.1186/s40168-020-00867-0

King, A. M. (2011). Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Netherlands: Elsevier Science.

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Kuhn, J. H. (2020). Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and Molecular Biology Reviews : MMBR*, *84*(2), e00061-19. https://doi.org/10.1128/MMBR.00061-19

Koyama, S., Sakai, C., Thomas, C. E., Nunoura, T., and Urayama, S. I. (2016). A new member of the family Totiviridae associated with arboreal ants (*Camponotus nipponicus*). *Archives of Virology*. 161, 2043–2045. https://doi.org/10.1007/s00705-016-2876-x

Koyama, S., Urayama, S., Ohmatsu, T., Sassa, Y., Sakai, C., Takata, M., Hayashi, S., Nagai, M., Furuya, T., Moriyama, H., Satoh, T., Ono, S., & Mizutani, T. (2015). Identification, characterization and full-length sequence analysis of a novel dsRNA virus isolated from the arboreal ant Camponotus yamaokai. *The Journal of General Virology*, *96*(Pt 7), 1930–1937. https://doi.org/10.1099/vir.0.000126

Lang, A. S., Vlok, M., Culley, A. I., Suttle, C. A., Takao, Y., Tomaru, Y., & Ictv Report Consortium (2021). ICTV Virus Taxonomy Profile: *Marnaviridae* 2021. *The Journal of General Girology*, *102*(8), 001633. https://doi.org/10.1099/jgv.0.001633

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 1-16. https://doi.org/10.1186/1471-2105-12-323

Liang, Y., Wang, L., Wang, Z., Zhao, J., Yang, Q., Wang, M., Yang, K., Zhang, L., Jiao, N., & Zhang, Y. (2019). Metagenomic Analysis of the Diversity of DNA Viruses in the Surface and Deep Sea of the South China Sea. *Frontiers in Microbiology*, *10*, 1951. https://doi.org/10.3389/fmicb.2019.01951

Lipcius, R. N., Eggleston, D. B., Heck, K. L., Seitz, R. D., and Montfrans, J. V. (2007). Post-Settlement Abundance, Survival, and Growth of Postlarvae and Young Juvenile Blue Crabs in Nursery Habitats. In The Blue Crab *Callinectes sapidus*. Maryland Sea Grant College, College Park, Maryland, pp.535-564.

Louisiana Direct Seafood. (n.d.). *Crab*. Retrieved April 29, 2025, from https://louisianadirectseafood.com/crab/

Maddison, Wayne P., and David R. Maddison. (2023). *Mesquite: A Modular System for Evolutionary Analysis*. Version 3.81. https://www.mesquiteproject.org.

Mancinelli, G., Bardelli, R., & Zenetos, A. (2021). A global occurrence database of the Atlantic blue crab *Callinectes sapidus*. *Scientific Data*, *8*(1), 111. https://doi.org/10.1038/s41597-021-00888-w

Maryland Department of Natural Resources. (2025). *Blue crabs – Chesapeake Bay*. eRegulations. https://www.eregulations.com/maryland/fishing/blue-crabs-chesapeake-bay

Messick, G. A., & Sindermann, C. J. (1992). *Synopsis of principal diseases of the blue crab, Callinectes sapidus* (NOAA Technical Memorandum NMFS-F/NEC-88). U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service. https://repository.library.noaa.gov/view/noaa/6066

Messyasz, A., Rosales, S. M., Mueller, R. S., Sawyer, T., Correa, A. M., 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

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mojica, K. D., & Brussaard, C. P. (2014). Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiology Ecology*, *89*(3), 495–515. https://doi.org/10.1111/1574-6941.12343

Moksnes, P. O., Lipcius, R. N., Pihl, L., & Van Montfrans, J. (1997). Cannibal–prey dynamics in young juveniles and postlarvae of the blue crab. *Journal of Experimental Marine Biology and Ecology*, *215*(2), 157-187. https://doi.org/10.1016/S0022-0981(97)00052-X

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

NOAA Fisheries. (2025, February 20). *Blue crab*. National Oceanic and Atmospheric Administration. Retrieved April 29, 2025, from https://www.fisheries.noaa.gov/species/blue-crab

NOAA Fisheries. (2022). *Fisheries One Stop Shop (FOSS): Landings*. National Oceanic and Atmospheric Administration. Retrieved October 4, 2023, from https://www.fisheries.noaa.gov/foss/f?p=215:200:14532189465770

Nunan, L. M., & Lightner, D. V. (2011). Optimized PCR assay for detection of white spot syndrome virus (WSSV). *Journal of Virological Methods*, *171*(1), 318–321. https://doi.org/10.1016/j.jviromet.2010.11.015

Orth, R., Montfrans, J. van, and Fishman, J. (1999). A Preliminary Study of Predation on Blue Crabs by Three Fish Predators in a Seagrass Bed. https://doi.org/10.25773/8gt7-n149

Papadopoulos, J. S., & Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, *23*(9), 1073–1079. https://doi.org/10.1093/bioinformatics/btm076

Paul-Pont, I., Dhand, N. K., & Whittington, R. J. (2013). Spatial distribution of mortality in Pacific oysters Crassostrea gigas: reflection on mechanisms of OsHV-1 transmission. *Diseases of Aquatic Organisms*, *105*(2), 127–138. https://doi.org/10.3354/dao02615

Perry, H. M., & VanderKooy, S. (Eds.). (2015). *The blue crab fishery of the Gulf of Mexico, United States: A regional management plan (2015 revision)* (Publication No. 243). Gulf States Marine Fisheries Commission.

Poulos, B. T., Tang, K. F., Pantoja, C. R., Bonami, J. R., & Lightner, D. V. (2006). Purification and characterization of infectious myonecrosis virus of penaeid shrimp. *Journal of General Virology*, *87*(4), 987-996. https://doi.org/10.1099/vir.0.81127-0

Proctor, L. M., & Fuhrman, J. A. (1990). Viral mortality of marine bacteria and cyanobacteria. *Nature*, *343*(6253), 60-62. https://doi.org/10.1038/343060a0

Protein [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] - [cited 2025 Apr 29]. Available from: https://www.ncbi.nlm.nih.gov/protein/

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rambaut, A. (2018). *FigTree v1.4.4* [Computer software]. Institute of Evolutionary Biology, University of Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Ren, J., Song, K., Deng, C., Ahlgren, N. A., Fuhrman, J. A., Li, Y., Xie, X., Poplin, R., & Sun, F. (2020). Identifying viruses from metagenomic data using deep learning. *Quantitative Biology*, *8*(1), 64–77. https://doi.org/10.1007/s40484-019-0187-4

Rhodes, C. J., Truscott, J. E., & Martin, A. P. (2008). Viral infection as a regulator of oceanic phytoplankton populations. *Journal of Marine Systems*, *74*(1-2), 216-226. https://doi.org/10.1016/j.jmarsys.2008.01.005

Rochon, D., Rubino, L., Russo, M., Martelli, G. P., & Lommel, S. (2012). Tombusviridae. In A. M. Q. King, M. J. Adams, E. B. Carstens, & E. J. Lefkowitz (Eds.), *Virus taxonomy: Classification and nomenclature of viruses: Ninth report of the International Committee on Taxonomy of Viruses* (pp. 1111–1138). Elsevier.

Rogers, H. A., Taylor, S. S., Hawke, J. P., & Lively, J. A. A. (2015). Variations in prevalence of viral, bacterial, and rhizocephalan diseases and parasites of the blue crab (*Callinectes sapidus*). *Journal of Invertebrate Pathology*, *127*, 54-62. https://doi.org/10.1016/j.jip.2015.03.002

Rosani, U., Shapiro, M., Venier, P., & Allam, B. (2019). A Needle in A Haystack: Tracing Bivalve-Associated Viruses in High-Throughput Transcriptomic Data. *Viruses*, *11*(3), 205. https://doi.org/10.3390/v11030205

Sakaguchi, S., Urayama, S. I., Takaki, Y., Hirosuna, K., Wu, H., Suzuki, Y., ... & Nakagawa, S. (2022). NeoRdRp: A comprehensive dataset for identifying RNA-dependent RNA polymerases of various RNA viruses from metatranscriptomic data. *Microbes and Environments*, *37*(3), ME22001. https://doi.org/10.1264/jsme2.ME22001

Saxowsky, D. (n.d.). *Purpose of food law (Section 2)*. North Dakota State University. Retrieved April 29, 2025, from https://www.ndsu.edu/pubweb/~saxowsky/aglawtextbk/chapters/foodlaw/470-6704.htm

Seitz, R. D., Lipcius, R. N., & Seebo, M. S. (2005). Food availability and growth of the blue crab in seagrass and unvegetated nurseries of Chesapeake Bay. *Journal of Experimental Marine Biology and Ecology*, *319*(1-2), 57-68. https://doi.org/10.1016/j.jembe.2004.10.013

Shen, W., & Ren, H. (2021). TaxonKit: A practical and efficient NCBI taxonomy toolkit. *Journal of Genetics and Genomics*, *48*(9), 844-850. https://doi.org/10.1016/j.jgg.2021.03.006

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., & Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*, *540*(7634), 539–543. https://doi.org/10.1038/nature20167

Shields, J. D., & Overstreet, R. M. (2003). The blue crab: diseases, parasites and other symbionts. *Faculty Publications from the Harold W. Manter Laboratory of Parasitology*, 426. https://digitalcommons.unl.edu/parasitologyfacpubs/436

Sõmera, M., Fargette, D., Hébrard, E., Sarmiento, C., & ICTV Report Consortium. (2021). ICTV Virus Taxonomy Profile: Solemoviridae. *Journal of General Virology*, *102*(1), 001707. https://doi.org/10.1099/jgv.0.001707

Spitznagel, M. I., Small, H. J., Lively, J. A., Shields, J. D., & Schott, E. J. (2019). Investigating risk factors for mortality and reovirus infection in aquaculture production of soft-shell blue crabs (*Callinectes sapidus)*. *Aquaculture*, *502*, 289-295. https://doi.org/10.1016/j.aquaculture.2018.12.051

Stentiford, G. D., Neil, D. M., Peeler, E. J., Shields, J. D., Small, H. J., Flegel, T. W., Vlak, J. M., Jones, B., Morado, F., Moss, S., Lotz, J., Bartholomay, L., Behringer, D. C., Hauton, C., & Lightner, D. V. (2012). Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. *Journal of Invertebrate Pathology*, *110*(2), 141–157. https://doi.org/10.1016/j.jip.2012.03.013

Suttle, C. A., Chan, A. M., & Cottrell, M. T. (1990). Infection of phytoplankton by viruses and reduction of primary productivity. *Nature*, *347*(6292), 467-469. https://doi.org/10.1038/347467a0

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

SIB Swiss Institute of Bioinformatics. (n.d.). *Narnaviridae*. ViralZone. Retrieved May 5, 2025, from https://viralzone.expasy.org/46

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232-W235. https://doi.org/10.1093/nar/gkw256

U.S. Congress. (n.d.). *21 U.S. Code § 342 – Adulterated food*. U.S. House of Representatives Office of the Law Revision Counsel. Retrieved April 29, 2025, from https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title21-section342&num=0&edition=prelim

U.S. Environmental Protection Agency. (2024, November 15). *Ocean dumping of fish wastes*. https://www.epa.gov/ocean-dumping/ocean-dumping-fish-wastes

U.S. Food and Drug Administration. (2023, January 18). *Food Code 2022* (10th ed.). https://www.fda.gov/media/164194/download?attachment

van der Loos, L. M., De Coninck, L., Zell, R., Lequime, S., Willems, A., De Clerck, O., & Matthijnssens, J. (2023). Highly divergent CRESS DNA and picorna-like viruses associated with bleached thalli of the green seaweed *Ulva*. *Microbiology Spectrum*, *11*(5), e00255-23. https://doi.org/10.1128/spectrum.00255-23

Virginia Marine Resources Commission. (2024, December 1). *2024 Annual Report on the Blue Crab Fishery Management Plan* (RD835). Commonwealth of Virginia. https://rga.lis.virginia.gov/Published/2024/RD835/PDF

Walters, E. A., Crowley-McIntyre, C. E., Gandy, R. L., Martin, C. W., Shea, C. P., & Behringer, D. C. (2025). Estimating discard mortality in Florida's blue crab *Callinectes sapidus* fishery using reflex impairment, injuries, and salinity. *Marine and Coastal Fisheries*, *17*(1), vtae001. https://doi.org/10.1093/mcfafs/vtae001

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wilkins D (2023). *gggenes: Draw Gene Arrow Maps in 'ggplot2'*. R package version 0.5.0, https://wilkox.org/gggenes/

Wolf, Y. I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., Krupovic, M., Fire, A., Dolja, V. V., & Koonin, E. V. (2020). Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nature Microbiology*, *5*(10), 1262–1270. https://doi.org/10.1038/s41564-020-0755-4

Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews*, *64*(1), 69-114. https://doi.org/10.1128/MMBR.64.1.69-114.2000

Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E. C., Li, W. X., & Ding, S. W. (2010). Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(4), 1606–1611. https://doi.org/10.1073/pnas.0911353107

Zhao, M., Behringer, D. C., Bojko, J., Kough, A. S., Plough, L., dos Santos Tavares, C. P., ... & Schott, E. J. (2020). Climate and season are associated with prevalence and distribution of trans-hemispheric blue crab reovirus (Callinectes sapidus reovirus 1). *Marine Ecology Progress Series*, *647*, 123-133. https://doi.org/10.3354/meps13405

Zhao, M., Plough, L. V., Behringer, D. C., Bojko, J., Kough, A. S., Alper, N. W., Xu, L., & Schott, E. J. (2023). Cross-Hemispheric Genetic Diversity and Spatial Genetic Structure of *Callinectes sapidus* Reovirus 1 (CsRV1). *Viruses*, *15*(2), 563. https://doi.org/10.3390/v15020563

Zhao, M., Xu, L., Bowers, H., & Schott, E. J. (2022). Characterization of Two Novel Toti-Like Viruses Co-infecting the Atlantic Blue Crab, *Callinectes sapidus*, in Its Northern Range of the United States. *Frontiers in Microbiology*, *13*, 855750. https://doi.org/10.3389/fmicb.2022.855750

Zhou, D., Liu, S., Guo, G., He, X., Xing, C., Miao, Q., Chen, G., Chen, X., Yan, H., Zeng, J., Zheng, Z., Deng, H., Weng, S., & He, J. (2022). Virome Analysis of Normal and Growth Retardation Disease-Affected Macrobrachium rosenbergii. *Microbiology Spectrum*, *10*(6), e0146222. https://doi.org/10.1128/spectrum.01462-22