

# **Ecology and Evolution of Giant Viruses in Deep Oceans Based on Long-term Multiomics Data and Global Metagenomic Data**

**Wenwen Liu**

## **Publications**

**Liu W**, Nagasaka K, Wu J, Ban H, Mimick E, Meng L, Neches RY, Moniruzzaman M, Yoshida T, Nishimura Y, Endo H, Okazaki Y, Ogata H. Giant viruses specific to deep oceans show persistent presence and activity. *mSystems* 10:e00932-25. (2025).

DOI: 10.1128/msystems.00932-25

The major findings presented in Chapters 2 and 3 of this thesis were previously published in the article listed above, which was published under the CC-BY 4.0 license (Creative Commons Attribution 4.0 International License). For detailed information about the CC-BY 4.0 license, please refer to the following link: <https://creativecommons.org/licenses/by/4.0/>

## **Open Access for this thesis**

This article is licensed under a Creative Commons Attribution 4.0 International license. This license permits copying and redistribution of the material in any medium or format, as well as remixing, transforming, and building upon the material for any purpose, provided that you comply with the license terms. Please refer to the following for details and to see if any changes were made: <https://creativecommons.org/licenses/by/4.0/>

## **Contributions**

Wenwen Liu conducted all sampling, laboratory experiments, and bioinformatics analyses. The global ocean dataset analysis builds on Komei Nagasaka's master thesis. Some of the sampling and analysis efforts were supported by co-authors of the associated publication listed above.

# Table of Contents

|                                                                                  |           |
|----------------------------------------------------------------------------------|-----------|
| <b>Table of Contents .....</b>                                                   | <b>2</b>  |
| <b>List of Figures .....</b>                                                     | <b>4</b>  |
| <b>List of Tables.....</b>                                                       | <b>6</b>  |
| <b>Abstract.....</b>                                                             | <b>7</b>  |
| <b>Chapter 1 General Introduction .....</b>                                      | <b>9</b>  |
| 1.1 The deep-sea ecosystem.....                                                  | 10        |
| 1.1.1 Microbial communities in deep oceans.....                                  | 10        |
| 1.1.2 Viruses and their ecological roles in deep oceans.....                     | 12        |
| 1.2 The discovery and characteristics of giant viruses.....                      | 13        |
| 1.2.1 Discovery of giant viruses.....                                            | 13        |
| 1.2.2 Host range and infection strategies of giant viruses.....                  | 14        |
| 1.2.3 Auxiliary metabolic genes and functional innovations in giant viruses..... | 16        |
| 1.3 Ecology and evolution roles of giant viruses .....                           | 17        |
| 1.3.1 Global and vertical distribution of giant viruses in the oceans .....      | 17        |
| 1.3.2 Ecological and evolution roles of giant viruses in the oceans.....         | 18        |
| 1.4 Methodological approaches to deep-sea studies .....                          | 19        |
| 1.4.1 Cultivation-dependent and cultivation-independent approaches .....         | 20        |
| 1.4.2 Deep-sea water sampling approaches .....                                   | 21        |
| 1.5 Objectives.....                                                              | 22        |
| <b>Chapter 2 Materials and Methods.....</b>                                      | <b>24</b> |
| 2.1 Sample collection .....                                                      | 24        |
| 2.2 Metabarcoding sequencing and analyses .....                                  | 26        |
| 2.3 Metagenomic sequencing and analyses.....                                     | 27        |
| 2.4 Metatranscriptomic sequencing and analyses .....                             | 29        |

|                                                                                                                |           |
|----------------------------------------------------------------------------------------------------------------|-----------|
| 2.5 Compiling the GVGR database and identifying deep-sea-specific GVMAGs .....                                 | 30        |
| 2.6 Protein prediction and annotation .....                                                                    | 32        |
| 2.7 Analyses of phylogenetic diversity and clade delineation .....                                             | 33        |
| <b>Chapter 3 Results and Discussion.....</b>                                                                   | <b>35</b> |
| 3.1 Data overview .....                                                                                        | 35        |
| 3.2 Mesopelagic microeukaryote community is distinct from that in the surface water at the sampling site ..... | 37        |
| 3.3 GVs exclusively detected in mesopelagic metagenomes show persistency at the study site .....               | 41        |
| 3.4 Mesopelagic-specific giant viruses are active across seasons.....                                          | 45        |
| 3.5 Mesopelagic GVMAGs encode specific gene functions.....                                                     | 48        |
| 3.6 Putative hosts of meso-exclusive GVMAGs.....                                                               | 52        |
| 3.7 Deep-sea-specific GVs distribute widely in the global ocean.....                                           | 56        |
| <b>Chapter 4 Conclusions and Perspectives .....</b>                                                            | <b>65</b> |
| 4.1 Summary .....                                                                                              | 65        |
| 4.2 Limitations and future plans.....                                                                          | 66        |
| <b>Data Availability.....</b>                                                                                  | <b>70</b> |
| <b>Acknowledgements.....</b>                                                                                   | <b>71</b> |
| <b>References .....</b>                                                                                        | <b>72</b> |

## List of Figures

|                                                                                                                                                    |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1 Schematic illustration of the ocean layers. ....                                                                                          | 10 |
| Figure 2 Timeline of major discoveries and methodological advances in giant virus research.<br>.....                                               | 13 |
| Figure 3 Sampling location and depth.....                                                                                                          | 24 |
| Figure 4 Pipeline for generating the giant virus genome reference (GVGR) database.....                                                             | 30 |
| Figure 5 Microeukaryote community composition of the seawater samples.....                                                                         | 37 |
| Figure 6 Comparison of mean relative abundance (logTPM+1) of taxa at class level<br>classification based on 18S rDNA/rRNA metabarcoding data. .... | 39 |
| Figure 7 Comparison of Levins' standardized niche breadth (BA) between surface and<br>mesopelagic ASVs.....                                        | 40 |
| Figure 8 Genome size distribution and relative abundance of Muroto GVMAGs in the<br>metagenomic and metatranscriptomic data. ....                  | 41 |
| Figure 9 Phylogenetic analysis of PolB from the Muroto samples.....                                                                                | 42 |
| Figure 10 Phylogenetic and taxonomic composition of Muroto GVMAGs. ....                                                                            | 43 |
| Figure 11 Dynamics of the presence and activity of Muroto GVMAGs. ....                                                                             | 44 |
| Figure 12 Gene expression profiles of six meso-exclusive GVMAGs. ....                                                                              | 46 |
| Figure 13 The relative abundance of Muroto GVMAGs between metagenomic and<br>metatranscriptomic datasets across seasons and depths. ....           | 47 |
| Figure 14 Top 20 highly expressed KEGG orthology (KO) terms for the Muroto GVMAGs.<br>.....                                                        | 50 |
| Figure 15 Assignment of eukaryotic-like protein sequences in the Muroto GVMAGs.....                                                                | 52 |
| Figure 16 Phylogenetic trees of eukaryotic-like protein sequences from meso-exclusive GV<br>and their homologs. ....                               | 53 |
| Figure 17 Co-occurring meso-exclusive GV and microeukaryotes. ....                                                                                 | 54 |
| Figure 18 Geographic distribution of 1,890 marine metagenomic datasets used in this study.<br>.....                                                | 56 |
| Figure 19 Distribution of all GVMAGs in the GVGR database and 101 deep-sea-specific<br>GVMAGs across global ocean depth layers. ....               | 57 |

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| Figure 20 Phylogenetic diversity, abundance, and biogeography of deep-sea-specific GVMAGs. .... | 59 |
| Figure 21 Functional enrichment of deep-sea-specific GVMAGs.....                                | 61 |
| Figure 22 Distribution of deep-sea-specific KOs across deep-sea-specific GVMAGs. ....           | 62 |

## **List of Tables**

|                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------|----|
| Table 1 Sampling methods used in this study .....                                                   | 25 |
| Table 2 Total DNA and RNA extracted from the Muroto samples collected at the mesopelagic layer..... | 35 |
| Table 3 Total DNA and RNA extracted from the Muroto samples collected at the surface layer .....    | 35 |
| Table 4 Environmental parameters measured at the Kochi Prefectural Deep Seawater Laboratory .....   | 36 |

## Abstract

Giant viruses (GVs) of the phyla *Nucleocytoviricota* and *Mirusviricota* are large double-stranded DNA viruses that infect diverse eukaryotic hosts and impact biogeochemical cycles. While their diversity and ecological roles have been extensively studied in the photic layer of the ocean, much less is known about their activity, population dynamics, and adaptive strategies in aphotic layers. However, deep-sea sampling poses significant challenges due to low microbial abundance and the difficulty of distinguishing resident viruses from those attached to sinking particles. To address these gaps, this dissertation combines time-series samplings from a mesopelagic coastal site with global metagenomic datasets to explore the ecology and evolution of GV in deep oceans.

First, I developed and implemented a deep-sea sampling protocol at the Kochi Prefectural Deep Seawater Laboratory in Muroto, Japan (the Muroto station). Eight seasonal time-series samplings were conducted for the surface and mesopelagic layers using the station's established deep-sea water pump system. I then integrated 18S rRNA gene metabarcoding, metagenomic, and metatranscriptomic data to investigate mesopelagic GV and their potential hosts. To further investigate the distribution and phylogenomic features of GV at a global scale across broader depths, I compiled 4,473 species-level GV genomes from the OceanDNA MAG project and other resources, and analyzed them using 1,890 marine metagenomes.

The analysis based on Muroto seawater identified 48 GV genomes, including six that were exclusively detected in the mesopelagic layer. Notably, these mesopelagic-specific GV showed persistent activity across seasons. Genes associated with the ubiquitin system were highly expressed in mesopelagic-specific GV, implying roles in modulating host cell death or protein turnover under low-energy conditions. Furthermore, co-occurrence and phylogenomic analyses linked mesopelagic-specific GV to multiple protist taxa, including Cercozoa, Picozoa, Syndiniales, dinoflagellates, and stramenopiles implying a diverse host range among heterotrophic or mixotrophic microeukaryotes.

The analysis based on the global ocean dataset revealed 101 deep-sea-specific GVs distributed across the GV phylogenetic tree, indicating that adaptation to deep-sea environments has occurred in multiple lineages. One clade enriched with deep-sea-specific GVs, including a mesopelagic-specific GV genome from the Muroto dataset, was found to be distributed across different ocean basins, highlighting their cosmopolitan nature. Comparative genomics revealed that 76 KEGG orthologs and 74 Pfam domains were specifically enriched in deep-sea-specific GVs, encompassing functions related to the ubiquitin system, energy metabolism, and nitrogen acquisition. These findings collectively support the hypothesis that distinct GV lineages may have evolved to adapt to deep-sea environments by acquiring genes to manipulate host cellular processes, thereby allowing them to thrive in the aphotic environment.

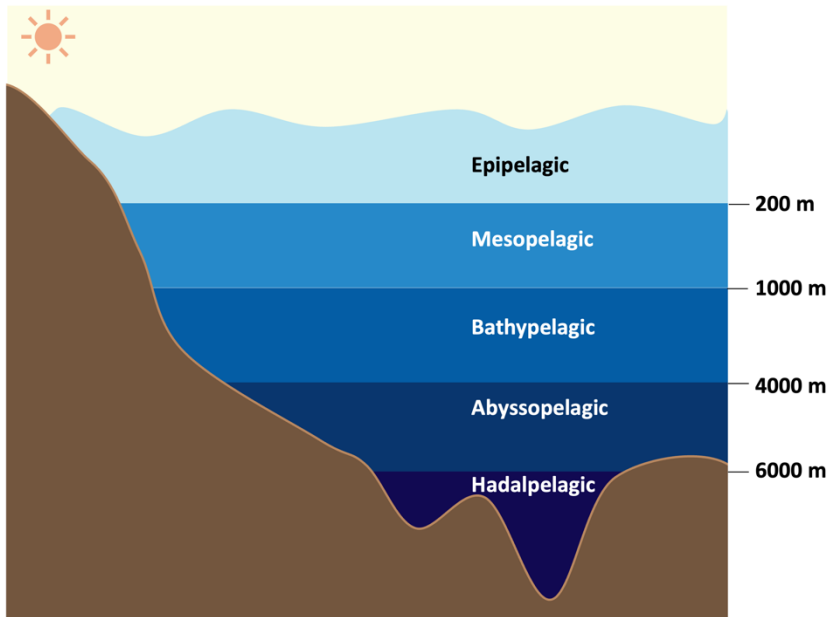
Collectively, this dissertation establishes a methodological framework for sampling and analyzing deep-sea GVs, identifies stable and active mesopelagic GVs likely infecting microeukaryotes endemic to the deep ocean, and reveals that multiple GV lineages are broadly distributed and have independently adapted to the deep-sea environments.

# Chapter 1 General Introduction

The oceans cover over 70% of the Earth's surface and contain approximately 97% of the planet's water, forming the largest and most dynamic ecosystem on Earth. Marine ecosystems are dominated by unicellular organisms—including archaea, bacteria, and eukaryotes—that play essential roles in global biogeochemical cycles, such as carbon fixation and nutrient recycling (Worden et al., 2015).

The deep ocean—defined as waters below 200 meters—represents approximately 95% of the global ocean volume (Danovaro et al., 2017). Although microorganisms and viruses are abundant in marine environments, their roles in the deep sea remain only partially understood. Among those, giant viruses (GVs) have recently emerged as a fascinating group due to their unusually large size and complex genomes (Chung et al., 2025; Gaïa et al., 2023; Schulz et al., 2022). In recent decades, the discovery of GV has reshaped the understanding of viral diversity and evolution. While their ecological roles in the sunlit ocean have been increasingly recognized, little is known about their activity and adaptation in the deep ocean. This chapter introduces the current understanding of deep-sea ecosystems and their microbial communities, reviews the current knowledge of GV in marine environments, and outlines the methodologies employed to study GV. Finally, it presents the research questions and objectives of this dissertation. Throughout this chapter, existing knowledge gaps in the deep-sea ecosystem will be highlighted to motivate the research questions addressed in this dissertation.

## 1.1 The deep-sea ecosystem



**Figure 1** Schematic illustration of the ocean layers.

The deep ocean is commonly divided into zones by depth: the mesopelagic (200–1,000 m), bathypelagic (1,000–4,000 m), abyssal (4,000–6,000 m), and hadal zones (>6,000 m) (Figure 1). These vast regions are characterized by the absence of sunlight, low temperature, high hydrostatic pressure, and high inorganic nutrient concentrations (Aristegui et al., 2009). As sunlight no longer penetrates below the mesopelagic zone, photosynthesis ceases entirely, and the ecosystem relies primarily on the vertical flux of organic matter and *in situ* chemoautotrophic production using reduced inorganic compounds (e.g., ammonia, carbon monoxide) (Dick et al., 2013). The deep ocean plays a critical role in global biogeochemical cycles, such as carbon sequestration and nutrient regeneration for sustained oceanic productivity (Herndl & Reinthaler, 2013; Thurber et al., 2014).

### 1.1.1 Microbial communities in deep oceans

Deep-sea microbial communities show distinct vertical stratification (Cordier et al., 2022) and often exhibit slow metabolic rates and novel strategies to utilize scarce resources. These reactions have an important effect on global carbon sequestration in the oceans (Jiao

& Zheng, 2011), with prokaryotes (bacteria and archaea) serving as entry points of carbon for the food web of the dark ocean. Many deep-sea prokaryotes can degrade recalcitrant dissolved organic matter or utilize alternative energy sources (Herndl et al., 2023). For example, Thaumarchaeota proliferate in the mesopelagic and deeper layers, where they oxidize ammonium to nitrite (Shi et al., 2023).

In contrast, deep-sea microeukaryotes (primarily protists) are less well studied due to low abundance, complex genomes, and diverse lifestyles. They occupy crucial ecological roles as grazers, predators, symbionts, and potential viral hosts, influencing microbial food webs and biogeochemical processes (Worden et al., 2015). Unlike the protist community, which showed a dynamic seasonality in the epipelagic layer, deep-sea microeukaryotes are more stable over time but distinct in composition, reflecting the relatively constant physical conditions of the deep ocean (Blanco-Bercial et al., 2022). Many deep-sea protists exhibit metabolic plasticity, switching between heterotrophy and mixotrophy to survive in energy-limited environments (Worden et al., 2015).

Although protist abundances decline with depth—from an average of approximately 72 cells/mL in mesopelagic waters to about 11 cells/mL in bathypelagic zones (Pernice et al., 2014), studies have revealed unexpectedly high species richness and compositional distinctiveness in the deep ocean (Cordier et al., 2022). The protists in the deep ocean are not merely passive or dormant components of the microbial community. Approximately 14.1% of amplicon sequence variants (ASVs) identified from deep-sea protists are unique to aphotic layers (Cordier et al., 2022). 18S rDNA/rRNA-based analyses suggest that while some taxa may be dormant, others like ciliates remain transcriptionally active (Giner et al., 2020), highlighting their functional relevance. Through associations with particles and interactions with viruses and prokaryotes, protists also serve as key components of the deep-sea microbial loop and likely hosts for large DNA viruses.

### 1.1.2 Viruses and their ecological roles in deep oceans

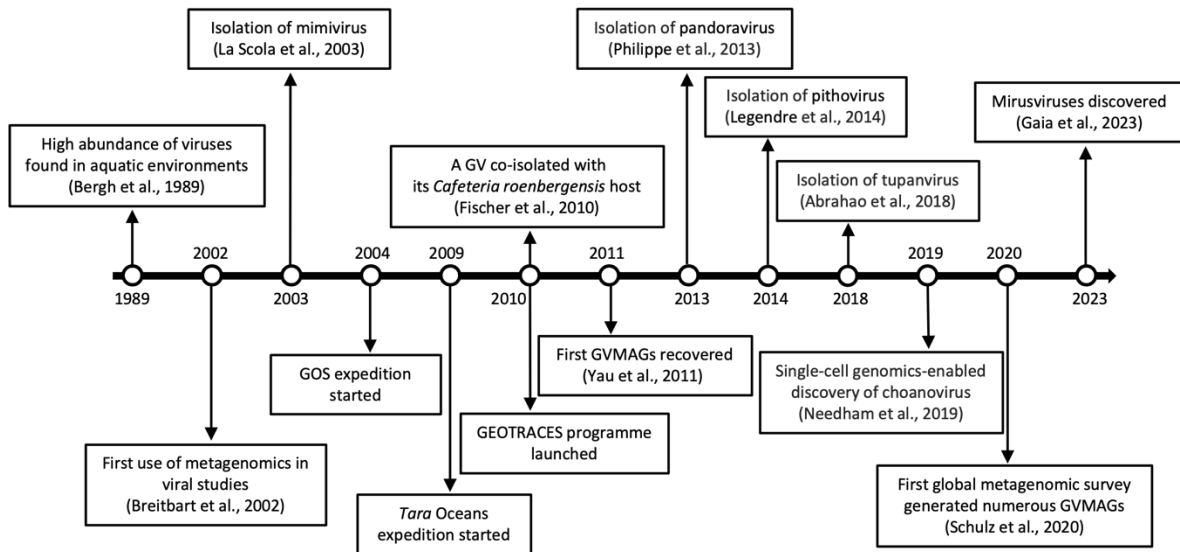
Viruses are the most abundant biological entities in the ocean and play pivotal roles in microbial ecosystems (Suttle, 2007). By infecting microbial hosts and inducing cell lysis, viruses regulate microbial community structure, mediate horizontal gene transfer, and recycle nutrients, eliminating up to 20% of microbial biomass in the ocean each day (Suttle, 2007). In the deep ocean, although energy input and host density are reduced, viral abundance remains substantial ( $0.6\text{--}60\times 10^8$  particles/L), around 0.5 to 2 orders of magnitude lower than in surface waters (Nagata et al., 2010).

Viral community composition in the deep sea varies with depth, region, and environmental factors such as particle flux and hydrostatic pressure (Han et al., 2024; Middelboe et al., 2006; Thomas et al., 2021). Deep-sea viruses often favor lysogeny—integrating into host genomes and persisting in a dormant state—as a survival strategy under low host density (Jian et al., 2021; Mizuno et al., 2016). Nonetheless, lytic cycles can still dominate in some bathypelagic regions, where virus-induced prokaryotic mortality significantly exceeds protistan grazing mortality (Lara et al., 2017). Viral lysis in nutrient-poor environments also plays a key role in sustaining deep-sea food webs by releasing labile dissolved organic matter that fuels microbial metabolism and nutrient turnover (Yakimov et al., 2007). Beyond mortality, viruses influence host physiology through auxiliary metabolic genes (AMGs), which can reprogram host metabolism to enhance survival. For example, some cyanophage genomes encode more than 20 AMGs that can alter the electron transport chain or enhance the carbon metabolism of their hosts (Heyerhoff et al., 2022).

Recently, metagenomic surveys have uncovered the deep ocean as a reservoir of novel DNA viruses, with unique genomic repertoires potentially shaped by deep-sea ecosystem properties (T. He et al., 2024). Yet, a major unresolved question is whether these viruses are truly endemic to deep waters or descend from surface layers via sinking particles. Disentangling these possibilities is critical for understanding the ecological role and evolutionary trajectory of deep-sea viruses.

## 1.2 The discovery and characteristics of giant viruses

### 1.2.1 Discovery of giant viruses



**Figure 2 Timeline of major discoveries and methodological advances in giant virus research.**

This timeline highlights key milestones in the study of aquatic viruses and giant viruses, including early recognition of the high abundance of viruses in aquatic environments (Bergh et al., 1989), the first application of metagenomics to viral ecology (Breitbart et al., 2002), and the discovery and isolation of multiple giant virus lineages (Abrahão et al., 2018; Fischer et al., 2010; Gaia et al., 2023; La Scola et al., 2003; Legendre et al., 2014; Philippe et al., 2013). Major oceanographic and metagenomic initiatives, such as the Global Ocean Sampling (GOS) expedition, the *Tara* Oceans expedition, and the GEOTRACES program, are shown alongside technological advances including viral metagenomics, single-cell genomics (Needham et al., 2019), and the recovery of giant virus metagenome-assembled genomes (GVMAGs) (Schulz, Roux, et al., 2020; Yau et al., 2011).

For over a century, viruses were assumed to be small and genetically simple. This assumption was overturned in 2003 with the discovery of mimivirus, isolated from an amoeba (La Scola et al., 2003) (Figure 2). Mimivirus measures approximately 700 nm in diameter and possesses an approximate 1.2 Mb double-stranded DNA genome encoding nearly 1,000 genes, many of which were previously thought to be exclusive to cellular organisms (Raoult et al., 2004). This milestone prompted a surge of interest in large DNA viruses and redefined the upper limits of viral size and complexity. In 2013, pandoravirus

was isolated from marine and freshwater sediments, harboring the largest viral genome recorded to date (2.5 Mb encoding nearly 2500 proteins) (Philippe et al., 2013). In 2014, pithovirus was reported to have the largest virion (approximately 1.5  $\mu\text{m}$ ), exceeding the size of many bacterial cells (Legendre et al., 2014). The isolation of tupanvirus has further expanded the known diversity of GVs. Discovered in 2018 from a Brazilian soda lake and deep-sea sediment, tupanvirus was found to carry an extensive array of translation-related genes and exhibited a unique tail-like morphology (Abrahão et al., 2018). Despite their structural and genomic diversity, these viruses cluster phylogenetically within the phylum *Nucleocytoviricota*, a group now encompassing most known GVs (Koonin et al., 2020).

Recent environmental metagenomic initiatives have significantly accelerated GV discovery. Global surveys such as *Tara* Oceans have recovered hundreds of novel GV genomes from marine samples, shedding light on their diversity, ecology, and evolution (Moniruzzaman et al., 2020; Schulz, Roux, et al., 2020). One particularly notable addition is the recently proposed phylum *Mirusviricota*, whose members possess capsid genes similar to those of *Herpesvirales* while clustering with *Nucleocytoviricota* based on other genomic features, suggesting a mosaic evolutionary origin (Gaïa et al., 2023).

Together, these findings reveal the unexpected genomic and morphological complexity of GVs and highlight the power of environmental genomics to uncover the megavirome, a previously unseen portion of the virosphere with broad implications for understanding virus–host coevolution and ecosystem dynamics.

### **1.2.2 Host range and infection strategies of giant viruses**

GVs infect an extraordinarily broad range of eukaryotes, from unicellular protists to multicellular animals (Schulz, Roux, et al., 2020). Early discoveries revealed that members of *Poxviridae*, *Ascoviridae*, *Iridoviridae*, and *Asfarviridae* infect various animals, while algae-associated GVs have been identified infecting diverse microalgae, including chlorophytes, haptophytes, pelagophytes, brown algae, and dinoflagellates (Sun et al., 2020). In laboratory systems, the heterotrophic protist *Acanthamoeba* has served as a key model

host, enabling the isolation and characterization of diverse GV lineages such as *Megamimivirinae*, *Marseilleviridae*, pandoraviruses, molliviruses, and medusaviruses (Legendre et al., 2015; Pagnier et al., 2013; Philippe et al., 2013; Yoshikawa et al., 2019). Recent large-scale host prediction efforts based on sequencing data suggest that Dikarya, Oomycetes, Streptophytina, and fungi emerged as putative hosts for a wide array of GV lineages, despite the absence of direct GV isolation from these taxa (Schulz, Roux, et al., 2020). Additionally, giant endogenous viral elements (GEVEs) identified in eukaryotic genomes—particularly those from mirusviruses—indicate past infections across a diverse range of protists, from algae to fungal-like organisms, thereby expanding the known host spectrum of GVs (Zhao et al., 2024).

GVs display diverse infection strategies and replication mechanisms vary by lineage (Schulz et al., 2022). Mimiviruses and related GVs typically enter host cells via phagocytosis, whereas others, such as marseilleviruses, utilize endocytosis or macropinocytosis (Rodrigues et al., 2016). Once inside the host, GVs establish viral factories that facilitate genome replication and virion assembly, followed by virion release through lysis or exocytosis (Schulz et al., 2022). Notably, a recent study demonstrated that a GEVE in *Chlamydomonas reinhardtii* can be reactivated to produce viral particles without causing apparent host damage, suggesting cryptic infection as a potential alternative interaction mode in GV-host systems (Erazo-Garcia et al., 2025).

In sum, marine GVs have extensive and intimate interactions with eukaryotic hosts. Their infections probably occur in all major eukaryotic lineages (Schulz, Roux, et al., 2020) and drive long-term co-evolution, prompting GVs to evolve sophisticated mechanisms for host manipulation. The next subsection explores one of the outcomes of this co-evolution: the incorporation of host-derived genes into viral genomes, yielding the so-called AMGs that grant GVs unique functional capabilities.

### 1.2.3 Auxiliary metabolic genes and functional innovations in giant viruses

AMGs are not strictly required for viral replication but enable GVs to manipulate host metabolic pathways, effectively transforming the host into a virus-driven metabolic system, or “virocell” (Rosenwasser et al., 2016). These genes are often acquired via horizontal gene transfer and retained for their fitness benefits during infection. GVs encode a wide array of metabolic and informational genes, covering pathways such as central carbon metabolism, lipid biosynthesis, nutrient transport, oxidative stress response, and even parts of the cytoskeleton and translation machinery (Barrado-Gil et al., 2017; Ha et al., 2021; Kijima et al., 2021; Moniruzzaman et al., 2020).

Tupanvirus, for example, harbors the most comprehensive set of translational genes among viruses, including parts of the ribosomal protein synthesis system (e.g., dozens of tRNAs, aminoacyl-tRNA synthetases, and translation factors). This complex genetic toolkit may enable better exploitation of hosts or facilitate survival under extreme conditions (Abrahão et al., 2018). *Emiliana huxleyi* viruses (EhV) carry genes for glycosphingolipid and sterol synthesis that reprogram host membranes, influence viral assembly, and even trigger programmed cell death in the host, facilitating efficient virion release (Wilson et al., 2005). One chlorovirus encodes a potassium ion channel, which depolarizes the host cell membrane immediately upon infection, likely aiding the early stages of viral entry and gene expression (Frohns et al., 2006; Kang et al., 2005). Beyond metabolism, GVs also encode actin-related proteins which might assist in reorganizing the host cell’s cytoskeleton to facilitate virus assembly or movement within the cell (Da Cunha et al., 2022). Recent findings show that GV-encoded aminoacyl-tRNA synthetases may trace back to proto-eukaryotic origins (Kijima et al., 2024).

Altogether, AMGs expand the functional potential of GVs, enabling precise control over host processes. Their presence illustrates the evolutionary entanglement between viruses and host metabolism and highlights the unique biology of GVs, particularly in challenging environments like the deep ocean.

## 1.3 Ecology and evolution roles of giant viruses

### 1.3.1 Global and vertical distribution of giant viruses in the oceans

Marine metagenomic studies have revealed the high abundance (Hingamp et al., 2013), activity (Carradec et al., 2018; Ha et al., 2021), diversity (Li et al., 2018; Mihara et al., 2018; Schulz, Roux, et al., 2020), and widespread geographic GV distribution in the ocean (Bäckström et al., 2019; Endo et al., 2020; Meng et al., 2023).

Surveys based on single marker genes indicated the widespread presence of viruses from the families *Phycodnaviridae* and *Mimiviridae* in diverse marine metagenomes, with these viruses being more abundant than eukaryotes in the photic layer (Hingamp et al., 2013). Subsequent efforts assembled complete GV genomes from *Tara* Oceans data, further confirming that GVs are widespread in marine ecosystems (Schulz, Roux, et al., 2020). Indeed, one survey reported that the richness of GV clades (specifically *Mimiviridae*) in surface oceans exceeded the richness of bacterial clades in the same samples (Mihara et al., 2018).

To date, studies have predominantly focused on the ocean's photic layer, while aphotic layers remain comparatively underexplored owing to the inherent difficulty in accessing these environments. Of note, all marine GVs have been isolated from the photic layer, except for one case from deep-sea sediment (Abrahão et al., 2018). Several previous studies nevertheless detected the presence of GV-related sequences in deep-sea environments, including genomic contigs (Jian et al., 2021; Mizuno et al., 2016; Sheam & Luo, 2025) and conserved marker genes (Acinas et al., 2021; Endo et al., 2020). Another study reported the transcription signals from mirusvirus genomes in mesopelagic waters based on *Tara* Oceans data (Gaïa et al., 2023). More recently, five GV metagenome assembled genomes (GVMAGs) were shown to be unique to water below 150 m (Ha et al., 2023) and a few GVMAGs were specifically and highly abundant in the deep layers of the North Pacific Subtropical Gyre (Sheam et al., 2025). These results clearly indicate the existence of GVs in deep oceans. However, whether these GVs represent deep-sea residents actively infecting

sympatric hosts or are just passively transported from the surface community through sinking remains unclear. For example, Endo *et al.* showed that the GV lineages detected in the mesopelagic layer are mostly (99%) also present in the photic layer and suggested that they are largely settled from the photic layer (Endo *et al.*, 2020). Sheam *et al.* also provided evidence that some GVs detected in sediment trap samples (4,000 m depth) were vertically transported from the surface layer (Sheam *et al.*, 2025). Nonetheless, a recent study revealed the existence of a GV community specific to an aphotic layer (65 m depth) of a freshwater lake (Zhang *et al.*, 2024), implying the existence of marine GVs specifically inhabiting dark environments.

Environmental factors such as temperature, host distribution, and nutrient availability are likely drivers of GV biogeography. For instance, a recent study found that GVs in polar oceans show genomic signs of adaptation to cold, polar conditions (Meng *et al.*, 2023). Similarly, GVs infecting *Micromonas* (a widespread picophytoplankton) were found to phylogenetically partition into several groups with cryophile or cryo-mesophile preferences, suggesting temperature-driven biogeography (Demory *et al.*, 2025). Depth may be another factor that can influence the GV communities. These findings suggest that GVs have undergone evolutionary diversification, enabling them to occupy a variety of ecological niches, including extreme environments such as the deep ocean.

### **1.3.2 Ecological and evolution roles of giant viruses in the oceans**

In marine environments, GVs affect the population dynamics of their hosts (Funaoka *et al.*, 2023; Hevroni *et al.*, 2023; Wang *et al.*, 2023), alter host metabolism via virus-induced metabolic reprogramming (Monier *et al.*, 2009; Rosenwasser *et al.*, 2016), and consequently impact carbon and nutrient cycles (Kaneko *et al.*, 2020). Their community structure closely mirrors that of microeukaryotes in time (Fang *et al.*, 2025; Prodinger *et al.*, 2022) and space (Xia *et al.*, 2022), suggesting strong ecological coupling.

These effects are particularly evident during algal blooms. For example, EhV infection significantly increases the release of acidic polysaccharides and particulate inorganic carbon,

enhancing carbon export to the deep ocean (Vincent et al., 2023). A global study further demonstrated that the composition of the eukaryotic virus community (dominated by GVs) is a strong predictor of carbon export efficiency from the surface ocean (Kaneko et al., 2020), reinforcing their role in carbon cycling.

Beyond ecology, GV–host interactions reflect evolutionary arms races. Hosts evolve resistance mechanisms, while viruses adapt to overcome them. Recent studies revealed that GVs maintain stable infections by producing polymorphic variants, allowing continued exploitation of bloom-forming dinoflagellates (Wang et al., 2023). One well-known outcome of this co-evolution is virus-driven metabolic reprogramming of the host cell. The high demand for metabolic building blocks by GVs may result in an environmental trade-off between susceptibility to viral infection and acclimation to abiotic environmental stresses. For example, viral infection weakened lipid remodeling process under phosphorus-limited conditions (Maat et al., 2016). GVs also mediate horizontal gene transfer. Their large genomes often harbor AMGs acquired from hosts, enabling viruses to reprogram host metabolism for viral benefit. GEVEs found in eukaryotic genomes further illustrate this long-term virus–host genetic exchange.

Altogether, GVs are emerging as key players in marine microbial dynamics and evolution, with significant roles in ecosystem function and genetic innovation.

## **1.4 Methodological approaches to deep-sea studies**

Investigating GVs in aphotic layers is challenging because of methodological and logistical difficulties (Wu et al., 2011), as GVs have lower abundance than bacteria in the same size fraction, especially in deeper layers (Hingamp et al., 2013). Therefore, researchers need to develop specialized approaches to sample, detect, identify, and analyze these viruses. This section outlines the key methodologies employed, focusing on culture-dependent or independent techniques and deep-sea sampling strategies. These approaches set the stage for the experimental design of the present study.

### 1.4.1 Cultivation-dependent and cultivation-independent approaches

Cultivation-dependent isolation and cultivation-independent metagenomics are the two primary methodologies that have driven the discovery and characterization of GVs.

Cultivation-based techniques, especially co-culture with amoebae, remain essential for isolating viral particles and characterizing their genomes, proteomes, and infection cycles. For instance, classical lysis plaque assays (Boughalmi et al., 2013) and transmission electron microscopy (TEM) based visualization (Athira et al., 2023) have led to the recovery of many GVs. More recently, automated systems such as flow cytometry-based detection (Khalil et al., 2016) and high-content screening microscopy (Francis et al., 2019) have accelerated virus detection. Despite automation, these approaches remain more labor-intensive than sequencing. As of now, more than 200 GV genomes have been recovered from isolates (Schulz et al., 2022), representing only a tiny fraction compared to environmental sequencing results. Moreover, these methods are restricted to viruses with culturable hosts, thereby excluding most deep-sea protists, and laboratory conditions often fail to replicate natural viral infection dynamics.

In contrast, cultivation-independent methods such as metagenomics and metatranscriptomics have vastly expanded our view of GV diversity and activity. These approaches have revealed the presence of GVs across nearly all sampled environments including the deep ocean. Metagenomic binning enables near-complete genome recovery without host isolation, although challenges such as misclassification and non-detection remain (Schulz, Andreani, et al., 2020). Large-scale projects such as the *Tara* Oceans and GEOTRACES have helped to build a comprehensive database like the Global Ocean Eukaryotic Viral database (GOEV), which includes 1,897 GVMAGs (Gaïa et al., 2023). Importantly, metatranscriptomics identifies transcriptionally active viruses, offering insight into ongoing infections and enabling researchers to distinguish active residents from passively transported particles—especially critical in deep-sea environments. Single-cell metagenomics and metatranscriptomics further enable the identification of native hosts (Fromm et al., 2024; Wilson et al., 2017). Additionally, community profiling through 18S

rRNA gene sequencing can be used in conjunction with co-occurrence network analysis to predict virus–host interactions (Meng et al., 2021). Time-series datasets also provide temporal context for viral succession and host–virus dynamics (Fang et al., 2025; Proding et al., 2022).

Integration of multiomics data, including metagenomics, metatranscriptomics, and marker gene sequencing, offers a powerful framework for enhancing functional resolution and virus–host linkage in deep-sea environments.

#### **1.4.2 Deep-sea water sampling approaches**

Sampling microbial communities from the deep ocean poses significant challenges due to low biomass, high pressure, and difficulties in preserving labile nucleic acids. Advances in deep-sea sampling techniques have improved the ability to study viruses in aphotic environments.

The most common approach for collecting deep-sea water is via conductivity-temperature-depth (CTD) rosette systems equipped with Niskin bottles, which can be triggered at designated depths to retrieve water samples ranging from tens to hundreds of liters. Global ocean surveys such as *Tara* Oceans (Hingamp et al., 2013), Malaspina Expedition (2010–2011) (Sánchez et al., 2024), and Bio-GEOTRACES have used this method to sample mesopelagic and bathypelagic waters, often followed by on-deck filtration for downstream DNA or RNA extraction. Time-series observatories such as HOT (Hawaii Ocean Time-Series) (Mende et al., 2017) and BATS (Bermuda Atlantic Time-Series) (He et al., 2024) have provided depth-resolved metagenomic datasets. Recently, transcriptomic studies in the deep ocean have also emerged. For example, metatranscriptomic data were collected at a depth of approximately 10,000 m in the Mariana Trench using a CTD–Niskin bottle system (Dang et al., 2024).

To detect GV and assess viral activity in the deep ocean, large volumes, often hundreds of liters, are needed for filtration and concentration. Moreover, transcriptomic analyses require rapid sampling and preservation to avoid RNA degradation and physiological

changes in microbial cells (Steiner et al., 2019). *In situ* large-volume filtration systems attached to moorings or submersibles allow high-throughput filtering at depth, minimizing sample degradation.

In Japan, infrastructure such as deep-sea water suction pipes has similarly supported deep-sea research, with such systems installed at seven locations by 2003 (Yamaguchi et al., 2003). Among these, the Kochi Prefectural Deep Seawater Laboratory in Muroto, Kochi Prefecture, Japan, provides access to oligotrophic coastal water influenced by the Kuroshio Current (Guo Y.J., 1991). The laboratory has the capability to continuously draw untreated surface (0.5 m) and mesopelagic (320 m) seawater (Yamaguchi et al., 2003), and supports scientific research on deep-sea water and various industrial applications such as food processing and aquaculture (Kang et al., 2004). This facility continuously pumps untreated mesopelagic water (320 m depth) to the surface, enabling rapid collection of deep water for virus concentration and RNA extraction. This setup allows for real-time access to deep-sea water, facilitating prompt viral concentration and RNA extraction for gene expression studies.

## 1.5 Objectives

In deep-sea studies, it is essential not only to detect the presence of viruses but also to determine whether they are actively infecting hosts. This distinction is particularly important in aphotic environments, where viruses may be transported from surface waters and persist as inactive particles. Metatranscriptomics provides critical insight into viral activity by identifying transcriptionally active viruses, enabling the differentiation between active infections and passive presence. However, such data from deep oceans remain scarce, leaving substantial gaps in understanding the ecology and evolution of GVs in these environments.

To address this knowledge gap, this thesis pursues two main objectives:

The first objective is to investigate the existence and activity of GVs, as well as the sympatric potential host community, in the mesopelagic layer. Eight seasonal sampling campaigns at the Muroto station provided paired surface and mesopelagic samples for 18S

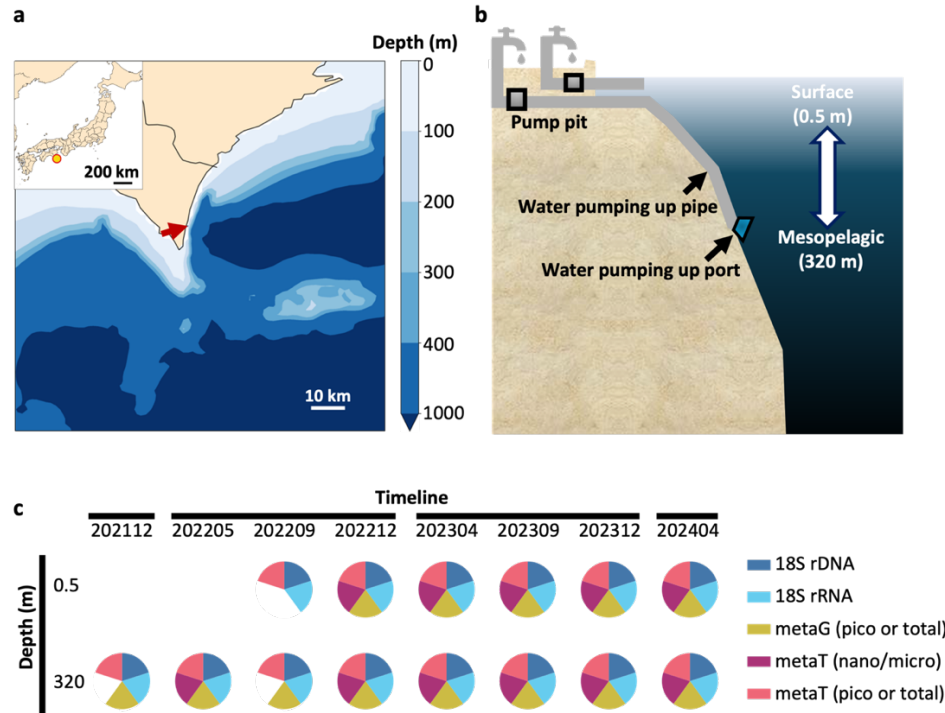
rRNA gene metabarcoding, metagenomic, and metatranscriptomic analyses. These data were used to reconstruct GVMAGs, track their abundance and activity over time, and infer virus–host associations via homology searches, phylogenomic analyses, and co-occurrence networks. This approach moves beyond simple presence–absence surveys to uncover whether GVs are persistent and transcriptionally active, and to identify their potential hosts in the mesopelagic layer.

The second objective is to identify the distribution and adaptation of GVs in deep oceans at a global scale. A GV genome reference (GVGR) dataset was compiled from publicly available data (Collier et al., 2023; Fang et al., 2025; Gaïa et al., 2023; Nishimura & Yoshizawa, 2022; Zhang et al., 2024; Zhao et al., 2024) and analyzed using 1,897 metagenomic samples from the global ocean. This global analysis identified deep-sea-specific GV lineages, distinguished them from cosmopolitan types, and examined their potential adaptations to aphotic environments using phylogenetic inference and AMG profiling.

This thesis integrates insights from both the Muroto time-series data and the global ocean dataset to evaluate the ecological and evolutionary roles of GVs in deep oceans, while also discussing current methodological limitations and proposing future directions.

## Chapter 2 Materials and Methods

### 2.1 Sample collection



**Figure 3 Sampling location and depth.**

(a) Location of the sampling site at the Kochi Prefectural Deep Seawater Laboratory. (b) Schematic illustration of surface and mesopelagic seawater intake equipment. (c) Diagram showing the sampling year and month as well as depth, either surface (0.5 m) or mesopelagic (320 m). 18S rDNA: DNA-based 18S rRNA metabarcoding; 18S rRNA: cDNA-based 18S rRNA metabarcoding; metaG (pico or total): metagenomic sequencing of the pico- or total-fraction; metaT (nano/micro): metatranscriptomic sequencing of the nano/micro-fraction; metaT (pico or total): metatranscriptomic sequencing of the pico- or total-fraction. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Eight time-series seawater samples were collected from two depths (surface: 0.5 m, mesopelagic: 320 m) at the Kochi Prefectural Deep Seawater Laboratory, Kochi Prefecture, Japan (Figure 3a), from December 2021 to April 2024. The seawater was pumped directly

from the ocean and obtained via the deep-sea water intake system at the facility (Yamaguchi et al., 2003).

**Table 1 Sampling methods used in this study**

This table is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

| Sampling date | Depth (m) | Pore size (μm) | Filtered water (L) | Method                                                                                             |
|---------------|-----------|----------------|--------------------|----------------------------------------------------------------------------------------------------|
| 2021.12.17    | 320       | 0.2–3          | 30                 | Water filtered through 3 μm membrane, followed by 0.2 μm Sterivex filter, using a peristaltic pump |
| 2021.12.17    | 320       | 3–144          | 220                | Water filtered through 3 μm membrane, using an air compact pump                                    |
| 2022.05.26    | 320       | 0.2–5          | 42                 | Water filtered through 5 μm membrane, followed by 0.2 μm Sterivex filter, using a peristaltic pump |
| 2022.05.26    | 320       | 5–144          | 1110               | Water filtered through 5 μm plankton net                                                           |
| 2022.09.29    | 320       | 0.2–5          | 49                 | Water filtered through 5 μm membrane, followed by 0.2 μm Sterivex filter, using a peristaltic pump |
| 2022.09.29    | 320       | 5–150          | 600                | Water filtered through 5 μm plankton net                                                           |
| 2022.12.21    | 320       | 0.2–150        | 33.2               | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2022.12.21    | 320       | 3–150          | 50.55              | Water filtered through 3 μm membrane, using an air compact pump                                    |
| 2023.04.19    | 320       | 0.2–150        | 69.00              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2023.04.19    | 320       | 3–150          | 185.80             | Water filtered through 3 μm membrane, using tap water pressure                                     |
| 2023.09.22    | 320       | 0.2–150        | 56.25              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2023.09.22    | 320       | 3–150          | 284.47             | Water filtered through 3 μm membrane, using tap water pressure                                     |
| 2023.12.15    | 320       | 0.2–150        | 56.80              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2023.12.15    | 320       | 3–150          | 263.60             | Water filtered through 3 μm membrane, using tap water pressure                                     |
| 2024.04.25    | 320       | 0.2–150        | 43.95              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2024.04.25    | 320       | 3–150          | 297.90             | Water filtered through 3 μm membrane, using tap water pressure                                     |
| 2022.09.29    | 0.5       | 0.2–5          | 12.6               | Water filtered through 5 μm membrane, followed by 0.2 μm Sterivex filter, using a peristaltic pump |
| 2022.09.29    | 0.5       | 5–150          | 10.6               | Water filtered through 5 μm membrane, using an air compact pump                                    |
| 2022.12.21    | 0.5       | 0.2–150        | 13.7               | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2022.12.21    | 0.5       | 3–150          | 26.8               | Water filtered through 3 μm membrane, using an air compact pump                                    |
| 2023.04.19    | 0.5       | 0.2–150        | 55.60              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2023.04.19    | 0.5       | 3–150          | 31.90              | Water filtered through 3 μm membrane, using an air compact pump                                    |
| 2023.09.22    | 0.5       | 0.2–150        | 42.90              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2023.09.22    | 0.5       | 3–150          | 38.65              | Water filtered through 3 μm membrane, using an air compact pump                                    |
| 2023.12.15    | 0.5       | 0.2–150        | 49.05              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2023.12.15    | 0.5       | 3–150          | 38.85              | Water filtered through 3 μm membrane, using an air compact pump                                    |
| 2024.04.25    | 0.5       | 0.2–150        | 30.00              | Water filtered through 0.2 μm Sterivex filter, using a peristaltic pump                            |
| 2024.04.25    | 0.5       | 3–150          | 37.55              | Water filtered through 3 μm membrane, using an air compact pump                                    |

The filtration system with membranes directly connected to the mesopelagic water tap was built to ensure that the water remained in its original condition (Figure 3b, Table 1). After pre-filtration through a 150-μm-pore-size nylon mesh to remove large organisms, a large volume of seawater, especially for mesopelagic samples (30–1110 L) (Table 1), was

filtered through 3- $\mu\text{m}$ - or 5- $\mu\text{m}$ -pore-size polycarbonate membranes (Merck, Germany) and 0.2- $\mu\text{m}$ -pore-size Sterivex-GP PES filters (Merck). Three groups of size fractionations were performed. The first group (3.0–150  $\mu\text{m}$  and 5.0–150  $\mu\text{m}$ ) mainly corresponds to nano- and microplankton and is referred to as the “nano/micro-fraction”. The second group (0.2–150  $\mu\text{m}$ ) covers pico-, nano-, and microplankton and is referred to as the “total-fraction”. The last group (0.2–3.0  $\mu\text{m}$  and 0.2–5.0  $\mu\text{m}$ ) mainly corresponds to picoplankton and is referred to as the “pico-fraction”. I collected pico-fraction samples and nano/micro-fraction samples in early samplings (until September 2022), and total-fraction samples and nano/micro-fraction samples in later samplings. All filtrated membranes were immediately frozen and transported in a cryo-shipper at liquid nitrogen temperatures to the laboratory and stored at  $-80^{\circ}\text{C}$  until DNA and RNA were extracted. Filtration and cryo-preservation of each filter were completed within 25 minutes.

Total DNA and RNA were extracted using the AllPrep RNA/DNA kit (Qiagen, Germany) following the protocol described by Okazaki *et al.* (Okazaki *et al.*, 2023). The DNA/RNA quantity was assessed using a Qubit fluorometer (Thermo Fisher Scientific, USA). The extracted DNA/RNA was used for 18S rDNA metabarcoding (28 samples), 18S rRNA metabarcoding (28 samples), metagenomic (13 samples), and metatranscriptomic (25 samples) sequencing (Figure 3c). Environmental variables, including temperature, salinity, and pH were recorded continuously by the Kochi Prefectural Deep Seawater Laboratory.

## 2.2 Metabarcoding sequencing and analyses

Microeukaryote amplicon sequencing was performed for both DNA and RNA-derived cDNA (rDNA and rRNA sets, respectively). For RNA, genomic DNA removal and first-strand cDNA synthesis were conducted using the SuperScript IV VILO Master Mix with the ezDNase enzyme kit (Thermo Fisher Scientific) following the manufacturer’s protocol. The V4 region of the 18S rRNA gene from both cDNA and DNA was amplified by the KAPA HiFi HotStart ReadyMix (Roche) using the universal eukaryotic primer set E572F/E1009R (Comeau *et al.*, 2011) attaching Illumina overhang adapters. The PCR conditions were as

follows: initial pre-denaturation at 95°C for 2 minutes, followed by 30 cycles of denaturation at 98°C for 20 seconds, annealing at 61°C for 15 seconds, and extension at 72°C for 30 seconds, with a final extension at 72°C for 2 minutes. Triplicate PCR products were pooled and then purified using VAHTS DNA Clean Beads (Vazyme). The amplicon libraries were sequenced on the Illumina MiSeq platform to generate paired-end reads ( $2 \times 300$  bp).

Raw sequencing reads were processed using QIIME2 v2024.10 (Bolyen et al., 2019) with the DADA2 plugin (Callahan et al., 2016). The “dada2 denoise-paired” command was used for adapter removal, primer removal, low-quality read trimming, dereplication, chimera removal, and amplicon sequence variant (ASV) identification. Rare ASVs that appeared in less than two sequences across all samples were excluded. Taxonomic classification was performed on the remaining ASVs based on a pre-trained naive Bayes classifier trained against the PR2 reference database v5.0.0 using the “feature-classifier classify-sklearn” plugin (Bokulich et al., 2018). ASVs classified as non-protist lineages (i.e., metazoa, fungi) were removed. Finally, 1,616,977 reads from 56 samples were grouped into 5,031 ASVs. To enable comparisons between samples, the ASV table was rarefied to the minimum read depth per sample (7,581 reads), resulting in a final dataset of 4,981 ASVs and 424,536 reads. Statistical differences among sample groups (e.g., depth and season) were tested by PERMANOVA (9,999 permutations) on the Bray–Curtis dissimilarity calculated from the Hellinger-transformed ASV abundance table, using the Vegan v2.6.10 package in R (R Core Team, 2024). The standardized Levins index (BA) (Krebs, 1989; Levins, 1968), representing the niche breadth, was calculated using the R package “spaa”. For each depth and size fraction, BA was estimated from the relative abundance of ASVs in the surface and mesopelagic samples collected on the same day.

## **2.3 Metagenomic sequencing and analyses**

Metagenomic shotgun sequencing libraries were constructed using the Illumina DNA Prep kit (Illumina) and sequenced on the Illumina NovaSeq 6000 platform, generating 150-bp paired-end reads with an average of 58 Gbp per sample. Raw reads were quality-controlled

using fastp v0.23.4 (Chen et al., 2018) with default parameters. For each metagenome, trimmed reads were assembled into contigs using MEGAHIT v1.2.9 (Li et al., 2015) in “meta-sensitive” mode. Contigs over 2.5 kb were retained for downstream analysis. Trimmed reads from all samples were primarily used for cross-mapping against contigs assembled from individual samples by Bowtie2 v2.4.5 (Langmead et al., 2019). Additionally, a co-assembly was performed by pooling reads from all mesopelagic samples. Contigs were clustered into bins using MetaBAT2 v2.15.15 (Kang et al., 2019). Metagenomic bins were generated for each of the 13 metagenomes and for the pooled mesopelagic metagenomes. Genes in MAGs were predicted using Prodigal-gv v2.11.0 (Camargo et al., 2023; Hyatt et al., 2010) with “meta” mode for marker gene identification and functional annotation.

NCV-MAGs and MV-MAGs were reconstructed separately following previously published methods (Fang et al., 2025; Zhang et al., 2024). To identify NCV-MAGs, I used a custom pipeline named hedera v.0.0.5 (<https://github.com/banhbio/hedera>). Briefly, the pipeline employs a gene density index [NCLDV index (Fang et al., 2025)] calculated using the genome size and the presence of 20 marker genes selected from the Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOG) (Yutin et al., 2009). MAGs with an NCLDV index over 5.75 are identified as potential NCV-MAGs. The NCV verification and decontamination is performed using the results of Viralrecall v2.1 (score > 0) (Aylward & Moniruzzaman, 2021), Virsorter2 v2.2.3 (max score group “NCLDV”) (Guo et al., 2021), CAT v5.2.3 (“Nucleocytoviricota”) (Von Meijenfeldt et al., 2019), and hidden Markov models (HMMs) built with 149 NCVOGs (E-value <  $1.0 \times 10^{-3}$ ) (Fang et al., 2025). Bins selected by hedera were further investigated for the possibility of chimeric bins (e.g., multiple viral genomes) by manual inspection using Anvi’o v8 (Eren et al., 2015). First, I primarily focused on deep branching clades that showed markedly different occurrence patterns or GC contents from other clades. If such clades showed the seven core marker genes of GVs (PolB, A32 packaging enzyme, superfamily II helicase, VLTF3 transcriptional factor, topoisomerase family II, transcription factor IIB, and RNA polymerase second largest subunit [RNAPS]), they were retained as a different bin, otherwise, they were discarded. Second, for the remaining clades, if I identified clearly different patterns of occurrence or GC content, I split

the bins into two or more sub-bins. Finally, the modified bins were re-assessed for the NCLDV index and only those with an NCLDV index over 5.75 were retained. For MV-MAGs, bins were identified as a mirusvirus if the HK97 MCP gene was detected using HMMER v3.4 (E-value <  $1.0 \times 10^{-3}$ , bit score > 100) (Mistry et al., 2013). After manual inspection with Anvi'o, MAGs containing a contig encoding HK97-MCP were retained. Decontamination was performed using CheckV v1.0.1 (database v1.4) (Nayfach et al., 2021) by concatenating all contigs of each MAG and using “end\_to\_end” mode to remove prokaryotic contigs and provirus segments (Zhang et al., 2024). Dereplication was conducted using dRep v3.5.0 (Olm et al., 2017) at an average nucleotide identity (ANI) of 95% with the parameters “--ignoreGenomeQuality --S\_algorithm ANImf -sa 0.95 --clusterAg single -sizeW 1”. The resulting non-redundant Muroto GVMAGs (NCV-MAGs and MV-MAGs) were species-level representatives. GVMAG names were assigned as follows: “NCV” or “MV”, followed by a serial number indicating the rank of genome size from large to small (e.g., NCV10, MV5).

To determine the relative abundances of Muroto GVMAGs, the coverage and transcripts (or reads) per million (TPM) (Mortazavi et al., 2008) of each MAG were calculated by mapping all quality-controlled trimmed reads to Muroto GVMAGs using CoverM v0.6.1 (Aroney et al., 2025).

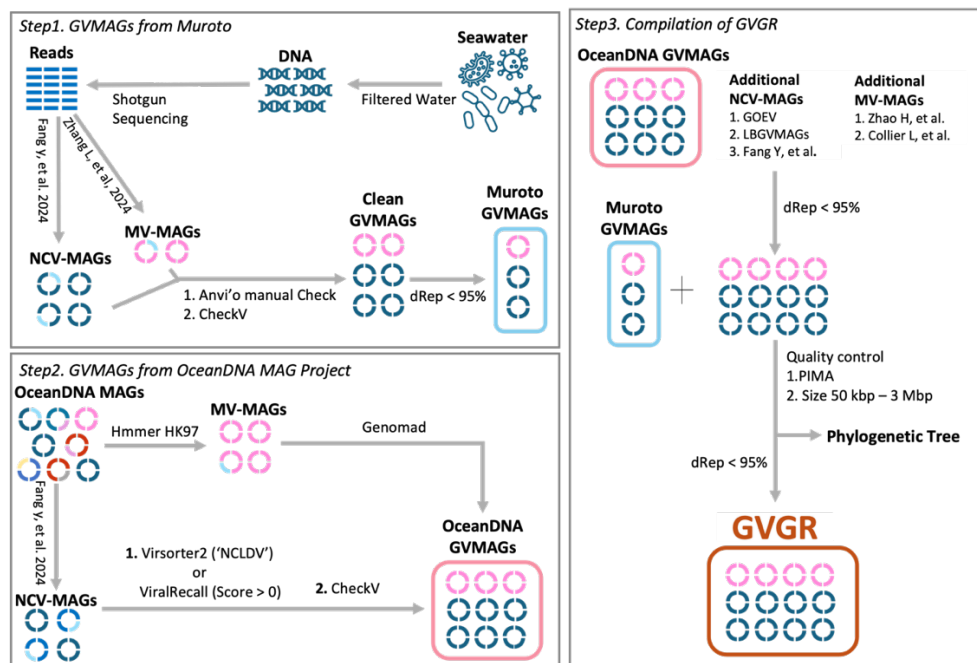
## 2.4 Metatranscriptomic sequencing and analyses

RNA samples were subjected to metatranscriptomic sequencing (Figure 3c). Poly-A strand-specific RNA sequencing libraries were constructed using the NEBNext® Poly(A) mRNA Magnetic Isolation Module and NEBNext® Ultra™ II Directional RNA Library Prep Kit (New England Biolabs, USA). Libraries were sequenced on the Illumina NovaSeq 6000 platform and 150-bp paired-end reads were generated. An average of 24 Gbp of sequence per sample was obtained. Raw reads were quality-controlled using fastp v0.23.4 with the “-l 100” parameter to remove low-quality reads and reads shorter than 100 bp. SortMeRNA v4.3.6 (Kopylova et al., 2012) was used to filter out rRNA reads against the Silva v138.1 LSU

NR99, Silva v138.1 SSU NR99, RFAM 5s, and RFAM 5.8s databases using default settings. Filtered mRNA reads were mapped against the Muroto GVMAGs using Salmon v1.10.1 (Patro et al., 2017) with the parameters “--meta --minScoreFraction 0.95 --validateMappings”. Transcript abundance was normalized as TPM. The results were used to quantify the relative expression level of each gene and MAG.

## 2.5 Compiling the GVGR database and identifying deep-sea-specific GVMAGs

To investigate the global deep-sea GV distribution and adaptation, a comprehensive giant virus genome reference was compiled, named GVGR, by integrating the Muroto GVMAGs with publicly available GV metagenomic datasets (Figure 4).



**Figure 4 Pipeline for generating the giant virus genome reference (GVGR) database.**

The pipeline consists of three steps: (1) manually curated GVMAGs (Muroto GVMAGs) generated from Muroto metagenomic data in this study; (2) GVMAGs (OceanDNA GVMAGs) extracted from the OceanDNA MAG project; and (3) integration of Muroto GVMAGs, OceanDNA GVMAGs, and previously published GVMAG datasets to establish the comprehensive GVGR database. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

NCV-MAGs from metagenomic bins generated using MetaBAT2 by the OceanDNA MAG project were extracted (Nishimura et al., 2022). Metagenomic bins with an NCLDV index over 5.75 were considered as potential NCV-MAGs. These MAGs were decontaminated through retaining contigs identified as “NCLDV” by Virsorter2 v2.2.4 or those with a ViralRecall v2.1 score greater than 0. Host-derived sequences and proviral segments were removed using CheckV v1.0.1, as mentioned above. For MV-MAG classification, the initial screening was consistent with the process described for Muroto MV-MAGs. Contigs classified as “chromosome” or “plasmid” using geNomad v1.8.0 (Camargo et al., 2023) were excluded. MAGs with a size between 50 kbp–3 Mbp in length were retained as GVMAGs.

Additionally, I incorporated publicly available GV genomes from multiple sources, including GOEVdb (Gaia et al., 2023), 1,065 GVMAGs from Uranouchi Inlet, Japan (Fang et al., 2025), 293 LBGVMAGs from Lake Biwa (Zhang et al., 2024), and four circular endogenous *Mirusviricota* (Collier et al., 2023; Zhao et al., 2024). All collected MAGs were pooled and dereplicated using dRep v3.5 at an ANI of 95% as previously mentioned. Within the ANI-based cluster, the MAGs with the highest N50 value among the top 30% largest sized genomes were selected as representative.

To assess MAG quality, a phylogeny-informed MAG assessment (PIMA) based on a phylogenetic tree and orthologous groups (OGs) was performed on the above-mentioned dataset and Muroto GVMAGs (Fang et al., 2025). For phylogenetic tree reconstruction, seven conserved marker genes were identified using the script “ncldv\_markersearch.py”(Moniruzzaman et al., 2020). For MV-MAGs, four marker genes (HK97-MCP, RNAPS, RNA polymerase largest subunit [RNAPL], and PolB) were identified using HMMER v3.4 (Mistry et al., 2013) with pairwise alignment against the reference HMM profiles. These marker genes were aligned and concatenated using Clustal Omega v1.2.4 (Sievers et al., 2014), trimmed with trimAl v1.2.1 (-gt 0.1) (Capella-Gutiérrez et al., 2009) and a phylogenetic tree was built with FastTree v2.1.11 (Price et al., 2010) using default parameters, serving as an input for PIMA of NCV-MAGs and MV-MAGs, separately.

OGs were classified using GVOG HMM profiles (Aylward et al., 2021) using HMMER v3.4 (E-value  $< 1.0 \times 10^{-10}$ ) (Mistry et al., 2013). If multiple OGs were assigned to a single sequence, the OG with the lowest E-value was retained.

PIMA was applied at a relative evolutionary divergence threshold of 0.65, corresponding to genus- or family-level classification (Fang et al., 2025). MAG consistency was evaluated based on the proportion of core genes present, while redundancy was calculated as the proportion of duplicated core genes exceeding the mode copy number observed across all MAGs within each lineage. MAGs exhibiting redundancy over 50% were removed. A second round of dRep dereplication was subsequently performed to integrate Muroto GVMAGs, producing the final GVGR dataset. The taxonomy of GVMAGs in the GVGR dataset was validated using TIGTOG (Ha et al., 2024). Gene prediction and abundance analyses in all 1,890 global samples (Nishimura et al., 2022) were performed using methods identical to those described above.

Each of these GVMAGs was categorized as “deep-sea-specific” or “non-deep-sea-specific”. Genomes showing overrepresentation in deep-sea samples (depth  $> 200$  m) were considered as deep-sea-specific GVMAGs. The overrepresentation was ascertained using Mann–Whitney U tests with Benjamini–Hochberg (BH) corrected  $P$  values ( $P$ -value  $< 0.05$ ) (Benjamini et al., 1995). Genomes with signals only in deep-sea samples (TPM  $> 0$ ) were also considered as deep-sea-specific GVMAGs. Other GVMAGs, not assigned to deep-sea-specific category, were referred to as “non-deep-sea-specific”.

## 2.6 Protein prediction and annotation

Proteins predicted from Muroto GVMAGs and the GVGR database were annotated with KEGG orthology (KO) using KofamScan v1.3.0 (Aramaki et al., 2020). KO annotations with the lowest E-values were retained. Furthermore, Pfam annotations of the genomes using AnnoMazing (<https://github.com/BenMinch/AnnoMazing>) were generated, a pipeline to annotate proteins based on HMM profiles. The annotations were performed using the Pfam database (Mistry et al., 2021) with an E-value cut-off of  $1.0 \times 10^{-5}$ .

To identify KO terms and Pfam domains significantly enriched in deep-sea-specific GVMAGs, Fisher's exact test was conducted to compare the frequency of each annotated KO term in deep-sea-specific GVMAGs and non-deep-sea specific GVMAGs. KO terms with a total frequency  $\leq 2$  were excluded from further analysis.

To infer the potential eukaryotic hosts of GVs, protein homology searches were performed using BLASTP in DIAMOND v2.1.10 (E-value  $< 1.0 \times 10^{-5}$  and identity  $> 50\%$ ) (Buchfink et al., 2021) against the MarFERReT ver1.1.1 database (Groussman et al., 2023). Hits with the lowest E-values were retained and used to assign candidate host taxonomy. For each retained protein, up to 100 homologs were further retrieved from a combined MarFERReT and NR database using DIAMOND BLASTP (E-value  $< 1.0 \times 10^{-5}$ , identity  $> 50\%$ ). These homologous sequence sets were subsequently included in the phylogenetic analyses described below to assess putative host associations suggested by horizontal gene transfer.

## 2.7 Analyses of phylogenetic diversity and clade delineation

A phylogenetic tree was first constructed based on PolB sequences ( $> 500$  amino acids) extracted from Muroto contigs ( $> 1$  kbp). The GOEV database and a wide range of eukaryotic and additional viral lineages (Gaïa et al., 2023; Kazlauskas et al., 2020) were used as a reference set of PolB sequences to evaluate the completeness of GV recovery in Muroto GVMAGs. Another tree for *Mirusviricota* was generated using HK97-MCP sequences from Muroto MV-MAGs, combined with previously reported *Mirusviricota* genomes from the GOEV database, LBGVMAGs, and circular endogenous MV-MAGs. For *Nucleocytoviricota*, a concatenated phylogenetic tree was constructed using the seven core marker genes (Aylward et al., 2021) for Muroto NCV-MAGs, along with 220 reference genomes from culture (Gaïa et al., 2023). Another concatenated phylogenetic tree was generated using the same seven core marker genes, incorporating all deep-sea-specific MAGs, Muroto meso-exclusive MAGs, and the GVGR dataset redundant by dRep v3.5.0 at an ANI of 85%. The marker gene search, alignment, concatenation, and trimming methods

were the same as those described for the PIMA above. The tree was inferred using IQ-TREE v2.2.0 (Nguyen et al., 2015) with the ModelFinder Plus option to determine the best-fitting model. Support values were inferred using 1,000 ultrafast bootstraps (Hoang et al., 2018). Tree visualization and rooting were carried out using iTOL v7 (Letunic et al., 2024).

## Chapter 3 Results and Discussion

### 3.1 Data overview

The pump systems at the study site (i.e., off the coast of Muroto; Figure 3) enabled me to collect and filter a large volume of water from two depths (0.5 m and 320 m) in a short time (Table 1). Size fractionations were performed to concentrate different microorganisms: pico-fraction (0.2–3.0  $\mu\text{m}$  or 0.2–5.0  $\mu\text{m}$ ), total-fraction (0.2–150  $\mu\text{m}$ ), and nano/micro-fraction (3.0–150  $\mu\text{m}$  or 5.0–150  $\mu\text{m}$ ).

**Table 2 Total DNA and RNA extracted from the Muroto samples collected at the mesopelagic layer**

This table is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

| Sampling date | Total DNA/RNA yield and sampled water volume from mesopelagic samples |                     |                   |                     |
|---------------|-----------------------------------------------------------------------|---------------------|-------------------|---------------------|
|               | nano/micro                                                            |                     | pico or total     |                     |
|               | DNA                                                                   | RNA                 | DNA               | RNA                 |
| 2021.12.17    | 300 ng (110 L)                                                        | <10 ng (110 L)      | 1422 ng (50 L)    | 31 ng (50 L)        |
| 2022.05.26    | 238 ng (1110 L)                                                       | 1778 ng (1110)      | 3099 ng (42 L)    | 992 ng (42 L)       |
| 2022.09.29    | 19 ng (400 L)                                                         | 79 ng (400 L)       | 2478 ng (56 L)    | 1077 ng (56 L)      |
| 2022.12.21    | 49 ng (51 L)                                                          | 242 ng (51 L)       | 251 ng (16 L)     | 415 ng (16 L)       |
| 2023.04.19    | 1765 ng (104.95 L)                                                    | 1026 ng (104.95 L)  | 6625 ng (68.55 L) | 3320 ng (68.55 L)   |
| 2023.09.22    | 405 ng (178.52 L)                                                     | 314 ng (178.52 L)   | 1750 ng (56.25 L) | 686.5 ng (56.25 L)  |
| 2023.12.15    | 1048.2 ng (186.65 L)                                                  | 422.9 ng (186.65 L) | 1967 ng (56.8 L)  | 1074.25 ng (56.8 L) |
| 2024.04.25    | 5214 ng (298 L)                                                       | 2644 ng (298 L)     | 4350 ng (44 L)    | 2720 ng (44 L)      |

**Table 3 Total DNA and RNA extracted from the Muroto samples collected at the surface layer**

This table is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

| Sampling date | Total DNA/RNA yield and sampled water volume from surface samples |                  |                   |                   |
|---------------|-------------------------------------------------------------------|------------------|-------------------|-------------------|
|               | nano/micro                                                        |                  | pico or total     |                   |
|               | DNA                                                               | RNA              | DNA               | RNA               |
| 2022.09.29    | 34 ng (8 L)                                                       | 154 ng (8 L)     | 228 ng (16 L)     | 664 ng (56 L)     |
| 2022.12.21    | 589 ng (20 L)                                                     | 868 ng (20 L)    | 144 ng (7 L)      | 340 ng (7 L)      |
| 2023.04.19    | 2370 ng (31.9 L)                                                  | 1300 ng (31.9 L) | 12030 ng (55.6 L) | 8650 ng (55.6 L)  |
| 2023.09.22    | 851 ng (38.65 L)                                                  | 978 ng (38.65 L) | 4945 ng (42.9 L)  | 7265 ng (42.9 L)  |
| 2023.12.15    | 1050 ng (38.85 L)                                                 | 860 ng (38.85 L) | 6522 ng (49.05 L) | 8700 ng (49.05 L) |
| 2024.04.25    | 940 ng (38 L)                                                     | 937 ng (38 L)    | 4015 ng (30 L)    | 5960 ng (30 L)    |

From these samples, I successfully extracted adequate quantities (> 500 ng) of high-quality DNA and RNA for sequencing (Table 2, Table 3). 18S rDNA and rRNA metabarcodes (4,981 ASVs; nano/micro-fraction and either pico- or total-fraction), metagenomic data (757 Gbp; pico- or total-fraction), and metatranscriptomic data (598 Gbp; nano/micro-fraction and either pico- or total-fraction) were generated.

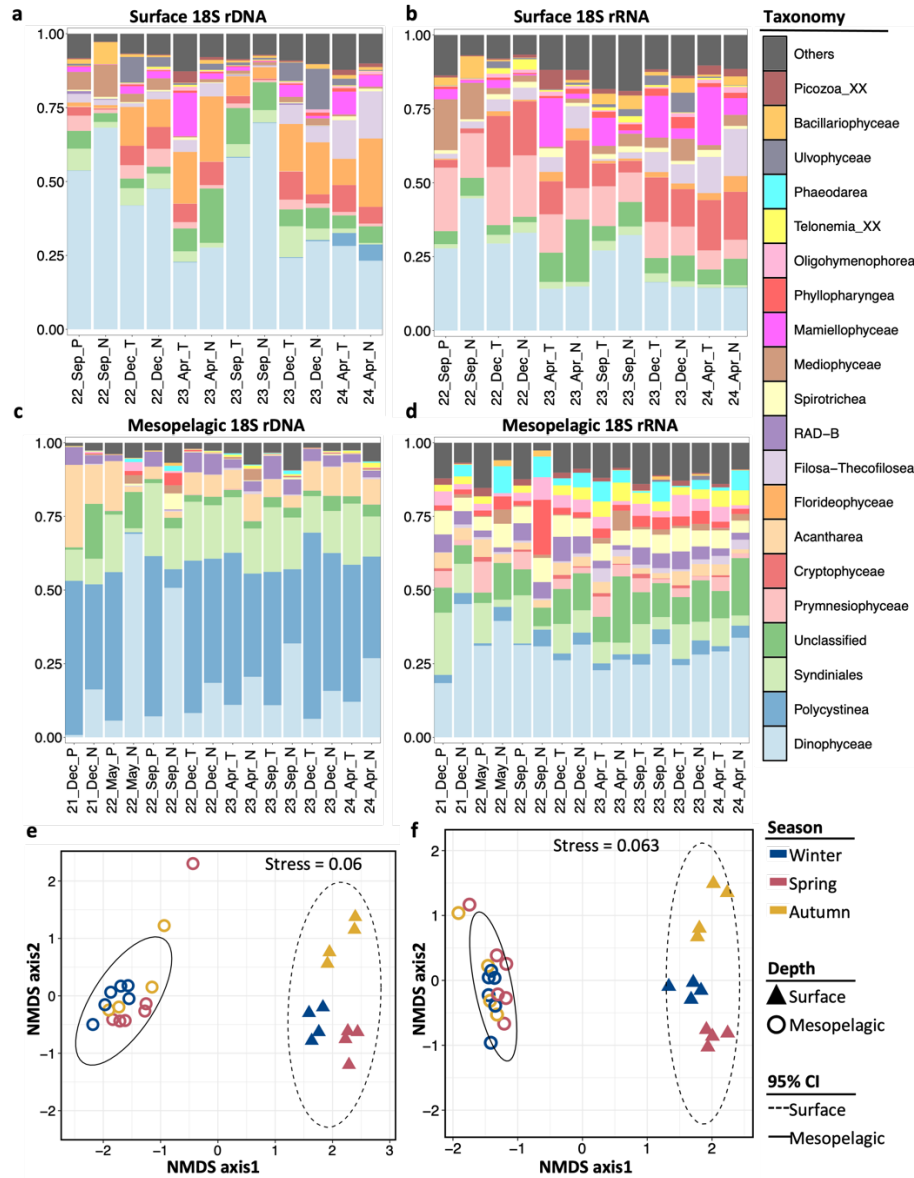
**Table 4 Environmental parameters measured at the Kochi Prefectural Deep Seawater Laboratory**

This table is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

| Sampling date | Deep-sea water temperature (°C) | Surface water temperature (°C) | Air temperature (°C) | Deep-sea water pH | Surface water pH | Deepsea water salinity ‰ | Surface water salinity ‰ |
|---------------|---------------------------------|--------------------------------|----------------------|-------------------|------------------|--------------------------|--------------------------|
| 2022.12.17    | 14.07                           | 19.36                          | 8.96                 | 7.82              | 8.18             | 33.12                    | 33.85                    |
| 2022.05.26    | 13.22                           | 21.93                          | 21.42                | 7.90              | 8.19             | 33.33                    | 32.65                    |
| 2022.09.29    | 14.71                           | 27.02                          | 25.24                | 7.86              | 8.13             | 33.72                    | 32.39                    |
| 2022.12.21    | 12.28                           | 18.43                          | 11.65                | 7.80              | 8.20             | 33.51                    | 33.79                    |
| 2023.04.19    | 11.79                           | 17.46                          | 20.55                | 7.83              | 8.21             | 33.65                    | 33.69                    |
| 2023.09.22    | 13.98                           | 27.41                          | 27.63                | 7.85              | 8.11             | 33.81                    | 32.89                    |
| 2023.12.15    | 13.08                           | 19.57                          | 18.96                | 7.84              | 8.16             | 33.49                    | 33.11                    |
| 2024.04.25    | 12.01                           | 19.69                          | 19.61                | 7.91              | 8.25             | 33.83                    | 33.16                    |

Environmental variables, including temperature, salinity, and pH, were recorded continuously by the Kochi Prefectural Deep Seawater Laboratory (Table 4). The pH in most deep-sea locations previously studied ranges from 7.8 to 7.9 (Park, 1966; Tan et al., 2012), and the observed average pH of 7.85 at this site falls within that typical deep-sea range.

### 3.2 Mesopelagic microeukaryote community is distinct from that in the surface water at the sampling site



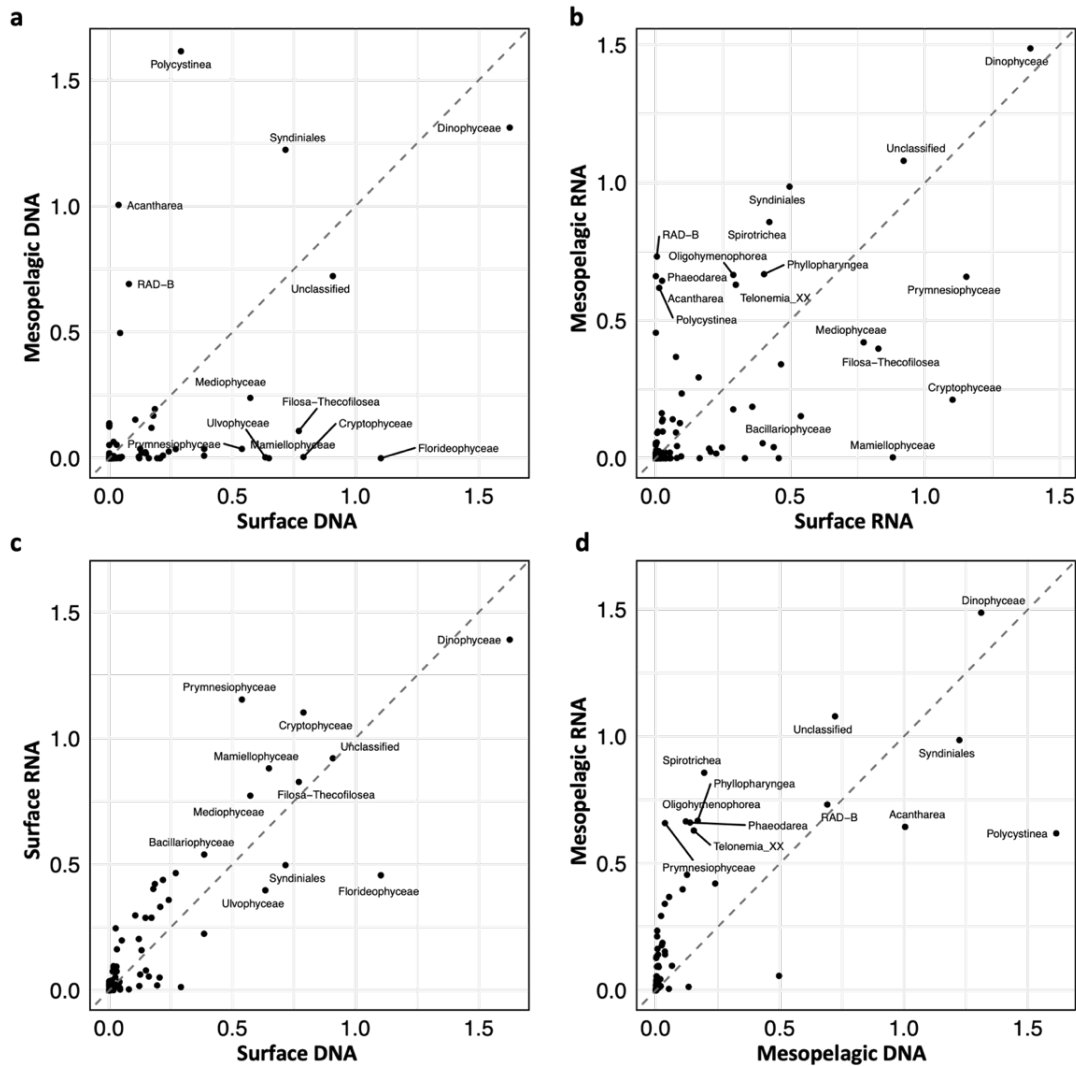
**Figure 5 Microeukaryote community composition of the seawater samples.**

(a–d) Relative abundance of taxa at the class level based on the metabarcoding data for surface 18S rDNA (a), surface 18S rRNA (b), mesopelagic 18S rDNA (c), and mesopelagic 18S rRNA (d). X-axis labels indicate the sampling date and size fractions (P: pico-fraction; T: total-fraction; N: nano/micro-fraction). (e,f) Non-metric multidimensional scaling (NMDS) plots showing differences in community compositions between surface and

mesopelagic samples based on 18S rDNA (e) and 18S rRNA (f). CI denotes confidence interval. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

In this study, 18S rDNA metabarcodes were used to measure the relative abundance of microeukaryotes at the ASV level, although the abundance of 18S rDNA is biased by the copy number of rRNA genes, which varies among organisms (Prokopowich et al., 2003; Zhu et al., 2005). 18S rRNA metabarcodes were used as a proxy for the metabolic activity of organisms (Giner et al., 2020). The microeukaryote community in the mesopelagic layer substantially differed from that of the surface layer at the study site (Figure 5a–d), suggesting that the two layers harbor distinct host candidates for GVs. Furthermore, the mesopelagic microeukaryote community characterized by 18S rRNA differed from that of 18S rDNA

metabarcodes, suggesting that some microeukaryotes are active in the mesopelagic layer, while others are inactive (i.e., in a dormant state or dead cells settled from the surface).



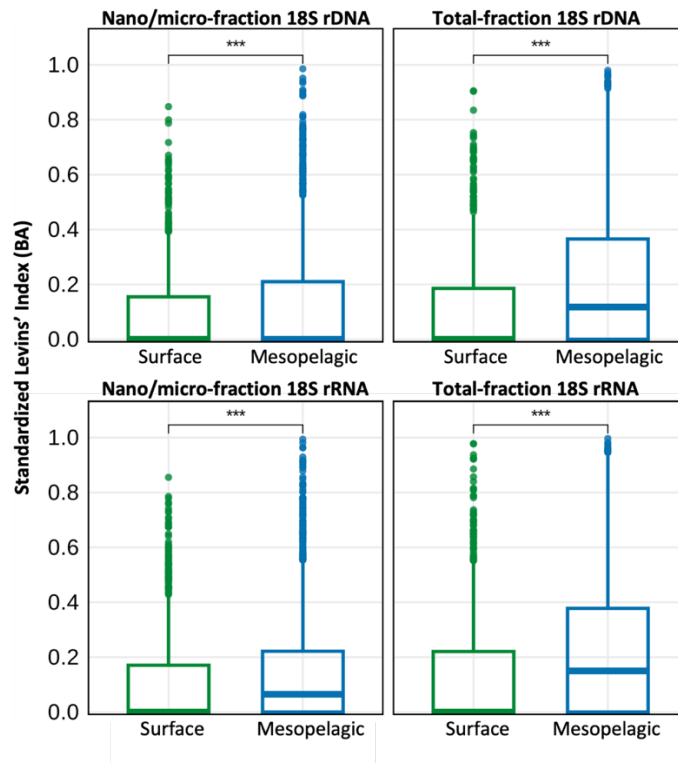
**Figure 6 Comparison of mean relative abundance (logTPM+1) of taxa at class level classification based on 18S rDNA/rRNA metabarcoding data.**

The taxa with logTPM+1 > 0.5 are labeled. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

For instance, Polycystinea had the highest abundance but was ranked eleventh in activity (Figure 6). Spirotrichea did not show high abundance but ranked third in activity (Figure 6).

These results suggest that the mesopelagic communities are a mixture of active and inactive organisms.

Non-metric multidimensional scaling (NMDS) ordination confirmed a clear separation between the surface and mesopelagic microeukaryote communities, regardless of whether 18S rDNA (Figure 5e;  $R^2 = 0.25$ , PERMANOVA,  $P$ -value  $< 0.01$ ) or 18S rRNA (Figure 5f;  $R^2 = 0.24$ , PERMANOVA,  $P$ -value  $< 0.01$ ) metabarcodes were used. Furthermore, the mesopelagic microeukaryote communities showed less enhanced seasonal changes than the surface communities.

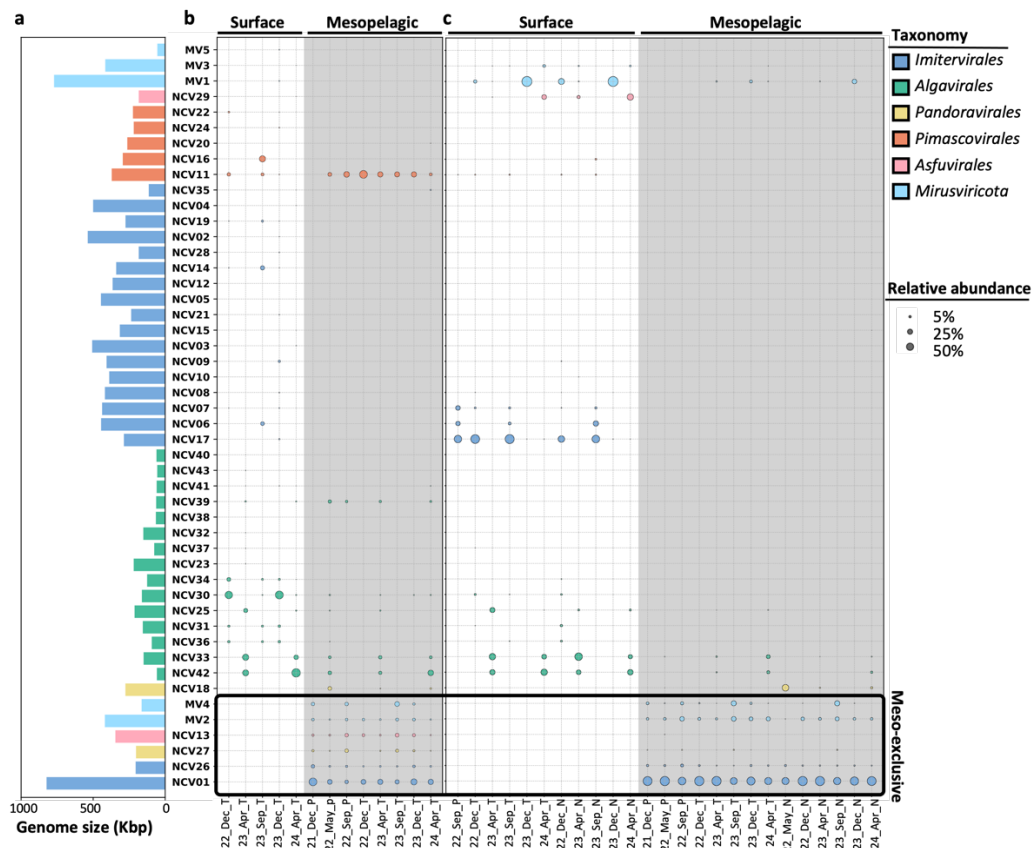


**Figure 7 Comparison of Levins' standardized niche breadth (BA) between surface and mesopelagic ASVs.**

Boxes show medians and interquartile ranges; whiskers extend to  $1.5 \times \text{IQR}$ ; points denote outliers. The Wilcoxon rank-sum test was used for statistical tests. Asterisks (\*\*\*) indicate  $P < 0.001$ . This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

To compare the persistence of ASVs between the layers, I calculated Levins' standardized niche breadth (BA) using only the surface and mesopelagic samples collected on the same day. In both the 18S rDNA and 18S rRNA datasets, the BA values were significantly higher for the mesopelagic than the surface ASVs in both size fractions (Figure 7,  $P$ -value  $< 0.001$ ), indicating higher persistence for the mesopelagic ASVs. These results suggest that the mesopelagic environment may provide a unique and stable host community for GVs.

### 3.3 GVs exclusively detected in mesopelagic metagenomes show persistency at the study site

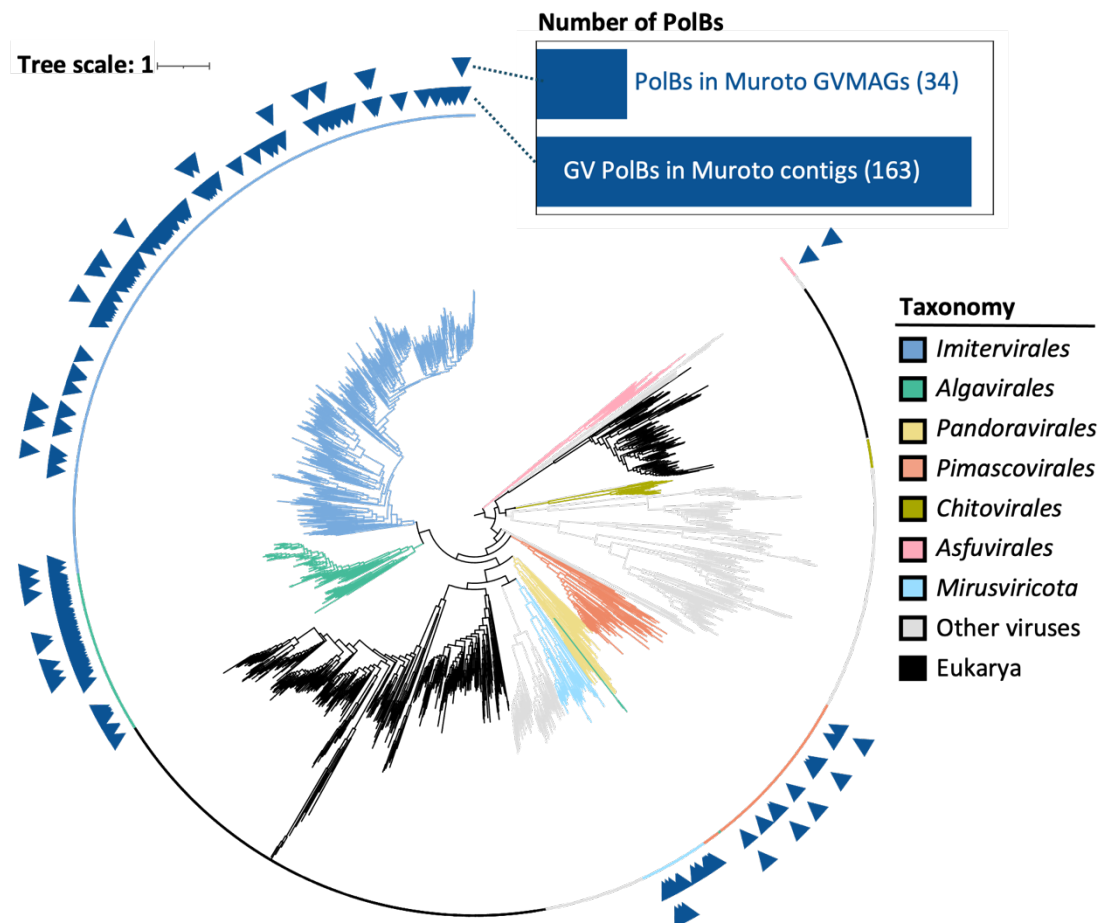


**Figure 8 Genome size distribution and relative abundance of Muroto GVMAGs in the metagenomic and metatranscriptomic data.**

(a) Genome size distribution of the 48 Muroto GVMAGs. (b) Relative abundance of Muroto GVMAGs in the

metagenomic data. (c) Relative abundance of Muroto GVMAGs in the metatranscriptomic data. In (b) and (c), samples from the mesopelagic layer are shown with a gray background. Circle size (i.e., area) represents relative abundance, and circle color indicates taxonomic classification. GVMAGs enclosed by a black box are mesopelagic-exclusive GVMAGs. X-axis labels indicate the sampling date and size fractions (P: pico-fraction; T: total-fraction; N: nano/micro-fraction). This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

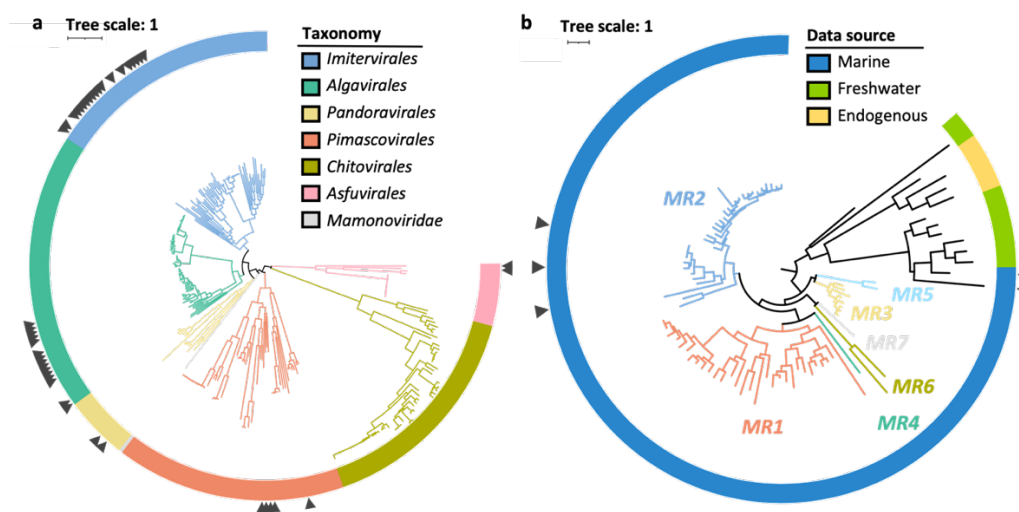
Forty-eight GVMAGs recovered from the Muroto metagenomic data met my quality criteria (Figure 4; see Materials and Methods), including 43 *Nucleocytoviricota* MAGs (NCV-MAGs) and five *Mirusviricota* MAGs (MV-MAGs) (Figure 8a). These GVMAGs exhibited genome sizes ranging from 54 kbp to 824 kbp and GC contents ranging from 24% to 53%.



**Figure 9** Phylogenetic analysis of PolB from the Muroto samples.

Maximum-likelihood phylogenetic tree of PolB from Muroto metagenomic contigs (>2.5 kbp) and reference sequences (Gaia et al., 2023; Kazlauskas et al., 2020). Branch colors and the innermost ring indicate viral classification. The two surrounding rings with triangles indicate the distribution of PolB sequences: the inner ring shows PolB sequences likely to be of GV origin and found in the Muroto contigs (>2.5 kbp), and the outer ring represents PolB sequences found in qualified Muroto GVMAGs. The accompanying bar plot shows the number of PolB sequences corresponding to each ring. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

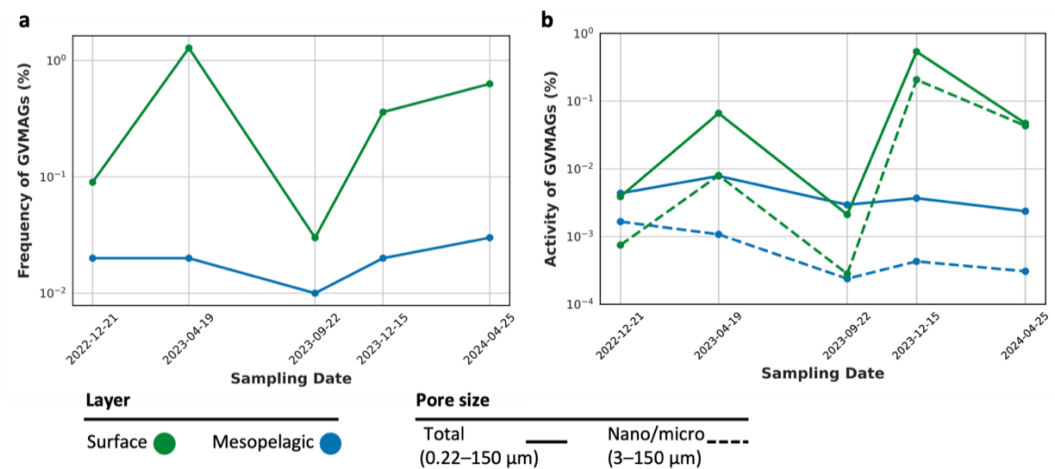
To assess the recovery level of GV lineages by GVMAGs, I investigated family-B DNA polymerase (PolB) sequences in the metagenomic dataset. Through phylogenetic analysis (Figure 9), I identified 163 PolB sequences predicted to be of GV origin. Of these PolB sequences, 34 (21%) were found in the Muroto GVMAGs, suggesting that many GVs in the study site were not represented in my GVMAG dataset.



**Figure 10 Phylogenetic and taxonomic composition of Muroto GVMAGs.**

(a) Maximum-likelihood phylogenetic tree of Muroto NCV-MAGs and 220 reference genomes inferred with IQ-TREE using seven conserved marker genes. Ring and branch colors indicate taxonomic classifications. (b) Phylogenetic tree of Muroto MV-MAGs and reference genomes inferred with IQ-TREE using the HK97 MCP gene. The ring indicates the MAG source, and branch colors represent clade classifications. MR1 to MR6 mark previously described marine mirusvirus clades. Gray triangles indicate Muroto GVMAGs. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Phylogenetic analysis of seven conserved NCV marker genes revealed that the 43 NCV-MAGs belonged to five orders of *Nucleocyotviricota*: 19 imiterviruses, 15 algaviruses, five pimascoviruses, two pandoraviruses, and two asfuviruses (Figure 10a). Two NCV-MAGs belonging to the order *Pandoravirales* were closely related to *Emiliana huxleyi* viruses (average nucleotide identity, ~85%). The five MV-MAGs were classified based on HK97 major capsid protein (MCP) phylogeny, including three in the MR2 marine clade and two in a clade composed of mirusvirus MAGs from a freshwater lake (Figure 10b) (Zhang et al., 2024).



**Figure 11 Dynamics of the presence and activity of Muroto GVMAGs.**

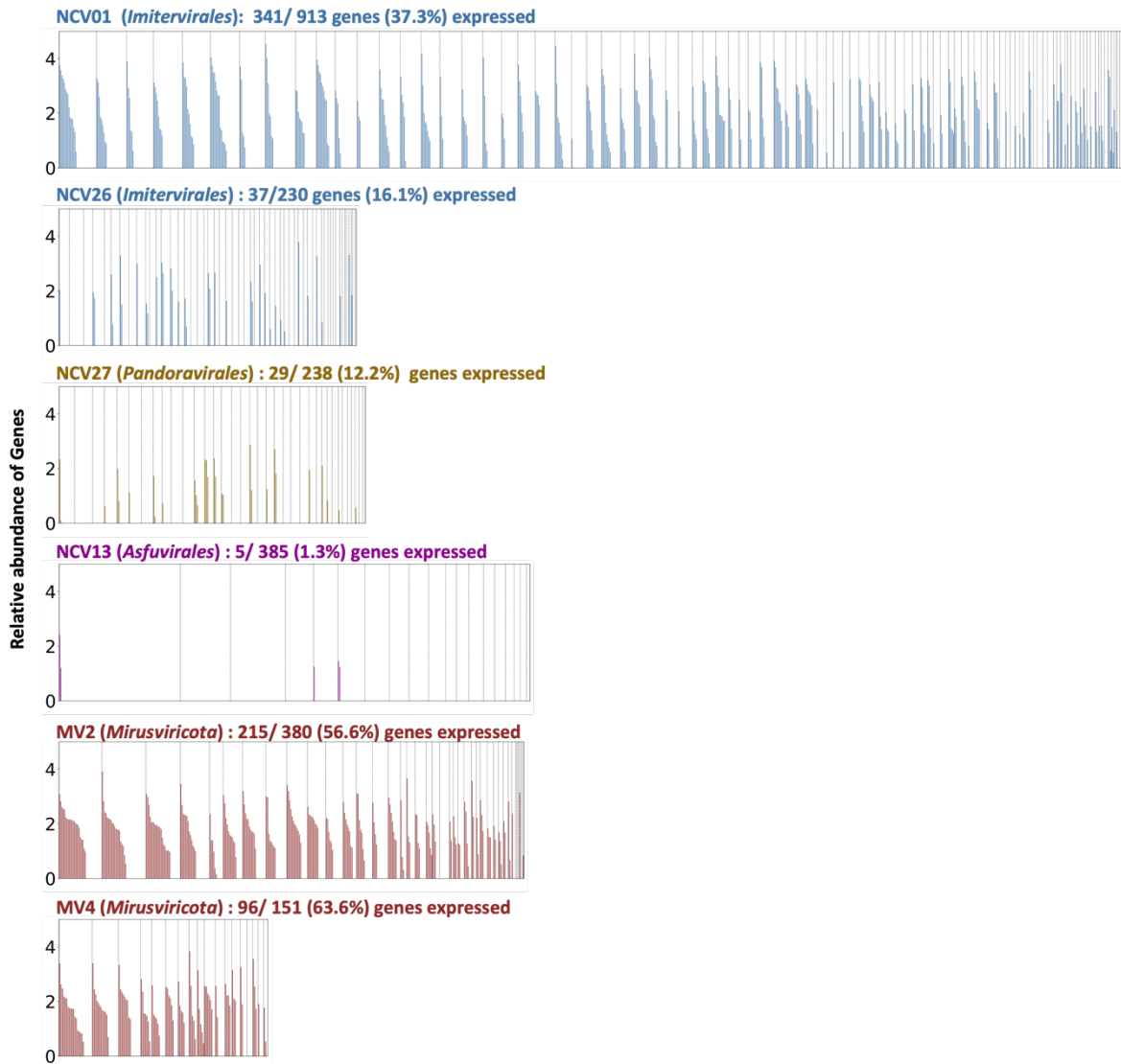
(a) Percentage of metagenomic reads mapped to the 48 Muroto GVMAGs over time. (b) Percentage of metatranscriptomic reads mapped to the 48 Muroto GVMAGs over time. Blue lines represent mesopelagic samples, while green lines represent surface samples. Solid lines indicate total-fraction (0.2–150  $\mu\text{m}$ ), whereas dashed lines correspond to nano/micro-fraction (5–150  $\mu\text{m}$ ). Time points are plotted based on actual sampling intervals, thus the spacing on the X-axis reflecting real temporal distances. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Mapping of the metagenomic reads on the 48 GVMAGs revealed that these MAGs represent 0.028%–1.28% (0.48% on average) of the total reads for the surface data compared with 0.01%–0.028% (0.019% on average) for the mesopelagic data (Figure 11a). The lower relative abundance of GVs in the mesopelagic samples explains the lower level of GVMAG recovery from these samples. Among the 48 GVMAGs, 17 GVMAGs (15 NCV-MAGs, 2

MV-MAGs) were detected in the mesopelagic metagenomes, including six (NCV01, 13, 26, 27; MV2, 4) that were exclusively found in the mesopelagic samples (Figure 8b). These “meso-exclusive” GVMAGs tended to show persistent presence in the mesopelagic layer across different seasons (Figure 8b). The other 11 GVMAGs detected in the mesopelagic metagenomic data were also detected in the surface metagenomes (Figure 8b). In sharp contrast with the six meso-exclusive GVMAGs, many of these GVMAGs showed strong seasonal dynamics (e.g., NCV30, 42) akin to the pattern observed for GVs in a shallow coastal area (Fang et al., 2025). Endo *et al.* previously showed that the community structure of GVs in the mesopelagic layer (200–1,000 m) is different from that in the photic layer based on the *Tara* Oceans data (Endo et al., 2020). My results confirmed not only the distinct community structure but also the difference in the community dynamics between the surface and mesopelagic layers.

### **3.4 Mesopelagic-specific giant viruses are active across seasons**

Mapping of the metatranscriptomic reads on the GVMAGs revealed a similar trend to the metagenomic data (Figure 8c, Figure 11), with the meso-exclusive GVMAGs showing seasonally stable activity and the surface layer community showing strong seasonal dynamics. Seventeen GVMAGs showed transcriptional activity in the mesopelagic layer. These included all six meso-exclusive GVMAGs (Figure 8c).

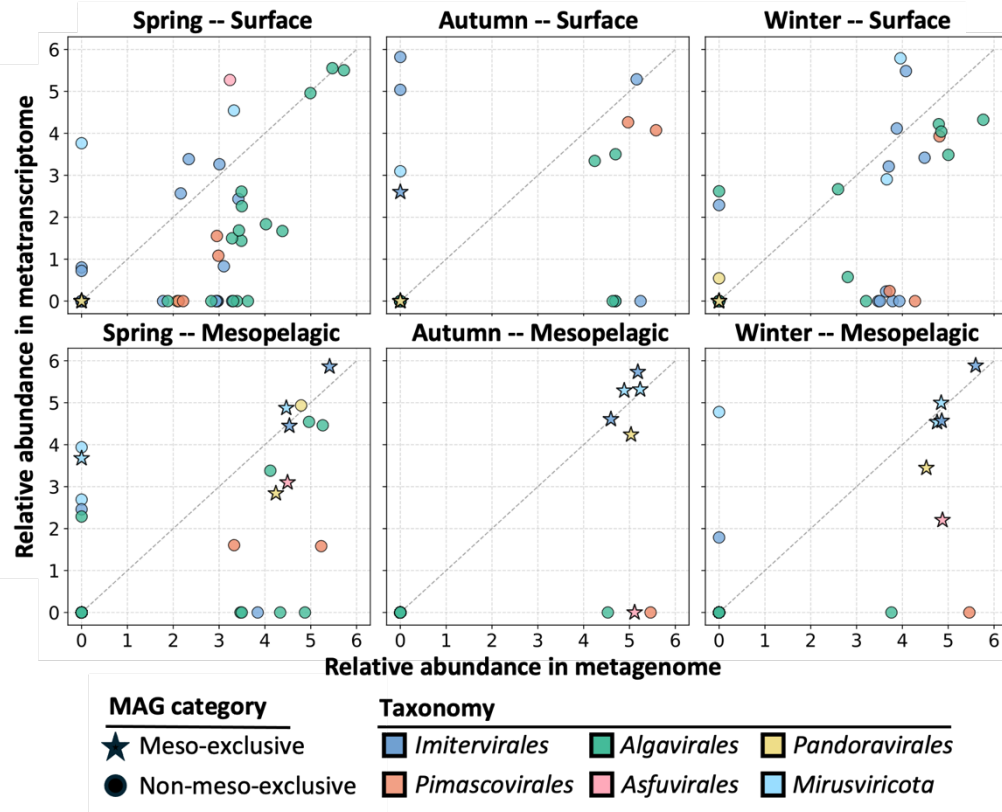


**Figure 12 Gene expression profiles of six meso-exclusive GVMAGs.**

Each horizontal panel represents one meso-exclusive GVMAG, with the X-axis showing individual genes and the Y-axis indicating mean relative levels of transcription (logTPM+1). Bar colors distinguish different GV orders. Vertical dashed lines indicate the boundary between contigs. Within individual contigs, genes are ordered by their expression levels. The total number of genes and the percentage of expressed genes are annotated for each GVMAG. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

For five of these GVMAGs (NCV01, 26, 29; MV2, 4), transcripts were detected for over 10% of genes encoded in the genomes (Figure 12 ), with NCV01, MV2, and MV4

expressing up to 37.3%–63.6% of their genes (Figure 12 ). This result suggests ongoing viral replication in the mesopelagic environment.



**Figure 13** The relative abundance of Muroto GVMAGs between metagenomic and metatranscriptomic datasets across seasons and depths.

Scatterplots show the relationship between average relative abundance in metagenomic (X-axis, logTPM+1) and metatranscriptomic (Y-axis, logTPM+1) datasets for Muroto GVMAGs. Columns represent samples from different seasons (spring, autumn, winter), and rows correspond to samples from different depths (surface and mesopelagic). Point color indicates taxonomic classification, and shape denotes MAG categories: stars represent meso-exclusive GVMAGs, while circles indicate non-meso-exclusive GVMAGs. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Furthermore, the transcriptional activities of many of these meso-exclusive GVMAGs were relatively constant over time (e.g., NCV01, NCV26, MV2, MV4) (Figure 8c, Figure 13). To the best of my knowledge, this is the first report of the transcriptional activity of GVs exclusively detected in an aphotic environment. The stable community structure and activity

of meso-exclusive GVs across seasons parallel the stability of the microeukaryotic community in the mesopelagic zone at the study site (Figure 5). In contrast to the meso-exclusive GVMAGs, several GVMAGs that were detected in both surface and mesopelagic metagenomes showed transcriptional activity only in the surface samples (e.g., NCV11, NCV30). These GVs in the mesopelagic layer may have originated from the surface through sinking processes, as previously suggested (Endo et al., 2020; Sheam et al., 2025). Overall, my findings support the existence of active and stable GV populations specific to the mesopelagic zone. Virus–host interactions may be less influenced by seasonal fluctuations in the relatively stable ecological conditions in this layer.

The difference in mapping depth for the metatranscriptome data between the surface and mesopelagic layers was less pronounced than the case for the metagenome data, with some sampling days showing comparable (or even higher) relative transcriptional activity of GVMAGs in the mesopelagic than the surface data. This suggests that even though fewer GVMAGs were reconstructed from the mesopelagic data, they may be highly active. For the total-fraction, the proportions of mapped transcriptomic reads were 0.0021%–0.54% (0.13% on average) for the surface samples and 0.0024%–0.0079% (0.0043% on average) for the mesopelagic samples. For the nano/micro-fraction, the proportions of mapped transcriptomic reads were 0.00028%–0.21% (0.052% on average) for the surface samples and 0.00024%–0.0017% (0.00074% on average) for the mesopelagic samples. The higher proportions of mapped mesopelagic transcriptomic reads in the total-fraction than in the nano/micro-fraction suggest that active infection of GVs in the mesopelagic layer is primarily associated with smaller host cells included in this fraction.

### **3.5 Mesopelagic GVMAGs encode specific gene functions**

The six meso-exclusive GVMAGs recovered in this study contained between 51 and 913 predicted genes per genome, with 880 genes (35.5%) being assigned to 594 KO terms. The six meso-exclusive GVMAGs showed distinct gene compositions enriched in functions

related to virus–host interactions and signaling, including components of the ubiquitin system, cytoskeletal proteins, and signal transduction regulators.

Among genes in the ubiquitin system, E3 ubiquitin–protein ligase RNF181 was found in 23 GVMAGs. However, other ubiquitin system genes, such as Ariadne-1 (NCV01, NCV26, NCV27), E3 ubiquitin-protein ligase RNF13 (NCV01, NCV26), and NEDD4-binding protein 2 (NCV01, NCV26), were specific to meso-exclusive GVMAGs. Actin, myosin, and flagellar basal-body rod protein FlgG were exclusively present in NCV01 and NCV26 and could benefit viruses by manipulating the localization of viral replication machinery during infection (Da Cunha et al., 2022; Ha et al., 2021; Kijima et al., 2024). NCV01 also encoded ninein-like protein, which is a member of the  $\gamma$ -tubulin complexes binding proteins (Li et al., 2011) and may function to modulate cytoskeleton. Additionally, phosphoinositide 3-kinase (PI3K), a regulator of membrane dynamics and endocytic trafficking, was detected in 4 of 6 meso-exclusive GVMAGs. PI3K has been implicated in supporting the formation of replication organelles of enteroviruses and may similarly support GV replication (Wu et al., 2024).



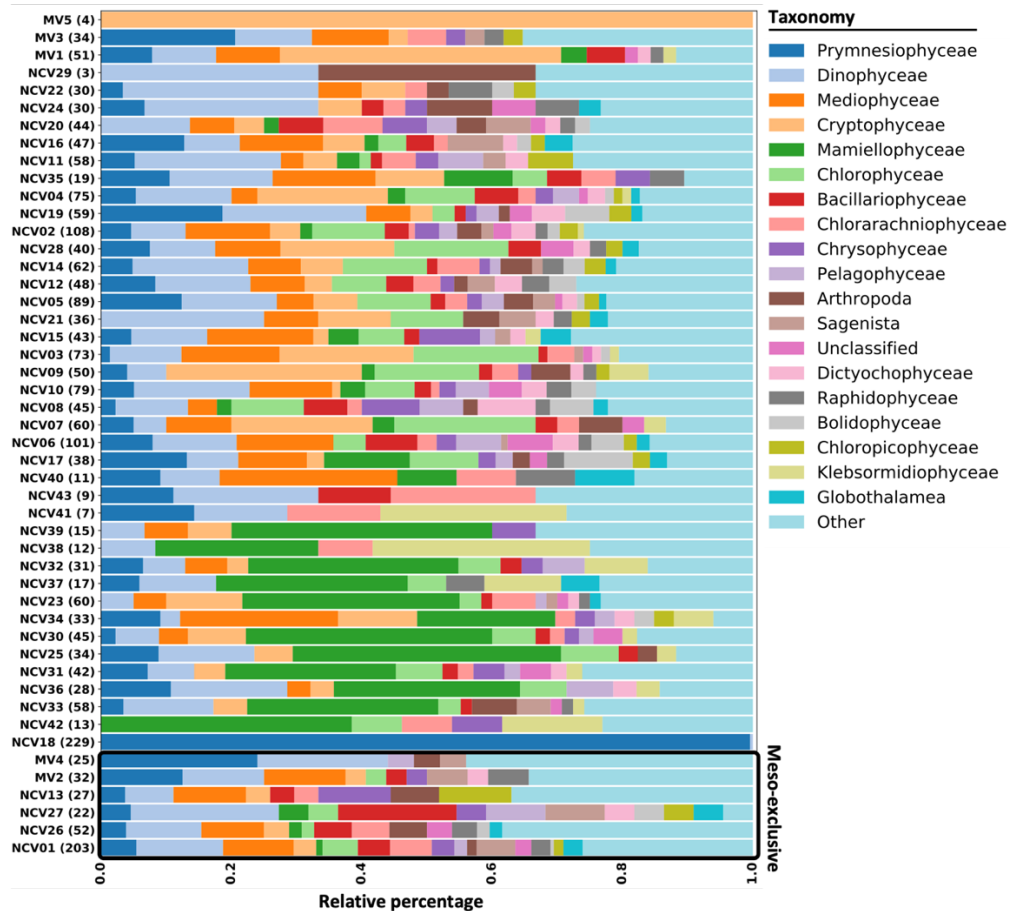
**Figure 14** Top 20 highly expressed KEGG orthology (KO) terms for the Muroto GVMAGs.

Heatmaps display the mean relative abundance ( $\log(\text{TPM}+1)$ ) of the top 20 most highly expressed KO terms of 48 Muroto GVMAGs in (a) surface and (b) mesopelagic metatranscriptomic datasets. Columns represent GVMAGs, while rows correspond to KO terms. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Notably, among the top 20 transcriptionally active KO terms in the mesopelagic samples, 7 were related to the ubiquitin system (e.g., Ariadne-1, NEDD4-binding protein 2, ubiquitin B, and E3 ubiquitin-protein ligases RNF181, EDD1, SIAH1, RNFT1) (Figure 14b), compared with only 1 in the surface samples (Figure 14a). The sequential action of ubiquitin systems conjugates ubiquitin to proteins and target them to proteasomes for degradation (Hershko, 2005). Eukaryotic viruses can hijack the ubiquitin-proteasome system to aid in various stages of viral propagation (Tang et al., 2018). Since normal protein folding is hindered by low temperature or high pressure (Kozuch et al., 2019; Ramanujam et al., 2020), the ubiquitin-related genes of GVs may function to modulate the host ubiquitin-proteasome

system to quality control viral and host proteins in harsh conditions in the mesopelagic layer, as previously hypothesized for GVs in cold environments (Buscaglia et al., 2024). The ubiquitin-proteasome system is also known to regulate apoptosis (Abbas et al., 2021). For example, SIAH1 mediates polyubiquitination and degradation of the tumor suppressor Homeodomain-interacting protein kinase 2 (HIPK2), thereby modulating cellular apoptosis (Winter et al., 2008). Thus, the expression of GV ubiquitin-related genes may regulate programmed cell death of the infected host cells. In the mesopelagic layer, which is characterized by relatively low host biomass (or encounter rates), the ability to suppress apoptosis might represent an adaptive strategy to prolong the life of the infected host cells, thereby allowing for greater viral production before host lysis and ensuring the continued propagation of the GV in resource-scarce environments.

### 3.6 Putative hosts of meso-exclusive GVMAGs

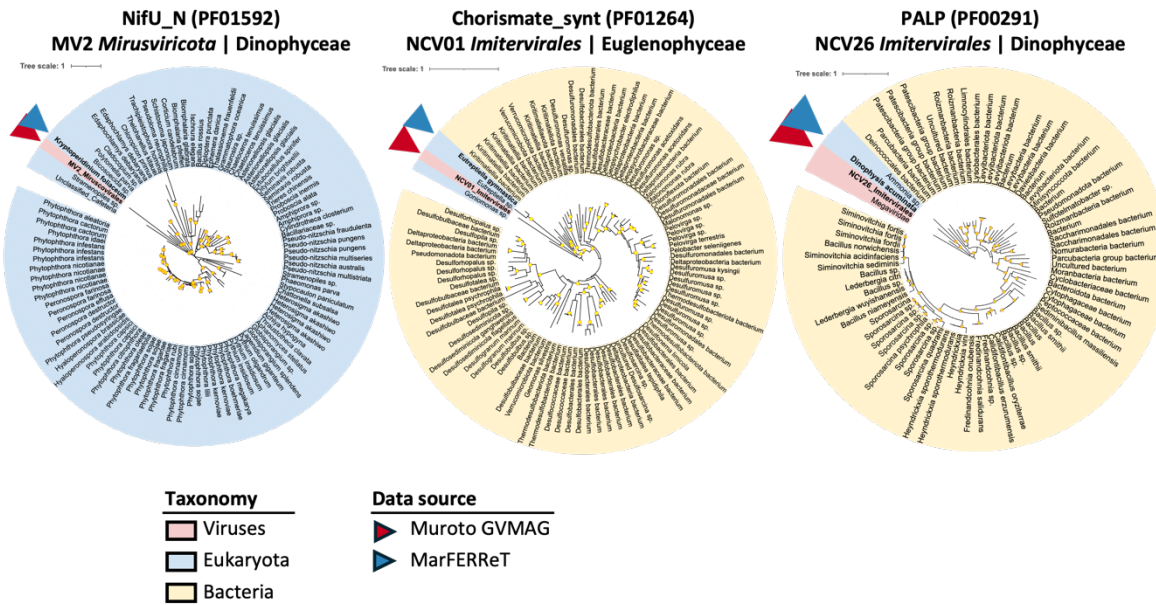


**Figure 15 Assignment of eukaryotic-like protein sequences in the Muroto GVMAGs.**

Each row represents a Muroto GVMAG. Numbers in parentheses following GVMAG identifiers indicate the total number of genes with hits to eukaryotic taxa in the MarFERRet database. The X-axis shows the relative percentage of gene hits to corresponding eukaryotic classes. GVMAGs enclosed by a black box indicate meso-exclusive GVMAGs. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

To gain insights into the potential hosts of the 48 GVMAGs, I first performed homology searches against the MarFERReT database (Figure 15). The best eukaryotic matches differed between meso-exclusive and other GVMAGs, with the former displaying fewer matches to green algae (i.e., Mamiellophyceae and Chlorophyceae). Meso-exclusive GVMAGs were enriched in matches to Prymnesiophyceae, Dinophyceae, and Mediophyceae. Of other

GVMAGs, MV5 and NCV18 showed many matches to a single taxon (Figure 15). MV5 showed four genes with best matches to Cryptophyceae, which was previously predicted as a host group of mirusvirus (Zhao et al., 2024). NCV18, which is closely related to *Emiliana huxleyi* viruses, showed many matches to *Emiliana huxleyi* (class Prymnesiophyceae) probably due to the inclusion of viral sequences or homologs in the MarFERReT database.

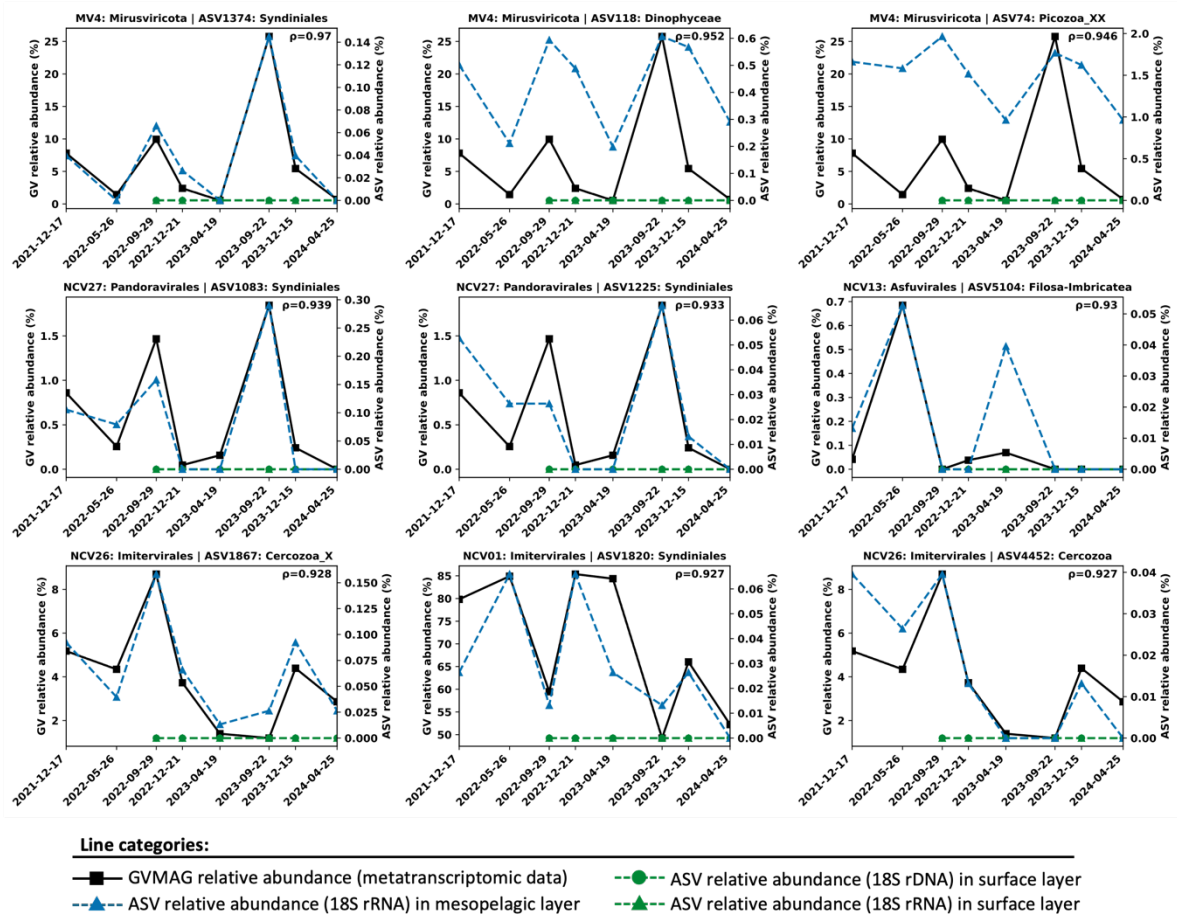


**Figure 16 Phylogenetic trees of eukaryotic-like protein sequences from meso-exclusive GV sequences and their homologs.**

Phylogenetic trees were inferred from protein sequences identified in Muroto GVMAGs and their close homologs retrieved from a combined MarFERReT and NR database. Bootstrap support values >85% are shown. Tip colors denote taxonomic groups (bacteria, eukaryotes, viruses). Blue triangles denote best MarFERReT matches, and red labels highlight proteins from Muroto GVMAG. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Phylogenetic analyses of protein sequences from meso-exclusive GV sequences were conducted that showed significant similarities with eukaryotic protein sequences (Figure 16). I found three cases that supported the predictions by the above MarFERReT-based analysis. The NifU-like N terminal domain (NifU\_N) of MV2 was clearly related to eukaryotic homologs with the closest relationship with sequences from dinoflagellates (*Kryptoperidinium*) and

stramenopiles. Two other cases suggested relationships between meso-exclusive GV and specific eukaryotes. However, interpreting these two trees as a support for specific virus-host relationships is not straightforward as most of other related sequences were bacterial sequences.



**Figure 17 Co-occurring meso-exclusive GV and microeukaryotes.**

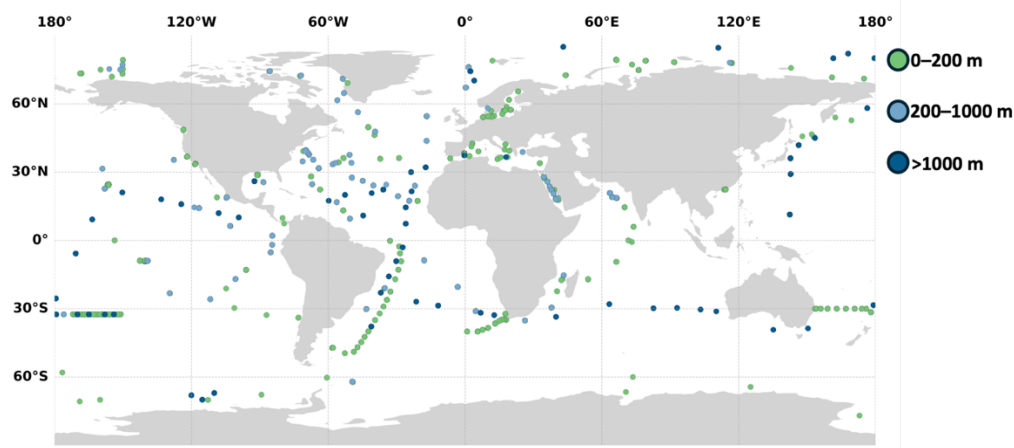
Spearman's correlation coefficients were calculated between the relative abundances of meso-exclusive GV and ASVs in pico- or total-fraction using metatranscriptome and 18S rRNA data from mesopelagic samples (8 paired samples). Significant associations were defined as adjusted  $P$ -value  $< 0.05$ . The relative abundance of ASVs in the surface layers (green line) was added for comparison. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

Furthermore, host inference by examining co-occurrence patterns between meso-exclusive GV and microeukaryotes was conducted. Using metatranscriptome and 18S rRNA

data from mesopelagic samples, I calculated Spearman's correlation coefficients for the relative abundance between meso-exclusive GVs and ASVs in pico- or total-fraction. Nine ASVs exhibited statistically significant positive correlations with five meso-exclusive GVs ( $P$ -value  $< 0.05$ ; Figure 17). The predicted potential hosts were heterotrophs (or potentially mixotrophs), belonging to Cercozoa (3 ASVs; Cercozoa, Cercozoa\_X, Filosa-Imbricatea; amoeboids or flagellates of Rhizaria), Picozoa (1 ASV; heterotrophic picoeukaryotes), Dinophyceae (1 ASV), and Syndiniales (4 ASVs; early branching dinoflagellates). These potential hosts were undetected in the 18S rRNA/rDNA data from the surface layer, suggesting mesopelagic or deeper layers as their main habitat. As members of Syndiniales parasitize dinoflagellates, their co-occurrence with GVs may reflect shared hosts rather than host-virus relationships. Cercozoa, Picozoa, and Dinophyceae were reported in the mesopelagic layer, with Cercozoa in particular showing the highest activity among all depths (Giner et al., 2020). Host-virus associations have been suggested for Cercozoa and Dinoflagellata, based on single-cell genomics data from induced algal bloom samples (Fromm et al., 2024).

Although these analyses point to potential hosts of meso-exclusive GVs, the limited representation of mesopelagic protist genomes in current reference databases, the small number of samples in this study, and the low abundance of GVs in the mesopelagic layer mean that these host predictions should be considered preliminary. Other approaches such as single-cell genomics may directly link GVs to their hosts, offering validation beyond sequence homology and co-occurrence (Fromm et al., 2024).

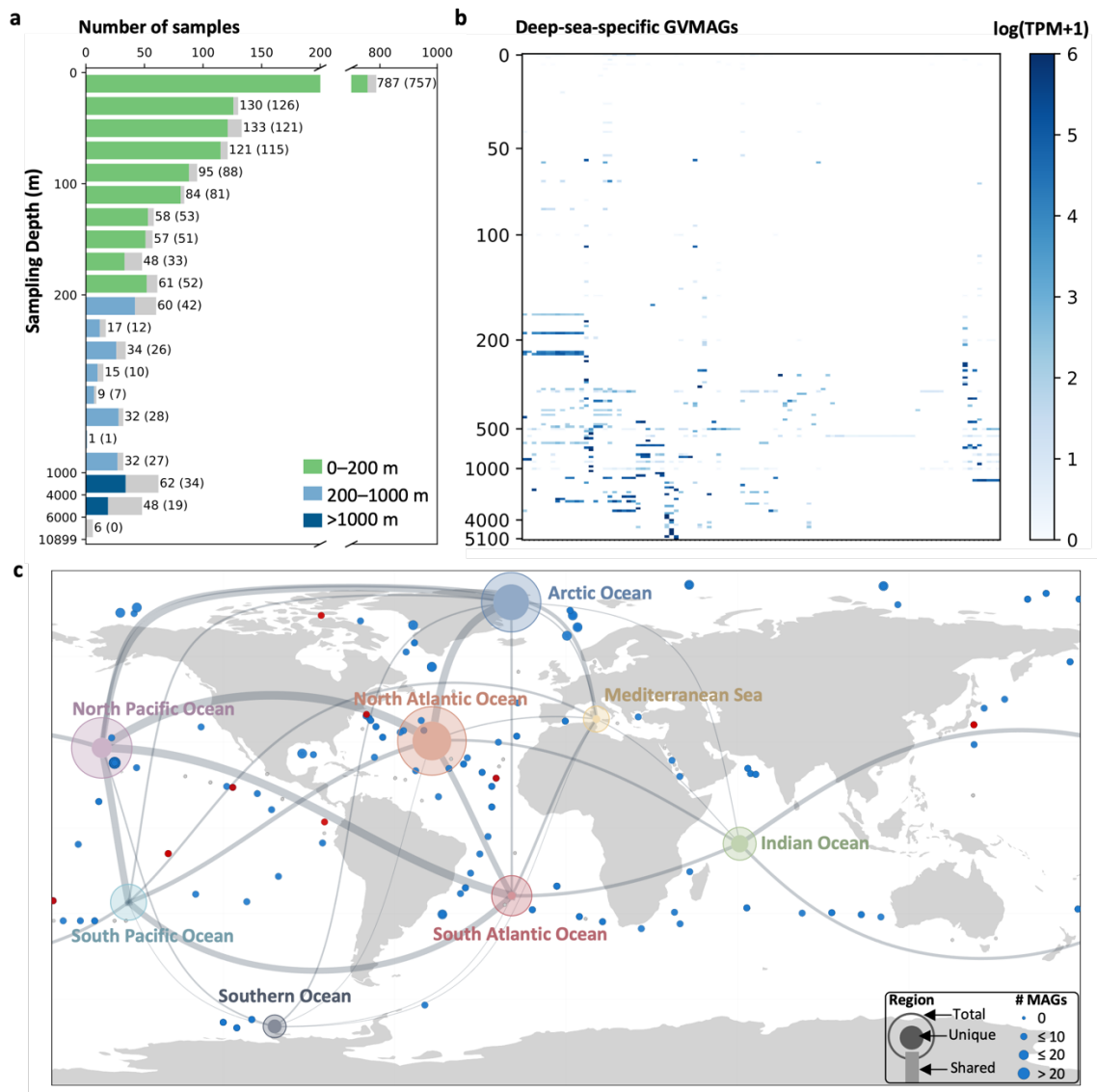
### 3.7 Deep-sea-specific GVs distribute widely in the global ocean



**Figure 18 Geographic distribution of 1,890 marine metagenomic datasets used in this study.**

Dot colors indicate different depth categories. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

My discovery of meso-exclusive GVs at a coastal site in Japan raises questions regarding their distributions and adaptation across different oceans and depth ranges. To address this issue and to catalog GV genomes specific to aphotic layers of the ocean, the GVGR database (see Materials and Methods) was constructed, containing 4,473 GVMAGs and analyzed 1,890 publicly available marine metagenomes collected from depths ranging from 0 m to 10,899 m (Figure 18).

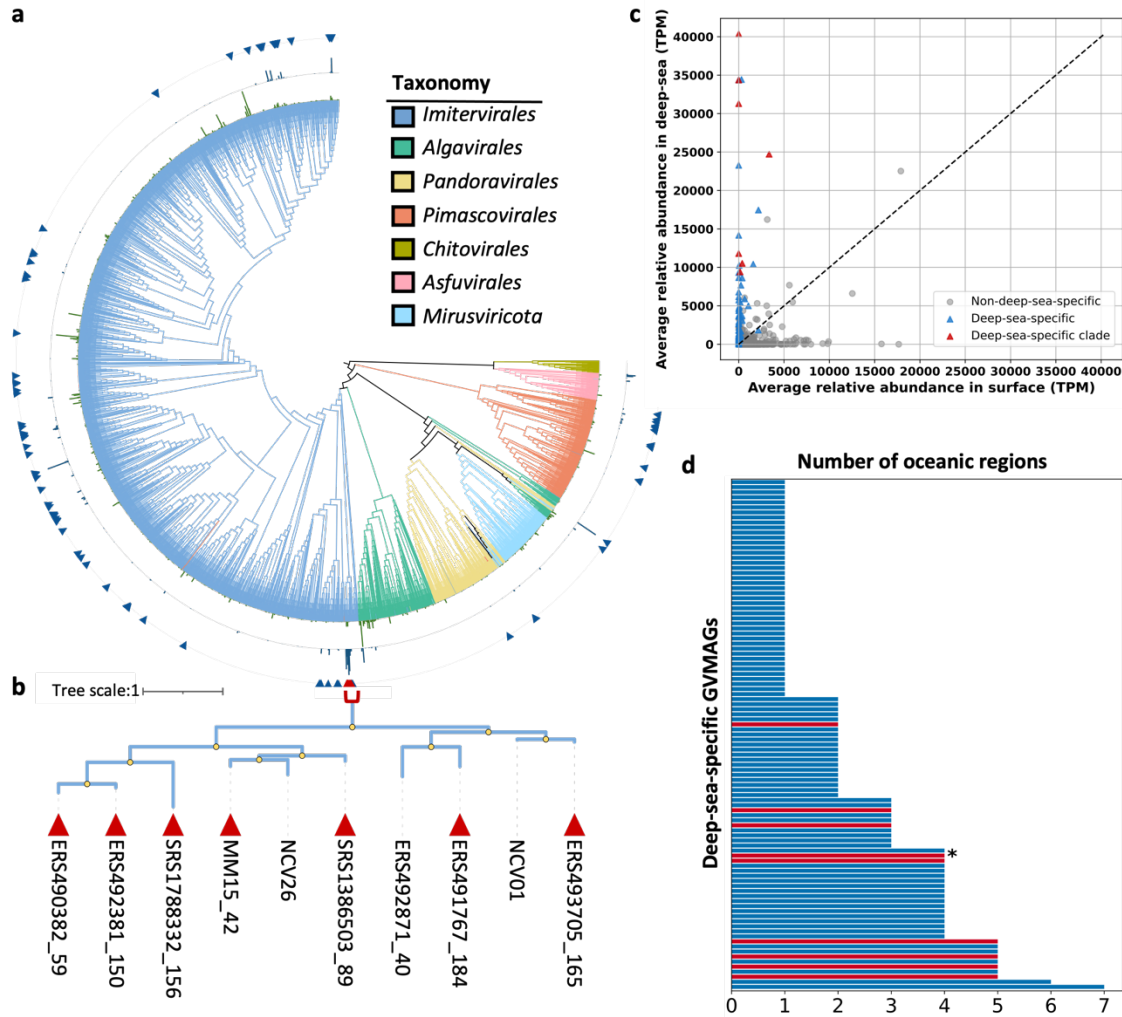


**Figure 19 Distribution of all GVMAGs in the GVGR database and 101 deep-sea-specific GVMAGs across global ocean depth layers.**

(a) Total number of samples and number of samples with GVMAGs detected across different depths. Bars are color-coded by depth category. Numbers at the end of each bar indicate the total number of samples in that depth range; values in parentheses represent the number of samples in which GVMAGs from the GVGR database were detected. (b) Relative abundance of 101 deep-sea-specific GVMAGs identified in the GVGR database across all depth layers. Each row represents a depth, and each column represents one deep-sea-specific GVMAG, with darker blue indicating higher abundance levels. (c) Distribution of deep-sea-specific GVMAGs across global deep-sea samples. Each dot on the map represents a deep-sea sample, with its size indicating the

number of deep-sea-specific GVMAGs detected. Gray points indicate samples without deep-sea-specific GVMAGs; blue points indicate samples containing deep-sea-specific GVMAGs; red points indicate samples in which ERS493705\_165 (representing Muroto meso-exclusive NCV01 after 95% sequence identity clustering) was detected. Eight large circles correspond to oceanic regions; their sizes reflect the total number of deep-sea-specific GVMAGs, while concentric rings indicate the proportions of GVMAGs unique to the region. Lines connecting the circles represent GVMAGs shared between regions, with line thickness proportional to the extent of sharing. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

The GVMAGs were found to have a wide vertical distribution, extending to a depth of 5,601 m (Figure 19a). Based on the occurrence of GVMAGs across depths, 101 deep-sea-specific GVMAGs were identified as being exclusively or predominantly distributed in the mesopelagic or deeper layers (Figure 19b) (see Materials and Methods). To further investigate their biogeographic patterns, I analyzed the distribution of deep-sea-specific GVMAGs across eight geographic regions and quantified their geographic ranges. The total number of deep-sea-specific GVMAGs was relatively high in the North Atlantic Ocean, North Pacific Ocean and Arctic Ocean, ranging from 44 to 53 (Figure 19c). Only eight were detected in the Southern Ocean, which may be due to limited sampling efforts.



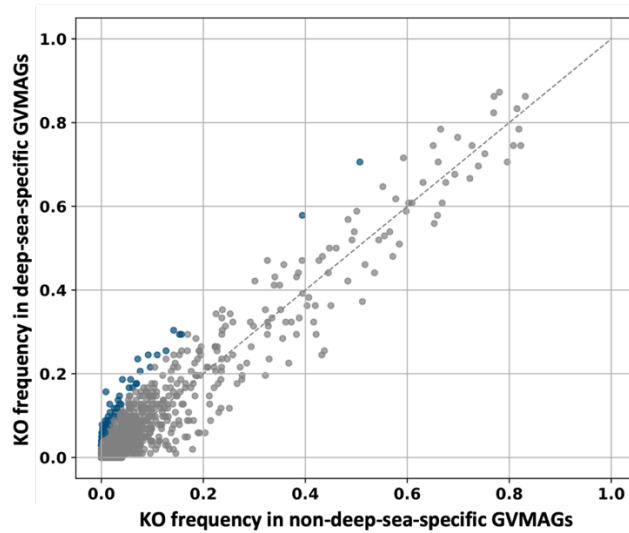
**Figure 20** Phylogenetic diversity, abundance, and biogeography of deep-sea-specific GVMAGs.

(a) Maximum-likelihood phylogenetic tree constructed from dereplicated GVGR genomes with 85% ANI, supplemented with 101 deep-sea-specific GVMAGs and Muroto meso-exclusive GVMAGs. Branch colors indicate taxonomic classifications. Green bars in the inner ring indicate the average relative abundances (TPM) of the GVMAGs in surface samples, and dark blue bars in the second ring indicate their average relative abundances in deep-sea samples. Triangles on the third ring indicate deep-sea-specific GVMAGs. The red brace highlights a clade composed predominantly of deep-sea-specific GVMAGs, as described in the main text. (b) Enlarged view of the deep-sea clade in the phylogenetic tree. Branches with ultrafast bootstrap support values > 95% are marked with yellow circles. The full identifier for “MM15\_42” is Moniruzzaman\_MM15\_ERX556043\_42\_dc. NCV01 and NCV26 are Muroto meso-exclusive GVMAGs, with NCV01 being dereplicated with the deep-sea-specific ERS493705\_165 and NCV26 not recognized as deep-sea-specific. (c) Comparison of average relative abundances of GVMAGs in surface and deep-sea metagenomes.

Blue triangles denote deep-sea-specific GVMAGs, red triangles indicate those in the deep-sea clade (highlighted in (b)), and gray circles represent non-deep-sea-specific GVMAGs. (d) Number of oceanic regions in which each deep-sea-specific GVMAG was present. Red bars mark MAGs from the deep-sea clade, and the asterisk marks ERS493705\_165 (representing Muroto NCV01). This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

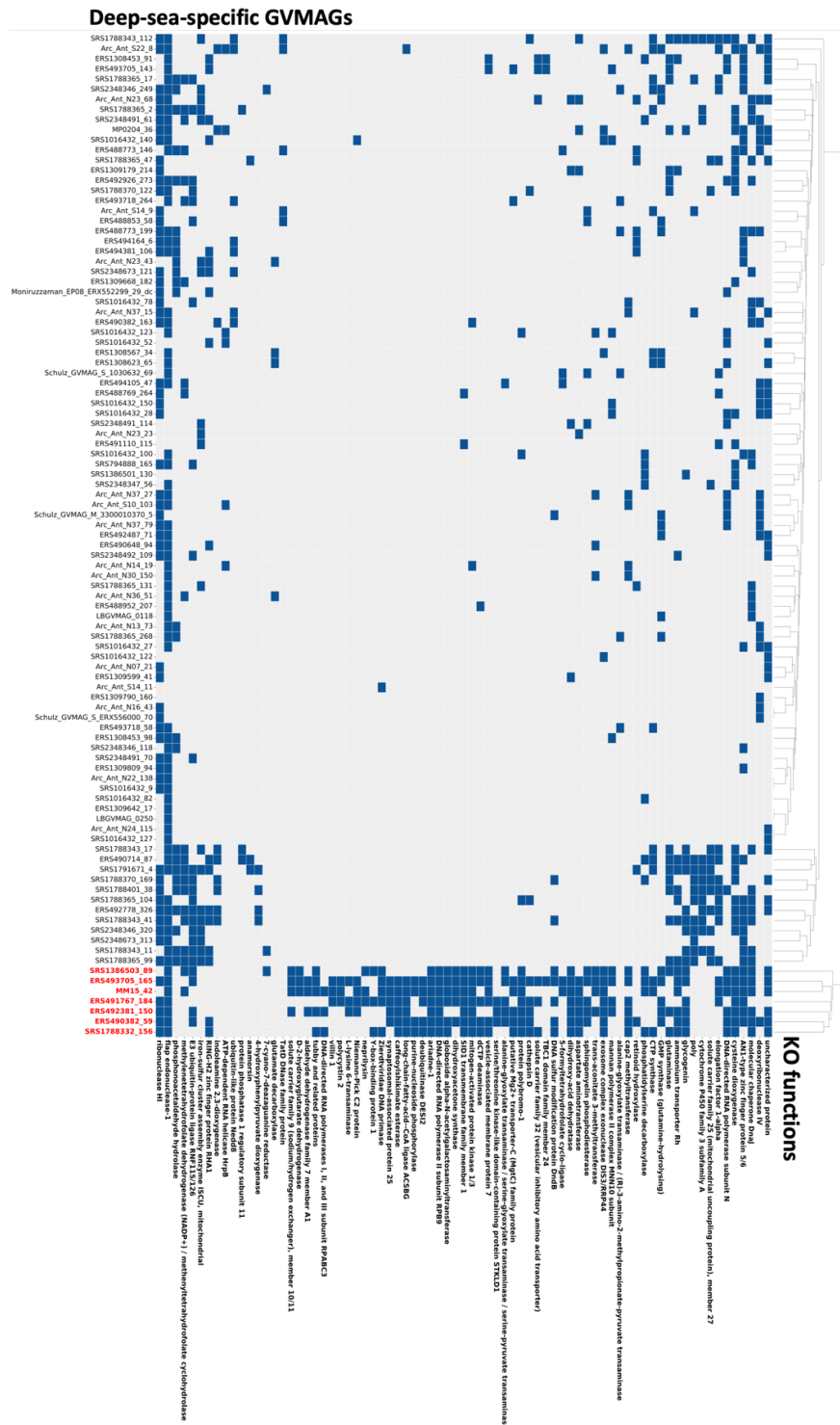
The connectivity map revealed sharing of GVMAGs between multiple oceanic regions (Figure 19c), suggesting the existence of widely distributed deep-sea-specific GVMAGs. Indeed, 55.4% of the deep-sea-specific GVMAGs were detected in two or more regions, with 10 GVs detected in five or more regions (Figure 20d). The Arctic Ocean and Southern Ocean exhibited a large proportion of unique deep-sea-specific GVMAGs (36.4% and 37.5%, respectively). These polar oceans thus represent a “hotspot” of endemic deep-sea-specific GVMAGs, consistent with previous findings (Endo et al., 2020), and may be explained by limited water mass exchanges and steep environmental gradients that form strong ecological barriers between high and lower latitude regions.

Phylogenetic analysis revealed that identified deep-sea-specific GVMAGs are scattered across the tree (Figure 20a), reminiscent of the phylogenetic distribution of GVs specific to cold environments (Meng et al., 2023). One clade within the order *Imitervirales* was enriched with deep-sea-specific GVMAGs (Figure 20b); two Muroto meso-exclusive GVMAGs (NCV01 and NCV26) were also placed within this clade. The members of this clade showed high relative abundances (Figure 20c) and wide geographic distributions (Figure 20d). The Muroto meso-exclusive NCV01 (represented by ERS493705\_165 after dereplication) was distributed in four oceanic regions (Figure 19c, Figure 20d).



**Figure 21 Functional enrichment of deep-sea-specific GVMAGs.**

Scatterplot comparing mean KO frequencies between deep-sea-specific and other GVMAGs. Each dot represents a KO term, with those enriched in deep-sea-specific GVMAGs highlighted in blue. The dashed diagonal line indicates the 1:1 ratio. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.



**Figure 22 Distribution of deep-sea-specific KOs across deep-sea-specific GVMAGs.**

Heatmap showing the presence (blue) or absence (gray) of KO terms enriched in deep-sea-specific GVMAGs.

Columns represent KO terms, and rows represent deep-sea-specific GVMAGs. MAG names shown in red indicate GVMAGs in the deep-sea-specific GVMAGs clade described in the main text. This figure is reproduced from Liu et al. (2023) under the CC-BY 4.0 license.

To investigate potential functional adaptations of deep-sea-specific GVMAGs, I identified significantly enriched KO terms and Pfam domains in deep-sea-specific GVMAGs using Fisher's exact test with FDR correction ( $P$ -value  $< 0.05$ ) (see Materials and Methods). Seventy-six KO terms were found to be enriched in deep-sea-specific GVMAGs (Figure 21, Figure 22). Consistent with my results from the Muroto data, genes related to the ubiquitin system were found as deep-sea-specific KOs. Ammonium transporters and glutaminase were also deep-sea-specific. A previous study also revealed an enrichment of ammonium transporters in GVs detected below 200 m and suggested that these transporters may enhance nitrogen acquisition by their hosts, possibly in competition with Thaumarchaeota (Sheam & Luo, 2025). The above mentioned *Imitervirales* deep-sea clade showed the highest number of deep-sea-specific KO terms (Figure 22), suggesting lineage-specific specialization. Pfam domain analysis revealed 74 domains enriched in deep-sea-specific GVMAGs, including homologs of cytochrome P450, CTP synthases, S-adenosylmethionine decarboxylase, and NifU\_N. P450 has been previously reported in many GV genomes and has been proposed to be functionally connected to the 2-oxoglutarate and Fe (II)-dependent dioxygenase (Lamb et al., 2019), which was also enriched in deep-sea-specific GVMAGs. The function of P450 in GVs remains unknown; however, it has been hypothesized that these genes might modulate viral or host lipid pools to aid energy production (Lamb et al., 2025). Additionally, NifU is involved in Fe-S cluster formation, which plays roles in diverse cellular processes. NifU has been previously reported in eukaryotes, although its specific function remains unresolved (Hwang et al., 1996). Two KO terms and 22 Pfam domains were enriched in the non-deep-sea-specific GVMAGs. These included domains for DNA damage repair (UV damage endonuclease and 5'-3' exonucleases), rhodopsins, phosphate transport associated proteins (PhoH), and sulfotransferases. The enrichment of PhoH suggests adaptation to phosphate-limited conditions (Kim et al., 1993; Moniruzzaman et al., 2020).

Taken together, the KO and Pfam enrichment patterns revealed that GVs in the deep-sea have evolved to encode several genes of unique functions, potentially enabling them to maintain replicative efficiency by modulating host metabolic programs under environmental constraints such as limited light availability and scarce organic compounds.

## Chapter 4 Conclusions and Perspectives

### 4.1 Summary

This doctoral work addresses the knowledge gaps in the ecology and evolution of GVs in deep oceans by integrating local time-series observations and global data analysis.

In the first stage, a high-fidelity sampling workflow was established optimized for mesopelagic waters. Given the low viral abundance and the rapid degradation of nucleic acids in deep-sea environments, this workflow was designed to maximize recovery and minimize contamination. The filtration system was directly connected to the mesopelagic seawater tap, enabling faster and cleaner processing. The sequential filtration and *in situ* preservation protocol was coupled with a computational pipeline for genome assembly and binning, enabling the recovery of high-quality GVMAGs from both newly collected samples at the Muroto station and public datasets. This framework contributed to the construction of the GVGR database, which includes 4,473 species-level GV genomes, thereby providing a comprehensive reference for downstream ecological analyses.

The second stage involved year-round, depth-resolved sampling at the Muroto station. Eight seasonal campaigns produced paired surface and mesopelagic samples for 18S rRNA gene metabarcoding, metagenomics, and metatranscriptomics. A total of 48 GVMAGs were reconstructed, including six detected exclusively in mesopelagic metagenomes, many of which showed persistent transcriptional activity across seasons. Conversely, some GVMAGs detected in the mesopelagic layer were active in the surface layer but inactive in the mesopelagic layer, suggesting that GV populations detected among mesopelagic metagenomes may partially originate from the surface through vertical transport. Transcripts of hallmark replication genes and AMGs were consistently detected in mesopelagic samples. Genes associated with the ubiquitin system were highly expressed in meso-specific GVs, implying roles in modulating host cell death or protein turnover under low-energy conditions. Host predictions indicated that Cercozoa, Picozoa, Syndiniales, dinoflagellates, and

stramenopiles are potential hosts for meso-specific GVs, implying a diverse host range among heterotrophic or mixotrophic microeukaryotes.

The third stage expanded the analysis to a global scale by examining 1,890 marine metagenomes. Using the GVGR database, the distribution and phylogenomic features of GVs across broader depths were assessed. This analysis identified 101 deep-sea-specific GVMAGs. These viruses were found across multiple oceanic regions and spanned diverse phylogenetic lineages, suggesting that deep-sea adaptation has occurred independently in multiple GV clades. Genomic differences between deep-sea-specific and other GVs were revealed, including 76 KO terms and 74 Pfam domains enriched in the deep-sea-specific GVMAGs. In addition to ubiquitin-related genes identified in Muroto samples, other enriched genes included homologs of cytochrome P450, CTP synthase, S-adenosylmethionine decarboxylase, and NifU\_N. These findings collectively support the hypothesis that multiple GV lineages have evolved distinct genetic repertoires to adapt to deep-sea environments by modulating host cellular processes to survive in the aphotic environment.

## **4.2 Limitations and future plans**

Although this thesis provides new insights into deep-sea GVs, several limitations remain that constrain the breadth and depth of conclusions, suggesting avenues for future research.

First, the spatial and temporal scope of sampling in this study was limited, which constrains the generality of the conclusions. Time-series samplings were collected from a single mesopelagic station (Muroto, 320 m depth) and over a relatively short period, so patterns observed may not hold across other regions or times. The deep ocean is vast and heterogeneous, and indeed viruses in other extreme environments can differ markedly. For example, nearly all dominant viral sequences in the Mariana Trench are unclassified novel viruses (Zhou et al., 2022) unlike those from shallower waters. Moreover, deep-sea multiomics data remain scarce compared to upper-ocean datasets, hampering global comparisons and the detection of active infections. This data gap is only beginning to be

filled: researchers have recently demonstrated that viral metatranscriptomes can be obtained even from the deepest ocean (Dang et al., 2024). Future work should therefore expand sampling in both space and time to test whether the patterns observed here are universal or locally specific. Broader and longer-term surveys, combined with expanded deep-sea metagenomic and metatranscriptomic datasets, will facilitate more robust comparisons and improve the distinction between dormant viral sequences from actively expressed infections.

Second, linking deep-sea GVs to their specific hosts remains challenging, limiting ecological interpretation. In my analyses, host predictions relied on sequence homology, phylogeny, and co-occurrence patterns, but these are hindered by the lack of reference genomes for deep-sea protists. Few eukaryotic microbial genomes from the mesopelagic or bathypelagic zones have been sequenced (and even fewer protist isolates are available), leaving many GV genes without identifiable host matches. Indeed, numerous deep-ocean protistan lineages are known only from environmental DNA with no cultured representatives, which restricts host–virus assignments to broad taxonomic hints at best. To overcome this, emerging single-cell technologies offer a promising path. For example, a recent study applied single-cell RNA sequencing to natural communities and successfully paired GV transcripts with those of their protist hosts in individual cells (Fromm et al., 2024; Hevroni et al., 2023), effectively catching host–virus interactions *in situ*. Such approaches have validated suspected GV–protist associations that were predicted from metagenomes (Fromm et al., 2024). Applying these cutting-edge approaches in the deep ocean, along with efforts to culture deep-sea protists and their viruses under simulated *in situ* conditions (high pressure, low temperature), will greatly improve host identification. Such methods can provide direct, high-resolution evidence of infection (beyond correlative predictions) and reveal whether deep-sea GVs exhibit unique infection strategies (e.g., chronic or latent cycles) adapted to the sparse host densities at depth.

Third, technical and bioinformatic biases in the data analysis methodology may have influenced the results, underscoring the need for continued method development. Metagenomic assembly and binning tend to recover genomes of relatively abundant and less

complex viruses, so rare or highly divergent GVs might have escaped detection in my data. GV genomes also contain a high fraction of ORFan genes (novel open reading frames with no known homologs), and some lack even hallmark genes like the major capsid protein (Abergel et al., 2015). This makes them difficult to recognize and correctly annotate using standard pipelines that depend on known reference genes. Consequently, the reliance on homology and hallmark genes may bias discovery toward lineages already represented in databases, while truly novel GVs remain hidden. Future studies should integrate complementary genomic approaches to mitigate these biases. For instance, long-read sequencing could complement short-read metagenomic data to capture the full spectrum of GV diversity, including larger or more complex genomes. Adopting multiple, orthogonal detection methods will help ensure that unusual or low-abundance GVs are not overlooked in deep-sea virome surveys.

Forth, the functional characterization of deep-sea GV genes remains rudimentary, presenting both a limitation and an opportunity. A large fraction of the genes encoded by the GVs in my study have no known function or homolog, reflecting the immense genomic novelty in the deep ocean virosphere. GV genomes encode a wealth of metabolic and structural proteins, many of which remain uncharacterized. For instance, I identified many GV metabolic genes putatively involved in nutrient cycling or host interaction, but without experimental evidence their actual activities remain speculative. To address this, future research should employ functional genomics and other phenotype-driven approaches. Laboratory experiments (for example, expressing candidate viral enzymes) will be crucial to assigning functions to currently hypothetical genes. Exploring gene suites that are unique to aphotic lineages through experiments has biotechnological benefits.

Finally, an open question with significant implications is the role of deep-sea GVs in carbon and nutrient cycling. Connecting genomic findings to ecosystem function remains speculative because of the lack of direct measurements. It is unknown to what extent GV infections contribute to deep-sea carbon sequestration or remineralization, for instance, by lysing hosts and releasing organic matter or by altering particle sinking dynamics. Future

work should integrate viral ecology with geochemical measurements to quantify these effects. For example, time-series observations from deep-moored sediment traps (as at Station ALOHA) could be combined with viral community analyses to determine whether fluctuations in specific viruses correlate with pulses of sinking particulate organic carbon. Initial evidence suggests that GVs do track with carbon export events, hinting at a causal role (Sheam et al., 2025). Similarly, experimental approaches like stable isotope tracing could directly test whether viral lysis accelerates organic matter turnover at depth or, alternatively, whether viral infection induces the formation of larger, faster-sinking aggregates from microbial cells.

In conclusion, each limitation identified in this study presents an opportunity for deeper exploration of deep-sea GVs. Addressing these gaps—through broader sampling, expanded multi-omics resources, improved host identification, enhanced methodologies, functional gene characterization, and integration with biogeochemical cycles—will pave the way for a more comprehensive understanding of the ecological roles and evolutionary adaptations of GVs in the deep ocean. Collaborative, interdisciplinary research will be key to uncovering how these viruses shape life and elemental processes in Earth’s largest and least understood ecosystem.

## **Data Availability**

The data generated from the article, the processed data, and metadata are available in the DNA database of Japan (DDBJ) at [www.ddbj.nig.ac.jp](http://www.ddbj.nig.ac.jp), and can be accessed under accession numbers PRJDB35423 and PRJDB35424. The nucleotide and protein sequences of the Muroto GVMAGs and GVGR database constructed in this study are available at GenomeNet: [https://www.genome.jp/ftp/db/community/GVGR\\_MurotoGV](https://www.genome.jp/ftp/db/community/GVGR_MurotoGV). Other data used in this study include the Giant Virus Orthologous Groups (GVOGs) database (<https://faylward.github.io/GVDB/>).

## Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Hiroyuki Ogata. He gave me the opportunity to join his laboratory, where I was first introduced to and trained in bioinformatics and marine giant viruses field. His generous academic guidance taught me not only how to conduct research rigorously, but also how to think critically as a scientist. Beyond academic support, he provided help in daily life and encouragement during difficult times. I will always remember his words about having lived abroad himself and understanding the challenges of being a foreigner. I feel truly fortunate to have had such a dedicated and compassionate advisor throughout my doctoral journey.

I would also like to thank associate professor Hisashi Endo and assistant professor Yusuke Okazaki for their valuable guidance and advice, which greatly helped me in sampling techniques, data analysis, and paper writing. I am grateful to Dr. Russell Y. Neches and Dr. Lingjie Meng for their contributions and support during data analysis and interpretation. I would like to thank Yosuke Nishimura for supporting global data analysis. I also thank Ethan Mimick and Prof. Mohammad Moniruzzaman for their insightful input during the analysis and interpretation of the manuscript. Special thanks go to Komei Nagasaka for his collaboration and support in both sample processing and global data analysis, which formed the basis of the comparative part of this work. I also thank all lab members for their valuable assistance with sampling, especially Jun Xia, Komei Nagasaka, Hiroto Sasaki, Dr. Russell Y. Neches, Prof. Hisashi Endo, and Prof. Yusuke Okazaki. I thank Prof. Pascal Hingamp for encouraging me and documenting the sampling process, which has become a treasure memory for me.

I would like to acknowledge the Kyoto University Supercomputer System and Mr. Hideya Uehara for computational support. I am also grateful to the Kochi Prefectural Deep Seawater Laboratory for providing access to sampling and environmental data.

Finally, I sincerely thank my family for their continuous support and companionship during my studies.

## References

- Abbas, R., & Larisch, S. (2021). Killing by degradation: regulation of apoptosis by the ubiquitin-proteasome-system. *Cells*, 10(12), 3465. <https://doi.org/10.3390/cells10123465>
- Abergel, C., Legendre, M., & Claverie, J. M. (2015). The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiology Reviews*, 39(6), 779–796. <https://doi.org/10.1093/femsre/fuv037>
- Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., Assis, F., Boratto, P., Andrade, M., Kroon, E. G., Ribeiro, B., Bergier, I., Seligmann, H., Ghigo, E., Colson, P., Levasseur, A., Kroemer, G., Raoult, D., & La Scola, B. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nature Communications*, 9(1), 749. <https://doi.org/10.1038/s41467-018-03168-1>
- Acinas, S. G., Sánchez, P., Salazar, G., Cornejo-Castillo, F. M., Sebastián, M., Logares, R., Royo-Llonch, M., Paoli, L., Sunagawa, S., Hingamp, P., Ogata, H., Lima-Mendez, G., Roux, S., González, J. M., Arrieta, J. M., Alam, I. S., Kamau, A., Bowler, C., Raes, J., ... Gasol, J. M. (2021). Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology*, 4(1), 604. <https://doi.org/10.1038/s42003-021-02112-2>
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, 36(7), 2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>
- Aristegui, J., Gasol, J. M., Duarte, C. M., & Herndl, G. J. (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnology and Oceanography*, 54(5), 1501–1529. <https://doi.org/10.4319/lo.2009.54.5.1501>
- Aroney, S. T. N., Newell, R. J. P., Nissen, J. N., Camargo, A. P., Tyson, G. W., & Woodcroft, B. J. (2025). CoverM: read alignment statistics for metagenomics. *Bioinformatics*, 41(4), btaf147. <https://doi.org/10.1093/bioinformatics/btaf147>
- Athira, M. T., & Antony, S. P. (2023). The Tiny Giants: Overview of Giant Viruses. *Ecological Genetics and Genomics*, 29, 100210. <https://doi.org/10.1016/j.egg.2023.100210>
- Aylward, F. O., & Moniruzzaman, M. (2021). ViralRecall—a flexible command-line tool for the detection of giant virus signatures in 'omic data. *Viruses*, 13(2), 150. <https://doi.org/10.3390/v13020150>
- Aylward, F. O., Moniruzzaman, M., Ha, A. D., & Koonin, E. V. (2021). A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biology*, 19(10), e3001430. <https://doi.org/10.1371/journal.pbio.3001430>
- Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., Spang, A., Wolf, Y. I., Koonin, E. V., & Ettema, T. J. G. (2019). Virus genomes from deep sea

- sediments expand the ocean megavirome and support independent origins of viral gigantism. *mBio*, 10(2), e02497-18. <https://doi.org/10.1128/mBio.02497-18>
- Barrado-Gil, L., Galindo, I., Martínez-Alonso, D., Viedma, S., & Alonso, C. (2017). The ubiquitin-proteasome system is required for African swine fever replication. *PLoS One*, 12(12), e0189741. <https://doi.org/10.1371/journal.pone.0189741>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- 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>
- Blanco-Bercial, L., Parsons, R., Bolaños, L. M., Johnson, R., Giovannoni, S. J., & Curry, R. (2022). The protist community traces seasonality and mesoscale hydrographic features in the oligotrophic Sargasso Sea. *Frontiers in Marine Science*, 9, 897140. <https://doi.org/10.3389/fmars.2022.897140>
- Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, 6(1), 90. <https://doi.org/10.1186/s40168-018-0470-z>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Boughalmi, M., Saadi, H., Pagnier, I., Colson, P., Fournous, G., Raoult, D., & La Scola, B. (2013). High-throughput isolation of giant viruses of the Mimiviridae and Marseilleviridae families in the Tunisian environment. *Environmental Microbiology*, 15(7), 2000–2007. <https://doi.org/10.1111/1462-2920.12068>
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, 99(22), 14250–14255. <https://doi.org/10.1073/pnas.202488399>
- Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18(4), 366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Buscaglia, M., Iriarte, J. L., Schulz, F., & Díez, B. (2024). Adaptation strategies of giant viruses to low-temperature marine ecosystems. *The ISME Journal*, 18(1), wrae162. <https://doi.org/10.1093/ismejo/wrae162>

- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., & Kyrpides, N. C. (2023). Identification of mobile genetic elements with geNomad. *Nature Biotechnology*, 42(8), 1303–1312. <https://doi.org/10.1038/s41587-023-01953-y>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A., Engelen, S., Madoui, M. A., Méheust, R., Poulain, J., Romac, S., Richter, D. J., Yoshikawa, G., ... Wincker, P. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*, 9(1), 373. <https://doi.org/10.1038/s41467-017-02342-1>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chung, D., Brask, N., Matar, S., Gallot-Lavallée, L., Pringle, E. S., Duguay, B. A., ... & Archibald, J. M. (2025). Persistent mirusvirus infection in the marine protist *Aurantiochytrium*. *Nature Communications*, 16(1), 9922. <https://doi.org/10.1038/s41467-025-65172-6>
- Collier, J. L., Rest, J. S., Gallot-Lavallée, L., Lavington, E., Kuo, A., Jenkins, J., Plott, C., Pangilinan, J., Daum, C., Grigoriev, I. V., Filloramo, G. V., Novák Vanclová, A. M. G., & Archibald, J. M. (2023). The protist *Aurantiochytrium* has universal subtelomeric rDNAs and is a host for mirusviruses. *Current Biology*, 33(23), 5199–5207.e4. <https://doi.org/10.1016/j.cub.2023.10.009>
- Comeau, A. M., Li, W. K. W., Tremblay, J. É., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One*, 6(11), e27492. <https://doi.org/10.1371/journal.pone.0027492>
- Cordier, T., Angeles, I. B., Henry, N., Lejzerowicz, F., Berney, C., Morard, R., Brandt, A., Cambon-Bonavita, M. A., Guidi, L., Lombard, F., Arbizu, P. M., Massana, R., Orejas, C., Poulain, J., Smith, C. R., Wincker, P., Arnaud-Haond, S., Gooday, A. J., de Vargas, C., & Pawlowski, J. (2022). Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. *Science Advances*, 8(5), eabj9309. <https://doi.org/10.1126/sciadv.abj9309>
- Da Cunha, V., Gaia, M., Ogata, H., Jaillon, O., Delmont, T. O., & Forterre, P. (2022). Giant viruses encode actin-related proteins. *Molecular Biology and Evolution*, 39(2), msac022. <https://doi.org/10.1093/molbev/msac022>
- Dang, Y. R., Cha, Q. Q., Liu, S. S., Wang, S. Y., Li, P. Y., Li, C. Y., Wang, P., Chen, X. L., Tian, J. W., Xin, Y., Chen, Y., Zhang, Y. Z., & Qin, Q. L. (2024). Phytoplankton-derived

- polysaccharides and microbial peptidoglycans are key nutrients for deep-sea microbes in the Mariana Trench. *Microbiome*, 12(1), 77. <https://doi.org/10.1186/s40168-024-01789-x>
- Danovaro, R., Corinaldesi, C., Dell’Anno, A., & Snelgrove, P. V. R. (2017). The deep-sea under global change. *Current Biology*, 27(11), R461–R465. <https://doi.org/10.1016/j.cub.2017.02.046>
- Demory, D., Endo, H., Baudoux, A. C., Bigeard, E., Grimsley, N., Simon, N., Ogata, H., & Weitz, J. S. (2025). Temperature-driven biogeography of marine giant viruses infecting picoeukaryotes *Micromonas*. *ISME Communications*, 5(1), ycaf137. <https://doi.org/10.1093/ismeco/ycaf137>
- Dick, G. J., Anantharaman, K., Baker, B. J., Li, M., Reed, D. C., & Sheik, C. S. (2013). The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Frontiers in Microbiology*, 4, 124. <https://doi.org/10.3389/fmicb.2013.00124>
- Endo, H., Blanc-Mathieu, R., Li, Y., Salazar, G., Henry, N., Labadie, K., de Vargas, C., Sullivan, M. B., Bowler, C., Wincker, P., Karp-Boss, L., Sunagawa, S., & Ogata, H. (2020). Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nature Ecology and Evolution*, 4(12), 1639–1649. <https://doi.org/10.1038/s41559-020-01288-w>
- Erazo-Garcia, M. P., Sheyn, U., Barth, Z. K., Craig, R. J., Wessman, P., Jivaji, A. M., Ray, W. K., Svensson-Coelho, M., Cornwallis, C. K., Rengefors, K., Brussaard, C. P. D., Moniruzzaman, M., & Aylward, F. O. (2025). Cryptic infection of a giant virus in a unicellular green alga. *Science*, 388(6748), eads6303. <https://doi.org/10.1126/science.ads6303>
- Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, 3, e1319. <https://doi.org/10.7717/peerj.1319>
- Fang, Y., Meng, L., Xia, J., Gotoh, Y., Hayashi, T., Nagasaki, K., Endo, H., Okazaki, Y., & Ogata, H. (2025). Genome-resolved year-round dynamics reveal a broad range of giant virus microdiversity. *mSystems*, 10(1), e01168-24. <https://doi.org/10.1128/msystems.01168-24>
- Fischer, M. G., Allen, M. J., Wilson, W. H., & Suttle, C. A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proceedings of the National Academy of Sciences of the United States of America*, 107(45), 19508–19513. <https://doi.org/10.1073/pnas.1007615107>
- Francis, R., Ominami, Y., Bou Khalil, J. Y., & La Scola, B. (2019). High-throughput isolation of giant viruses using high-content screening. *Communications Biology*, 2(1), 216. <https://doi.org/10.1038/s42003-019-0475-6>
- Frohns, F., Käsmann, A., Kramer, D., Schäfer, B., Mehmel, M., Kang, M., Van Etten, J. L., Gazzarrini, S., Moroni, A., & Thiel, G. (2006). Potassium ion channels of Chlorella viruses cause rapid depolarization of host cells during infection. *Journal of Virology*, 80(5), 2437–2444. <https://doi.org/10.1128/JVI.80.5.2437-2444.2006>

- Fromm, A., Hevroni, G., Vincent, F., Schatz, D., Martinez-Gutierrez, C. A., Aylward, F. O., & Vardi, A. (2024). Single-cell RNA-seq of the rare virosphere reveals the native hosts of giant viruses in the marine environment. *Nature Microbiology*, 9(6), 1619–1629. <https://doi.org/10.1038/s41564-024-01669-y>
- Funaoka, Y., Hiromoto, H., Morimoto, D., Takahashi, M., Wada, K., & Nagasaki, K. (2023). Diversity in infection specificity between the bloom-forming microalga *Heterosigma akashiwo* and its dsDNA Virus, *Heterosigma akashiwo* Virus. *Microbes and Environments*, 38(2), ME23036. <https://doi.org/10.1264/jsme2.ME23036>
- Gaïa, M., Meng, L., Pelletier, E., Forterre, P., Vanni, C., Fernandez-Guerra, A., Jaillon, O., Wincker, P., Ogata, H., Krupovic, M., & Delmont, T. O. (2023). Mirusviruses link herpesviruses to giant viruses. *Nature*, 616(7958), 783–789. <https://doi.org/10.1038/s41586-023-05962-4>
- Giner, C. R., Pernice, M. C., Balagué, V., Duarte, C. M., Gasol, J. M., Logares, R., & Massana, R. (2020). Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *The ISME Journal*, 14(2), 437–449. <https://doi.org/10.1038/s41396-019-0506-9>
- Groussman, R. D., Blaskowski, S., Coesel, S. N., & Armbrust, E. V. (2023). MarFERReT, an open-source, version-controlled reference library of marine microbial eukaryote functional genes. *Scientific Data*, 10(1), 926. <https://doi.org/10.1038/s41597-023-02842-4>
- Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, 9(1), 37. <https://doi.org/10.1186/s40168-020-00990-y>
- Guo Y.J. (1991). The Kuroshio. Part II. Primary productivity and phytoplankton. *Oceanography and Marine Biology: An Annual Review*, 29, 155–189.
- Ha, A. D., & Aylward, F. O. (2024). Automated classification of giant virus genomes using a random forest model built on trademark protein families. *Npj Viruses*, 2(1), 9. <https://doi.org/10.1038/s44298-024-00021-9>
- Ha, A. D., Moniruzzaman, M., & Aylward, F. O. (2021). High transcriptional activity and diverse functional repertoires of hundreds of giant viruses in a coastal marine system. *mSystems*, 6(4), e00293-21. <https://doi.org/10.1128/msystems.00293-21>
- Ha, A. D., Moniruzzaman, M., & Aylward, F. O. (2023). Assessing the biogeography of marine giant viruses in four oceanic transects. *ISME Communications*, 3(1), 43. <https://doi.org/10.1038/s43705-023-00252-6>
- Han, Y., Gao, C., Liang, Y., McMinn, A., & Wang, M. (2024). Diversity and Ecological Roles of Deep-Sea Viruses. *Ocean-Land-Atmosphere Research*, 3, 0067. <https://doi.org/10.34133/olar.0067>
- He, C., Fuchich, D., Sosa, A., Wang, H., Kan, J., Liu, J., Xu, Y., Jiao, N., Gonsior, M., & Chen, F. (2024). Deep metagenomic sequencing unveils novel SAR202 lineages and their vertical

- adaptation in the ocean. *Communications Biology*, 7(1), 853. <https://doi.org/10.1038/s42003-024-06535-5>
- He, T., Jin, M., Cui, P., Sun, X., He, X., Huang, Y., Xiao, X., Zhang, T., & Zhang, X. (2024). Environmental viromes reveal the global distribution signatures of deep-sea DNA viruses. *Journal of Advanced Research*, 57, 107–117. <https://doi.org/10.1016/j.jare.2023.04.009>
- Herndl, G. J., Bayer, B., Baltar, F., & Reinthaler, T. (2023). Prokaryotic Life in the Deep Ocean's Water Column. *Annual Review of Marine Science*, 15, 461–483. <https://doi.org/10.1146/annurev-marine-032122-115655>
- Herndl, G. J., & Reinthaler, T. (2013). Microbial control of the dark end of the biological pump. *Nature Geoscience*, 6(9), 718–724. <https://doi.org/10.1038/ngeo1921>
- Hershko, A. (2005). The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. *Cell Death and Differentiation*, 12(9), 1191–1197. <https://doi.org/10.1038/sj.cdd.4401702>
- Hevroni, G., Vincent, F., Ku, C., Sheyn, U., & Vardi, A. (2023). Daily turnover of active giant virus infection during algal blooms revealed by single-cell transcriptomics. *Science Advances*, 9(41), eadf7971. <https://doi.org/10.1126/sciadv.adf7971>
- Heyerhoff, B., Engelen, B., & Bunse, C. (2022). Auxiliary Metabolic Gene Functions in Pelagic and Benthic Viruses of the Baltic Sea. *Frontiers in Microbiology*, 13, 863620. <https://doi.org/10.3389/fmicb.2022.863620>
- Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., Ferrera, I., Sarmiento, H., Villar, E., Lima-Mendez, G., Faust, K., Sunagawa, S., Claverie, J. M., Moreau, H., Desdevises, Y., Bork, P., Raes, J., De Vargas, C., Karsenti, E., ... Ogata, H. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal*, 7(9), 1678–1695. <https://doi.org/10.1038/ismej.2013.59>
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Hwang, D. M., Dempsey, A., Tan, K. T., & Liew, C. C. (1996). A modular domain of NiFu, a nitrogen fixation cluster protein, is highly conserved in evolution. *Journal of Molecular Evolution*, 43(5), 536–540. <https://doi.org/10.1007/bf02337525>
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119. <https://doi.org/10.1186/1471-2105-11-119>
- Jian, H., Yi, Y., Wang, J., Hao, Y., Zhang, M., Wang, S., Meng, C., Zhang, Y., Jing, H., Wang, Y., & Xiao, X. (2021). Diversity and distribution of viruses inhabiting the deepest ocean on Earth. *The ISME Journal*, 15(10), 3094–3110. <https://doi.org/10.1038/s41396-021-00994-y>

- Jiao, N., & Zheng, Q. (2011). The microbial carbon pump: from genes to ecosystems. *Applied and Environmental Microbiology*, 77(21), 7439–7444. <https://doi.org/10.1128/AEM.05640-11>
- Kaneko, H., Blanc-Mathieu, R., Endo, H., Chaffron, S., Delmont, T. O., Gaia, M., Henry, N., Hernández-Velázquez, R., Nguyen, C. H., Mamitsuka, H., Forterre, P., Jaillon, O., de Vargas, C., Sullivan, M. B., Suttle, C. A., Guidi, L., & Ogata, H. (2020). Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience*, 24(1), 102002. <https://doi.org/10.1016/j.isci.2020.102002>
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359. <https://doi.org/10.7717/peerj.7359>
- Kang, M., Dunigan, D. D., & Van Etten, J. L. (2005). Chlorovirus: a genus of Phycodnaviridae that infects certain chlorella-like green algae. *Molecular Plant Pathology*, 6(3), 213–224. <https://doi.org/10.1111/j.1364-3703.2005.00281.x>
- Kang, Y., Kawamoto, J., Kaneda, S., Aritome, K., & Sakurai, K. (2004). Characterization of selenium in the deep ocean water pumped up at Muroto, Japan. *Deep Ocean Water Research*, 5, 43–52.
- Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P., & Venclovas, C. S. (2020). Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Research*, 48(18), 10142–10156. <https://doi.org/10.1093/nar/gkaa760>
- Khalil, J. Y. B., Robert, S., Reteno, D. G., Andreani, J., Raoult, D., & La Scola, B. (2016). High-throughput isolation of giant viruses in liquid medium using automated flow cytometry and fluorescence staining. *Frontiers in Microbiology*, 7, 26. <https://doi.org/10.3389/fmicb.2016.00026>
- Kijima, S., Delmont, T. O., Miyazaki, U., Gaia, M., Endo, H., & Ogata, H. (2021). Discovery of viral myosin genes with complex evolutionary history within plankton. *Frontiers in Microbiology*, 12, 683294. <https://doi.org/10.3389/fmicb.2021.683294>
- Kijima, S., Hikida, H., Delmont, T. O., Gaia, M., & Ogata, H. (2024). Complex genomes of early nucleocytoviruses revealed by ancient origins of viral aminoacyl-tRNA synthetases. *Molecular Biology and Evolution*, 41(8), msae149. <https://doi.org/10.1093/molbev/msae149>
- Kim, S. K., Makino, K., Amemura, M., Shinagawa, H., & Nakata, A. (1993). Molecular analysis of the phoH gene, belonging to the phosphate regulon in Escherichia coli. *Journal of Bacteriology*, 175(5), 1316–1324. <https://doi.org/10.1128/jb.175.5.1316-1324.1993>
- 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*, 84(2): e00061-19. <https://doi.org/10.1128/MMBR.00061-19>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28(24), 3211–3217.

<https://doi.org/10.1093/bioinformatics/bts611>

- Kozuch, D. J., Stillinger, F. H., & Debenedetti, P. G. (2019). Low temperature protein refolding suggested by molecular simulation. *The Journal of Chemical Physics*, 151(18), 185101. <https://doi.org/10.1063/1.5128211>
- Krebs, C. J. (1989). *Ecological methodology*. HarperCollins College Division, New York.
- La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J. M., & Raoult, D. (2003). A giant virus in amoebae. *Science*, 299(5615), 2033. <https://doi.org/10.1126/science.1081867>
- Lamb, D. C., Follmer, A. H., Goldstone, J. V., Nelson, D. R., Warrilow, A. G., Price, C. L., True, M. Y., Kelly, S. L., Poulos, T. L., & Stegeman, J. J. (2019). On the occurrence of cytochrome P450 in viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 116(25), 12343–12352. <https://doi.org/10.1073/pnas.1901080116>
- Lamb, D. C., Goldstone, J. V., Belhaouari, D. B., Andréani, J., Farooqi, A., Allen, M. J., Kelly, S. L., La Scola, B., & Stegeman, J. J. (2025). Cytochrome b5 occurrence in giant and other viruses belonging to the phylum Nucleocytoviricota. *Npj Viruses*, 3(1), 8. <https://doi.org/10.1038/s44298-025-00091-3>
- Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*, 35(3), 421–432. <https://doi.org/10.1093/bioinformatics/bty648>
- Lara, E., Vaqué, D., Sà, E. L., Boras, J. A., Gomes, A., Borrull, E., Díez-Vives, C., Teira, E., Pernice, M. C., Garcia, F. C., Forn, I., Castillo, Y. M., Peiró, A., Salazar, G., Morán, X. A. G., Massana, R., Catalá, T. S., Luna, G. M., Agustí, S., ... Duarte, C. M. (2017). Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Science Advances*, 3(9), e1602565. <https://doi.org/10.1126/sciadv.1602565>
- Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M., Poirrot, O., Bertaux, L., Bruley, C., Couté, Y., Rivkina, E., Abergel, C., & Claverie, J. M. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), 4274–4279. <https://doi.org/10.1073/pnas.1320670111>
- Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., Alempic, J. M., Ramus, C., Bruley, C., Labadie, K., Shmakova, L., Rivkina, E., Couté, Y., Abergel, C., & Claverie, J. M. (2015). In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), E5327–E5335. <https://doi.org/10.1073/pnas.1510795112>
- Letunic, I., & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, 52(W1), W78–W82. <https://doi.org/10.1093/nar/gkae268>

- Levins, R. (1968). *Evolution in changing environments*. Princeton University Press, Princeton, New Jersey, USA.
- Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li, J., & Zhan, Q. (2011). The role of centrosomal Nlp in the control of mitotic progression and tumorigenesis. *British Journal of Cancer*, 104(10), 1523–1528. <https://doi.org/10.1038/bjc.2011.130>
- Li, Y., Hingamp, P., Watai, H., Endo, H., Yoshida, T., & Ogata, H. (2018). Degenerate PCR primers to reveal the diversity of giant viruses in coastal waters. *Viruses*, 10(9), 496. <https://doi.org/10.3390/v10090496>
- Luo, E., Leu, A. O., Eppley, J. M., Karl, D. M., & DeLong, E. F. (2022). Diversity and origins of bacterial and archaeal viruses on sinking particles reaching the abyssal ocean. *The ISME Journal*, 16(6), 1627–1635. <https://doi.org/10.1038/s41396-022-01202-1>
- Maat, D. S., Bale, N. J., Hopmans, E. C., Damsté, J. S. S., Schouten, S., & Brussaard, C. P. D. (2016). Increasing P limitation and viral infection impact lipid remodeling of the picophytoplankter *Micromonas pusilla*. *Biogeosciences*, 13(5), 1667–1676. <https://doi.org/10.5194/bg-13-1667-2016>
- Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M., & Delong, E. F. (2017). Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nature Microbiology*, 2(10), 1367–1373. <https://doi.org/10.1038/s41564-017-0008-3>
- Meng, L., Delmont, T. O., Gaïa, M., Pelletier, E., Fernández-Guerra, A., Chaffron, S., Neches, R. Y., Wu, J., Kaneko, H., Endo, H., & Ogata, H. (2023). Genomic adaptation of giant viruses in polar oceans. *Nature Communications*, 14(1), 6233. <https://doi.org/10.1038/s41467-023-41910-6>
- Meng, L., Endo, H., Blanc-Mathieu, R., Chaffron, S., Hernández-Velázquez, R., Kaneko, H., & Ogata, H. (2021). Quantitative assessment of nucleocytoplasmic large DNA virus and host interactions predicted by co-occurrence analyses. *mSphere*, 6(2), e01298-20. <https://doi.org/10.1128/mSphere.01298-20>
- Middelboe, M., Glud, R. N., Wenzhöfer, F., Oguri, K., & Kitazato, H. (2006). Spatial distribution and activity of viruses in the deep-sea sediments of Sagami Bay, Japan. *Deep Sea Research Part I: Oceanographic Research Papers*, 53(1), 1–13. <https://doi.org/10.1016/j.dsr.2005.09.008>
- Mihara, T., Koyano, H., Hingamp, P., Grimsley, N., Goto, S., & Ogata, H. (2018). Taxon richness of “Megaviridae” exceeds those of bacteria and archaea in the ocean. *Microbes and Environments*, 33(2), 162–171. <https://doi.org/10.1264/jsme2.ME17203>
- 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>

- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, 41(12), e121. <https://doi.org/10.1093/nar/gkt263>
- Mizuno, C. M., Ghai, R., Saghai, A., López-García, P., & Rodriguez-Valera, F. (2016). Genomes of abundant and widespread viruses from the deep ocean. *mBio*, 7(4), e00805-16. <https://doi.org/10.1128/mBio.00805-16>
- Monier, A., Pagarete, A., De Vargas, C., Allen, M. J., Read, B., Claverie, J. M., & Ogata, H. (2009). Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Research*, 19(8), 1441–1449. <https://doi.org/10.1101/gr.091686.109>
- Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R., & Aylward, F. O. (2020). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nature Communications*, 11(1), 1710. <https://doi.org/10.1038/s41467-020-15507-2>
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621–628. <https://doi.org/10.1038/nmeth.1226>
- Nagata, T., Tamburini, C., Arístegui, J., Baltar, F., Bochdansky, A. B., Fonda-Umani, S., Fukuda, H., Gogou, A., Hansell, D. A., Hansman, R. L., Herndl, G. J., Panagiotopoulos, C., Reinthaler, T., Sohrin, R., Verdugo, P., Yamada, N., Yamashita, Y., Yokokawa, T., & Bartlett, D. H. (2010). Emerging concepts on microbial processes in the bathypelagic ocean - Ecology, biogeochemistry, and genomics. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 57(16), 1519–1536. <https://doi.org/10.1016/j.dsr2.2010.02.019>
- Nayfach, S., Camargo, A. P., Schulz, F., Eloie-Fadrosch, 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>
- Needham, D. M., Yoshizawa, S., Hosaka, T., Poirier, C., Choi, C. J., Hehenberger, E., Irwin, N. A. T., Wilken, S., Yung, C. M., Bachy, C., Kurihara, R., Nakajima, Y., Kojima, K., Kimura-Someya, T., Leonard, G., Malmstrom, R. R., Mende, D. R., Olson, D. K., Sudo, Y., ... Worden, A. Z. (2019). A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proceedings of the National Academy of Sciences of the United States of America*, 116(41), 20574–20583. <https://doi.org/10.1073/pnas.1907517116>
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nishimura, Y., & Yoshizawa, S. (2022). The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Scientific Data*, 9(1), 305. <https://doi.org/10.1038/s41597-022-01392-5>

- Okazaki, Y., Nguyen, T. T., Nishihara, A., Endo, H., Ogata, H., Nakano, S. I., & Tamaki, H. (2023). A fast and easy method to co-extract DNA and RNA from an environmental microbial sample. *Microbes and Environments*, 38(1), ME22102. <https://doi.org/10.1264/jsme2.ME22102>
- Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), 2864–2868. <https://doi.org/10.1038/ismej.2017.126>
- Pagnier, I., Reteno, D. G. I., Saadi, H., Boughalmi, M., Gaia, M., Slimani, M., Ngounga, T., Bekliz, M., Colson, P., Raoult, D., & La Scola, B. (2013). A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. *Intervirology*, 56(6), 354–363. <https://doi.org/10.1159/000354556>
- Park, K. (1966). Deep-sea pH. *Science*, 154(3756), 1540–1542. <https://doi.org/10.1126/science.154.3756.1540>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- Pernice, M. C., Forn, I., Gomes, A., Lara, E., Alonso-Sáez, L., Arrieta, J. M., Del Carmen Garcia, F., Hernando-Morales, V., Mackenzie, R., Mestre, M., Sintes, E., Teira, E., Valencia, J., Varela, M. M., Vaqué, D., Duarte, C. M., Gasol, J. M., & Massana, R. (2014). Global abundance of planktonic heterotrophic protists in the deep ocean. *The ISME Journal*, 9(3), 782–792. <https://doi.org/10.1038/ismej.2014.168>
- Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J. M., & Abergel, C. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science*, 341(6143), 281–286. <https://doi.org/10.1126/science.1239181>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3), e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Prodinger, F., Endo, H., Takano, Y., Li, Y., Tominaga, K., Isozaki, T., Blanc-Mathieu, R., Gotoh, Y., Hayashi, T., Taniguchi, E., Nagasaki, K., Yoshida, T., & Ogata, H. (2022). Year-round dynamics of amplicon sequence variant communities differ among eukaryotes, Imitervirales and prokaryotes in a coastal ecosystem. *FEMS Microbiology Ecology*, 97(12), fiab167. <https://doi.org/10.1093/femsec/fiab167>
- Prokopyowich, C. D., Gregory, T. R., & Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46(1), 48–50. <https://doi.org/10.1139/g02-103>
- R Core Team. (2024). *R: a language and environment for statistical computing*. <https://www.R-project.org/>

- Ramanujam, V., Alderson, T. R., Pritišanac, I., Ying, J., & Bax, A. (2020). Protein structural changes characterized by high-pressure, pulsed field gradient diffusion NMR spectroscopy. *Journal of Magnetic Resonance*, 312, 106701. <https://doi.org/10.1016/j.jmr.2020.106701>
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., & Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science*, 306(5700), 1344–1350. <https://doi.org/10.1126/science.1101485>
- Rodrigues, R. A. L., Abrahão, J. S., Drumond, B. P., & Kroon, E. G. (2016). Giants among larges: how gigantism impacts giant virus entry into amoebae. *Current Opinion in Microbiology*, 31, 88–93. <https://doi.org/10.1016/j.mib.2016.03.009>
- Rosenwasser, S., Ziv, C., Creveld, S. G. van, & Vardi, A. (2016). Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean. *Trends in Microbiology*, 24(10), 821–832. <https://doi.org/10.1016/j.tim.2016.06.006>
- Sánchez, P., Coutinho, F. H., Sebastián, M., Pernice, M. C., Rodríguez-Martínez, R., Salazar, G., Cornejo-Castillo, F. M., Pesant, S., López-Alforja, X., López-García, E. M., Agustí, S., Gojobori, T., Logares, R., Sala, M. M., Vaqué, D., Massana, R., Duarte, C. M., Acinas, S. G., & Gasol, J. M. (2024). Marine picoplankton metagenomes and MAGs from eleven vertical profiles obtained by the Malaspina Expedition. *Scientific Data*, 11(1), 154. <https://doi.org/10.1038/s41597-024-02974-1>
- Schulz, F., Abergel, C., & Woyke, T. (2022). Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nature Reviews Microbiology*, 20(12), 721–736. <https://doi.org/10.1038/s41579-022-00754-5>
- Schulz, F., Andreani, J., Francis, R., Boudjemaa, H., Bou Khalil, J. Y., Lee, J., La Scola, B., & Woyke, T. (2020). Advantages and Limits of Metagenomic Assembly and Binning of a Giant Virus. *mSystems*, 5(3), e00048–20. <https://doi.org/10.1128/mSystems.00048-20>
- Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., McMahon, K. D., Konstantinidis, K. T., Eloë-Fadrosch, E. A., Kyrpides, N. C., & Woyke, T. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature*, 578(7795), 432–436. <https://doi.org/10.1038/s41586-020-1957-x>
- Sheam, M. M., & Luo, E. (2025). Vertical transport and spatiotemporal dynamics of giant viruses in the North Pacific Subtropical Gyre. *The ISME Journal*, 19(1), wraf094. <https://doi.org/10.1093/ismejo/wraf094>
- Shi, J., Wang, H., Zeng, Y., Fan, Y., Chen, H., Yuan, C., Li, Y., Huang, M., Shi, X., & He, P. (2023). Diversity and distribution of archaeal and bacterial nitrifiers in deep oceans. *Journal of Sea Research*, 193, 102389. <https://doi.org/10.1016/j.seares.2023.102389>
- Sievers, F., & Higgins, D. G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in Molecular Biology*, 1079, 105–116. [https://doi.org/10.1007/978-1-62703-646-7\\_6](https://doi.org/10.1007/978-1-62703-646-7_6)

- Steiner, P. A., De Corte, D., Geijo, J., Mena, C., Yokokawa, T., Rattei, T., Herndl, G. J., & Sintes, E. (2019). Highly variable mRNA half-life time within marine bacterial taxa and functional genes. *Environmental Microbiology*, 21(10), 3873–3884. <https://doi.org/10.1111/1462-2920.14737>
- Sun, T. W., Yang, C. L., Kao, T. T., Wang, T. H., Lai, M. W., & Ku, C. (2020). Host range and coding potential of eukaryotic giant viruses. *Viruses*, 12(11), 1337. <https://doi.org/10.3390/v12111337>
- 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>
- Tan, C., Ding, K., Jin, B., Seyfried, W. E., & Chen, Y. (2012). Development of an in situ pH Calibrator in deep sea environments. *IEEE/ASME Transactions on Mechatronics*, 17(1), 8–15. <https://doi.org/10.1109/TMECH.2011.2168419>
- Tang, Q., Wu, P., Chen, H., & Li, G. (2018). Pleiotropic roles of the ubiquitin-proteasome system during viral propagation. *Life Sciences*, 207, 350–354. <https://doi.org/10.1016/j.lfs.2018.06.014>
- Thomas, E., Anderson, R. E., Li, V., Rogan, L. J., & Huber, J. A. (2021). Diverse Viruses in Deep-Sea Hydrothermal Vent Fluids Have Restricted Dispersal across Ocean Basins. *mSystems*, 6(3), e00068-21. <https://doi.org/10.1128/mSystems.00068-21>
- Thurber, A. R., Sweetman, A. K., Narayanaswamy, B. E., Jones, B., Ingels, J., & Hansman, R. L. (2014). Ecosystem function and services provided by the deep sea. *Biogeosciences*, 11(14), 3941–3963. <https://doi.org/10.5194/bg-11-3941-2014>
- Vincent, F., Gralka, M., Schleyer, G., Schatz, D., Cabrera-Brufau, M., Kuhlisch, C., Sichert, A., Vidal-Melgosa, S., Mayers, K., Barak-Gavish, N., Flores, J. M., Masdeu-Navarro, M., Egge, J. K., Larsen, A., Hehemann, J. H., Marrasé, C., Simó, R., Cordero, O. X., & Vardi, A. (2023). Viral infection switches the balance between bacterial and eukaryotic recyclers of organic matter during coccolithophore blooms. *Nature Communications*, 14(1), 510. <https://doi.org/10.1038/s41467-023-36049-3>
- Von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., & Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology*, 20(1), 217. <https://doi.org/10.1186/s13059-019-1817-x>
- Wang, J., Li, L., & Lin, S. (2023). Active viral infection during blooms of a dinoflagellate indicates dinoflagellate-viral co-adaptation. *Applied Environmental Microbiology*, 89(11), e01156-23. <https://doi.org/10.1128/aem.01156-23>
- Wilson, W. H., Gilg, I. C., Moniruzzaman, M., Field, E. K., Koren, S., Lecleir, G. R., Martínez Martínez, J., Poulton, N. J., Swan, B. K., Stepanauskas, R., & Wilhelm, S. W. (2017). Genomic exploration of individual giant ocean viruses. *The ISME Journal*, 11(8), 1736–1745. <https://doi.org/10.1038/ismej.2017.61>
- Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T. G., Parkhill, J., Barrell, B. G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M. A., Price, C., Rabinowitsch, E.,

- Walker, D., Craigon, M., Roy, D., & Ghazal, P. (2005). Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science*, 309(5737), 1090–1092. <https://doi.org/10.1126/science.1113109>
- Winter, M., Sombroek, D., Dauth, I., Moehlenbrink, J., Scheuermann, K., Crone, J., & Hofmann, T. G. (2008). Control of HIPK2 stability by ubiquitin ligase Siah-1 and checkpoint kinases ATM and ATR. *Nature Cell Biology*, 10(7), 812–824. <https://doi.org/10.1038/ncb1743>
- Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, 347(6223), 1257594. <https://doi.org/10.1126/science.1257594>
- Wu, B., Fan, T., Chen, X., He, Y., & Wang, H. (2024). The class III phosphatidylinositol 3-kinase VPS34 supports EV71 replication by promoting viral replication organelle formation. *Journal of Virology*, 98(10), e00695-24. <https://doi.org/10.1128/jvi.00695-24>
- Wu, J., Gao, W., Zhang, W., & Meldrum, D. R. (2011). Optimization of whole-transcriptome amplification from low cell density deep-sea microbial samples for metatranscriptomic analysis. *Journal of Microbiological Methods*, 84(1), 88–93. <https://doi.org/10.1016/j.mimet.2010.10.018>
- Xia, J., Kameyama, S., Proding, F., Yoshida, T., Cho, K. H., Jung, J., Kang, S. H., Yang, E. J., Ogata, H., & Endo, H. (2022). Tight association between microbial eukaryote and giant virus communities in the Arctic Ocean. *Limnology and Oceanography*, 67(6), 1343–1356. <https://doi.org/10.1002/lno.12086>
- Yakimov, M. M., La Cono, V., Denaro, R., D’Auria, G., Decembrini, F., Timmis, K. N., Golyshin, P. N., & Giuliano, L. (2007). Primary producing prokaryotic communities of brine, interface and seawater above the halocline of deep anoxic lake L’Atalante, Eastern Mediterranean Sea. *The ISME Journal*, 1(8), 743–755. <https://doi.org/10.1038/ismej.2007.83>
- Yamaguchi, T., Inoue, T., Hirakawa, M., Abe, S., Ishii, K., Kagoura, T., & Fujiwara, M. (2003). Deep-sea water suction technology. *Furukawa Review*, (24), 75–80.
- Yau, S., Lauro, F. M., DeMaere, M. Z., Brown, M. V., Thomas, T., Raftery, M. J., Andrews-Pfannkoch, C., Lewis, M., Hoffman, J. M., Gibson, J. A., & Cavicchioli, R. (2011). Virophage control of antarctic algal host-virus dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), 6163–6168. <https://doi.org/10.1073/pnas.1018221108>
- Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., Ogata, H., & Takemura, M. (2019). Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. *Journal of Virology*, 93(8), e02130-18. <https://doi.org/10.1128/JVI.02130-18>
- Yutin, N., Wolf, Y. I., Raoult, D., & Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virology Journal*, 6(1), 223. <https://doi.org/10.1186/1743-422X-6-223>

- Zhang, L., Meng, L., Fang, Y., Ogata, H., & Okazaki, Y. (2024). Spatiotemporal dynamics of giant viruses within a deep freshwater lake reveal a distinct dark-water community. *The ISME Journal*, 18(1), wrae182. <https://doi.org/10.1093/ismejo/wrae182>
- Zhao, H., Meng, L., Hikida, H., & Ogata, H. (2024). Eukaryotic genomic data uncover an extensive host range of mirusviruses. *Current Biology*, 34(12), 2633-2643.e3. <https://doi.org/10.1016/j.cub.2024.04.085>
- Zhou, Y. L., Mara, P., Vik, D., Edgcomb, V. P., Sullivan, M. B., & Wang, Y. (2022). Ecogenomics reveals viral communities across the Challenger Deep oceanic trench. *Communications Biology*, 5(1), 1055. <https://doi.org/10.1038/s42003-022-04027-y>
- Zhu, F., Massana, R., Not, F., Marie, D., & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, 52(1), 79–92. <https://doi.org/10.1016/j.femsec.2004.10.006>