Novel metabarcoding method reveals pronounced seasonality and high turnover rate of *Imitervirales* in coastal communities

Kyoto University, Graduate School of Science

Division of Biological Science, Department of Biophysics

Florian Prodinger

2022

Publications

(1)

Prodinger, F., Endo, H., Gotoh, Y., Li, Y., Morimoto, D., Omae, K., Tominaga, K., Blanc-Mathieu, R., Takano, Y., Hayashi, T., et al. (2020). An Optimized Metabarcoding Method for *Mimiviridae*. Microorganisms 8, 506.

(2)

Prodinger, F., Endo, H., Takano, Y., Li, Y., Tominaga, K., Isozaki, T., Blanc-Mathieu, R., Gotoh, Y., Tetsuya, H., Taniguchi, E., et al. (2021). Year-round dynamics of amplicon sequence variant communities differ among eukaryotes, *Imitervirales*, and prokaryotes in a coastal ecosystem. *FEMS Microbiology Ecology*, fiab167 (2021), doi:10.1093/femsec/fiab167.

The major findings of chapter two and chapter three of this thesis have previously been published in the articles listed above.

Abstract

Imitervirales, an order of giant viruses of the phylum *Nucleocytoviricota*, are abundant in the ocean. In recent studies *Imitervirales* were further shown to be diverse and able to influence biogeochemical cycles. These findings highlight that *Imitervirales* are an important part of the microbial communities of the ocean and that it is necessary to study their community dynamics to understand the complex interaction of oceanic microorganisms. Yet, studies targeting their community dynamics and interactions with other microbial communities are still rare.

A recent study established a holistic metabarcoding approach named MEGAPRIMER to analyze *Imitervirales* communities in several oceanic areas in Japan. However, this method employed 82 degenerate primer pairs targeting a conserved gene, and thus necessitated 82 separate polymerase chain reaction (PCR) amplifications per sample. Therefore, its applicability was limited to studies with a small number of samples.

To overcome this limitation, I tried mixing MEGAPRIMER to "primer–pair–cocktails" in the first part of my thesis. I also investigated the possible bias of MEGAPRIMER with real time polymerase chain reaction (i.e., qPCR). I found that "primer–pair–cocktails" were an alternative to using primers separately. This new *Imitervirales* metabarcoding workflow reduced the amount of necessary sample DNA and preparation time while reproducing previous results.

In the second part of this thesis, I employed metabarcoding of different microbes (i.e., eukaryotes, prokaryotes, and *Imitervirales*) to study their diversity and community dynamics in a coastal area in Shikoku, Japan, based on a set of 43 samples collected during 20 months. The analysis resulted in three major discoveries. I showed that (1) all three microbial groups displayed similar seasonal cycles in their community dynamics. While the seasonality of eukaryotic and prokaryotic microbes was known, the seasonality of *Imitervirales* community has not been observed before. (2) The taxonomic units represented by amplicon sequence variants (ASVs)

exhibited systematic differences in their persistence and recurrence among microbial communities when compared weeks, months, or one year apart. (3) Abiotic factors (e.g., temperature, nutrients, or salinity) had low explanatory power over the community compositions of eukaryotes and *Imitervirales*.

From these observations I concluded that how quickly communities changed their members varies for different communities. *Imitervirales* showed the highest turnover rate of any microbial community. This may be due to the high selective pressure on viruses to adapt to the also rapidly changing host community. The differences in community dynamics observed here were interpreted by formulating the "community memory" hypothesis. This hypothesis states that current communities influence the formation of future communities, however this influence diminishes over time. How long a previous community can impact the composition of a future community is different for different microbes.

Table of Contents

1 Chapter 1: Introduction

1.1 Marine microbial ecosystem

The importance of marine ecosystems can hardly be overstated. For example, microbial photosynthetic production of O_2 in the ocean started roughly 3 billion years ago and has continuously provided the partial pressure of O_2 in the atmosphere ever since (Lyons et al., 2014). This enabled the emergence of multicellular life as we know it. Nowadays primary producers in the ocean, which are mostly unicellular algae, contribute roughly half (45%) of the total O_2 production (Field et al., 1998). Unicellular algae are also heavily involved in the cycling of other elements found in the ocean like nitrogen, phosphorus, and iron (Falkowski, 2001; Henley et al., 2020) and by removing CO_2 from the atmosphere (Hain et al., 2014).

Despite these important contributions to biogeochemical cycling of several elements, primary producers in marine environments make up only 1% of the photosynthetic biomass of the Earth (Falkowski and Raven, 2007). This is in contrast to terrestrial multicellular primary producers, that contribute roughly 80% of the global biomass (Bar-On et al., 2018). This unequal biomass distribution is rooted in the difference of the primary producer's adaption to their respective habitats (i.e., aquatic or terrestrial) (Bar-On and Milo, 2019; Falkowski, 2001).

It follows that the structure of interaction networks like the food web in marine environments is fundamentally different from that of terrestrial ecosystems (Bar-On and Milo, 2019). Photosynthetic cellular microbes are the basis of the food web in the ocean (Trombetta et al., 2020). The primary producers (photoautotrophs or mixotrophs) are directly or indirectly consumed by heterotrophs (including animals) or mixotrophs (Henley et al., 2020; Tynan, 1998) and thus their community exerts a bottom–up control on consumers (Menge, 2000). Paradoxically, consumers make up four times as much biomass (i.e., organic carbon) as their primary source of nutrition (Bar-On and Milo, 2019). This minority of primary producers can sustain the majority of

consumers due to their different community dynamics, specifically the high turnover rate of the unicellular primary producers (Bar-On and Milo, 2019).

This demonstrates that marine microbial communities, particularly cellular phytoplankton (Falkowski, 2001) and viruses (Brum et al., 2015), are important drivers of marine ecosystems and influence geochemical cycles (Falkowski, 1994). Characterizing the structure and dynamics of marine microbial communities is therefore an important task to better understand the mechanisms that govern marine ecosystems, the functional consequences of their dynamics, and the interactions of microbial communities with their environment.

The evolutionary origin of cellular photosynthesis and therefore primary production were prokaryotes, more precisely bacteria (Blankenship, 2010). They are the most abundant cellular microbes in the ocean and it has been estimated that the ocean contains a total of 10^{29} bacterial cells (Whitman et al., 1998). Prokaryotes are also heavily involved in carbon cycling (Ogawa et al., 2001) and cyanobacteria are ubiquitously present and the most abundant photosynthetic organisms in the ocean (Sukenik et al., 2009).

While prokaryotic phytoplankton outnumbers unicellular eukaryotic phytoplankton, the biomass and primary production of unicellular cellular eukaryotic phytoplankton in the ocean is larger (Bar-On et al., 2018; Li, 1995; Worden et al., 2004). The most abundant photosynthetic micro–eukaryotes in the size range of 0.8µm – 5µm in the ocean are Alveolata, Stramenopiles, and Rhizaria according to a metagenomics based study (Vargas et al., 2015). They are often grouped together as the SAR super–group (Adl et al., 2012). Previous studies used chlorophyll *a* and carotenoids to estimate the biomass of eukaryotic plankton and found that microbial communities of oligotrophic mid–ocean environments are mostly populated by green picoplankton (e.g., chlorophytes) (Falkowski and Knoll, 2007; Hirata et al., 2011) and haptophytes (Hirata et al., 2011; Liu et al., 2009).

The abundances of these different eukaryotic lineages also vary according to examined region of the ocean (Liu et al., 2009; Vargas et al., 2015). Haptophytes are diverse mixotrophic eukaryotic microalgae (Moon-van der Staay et al., 2001) that are present throughout the sunlit oligotrophic ocean (Hirata et al., 2011; Liu et al., 2009). They are especially abundant in tropical Atlantic and Pacific sites where they contribute from 20% to 50% of the standing stock of chlorophyll *a* (Liu et al., 2009). In nutrient rich coastal areas, the primary production of phototrophic eukaryotic phytoplankton is especially important, because eukaryotic single cellular algae together with cyanobacteria form the basis of the food web (Caron et al., 2012). The most prominent primary producers of the shelf–sea are Alveolata (mostly dinoflagellates), diatoms, as well as haptophytes (Falkowski and Knoll, 2007; Liu et al., 2009).

1.2 Viruses in marine microbial ecosystem

The third major microbial group of the ocean, besides prokaryotes and eukaryotes, are viruses. Marine prokaryote infecting viruses are 15 times more abundant than prokaryotes themselves (Proctor and Fuhrman, 1990; Suttle, 2007). Bacterial communities in a variety of environments were found to contain 0.4% to 8% of infected cells (Fuhrman, 1999). Heterotrophic bacteria show higher average infection rates with estimates of up to 20% of cells being infected at any given point in time (Suttle, 1994). Viral lysis contributes from 8% up to 43% of bacterial mortality (Fuhrman, 1999). The bacterial community is thus in part controlled by infection and subsequent lysis (Suttle, 2007).

Phages aid genetic diversification by transferring genes between members of the community (Arnold et al., 2021; Gregory et al., 2016) and through top–down regulation of the community (Koskella and Brockhurst, 2014). Top–down regulation includes the infection of exceedingly active or abundant members of the community, which is a common survival strategy of viruses (Winter et al., 2010). This strategy was named “killing the winner” (Winter et al., 2010).

Phages thereby moderate the community composition of prokaryotes by killing the winner (Koskella and Brockhurst, 2014). Aside from influencing community composition, bacterial viruses were shown to manipulate the metabolism of their hosts (Poranen et al., 2006) and to influence the geochemical cycles of carbon (Guidi et al., 2016), nitrogen, phosphorus, and other nutrients (Brussaard et al., 2008; Yoshida et al., 2018).

A major lineage of eukaryotic viruses of the ocean belongs to the phylum *Nucleocytoviricota* (Hingamp et al., 2013). *Nucleocytoviricota* were previously called NCLDV an abbreviation of the descriptive name "nucleocytoplasmic large DNA viruses" (Iyer et al., 2001). The phylum is monophyletic and includes several families, which all share a set of core genes (Iyer et al., 2001; Koonin et al., 2019). Two lineages of *Nucleocytoviricota* are especially important, namely the family *Phycodnaviridae* (of the order *Algavirales*) and the order *Imitervirales*, since they are abundantly present in the ocean (Hingamp et al., 2013). *Phycodnaviridae* are known as algae infecting viruses, hence the prefix "phyco" which means algae (Maruyama and Ueki, 2016; Nagasaki and Yamaguchi, 1997). My study focuses on *Imitervirales*, which have been discovered more recently and were not considered algal viruses at first (Hingamp et al., 2013; La Scola et al., 2003).

Nucleocytoviricota infecting bloom forming species like *E. huxleyi* were found to be as abundant as 10^7 viruses in a single milliliter of seawater after a bloom (Schroeder et al., 2003). The *Nucleocytoviricota* genome concentrations of the open ocean were estimated between 10^3 to 10^5 genomes per mL for samples from photic zones (5m depth) to deep chlorophyll maximum zones (ranging from 20m to 200m). Another study confirmed the presence of *Nucleocytoviricota* by showing that nearly 30,000 *Nucleocytoviricota* genes were transcribed in the ocean (Carradec et al., 2018). In coastal areas 145 *Nucleocytoviricota* genomes (nearly exclusively consisting of genomes of the order *Imitervirales* and the family *Phycodnaviridae*) were found to be

transcriptionally active, with the authors estimating that many viruses were missed due to their strict criteria for detection (Ha et al., 2021).

Recent studies addressed the influence of viruses on cellular microbes through host reprogramming. By extension viruses may impact geochemical cycling and cause heightened primary production. This hypothesis is supported by the fact that viral community composition can predict carbon export efficiency (Kaneko et al., 2020). Recently discovered *Nucleocytoviricota* in the order *Imitervirales* were also shown to encode a pathway potentially enabling photoheterotrophic capabilities (Needham et al., 2019) which may influence their host's metabolic activity (Schulz et al., 2020). Such metabolic reprogramming was previously discussed for phage–bacteria interactions (Dammeyer et al., 2008), *Ostreococcus tauri* and its viruses (Monier et al., 2017), and for viruses in the epipelagic ocean that transcribe auxiliary genes for sulfur and nitrogen cycling (Roux et al., 2016).

Suspected host reprogramming by *Imitervirales* occurred mostly for pathways of light dependent proton pumps, carbon fixation, photosynthesis, several substrate transport processes (Schulz et al., 2020), and metabolic pathways (Blanc-Mathieu et al., 2021). *Imitervirales* were also shown to exchange genes between each other (Kijima et al., 2021). Horizontal gene transfer from and to eukaryotes is also common (Cunha et al., 2020; Schulz et al., 2020). Thereby *Nucleocytoviricota* were found to be important players in the ocean's ecosystems (Schulz et al., 2020). Yet, the ecological roles of *Nucleocytoviricota* and especially *Imitervirales* have not been studied very extensively (Endo et al., 2020).

1.3 *Imitervirales*

1.3.1 Discovery of *Mimivirus*

The first discovered member of the *Imitervirales* was the *Acanthamoeba* infecting *Mimivirus* (APMV) which was isolated in 1992 and first described in 1997 as a gram–positive

bacterium due to its size. It was not recognized as a virus until 2003 (Raoult et al., 2007; La Scola et al., 2003). The reason for the long time it took to officially recognize APMV as a virus might originate from how little it fit the definition of "virus". The definition of viruses in 2003 focused on the size of the virion and other traits, like having a protein coat and existing outside its host cell (Forterre, 2013; Jacob and Wollman, 1961). Even now common dictionaries define virus based on particle size (merriam-webster.com/dictionary/virus, dictionary.cambridge.org/dictionary/english/virus, accessed 16.3.2021). *Mimivirus* was gram positive, had hair like filaments, was observable with a light microscope, and had a genome larger than that of several bacteria (Birtles et al., 1997; La Scola et al., 2003), thus it was not a typical virus and not immediately recognized as a virus. Any doubt was removed when tunneling electron microscopy images were made available (La Scola et al., 2003).

Further research continued to isolate new *Imitervirales* and it became clear that not only the size of *Imitervirales* was atypical for viruses, but also the content of their genome. These new studies showed how much *Imitervirales* deviated from the traits usually associated with viruses. Viruses were meant to be completely dependent on the translational machinery of their host, yet *Imitervirales* were found to encode part of their own translational machinery (Abergel et al., 2007; Abrahão et al., 2018). Viruses were known to have genomes smaller than cellular microbes, yet several *Imitervirales* were found to have genomes over 1 million basepairs, larger than that of some bacteria (Abrahão et al., 2018; Aherfi et al., 2016; Blanc-Mathieu et al., 2021; Wilhelm et al., 2017). These incongruences prompted a discussion about redefining viruses (Raoult and Forterre, 2008), or at least distinguishing viruses and virocells. The concept of a virocells was established by Forterre (Forterre, 2012, 2013). Forterre suggested that the virus is considered a living entity during its life cycle inside its host cell and called a virocells. Virocells produce virions instead of the usual reproductive cycle of the host cell (e.g., mitosis). Virions are considered an inactive particle

that need a host to become “living” virocells (Forterre, 2012, 2013). Some authors even suggested that *Imitervirales* might be a fourth domain of life aside bacteria, archaea, and eukaryotes (Boyer et al., 2010; Colson et al., 2018; Legendre et al., 2012). This hypothesis about a fourth domain of life seems unlikely to me and is indeed refuted by several authors (Forterre et al., 2014; Moreira and López-García, 2015).

The discovery of *Imitervirales* highlighted how little we still know about viruses. Their genome and phenotype (i.e., giant size) raised many questions and the discovery of new species keep adding to the uniqueness of *Imitervirales* through their virion shape (Abrahão et al., 2018), their genome size (Blanc-Mathieu et al., 2021), and by encoding genes that were thought to be exclusive to cellular microbes. This includes recently discovered *Imitervirales* genomes that contain myosin encoding genes (Kijima et al., 2021), eukaryotic actin homologues called “viractins” (Cunha et al., 2020), tRNA encoding genes (Schulz et al., 2017), metabolic genes (Blanc-Mathieu et al., 2021), fermentation genes (Schvarcz and Steward, 2018), and genes coding for a rhodopsin photosystem (Needham et al., 2019).

1.3.2 Biogeography and ecological roles of *Imitervirales*

It has been known that viruses are abundant in the ocean (Bergh et al., 1989; Torrella and Morita, 1979) and it was shown that viruses outnumber all other oceanic microbes (Suttle, 2007). However, it was not until metagenomic analysis became possible that large DNA viruses were also found to be abundant in the ocean, specifically *Nucleocytoviricota* were found to be more abundant than eukaryotes (Hingamp et al., 2013). An early ocean metagenomics study first showed that *Phycodnaviridae* are the most abundant and *Imitervirales* the second most abundant viruses of the phylum *Nucleocytoviricota* (Hingamp et al., 2013). Recent studies however also emphasized the abundance of *Imitervirales* and even suggested them to be the most abundant in

certain areas and the most diverse oceanic *Nucleocytoviricota* (Carradec et al., 2018; Endo et al., 2020; Kaneko et al., 2020).

It is likely that interactions between *Imitervirales* and single cellular eukaryotes occur frequently, since active *Imitervirales* infections were observed by transcriptomics in the whole ocean (Carradec et al., 2018) as well as coastal areas (Ha et al., 2021). This high transcriptional activity and the recently reported long latent period of certain haptophyte infecting *Imitervirales* (Blanc-Mathieu et al., 2021) may mean that eukaryotic communities contain *Imitervirales*-infected members at any given moment. That virocells are part of a cellular community was recently discovered for a chlorophyte population of *Ostreococcus* spp, in which up to 60% of the population was infected (Castillo et al., 2021).

Imitervirales and other giant viruses were shown to infect a wide range of hosts, exchange genes with them and possibly reprogram them after infection (Schulz et al., 2020). It is therefore not surprising that the composition of the *Imitervirales* community correlates with the composition in specific host groups (Endo et al., 2020). Even though these studies suggest a high impact of *Imitervirales* and other *Nucleocytoviricota* on the eukaryotic community or even the biogeochemical processes, studies specifically targeting *Imitervirales* and their communities are few and far between.

1.4 Marine microbial community dynamics

When microbial ecosystems were first studied the assumption was that the oligotrophic ocean would support only a few highly selected species due to the competitive exclusion principle (Hardin, 1960), however the opposite was the case and diverse oceanic microbes were observed (Hutchinson, 1961). This discovery made it obvious that oceanic environments were more complex than anticipated and that in depth analysis of microbial communities in time and space were necessary to better understand what drives marine ecosystems.

One such driver of marine ecosystems seems to be the change of seasons. Long term sampling studies of oligotrophic oceanic communities showed that marine bacterial communities possess a strong seasonality (Fuhrman et al., 2015) and are also heavily influenced by temperature (Ward et al., 2017). Eukaryotic phytoplankton of the oligotrophic ocean was also shown to change with the seasons (Choi et al., 2020), this change was attributed to temperature (Sunagawa et al., 2015). The viral communities attacking these marine microbes were also observed to show seasonal changes. Phages in particular were shown to possess a strong seasonality (Chow and Fuhrman, 2012; Ignacio-Espinoza et al., 2020; Sandaa et al., 2018). Studies about the seasonality of large eukaryotic viruses in the oligotrophic ocean are few, but two families in the phylum *Imitervirales* were shown to have dissimilar community compositions in colder months (i.e., winter and spring) and warmer months (i.e., summer) in the arctic ocean (Sandaa et al., 2018).

Coastal microbial communities are known to possess a clear seasonality. This has been shown for cellular microbes, such as bacteria (Chafee et al., 2018; Ward et al., 2017) and eukaryotes in cold (Gran-Stadniczeñko et al., 2019a) and warm climates (Giner et al., 2019). *Myoviridae* communities of a Norwegian fjord were also shown to vary seasonally, (Pagarete et al., 2013). Interestingly, a monthly sampling study was carried out during two years to analyze the community dynamics of large DNA viruses, yet a clear seasonality was not observed (Gran-Stadniczeñko et al., 2019b). The authors however pointed out that they anticipated the large dsDNA virus community to show seasonality, but could not prove it due to methodological limitations, namely that their analysis targeted the community not holistically (Gran-Stadniczeñko et al., 2019b).

Aside from another study that was able to show that the 1 of 5 large dsDNA viral communities was significantly different to the other 4 communities (Johannessen et al., 2017), I

was not able to find any further community composition or dynamics studies that included *Imitervirales*.

1.5 Current methods for characterization of community structures

Most microbes of the ocean are currently not cultured and may not be possible to keep in culture at all (Steen et al., 2019). Therefore, it is necessary to study microbial communities in the ocean directly from samples of the aquatic environment. Aquatic microbes were first observed through a microscope by Leeuwenhoek in 1677 (Lane, 2015). Techniques that involve DNA sequencing of conserved genes developed during the 1980s (Woese et al., 1984).

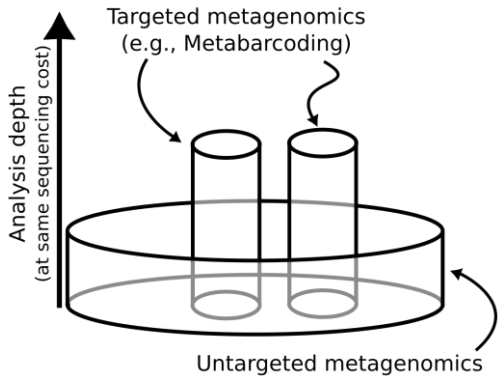

Figure 1-1: Different methods of taking a "snapshot" of microbial communities. The horizontal axis represents the obtained genetic information, while the vertical axis is the observed depth of the community composition. Metagenomics of primer amplified highly variant regions of conserved genes (i.e., DNA barcoding or metabarcoding) can reveal most of the community while the sequencing of digested sample DNA shows fewer members of the community, but reveals more of their genome. Figure recreated with updated terminology from Suenaga, 2012.

The recent advent of high throughput sequencing enabled rapid and cheap sequencing of many samples. It became common practice in marine microbial ecology to study the composition of microbial communities by sequencing a marker gene that is universal to cellular life (Olsen et al., 1986; Woese et al., 1984) (targeted metagenomics) or by sequencing (Anderson, 1981) all of the available cellular DNA of a seawater sample (Venter et al., 2004) (untargeted metagenomics). These methods have different merits and supply different information about a community and its members (Figure 1-1).

1.5.1 Untargeted metagenomics

Untargeted metagenomics rely on a “shotgun” approach to sequence all available DNA of a sample in short fragments and subsequently assembling the DNA to contigs or even genomes. In order to distinguish this analysis method to other sequencing methods for ecological analysis, this

specific method will be referred to as untargeted metagenomics throughout this thesis, most other modern studies simply refer to the same method as metagenomics.

Untargeted metagenomics comprehensively sample all genes of all organisms found in a single sample and thereby enable a variety of analysis, like the study of diversity and abundance of a microorganisms in the analyzed sample. This data can also be used in other creative ways. For instance, the *Tara* Oceans metagenes catalogue was curated using untargeted metagenomics on oceanic seawater samples and their data was used by Li et al., to design a metabarcoding method to study oceanic viruses (Li et al., 2018). The *Tara* Oceans metagenes catalogue also enabled matching virus–host pairs through horizontal genes transfer (Schulz et al., 2020), since whole genomes were recovered.

However, untargeted metagenomics studies that include lowly abundant members of the community of coastal or oceanic *Imitervirales* are expensive to perform. In untargeted metagenomics DNA sequences of all the available microbial DNA in a sample are generated proportionally to their abundance. Rare species can therefore only be observed by raising the sequencing depth, which raises sequencing cost proportionally. Targeted sequencing approaches specifically target certain communities (e.g., bacteria or large viruses) and can detect rare species, which were shown to disproportionally contribute to community dynamics (Giner et al., 2019; Ignacio-Espinoza et al., 2020; Lynch and Neufeld, 2015). *Imitervirales* are more diverse than bacteria (Mihara et al., 2018) but similar in size (Wilhelm et al., 2017) and far less abundant in the ocean (Hingamp et al., 2013).

Untargeted metagenomics studies have gained popularity due to their many advantages. However, concerning *Imitervirales* community analysis, it seems reasonable to first perform a highly sensitive, quick, and relatively cheap targeted metagenomics based analysis, and designing further studies using untargeted metagenomics on the outcome of the targeted analysis.

1.5.2 Targeted metagenomics

Targeted metagenomics refers to the practice of analyzing only a small part (the targeted region) of a genome. This is usually a variable region of a conserved gene and it is analyzed by creating multiple copies of the region with specific primers and a polymerase chain reaction (PCR) enzyme. The sequence of nucleotides of these short DNA copies is then determined and assigned to a more complete sequence by comparing it to entries of a sequence database. The practice of matching a short, highly variable region of a gene to a longer sequence to analyze the community composition in a sample is referred to as "DNA barcoding". The first genetic diversity analysis methods for cellular microbes based on DNA barcoding were established nearly 40 years ago (Pace et al., 1986; Woese et al., 1984). Targeted metagenomics methods have since been further refined to "metabarcoding". This approach can perform several DNA barcoding analyses in parallel with strictly evaluated universal primer pairs for both eukaryotic communities (Bradley et al., 2016) as well as prokaryotic communities (Takahashi et al., 2014).

Studying *Imitervirales* by metabarcoding is more difficult than studying cellular communities, because the applicable regions of the few conserved genes of *Imitervirales* tend to show higher variability (Gran-Stadniczeñko et al., 2019b; Johannessen et al., 2017). The first primer pairs targeting the DNA polymerase B gene (*polB*) of oceanic large dsDNA viruses infecting phytoplankton were designed in the 1990s (Chen and Suttle, 1995). A DNA barcoding approach was necessary, because large dsDNA viruses were not taxonomically distinguishable by electron microscopy, despite large genetic differences (Chen and Suttle, 1995). Five distinguished operational taxonomy units (OTUs) were identified with manual Sanger sequencing (Chen et al., 1996). "Taxonomy units" usually refer to genetically similar groups of organisms. In many studies they are "operational" since the definition of such a taxonomy unit can vary between different

studies. These operational taxonomy units (OTUs) contrast the traditional taxonomic units (e.g., "species") by relying on DNA sequences and gene content instead of grouping by phenotype.

After pyrosequencing and other modern sequencing approaches became viable options, further studies were conducted. They targeted different conserved regions of *Imitervirales* (sometimes including *Phycodnaviridae*), like the major capsid protein (Gran-Stadniczeñko et al., 2019b; Johannessen et al., 2017; Larsen et al., 2008; Moniruzzaman et al., 2016), the mismatch repair protein (Wilson et al., 2014), or the *polB* gene (Clerissi et al., 2014; Li et al., 2018). Due to their technological advantage these modern studies were able to show a dozen or more OTUs (Larsen et al., 2008; Wilson et al., 2014) up to around 150 OTUs (Johannessen et al., 2017; Moniruzzaman et al., 2016). A more recent publication was able to discover over 300 OTUs despite claiming that they were not able to detect the community holistically (Gran-Stadniczeñko et al., 2019b).

A few of these pioneering studies reported that their primers may be specific only to a fraction of the targeted group of viruses and not show the community holistically (Gran-Stadniczeñko et al., 2019b; Larsen et al., 2008). This problem was most likely caused by a lack of sequences to base primer design on as well as the scarcity of viral conserved genes (Legendre et al., 2012).Therefore, studies investigating the communities of *Imitervirales* are still rare compared to studies about cellular communities (Figure 1-2), for which different primer strategies were already proposed and validated (Bradley et al., 2016; Case et al., 2007; Pace et al., 1986).

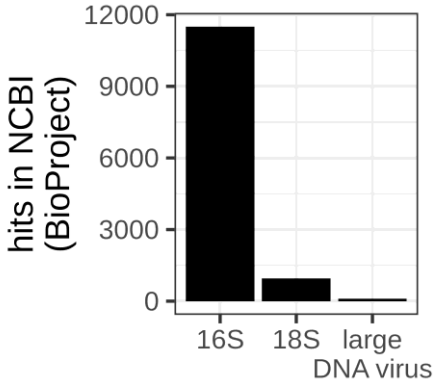

Figure 1-2: Database entries in March 2021 for prokaryotes, eukaryotes, and large viruses. The used search terms for prokaryotes and eukaryotes were "16S" and "18S", respectively. Therefore only metabarcoding database entries were found. The search terms for large DNA viruses were "Mimiviridae", "Megaviridae", "Imitervirales", and "NCLDV". This includes metabarcoding as well as metagenomics based entries. The search was conducted in NCBI's BioProject archive.

A recent metabarcoding study tackled this issue by embracing the aforementioned high variability of *Imitervirales* conserved genes (Li et al., 2018). Previous primers were designed by relying on specific isolates of the targeted species' genome and directly based the primer design on a chosen gene. Li et al. instead used the 17 known *Imitervirales polB* genes to perform a hidden–Markov–model search in *Tara* Oceans metagenes to detect a large set of roughly 1,000 environmental *Imitervirales polB* sequences (Li et al., 2018). Previous studies repeatedly found that a single degenerate primer pair was not able to cover the diversity of *Imitervirales* holistically (Gran-Stadniczeñko et al., 2019b; Larsen et al., 2008). Therefore, Li et al. used these 1,000 environmental *Imitervirales polB* sequences to design a set of 82 degenerate primer pairs that was theoretically able to amplify all of the *polB* genes found in the ocean (Li et al., 2018). This set of 82 primer pairs was named "MEGAPRIMER" and Li et al used it to reveal that up to 6,500 *Imitervirales* OTUs can be found in a single liter of sea water (Li et al., 2018, 2019).

Using such a holistic method on a time series of samples is a common approach to investigate what drives the oceanic communities. The time series of community change can be associated to physical factors (Gilbert et al., 2009; Sunagawa et al., 2015), biological factors (Lima-Mendez et al., 2015; Needham and Fuhrman, 2016), or sub communities within (Lynch and Neufeld, 2015; Shade et al., 2014). However, at the time of publication, Li et al.'s "MEGAPRIMER" method was not applicable to study community compositions of sets of samples, because the laborious protocol restricted the analysis to few samples (Li et al., 2018, 2019).

1.6 Objectives

The first objective is to improve the workflow of the MEGAPRMER method by Li et al. to allow the application of MEGAPRIMER to time series studies with many samples. Mixing primer pairs to cocktails as well as further validating MEGAPRIMER with multiple samples was already suggested by Li et al. themselves, however these experiments were not feasible without further

improving the sample preparation workflow. Hence, they considered technical improvements of the MEGAPRIMER analysis method a necessary extension of their work (Li et al., 2018). These suggestions are realized in the second chapter of this work by mixing the degenerate 82 primer pairs into “cocktails” and by evaluating the amplification conditions as well as the amplicon cleanup.

The second objective is to better understand the community dynamics of *Imitervirales* and connect them with eukaryotic and prokaryotic communities as well as environmental factors. This is achieved in chapter three by examining the seasonality and turnover rate of the *Imitervirales* community by using a set of 43 samples collected from a coastal environment during 20 months. The analyses of the community dynamics are based on traditional dissimilarity metrics, as well as more sophisticated statistical analysis. Recurrence, persistence and change in dissimilarity of the community over time are used as proxy for the actual turnover rate.

The fourth chapter concisely discusses the results of the second and third chapter. I will discuss major findings of this thesis and lay out how this thesis contributed to the field. Finally, I will propose further strategies to investigate the *Imitervirales* community.

2 Chapter 2: *Imitervirales* diversity analysis method optimization

2.1 Abstract

Imitervirales is an order (phylum: *Nucleocytoviricota*) of giant DNA viruses that are abundant in the ocean. Recent studies highlighted the ecological relevance of *Imitervirales* in marine environments through frequent discoveries of new *Imitervirales* and the influence of *Imitervirales* on carbon export. To further study *Imitervirales* communities several metabarcoding methods for ecological profiling have been developed. A metabarcoding approach published in 2018 implemented 82 degenerate primer pairs targeting the conserved, but highly variant *polB* gene of *Imitervirales* (i.e., MEGAPRIMER) and showed a hitherto undetected diversity of *Imitervirales* in a coastal seawater sample. The authors also discussed drawbacks of this method. Namely, that it required comparatively high amounts of sample DNA. Furthermore, the method was not viable for studies with large sample sets (e.g., time series sampling) due to a laborious amplicon preparation protocol. In this study, I tested various PCR conditions as well as amplicon purification protocols to improve the original method. Primer pair "cocktails" were introduced to reduce the amount of sample DNA and preparation time. Finally, quantitative real time PCR (qPCR) analysis of several OTUs (operational taxonomy units) was used to validate the new method. Even though the qPCR based quantification revealed possible amplification bias, the metabarcoding frequency profiles across samples were verified. Through this effort the MEGAPRIMER analysis protocol was streamlined by reducing analysis time while reproducing previous results. This new method facilitates a high though put analysis workflow for larger sample sets.

Contributions

Sea water sampling was performed by Yoshihito Takano and Keizo Nagasaki. DNA sequencing support was provided by Yasuhiro Gotoh, Tetsuya Hayashi, Daichi Morimoto, Kimiho Omae, Kento Tominaga, and Takashi Yoshida. I performed both wet experiments and dry analyses, and wrote the initial manuscript.

2.2 Introduction

Imitervirales are an order of large dsDNA viruses in the phylum *Nucleocytoviricota*. The largest *Imitervirales* have virion diameters up to 750 nm (AMPV , Tupanvirus) and genome sizes ranging from several hundred to well over a thousand kilo base pair (Abrahão et al., 2018; Fischer et al., 2010; La Scola et al., 2003). Recently isolated *Imitervirales* infect mostly unicellular algae in the ocean (Blanc-Mathieu et al., 2021; Needham et al., 2019; Schvarcz and Steward, 2018). Oceanic *Imitervirales* feature smaller particles 140 nm to 310 nm (Gallot-Lavallée et al., 2017; Johannessen et al., 2015) and most reported genomes are between 370 kb and 560 kb (Johannessen et al., 2015; Schvarcz and Steward, 2018). The largest photosynthetic eukaryote infecting *Imitervirales* genome to date is that of PkV RF01 with a size of 1.4 Mbp (Blanc-Mathieu et al., 2021). PkV RF01's genome is more than twice as large as the second largest genome of TetV with a size of 668 kbp (Schvarcz and Steward, 2018).

Imitervirales were associated with harmful algae blooms in coastal areas and shown to infect eukaryotes commonly found in the open ocean. AaV (*Aureococcus anophagefferens virus*) was detected alongside a bloom of its host (pelagophytes) during a brown tide (Moniruzzaman et al., 2014, 2016). *Prymnesium parvum* and *Haptolina ericina* (both bloom forming haptophytes) were found to be *Imitervirales* hosts as well (Gallot-Lavallée et al., 2015; Hansen et al., 1995; Sandaa et al., 2001; Wagstaff et al., 2017). Hosts of other taxa are *Tetraselmis* (a Chlorophyte), the heterotrophic protists *Cafeteria roenbergensis*, and a member of the Choanoflagellates. They are infected by Tetraselmis virus (TetV)(Schvarcz and Steward, 2018), Cafeteria-roenbergensis-Virus (CroV)(Fischer et al., 2010), and ChoanoVirus (Needham et al., 2019), respectively.

Recently, the *Imitervirales* community has become a focus of viral research since they infect a broad range of eukaryotic taxa (Schulz et al., 2020) and influence the carbon cycle (Kaneko et al., 2020). Furthermore, the *Imitervirales* community is as diverse as prokaryotes, (Mihara et al.,

2018) as abundant as eukaryotes (Hingamp et al., 2013), and transcriptionally active in the ocean (Carradec et al., 2018).

Ecological analysis of the *Imitervirales* community has been difficult due to a high variability in the few possible regions for diversity analysis (Gran-Stadniczeñko et al., 2019b; Li et al., 2018; Wilson et al., 2014). Li et al. proposed a new approach to metabarcoding–primer–design based on the *polB* genes of large DNA viruses found in the *Tara* Oceans metagenes (Li et al., 2018; Sunagawa et al., 2015). In their study, Li et al. combined a set of 82 degenerate primer pairs to holistically target the high nucleotide sequence variations of the conserved *polB* gene (Li et al., 2018). Thereby, they were able to identify a hitherto undetected richness of 5,595 non singleton *Imitervirales* OTUs (operational taxonomy units at 97% similarity) in a single liter of seawater (Li et al., 2018). In a second study, they showed the wide range of environments in which *Imitervirales* are present by analyzing fresh and sea water (i.e., samples from a hot spring, the mangroves, the Sea of Japan, and the Osaka Bay) (Li et al., 2019). Their approach was successful for studies with few samples, but the authors recommended to further evaluate and possibly improve their method (Li et al., 2018). The drawback of using 82 primers is the linear increase in time and sample DNA consumption, as well as a high risk of errors (e.g., pipetting mistakes or sample swaps) due to many PCR reactions per sample. Additionally, biases may be introduced during amplicon generation, since several samples can hardly be processed simultaneously. For instance, amplification of one sample with MEGAPRIMER takes up 82 PCR tubes of the 96 PCR tube well plate used in most thermal cyclers.

Here I aim to improve the original MEGAPRIMER protocol by introducing primer pair "cocktails". Instead of using each primer pair separately, several primer pairs were mixed together before performing the *polB* gene amplification. This strategy was tested on four seawater samples of different coastal environments and different seasons. As Li et al suggested in their original study

(Li et al., 2018), I further evaluated possible amplification biases by quantifying several OTUs using qPCR. Finally, I analyzed the eukaryotic host community of *Imitervirales* with an 18S rRNA gene metabarcoding analysis.

2.3 Methods

2.3.1 Seawater sampling and DNA extraction

Four seawater samples were analyzed, including the seawater sample of the original MEGAPRIMER study (Li et al., 2018) and 3 samples from the Uranouchi Inlet, Kochi Prefecture, Japan. The seawater sample (4L) of Li et al. were sampled by my collaborators from Kyoto university at a 5m depth at the entrance of the Osaka Bay, Japan (34°19’28”N, 135°7’15”E) on 30 October 2015 (Li et al., 2018). The Uranouchi Inlet samples were taken by my collaborators from Kochi University. They sampled 10L seawater of the Uranouchi Inlet at a 5m depth at 3 different locations and dates: Uranouchi station F (33°26’33.6”N 133°24’41.8”E) on 21 June 2017, Uranouchi station J (33°25’43.2”N 133°22’49.5”E) on 06 July 2017, Uranouchi station M (33°25’60.0”N 133°24’38.3”E) on 10 November 2017.

The Osaka Bay sample was filtered through a 3.0µm–pore polycarbonate membrane filter (diameter 142 mm, polycarbonate; Merck, Darmstadt, Germany) and 1L of the filtrate was further filtered with a 0.22µm–pore filtration unit (Durapore Membrane Filters, PVDF, Merck). The Uranouchi Inlet samples were filtered thrice. First with a 3.0µm–pore filter, the collected filtrated was further filtered with a 0.8µm–pore polycarbonate membrane filter (diameter 142 mm, polycarbonate; Merck). Finally 1L of the 0.8µm filtrate was filtered with a 0.22µm filtration unit (Sterivex, polycarbonate, Merck). The filters were stored at -80 °C until DNA extraction. I extracted DNA following the “Proteinase–K method” for the 0.22µm filtration units (Frias-Lopez et al., 2008) and the xanthogenate–sodium dodecyl sulfate method for the 0.8µm and 3µm filters (Yoshida et al., 2003). I stored the extracted DNA in solution at -20 °C.

2.3.2 *PolB* gene amplification, purification, and sequencing

MEGAPRIMER consists of 82 degenerate primers. The primer pairs were mixed to primer pair cocktails (mixtures of 5, 10, or 20 primer pairs) to reduce the number of PCR amplifications per sample. Which primers were mixed into cocktails is shown in tables 3-1 to 3-4. Primer pairs were also used separately. If primer pairs are used separately, 82 PCR reactions are necessary. If 5 primer pairs are mixed, 17 PCR amplifications are necessary (i.e., 1 amplification by PCR for each cocktail in Table 2-1) to generate amplicons from 1 sample. Mixtures of 10 primer pairs or 20 primer pairs, resulted in 9 (MP10.v1, Table 2-2), 8 (MP10.v2, Table 2-3) or 5 (MP20, Table 2-4) PCR amplifications per sample, respectively. Primers were mixed according to prevalence (i.e., frequency of presence in samples based on *in silico* PCR in *Tara* Oceans metagenes), except for MP10.v2 (Table 2-3), in which primer pairs were mixed to generate a similar average annealing temperature of around 47 °C in each cocktail.

Table 2-1: Five primer pair cocktails (i.e., MP5). Mixing table for cocktails with 5 primer pairs each. MP5 consists of 17 primer cocktails, each cocktail contains 3 to 5 primer pairs.

MP5	cocktail 1	cocktail 2	cocktail 3	cocktail 4	cocktail 5	cocktail 6	cocktail 7	cocktail 8	cocktail 9
primer1	PP28	PP57	PP16	PP41	PP53	PP1	PP66	PP20	PP27
primer2	PP25	PP83	PP15	PP2	PP39	PP21	PP10	PP7	PP54
primer3	PP70	PP18	PP26	PP55	PP42	PP44	PP12	PP40	PP49
primer4	PP47	PP34	PP14	PP45	PP38	PP32	PP22	PP64	PP36
primer5	PP61	PP11	PP24	PP43	PP37	PP9	PP29	PP46	PP19

MP5 (cont.)	cocktail 10	cocktail 11	cocktail 12	cocktail 13	cocktail 14	cocktail 15	cocktail 16	cocktail 17
primer1	PP17	PP4	PP56	PP67	PP77	PP13	PP75	PP82
primer2	PP50	PP58	PP23	PP3	PP59	PP76	PP6	PP80
primer3	PP69	PP8	PP35	PP71	PP78	PP81	PP33	PP30
primer4	PP60	PP74	PP31	PP48	PP63	PP79	PP51	
primer5		PP65	PP52	PP73	PP72	PP68	PP5	

Table 2-2: Ten primer pair cocktails (i.e., MP10.v1). MP10 version 1 consists of 9 primer cocktails, each cocktail contains either 2 or 10 primer pairs.

MP10.v1	cocktail 1	cocktail 2	cocktail 3	cocktail 4	cocktail 5	cocktail 6	cocktail 7	cocktail 8	cocktail 9
primer1	PP28	PP16	PP53	PP66	PP27	PP58	PP3	PP76	PP80
primer2	PP25	PP15	PP39	PP10	PP54	PP8	PP71	PP81	PP30
primer3	PP70	PP26	PP42	PP12	PP49	PP74	PP48	PP79	
primer4	PP47	PP14	PP38	PP22	PP36	PP65	PP73	PP68	
primer5	PP61	PP24	PP37	PP29	PP19	PP56	PP77	PP75	
primer6	PP57	PP41	PP1	PP20	PP17	PP23	PP59	PP6	
primer7	PP83	PP2	PP21	PP7	PP50	PP35	PP78	PP33	
primer8	PP18	PP55	PP44	PP40	PP69	PP31	PP63	PP51	
primer9	PP34	PP45	PP32	PP64	PP60	PP52	PP72	PP5	
primer10	PP11	PP43	PP9	PP46	PP4	PP67	PP13	PP82	

Table 2-3: Second 10 primer pair cocktail mixing table (i.e., MP10.v2). MP10 version 2 consists of 8 primer cocktails, each cocktail contains either 10 or 11 primer pairs.

MP10.v2	cocktail 1	cocktail 2	cocktail 3	cocktail 4	cocktail 5	cocktail 6	cocktail 7	cocktail 8
primer1	PP45	PP28	PP83	PP25	PP11	PP55	PP38	PP41
primer2	PP43	PP37	PP18	PP70	PP16	PP53	PP01	PP02
primer3	PP42	PP21	PP34	PP47	PP15	PP66	PP09	PP39
primer4	PP40	PP12	PP44	PP61	PP26	PP10	PP22	PP29
primer5	PP27	PP64	PP32	PP57	PP14	PP20	PP07	PP46
primer6	PP49	PP54	PP69	PP65	PP24	PP08	PP58	PP17
primer7	PP36	PP23	PP31	PP56	PP35	PP52	PP74	PP50
primer8	PP19	PP67	PP71	PP48	PP73	PP03	PP77	PP60
primer9	PP04	PP59	PP63	PP13	PP76	PP68	PP78	PP81
primer10	PP75	PP06	PP33	PP79	PP51	PP80	PP72	PP82
primer11	PP05	PP30						

Table 2-4: Twenty primer pair cocktails (i.e., MP20). MP20 consists of 5 primer cocktails, each cocktail contains either 2 or 20 primer pairs.

MP20	cocktail 1	cocktail 2	cocktail 3	cocktail 4	cocktail 5
primer1	PP28	PP53	PP27	PP3	PP80
primer2	PP25	PP39	PP54	PP71	PP30
primer3	PP70	PP42	PP49	PP48	
primer4	PP47	PP38	PP36	PP73	
primer5	PP61	PP37	PP19	PP77	
primer6	PP57	PP1	PP17	PP59	
primer7	PP83	PP21	PP50	PP78	
primer8	PP18	PP44	PP69	PP63	
primer9	PP34	PP32	PP60	PP72	
primer10	PP11	PP9	PP4	PP13	
primer11	PP16	PP66	PP58	PP76	
primer12	PP15	PP10	PP8	PP81	
primer13	PP26	PP12	PP74	PP79	
primer14	PP14	PP22	PP65	PP68	
primer15	PP24	PP29	PP56	PP75	
primer16	PP41	PP20	PP23	PP6	
primer17	PP2	PP7	PP35	PP33	
primer18	PP55	PP40	PP31	PP51	
primer19	PP45	PP64	PP52	PP5	
primer20	PP43	PP46	PP67	PP82	

I performed 8 sequencing runs, 7 of which were deep sequencing (i.e., only 1 sample for each sequencing run) and 1 shallow depth sequencing run (Table 2-5). In the shallow depth sequencing run MEGAPRIMER amplicons and cellular marker gene amplicons were sequenced together. The amount of barcoded amplicons in the final library was close to the highest possible number of simultaneously analyzable samples.

Table 2-5: Overview of the generated datasets, sequencing runs, and amplification protocols. Amplicon generation and purification was performed with several protocols and on several samples. The difference among protocols are detailed in table 3-6.

Dataset	Sequencing run number	Sampling location	Sampling date	Primer cocktail	Protocol number
D-OB-MP5-1	1	OB	2015.10.30	MP5	1
D-OB-MP10-1	2	OB	2015.10.30	MP10.v1	1
D-OB-MP20-1	3	OB	2015.10.30	MP20	1
D-OB-MP1-2	4	OB	2015.10.30	MP1 (no mix)	2
D-OB-MP5-2	5	OB	2015.10.30	MP5	1
D-OB-MP10-2	6	OB	2015.10.30	MP10.v1	1
D-OB-MP20-2	7	OB	2015.10.30	MP20	1
S-OB-MP10-1	8	OB	2015.10.30	MP10.v2	3
S-OB-MP10-2	8	OB	2015.10.30	MP10.v1	1
S-OB-MP10-3	8	OB	2015.10.30	MP10.v2	4
S-OB-MP10-4	8	OB	2015.10.30	MP10.v1	5
S-UF-MP10	8	UF	2017.6.21	MP10.v2	3
S-UJ-MP1	8	UJ	2017.7.6	MP1 (no mix)	3
S-UJ-MP10	8	UJ	2017.7.6	MP10.v2	3
S-UM-MP10	8	UM	2017.11.10	MP10.v2	3

I tested several different protocols for amplicon generation on different samples and generated 15 different datasets, listed in Table 2-5. The dataset names, shown in the "Dataset" column of Table 2-5, contain information about the metabarcoding workflow protocol used to generate the data. The names contain the sequencing depth (i.e., Deep or Shallow), the sampling location (e.g., OB for Osaka Bay), the primer cocktail method (e.g., MP1, MP5), and the replicate number (i.e., 1 to 4) separated by a hyphen. All replicates are number starting from 1, only D-OB-MP1-2 does not have a replicate number 1, instead Li et al.'s Osaka Bay dataset of the 2018 study

is considered a technical replicate. Li et al.'s 2018 dataset will henceforth be referred to as D-OB-MP1-0. The 8 shallow depth sequencing datasets were generated with mostly MP10 cocktail methods, except 1 experiment with no primer mixing (i.e., S-UJ-MP1). I tested different PCR conditions with the 7 shallow depth MP10 experiments. I varied the template and primer concentrations, MP10 mixtures, and cleanup protocols (Table 2-5 and Table 2-6). The amplicon cleanup step was performed with either Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea, CA) according to the protocol recommended by Illumina or ethanol precipitation with subsequent gel extraction (2% agarose gel in TAE buffer, extraction kit by Wizard SV, Promega, Madison, WI), because MEGAPRIMER amplification produced many unspecific amplicons with a length of >100 bp alongside the desired amplicons (Figure 2-1). The gel extraction cleanup was unique to protocol number 3 (Table 2-6), all other protocols included magnetic beads cleanup. Cleanup success was sporadically tested with a lab–on–a–chip electrophoresis system (2100 Bioanalyzer, Agilent Technologies, Santa Clara, CA) or agarose gel electrophoresis and subsequent gel staining (GelRed or GelGreen, Biotium, Hayward, CA).

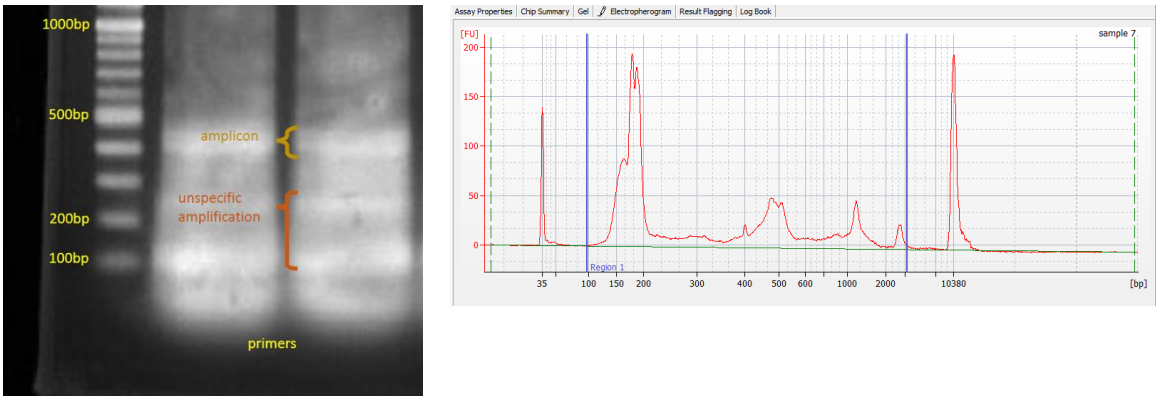

Figure 2-1: Amplification of the UJ sample with different MEGAPRIMER methods. The high concentration of unspecific amplification products of over >100 bp made sequencing impossible. (Left) The gel image shows a gel band during sample preparation with MP10v2 protocol 3 before a cleanup was performed. MEGAPRIMER produced unspecific amplifications. (Right) A chip based electrophoresis device shows the amplification product of the indexing PCR with MP10v1 and protocol 1. Both polB *amplicons and unspecific, short amplicons remained after the cleanup with magnetic beads (internal standards of the electrophoresis device: 35 bp and 10.4 kbp).*

Table 2-6: Overview of different PCR and sample cleanup protocols. All values in final concentrations. I varied the primer concentration, sample DNA concentration and tested different methods to remove unspecific amplifications.

protocol number	primer concentration [µmol.L^{-1}]	sample DNA concentration [ng.µL^{-1}]	amplicon cleanup
1	8	0.025	AMPure XP
2	8	0.04	AMPure XP
3	6	0.2	gel extraction
4	6	0.025	AMPure XP
5	6	0.2	AMPure XP

Aside from the variations shown in Table 2-6, the PCR conditions (e.g., PCR thermal cycling) were kept identical for all conducted experiments. I used 50% KAPA Hifi Hotstart ReadyMix (Roche, Basel, Switzerland) for all MEGAPRIMER based amplifications. The thermal cycle started with 3 min of 95 °C. The 3 step cycle was repeated 32 times with 30s per step. The steps were 94 °C for melting, 54 °C for annealing, and 72 °C for elongation. The last elongation step was 72 °C for 4 min.

After amplicon purification with either of the different cleanup protocols, the indexing PCR step was performed to attach the dual indices according to Illumina's library preparation protocol. Another magnetic bead purification step was performed and the indexed amplicons were mixed in equal volumes for the final library preparation. The final library concentration was 10 pmol L^{-1}. An internal standard was added, 25% PhiX for run 1-7, 50% PhiX for run 8 (Table 2-5).

2.3.3 Raw read filtering with Megaviridae Amplicon Processing System (MAPS)

Alongside the MEGAPRIMER method, Li et al. published an *Imitervirales polB* read processing pipeline. The pipeline works in consecutively performed steps. It removes low quality reads, trims primers, merges the forward and reverse reads, removes chimeric sequences, groups all identical reads, translates the reads for a subsequent steps, filters none *Imitervirales* reads

(with a similarity search of the translated *polB* gene and phylogenetic placement in a reference tree), and finally MAPS trims the amplicons at a common coding region of the *polB* gene.

After MAPS removed low quality and non *Imitervirales* reads, I clustered the sequences at 97% similarity for OTU generation with cd-hit (v 4.6.8) (Li and Godzik, 2006). The resulting OTU table was read with R (v 3.4.2) (www.r-project.org, accessed 23.3.2021) to generate figures and perform statistical analysis. The vegan package (2.5-6) (Oksanen et al., 2012) was used for read rarefication and Jaccard Dissimilarity calculation. The non–metric multidimensional scaling (NMDS) and hierarchical clustering was performed with the cmdscale function and vegan's hclust function, respectively. Plots were generated with ggplot2 (Wickham, 2016) or iNEXT (Hsieh et al., 2016). I created a phylogenetic tree using default Mafft (Katoh and Standley, 2013) and default FastTree (Price et al., 2010). Anvi'o was used to visualize the phylogenetic tree (Eren et al., 2015).

2.3.4 *Imitervirales* OTU quantification with qPCR

I selected the 10 most abundant OTUs from the 15 datasets. These OTUs overlapped since the 15 datasets were generated from only 4 seawater samples. In total 58 OTUs were chosen as the "most abundant OTUs" based on high relative read percentages. Each of these OTUs consisted of several different genotypes (i.e., slightly different variants of the DNA sequence of the *polB* amplicon). Only 43 OTUs showed a clear dominant genotype (i.e., a *polB* DNA sequence that generated at least half of all reads for that OTU) to target with qPCR. Primers for these 43 OTUs were designed with Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012). The optimal primer size was set to 20 bp and the product size range was set from 50 bp to 250 bp. Primer3 was able to generate 4 different sets of qPCR primers for 23 of the 43 OTUs. I tested these primers *in silico* by performing a BLAST search against RefSeq, which returned no hits (E-value < 10^{-4}). The primers were further tested by searching their sequence against all generated *Imitervirales* reads of this study. All primers that were able to target OTUs other than their intended target were

removed. After *in silico* specificity testing only 8 qPCR primer pairs, targeting 8 different OTUs, remained. I ordered these primer and target sequences from a vendor. The specificity of the 8 qPCR primer pairs was tested by amplifying the target region, 2 of the 8 primer pairs showed unspecific binding (multiple dissociation curves) and were discarded. The quantification of 6 *Imitervirales* OTUs was conducted using qPCR. I mixed 6.25 µL of TB Green Premix Ex Taq™ II (Takara Bio Inc., Shiga, Japan), 1 µL of 20µm reverse and forward primer (final concentration 1.6 µmol L^{-1}), and 1 µL of 1 ng μL^{-1} sample DNA (final concentration 0.08 ng μL^{-1}). The quantification was performed with 50 cycles with 95 °C, 55 °C, and 72 °C with 20s each. Fluorescence was recorded during the last step. The limit of quantification (LoQ) was determined according to Forootan et al.'s method (Forootan et al., 2017). The coefficient of variation (CV = 100.SD.$mean^{-1}$) is used to determine parameters for the results of the qPCR experiments (Forootan et al., 2017). The standard serial dilution (10–10^7 molecules) was used to calculate the CV. I conducted at least 3 measurements for every dilution. The limit of quantification (LoQ) was defined as either the concentration of standards where the CV is at least 50% of the measured molecules or 10 or less copies of DNA. The limit of detection (LoD) was defined as less than 1 copy of DNA on average of at least 3 measurements.

2.3.5 Metabarcoding analysis of eukaryotes and subsequent raw read processing

The DNA extract of the 3µm pore size filter was used for a eukaryotic community analysis with a commonly used marker gene amplification. The targeted region was the V8/V9 region of the 18S ribosomal RNA gene using a forward and reverse primer pair called "V8 F" and "1510", respectively (forward: MiSeq adapter + ATAACAGGTCTGTGATGCCCT and reverse: MiSeq adapter + CCTTCYGCAGGTTCACCTAC) (Bradley et al., 2016). The PCR was performed on a mixture of 5 µL (1 µmol L^{-1}) of each primer, 2.5 µL DNA (0.25 ng μL^{-1}) sample, and 12.5 µL 2× KAPA HiFi HotStart ReadyMix. The thermal cycle started at 98°C (3 min) followed by 25 cycles of 98°C (20s), 65°C

(15s), 72°C (15s), and ending on 72°C for 10 min. Purification was performed with magnetic beads and amplification success was confirmed with agarose gel electrophoresis and subsequent staining. The dual indices necessary for metabarcoding were attached with another PCR reaction according to the manufacturer's protocol and a library with a concentration of 2 nmol L^{-1} was prepared for sequencing (2 × 300 nucleotides) according to the manufacturer's protocol. The final concentration was 10 pM and a 25% phiX spike–in was added.

The subsequent bioinformatics analysis was performed with QIIME 2 (version 2018.11.10) (Bokulich et al., 2018; Caporaso et al., 2010) and visualization was performed with R's ggplot2 and python scripts. The raw files were parsed in QIIME 2, primers were removed using Cutadapt (Martin, 2011). Vsearch was used to merge reads (35 bp minimum overlap, 5 allowed mismatches) (Rognes et al., 2016), and after removing reads with a PHRED score below 10, dereplicated and chimera check (against SILVA 132) them. Eukaryotic OTUs were clustered at 99% nucleotide identity, singletons were removed. Taxonomic annotation was performed with QIIME 2's feature-classifier (Bokulich et al., 2018) with 99% identity against SILVA 132 majority database. Unassigned OTUs were removed. The resulting OTU table was exported from QIIME and processed in R for subsampling and NMDS generation in the same way the *Imitervirales* OTU tables were processed. The raw reads were uploaded to the DDBJ. The levels 4-6 of the SILVA annotation were used to classify the eukaryotic reads taxonomically into "lineages" to summarize and visualize the community compositions of eukaryotes of the 4 seawater samples.

2.4 Results

2.4.1 Different Primer Cocktail methods produced similar *Imitervirales* Community Profiles

In total, 36.8 million raw paired–end *polB* reads were generated across 15 datasets analyzing four samples taken at different locations and times. MAPS verified 5.7 million as *Imitervirales* reads. The proportion of *Imitervirales* reads among raw reads was 26% on average, but ranged from 8% to 57%. These reads were clustered into 6,045 none singleton *Imitervirales* OTUs (Table 2-7 and Figure 2-2).

Table 2-7: Raw and Imitervirales reads of this study.

Dataset	Number of raw reads	*Imitervirales* reads	Proportion of *Imitervirales* reads	Number of OTUs	Primer cocktail	Protocol number
D-OB-MP1-0 (Li et al., 2018)	16,677,495	8,432,837	51%	5,595	MP1 (58/82 primer pairs)	-
D-OB-MP5-1	5,078,212	992,088	20%	3,018	MP5	1
D-OB-MP10-1	5,995,548	1,916,193	32%	3,396	MP10.v1	1
D-OB-MP20-1	10,720,091	1,019,645	10%	3,110	MP20	1
D-OB-MP1-2	2,205,016	497,356	23%	2,608	MP1	2
D-OB-MP5-2	2,992,984	273,153	9%	2,426	MP5	1
D-OB-MP10-2	4,521,841	340,129	8%	2,912	MP10.v1	1
D-OB-MP20-2	4,752,035	452,365	10%	2,755	MP20	1
S-OB-MP10-1	60,348	5,258	9%	744	MP10.v2	3
S-OB-MP10-2	78,067	37,638	48%	1,487	MP10.v1	1
S-OB-MP10-3	34,860	11,942	34%	1,243	MP10.v2	4
S-OB-MP10-4	38,477	21,965	57%	1,388	MP10.v1	5
S-UF-MP10	96,149	29,275	30%	601	MP10.v2	3
S-UJ-MP1	67,990	19,151	28%	470	MP1	3
S-UJ-MP10	68,168	31,276	46%	539	MP10.v2	3
S-UM-MP10	82,516	18,911	23%	595	MP10.v2	3

The deep sequencing runs (run 1–7) generated 36.3 million reads (2.2 million to 10.7 million reads per dataset), 5.5 million of which were bone fine *Imitervirales* reads (0.27 million to 2.2 million reads per dataset), which clustered into 2,426 to 3,396 *Imitervirales* OTUs. The deep

sequencing accounted for 97% of validated reads of this study. The eight shallow sequencing datasets of a single sequencing run (run 8) generated 0.53 million raw reads, 175,416 of which were assigned to *Imitervirales*. The reads of the shallow depth sequencing datasets (run 8) were classified into 470 - 1,487 *Imitervirales* OTUs. As expected, the deep sequencing generated more reads (and thereby detected more OTUs) than the shallow depth sequencing.

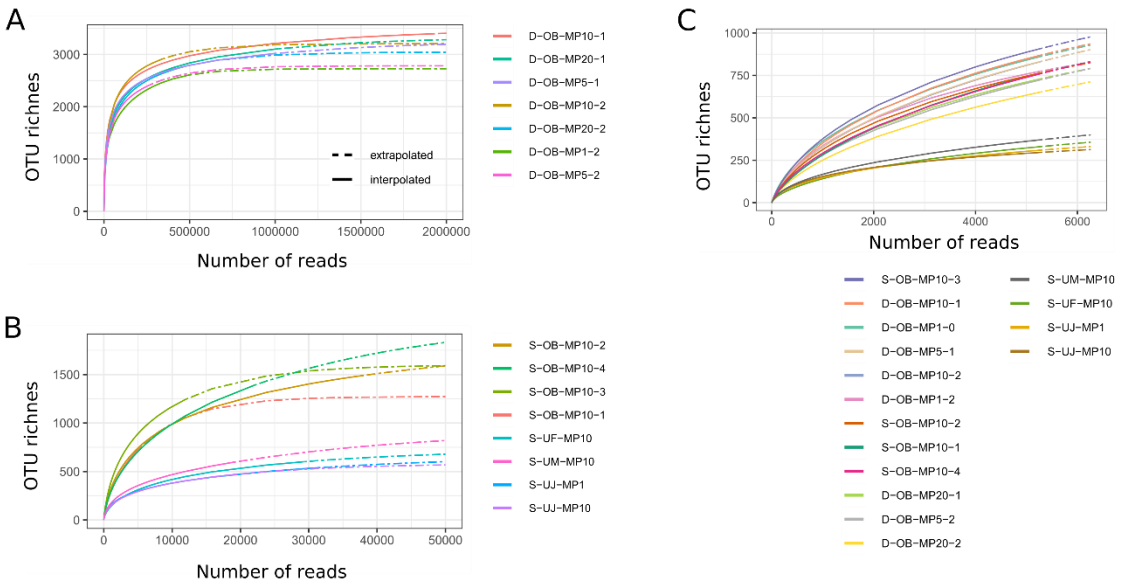

Figure 2-2: Rare faction curves of all conducted experiments. The different sequencing depth resulted in a different number of OTUs discovered. (A) The deep sequencing run's rare faction curves reached saturation. (B) The low depth sequencing run resulted in less reads and OTUs detected. (C) When subsampled at 5,000 reads, the deep sequencing and shallow depth sequencing run show similar rare faction curves. Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

While the number of discovered OTUs is dependent on the sequencing depth, the datasets can still be compared through either subsampling at a common sequencing depth (Figure 2-2C) or normalization of reads (Figure 2-3). Even though different cocktails methods were used (i.e., MP1, MP5, MP10.v1, MP10.v2, and MP20) and the sample preparation varied (i.e., primer concertation, DNA template concentration, and cleanup protocol), the OTU profile was similar for replicates of the same sample (i.e., 11 replicates of the OB sample, 2 replicates of the UJ sample) (Figure 2-3). Furthermore, after subsampling the number of discovered OTUs was comparable between low depth experiments and deep sequencing experiments (Figure 2-2C).

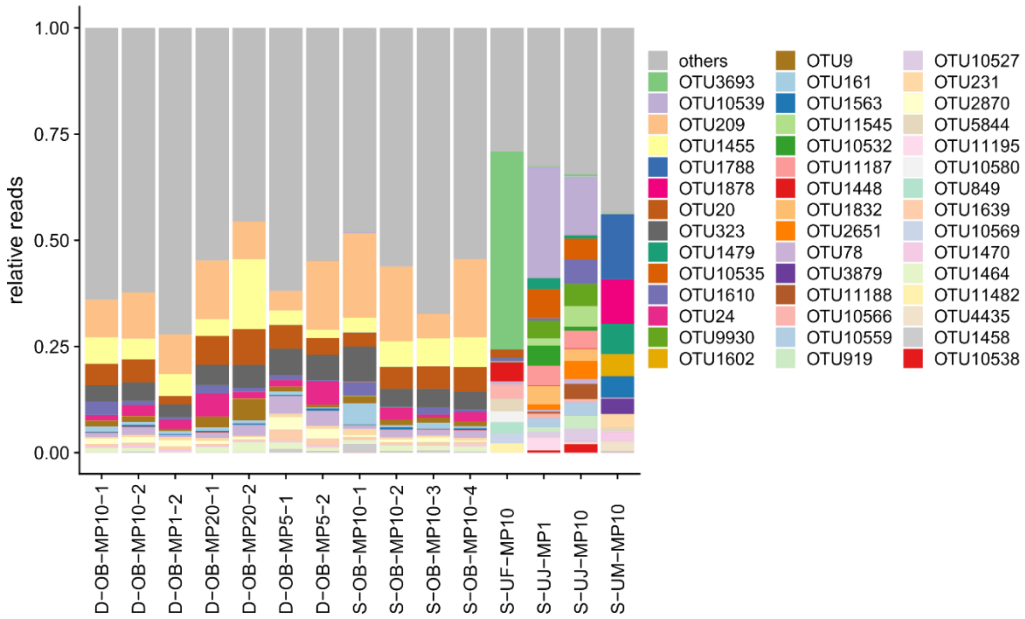

Figure 2-3: Relative frequencies of Imitervirales OTUs across samples and replicates. The 44 most abundant OTUs are displayed, all of them accumulated a relative reads frequency of at least 2% in at least one dataset. The remaining OTUs are grouped together in "others". Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

The previously generated dataset D-OB-MP1-0 used 58 of the 82 MEGARIMER primer pairs (Li et al., 2018). When the experiment was repeated and all 82 amplicons were used for sequencing, many primers were found to have produced bone fide *Imitervirales* reads (Figure 2-4). On the other hand, several primers that were manually selected previously, did not yield *Imitervirales* reads in D-OB-MP1-2 (Figure 2-4). The sample depth normalized D-OB-MP1-0 dataset showed a similar rare faction curve than other OB experiments (Figure 2-2) and clustered with the other OB sample datasets (Figure 2-5). Datasets generated from the same samples generally clustered close together (Figure 2-5).

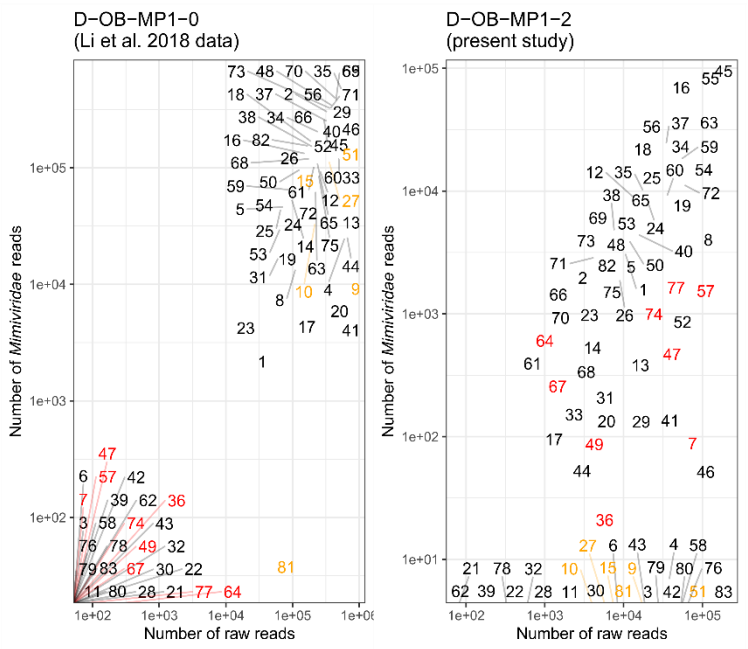

Figure 2-4: Comparison of MP1 results between the previous and present studies. The numbers in the plot refer to individual primer pairs. In the previous study (D−OB−MP1−0) the 24 primer pairs (red) were removed, nine of which produced Imitervirales reads in the present study. The six primer pairs shown in orange did not produce reads in the MP1 experiment of this study (D−OB−MP1−2). Axes are in logarithmic scale. Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

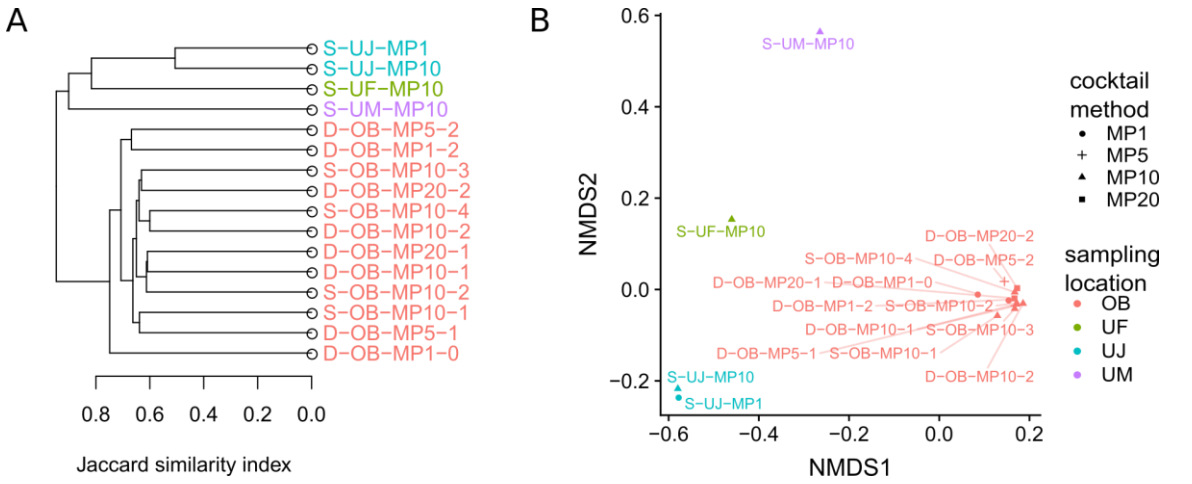

Figure 2-5: Jaccard Dissimilarity of the technical replicates of several samples. (A) The hierarchical clustering analysis of all Imitervirales libraries shows that the Imitervirales from OB is distinct from the UJ communities. (B) The same information visualized by non–metric multidimensional scaling (NMDS) ordination of all generated datasets. Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

I tested three different primer cocktails methods (i.e., MP5, MP10, and MP20) and a method in which every primer pair was used separately for amplification (i.e., MP1). The amount of generated reads varied, but the number of detected OTUs was similar between all methods (Table 2-7 and Figure 2-2), with MP10 resulting in slightly more reads and OTUs. The major differences of the cocktail methods stemmed from the necessary amount of sample DNA and the ratio of *Imitervirales* reads (i.e., reads after MAPS analysis) to raw reads (Figure 2-6).

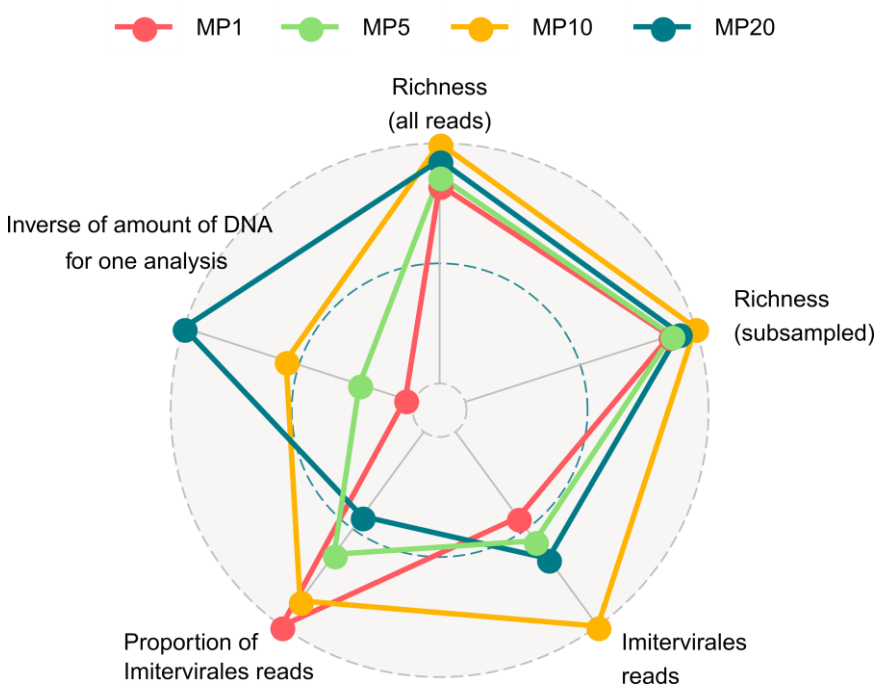

Figure 2-6: The radar chart shows the merits and disadvantages of different primer cocktail methods. The mean outcome of the deep sequencing runs of the Osaka Bay sample (i.e., runs 1–7) were used to create the plot. The five qualities I use to assess the methods were total richness, subsampled richness, amount of Imitervirales reads, proportion of Imitervirales reads to raw reads, and the inverse of the total amount of template DNA needed for one analysis. Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

Several different combinations of amplification protocols, cleanup protocols, and two different versions of MP10 were tested on the Osaka Bay sample and analyzed with shallow depth sequencing. The relative abundance of *Imitervirales* OTUs and the community composition in the each of the replicates of the Osaka Bay data sets were similar (Figure 2-3 and Figure 2-5). However, the amount of generated *Imitervirales* reads (5,528 – 37,638) and ratio of *Imitervirales* reads to raw reads (9% - 57%, Table 2-7) varied. S-OB-MP10-2 generated the most *Imitervirales*

reads and had a high ratio of *Imitervirales* reads (48%). The method I used to generate S-OB-MP10-2 was MP10 version 1 with protocol 1. However, when I employed the same method on the UJ sample, I was repeatedly unable to produce amplicons without impurities (e.g., unspecific amplifications or remaining primers, Figure 2-1). I applied an amplification method based on MP10.v2 and gel extraction (protocol 3) and I was able to generate datasets for the Osaka Bay sample (S-OB-MP10-1) and all of the Uranouchi Inlet samples (i.e., S-UF-MP10, S-UJ-MP10, and S-UM-MP10). The OTU richness of the Osaka Bay sample was higher than Uranouchi Inlet. The Osaka Bay dataset's richness was 788 OTUs on average after subsampling and the dataset that detected the fewest OTUs in the Osaka Bay found 651 OTUs. The richness of the Uranouchi Inlet communities was 334 OTUs on average (UJ: 303 OTUs, UF: 330 OTUs, UM: 370 OTUs).

2.4.2 Comparison of relative metabarcoding profiles against absolute quantification of OTUs

I selected six *Imitervirales* OTUs (i.e., OTU1610, OTU231, OTU5844, OTU1788, OTU323, and OTU1458) that were found using metabarcoding. As suggested by a previous study (Li et al., 2018), the relative read frequencies were not reliable for assessing the *polB* gene copy numbers. For example, OTU1610 had a relative read frequency of 1.4 ±1 % in the Osaka Bay sample, the qPCR inferred quantification showed that the OTU had a concentration of 756 ± 271 molecules mL^{-1} in the Osaka Bay sample. In contrast, OTU323 had a higher relative read frequency (4.9 ± 1.4%), but the concentration in the same sample was lower with 33 ± 7 molecules mL^{-1} (Figure 2-7).

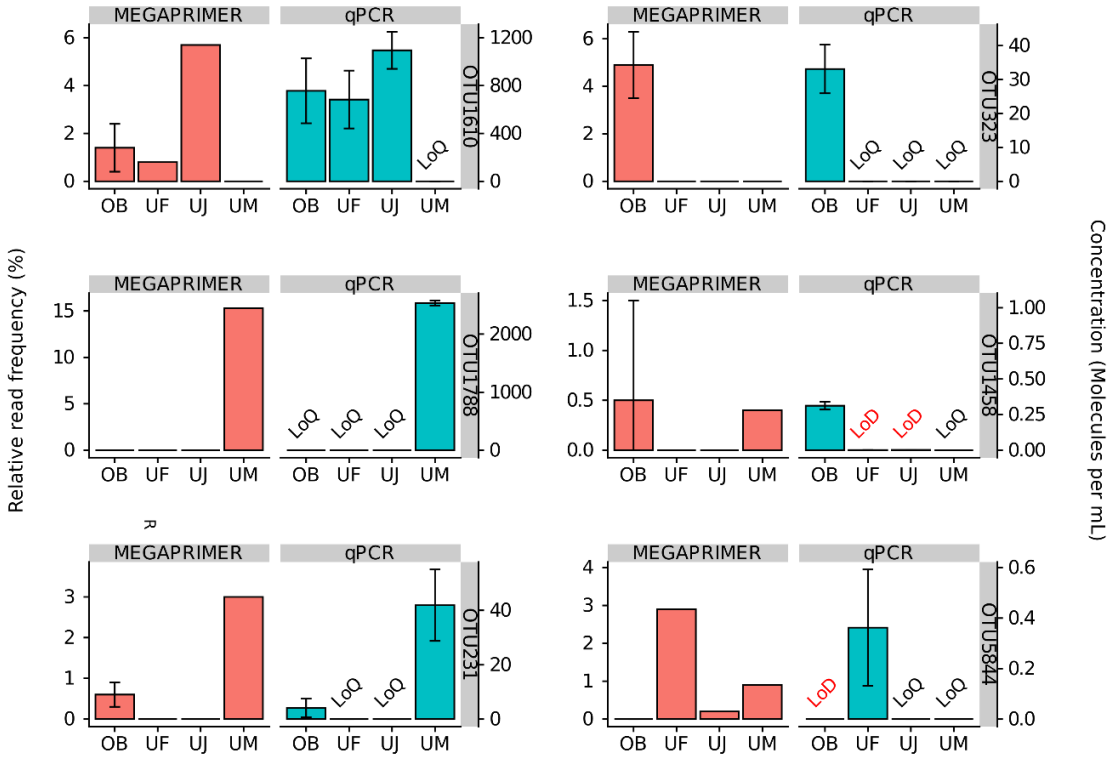

Figure 2-7: Similar OTU abundance profiles of MEGAPRIMER and qPCR. The relative read frequency of OTUs of the Osaka Bay sample were calculated by averging all realtive reads. Error bars show one standard deviation. The Uranouchi Inlet sample names are abbreviated: S-UF-MP10 (UF), S-UJ-MP10 (UJ), and S-UM-MP10 (UM). Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

Even though the relative read frequencies did not reliably indicate individual OTU concentration, the OTU abundance profiled obtained through both methods were similar (Figure 2-7). Furthermore, OTUs that showed higher concentrations (>1,000 molecules mL^{-1}) also had the highest relative read percentages 5.7% and 15.3% (10.5% on average). The seven OTUs that had a quantifiable concentration produced relative read counts between 0.49% and 4.9% (2.0% on average). Three OTUs were below LoD with qPCR and, if at all detected with MEGAPRIMER, showed low relative read percentages (<0.01%). A positive correlation between relative read percentages and qPCR inferred molecular concentrations was found (Pearson's r = 0.85, excluding LoD OTUs, $p < 10^{-3}$).

2.4.3 Eukaryotic community composition

Table 2-8: 18S Number of reads and OTUs of the 18S amplicon sequencing.

Sampling location	Number of raw reads	Taxonomically annotated reads	Number of OTUs
OB	67,028	44,727	439
UF	95,352	63,833	325
UJ	81,281	67,479	285
UM	80,237	54,845	528
Total	323,898	230,884	1,156

The *Imitervirales* community of the Osaka Bay had a higher richness than the Uranouchi Inlet communities (Figure 2-2) and was generally more dissimilar to the Uranouchi Inlet communities than they were to each other (Figure 2-5). I investigated if these differences in richness and community dissimilarities were also found in the eukaryotic communities. I analyzed the samples by 18S rRNA amplicons sequencing, which resulted in 323,898 raw paired–end reads after removing low quality, unmerged or not assigned reads, 230,884 eukaryotic reads were clustered into 1,156 non singleton OTUs (99% identity, Table 2-8). The richness of each of the four seawater samples ranged from 285 to 528 OTUs (Table 2-8 and Figure 2-8A) and while the most dominant lineages overlapped (Figure 2-8C), the Uranouchi Inlet samples were more similar to each other than to the Osaka Bay sample (Figure 2-8B).

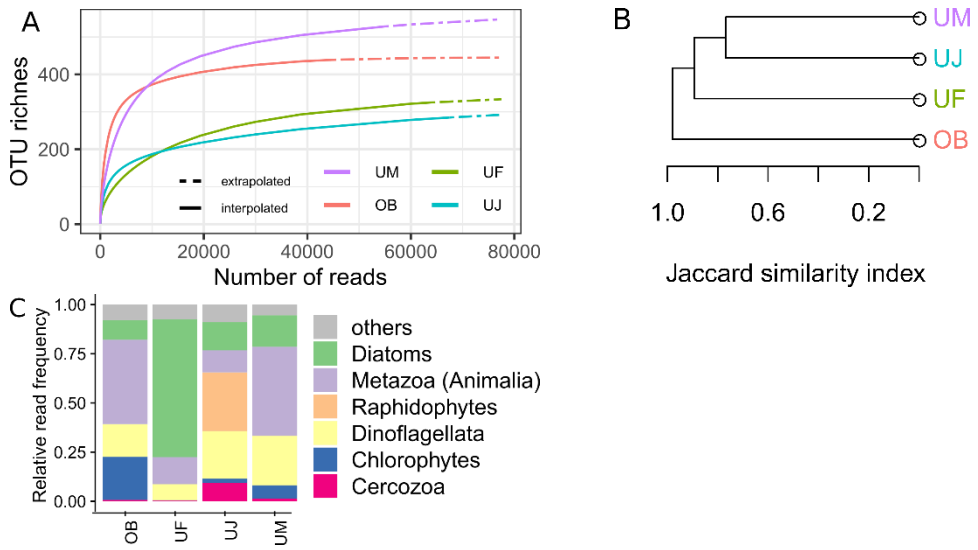

Figure 2-8: Details of the 18S amplicon sequencing. (A) The Uranouchi station M and the Osaka Bay dataset showed the higher richness. (B) The OB eukaryote community was more dissimilar than the Uranouchi Inlet communities were to each other. (C) Distribution of reads between the dominant linages was different for the all samples, but Metazoa, Diatoms, and Dinoflagelattes were abundantly present in each sample. Figure A and C taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

The most abundant members of the Osaka Bay community were a Metazoa OTU (Siphonophorae, 15.8%), a green algae OTU (*Coccomyxa* sp., 9.1%), a Dinophyceae OTU (7.1%), and a *Minutocellus* OTU (centric Diatom, 5.8%). Out of these 4 OTUs only the *Minutocellus* OTU was observed in the Uranouchi Inlet with relatively high read counts (UF sample: 9.5%). Several abundant OTUs were found in two or more Uranouchi Inlet samples. *Thalassiosira* OTUs (also a centric Diatom) were found in all Uranouchi Inlet samples with high relative abundance (UF: 44.7%; UJ: 12.9%, and UM: 7.2%). Other OTUs were found in two Uranouchi Inlet samples, namely a *Cyclotella choctawhatcheeana* OTU (UF: 12.0% and UM: 7.7%) and an OTU assigned to the bloom forming *K. Mikimotoi* (UJ: 16.3% and UM: 20.2%). The bloom forming Raphidophyte *Chattonella* sp. contributed over one fourth of the reads in UJ (29.6%) but was not present in the other samples (Figure 2-9).

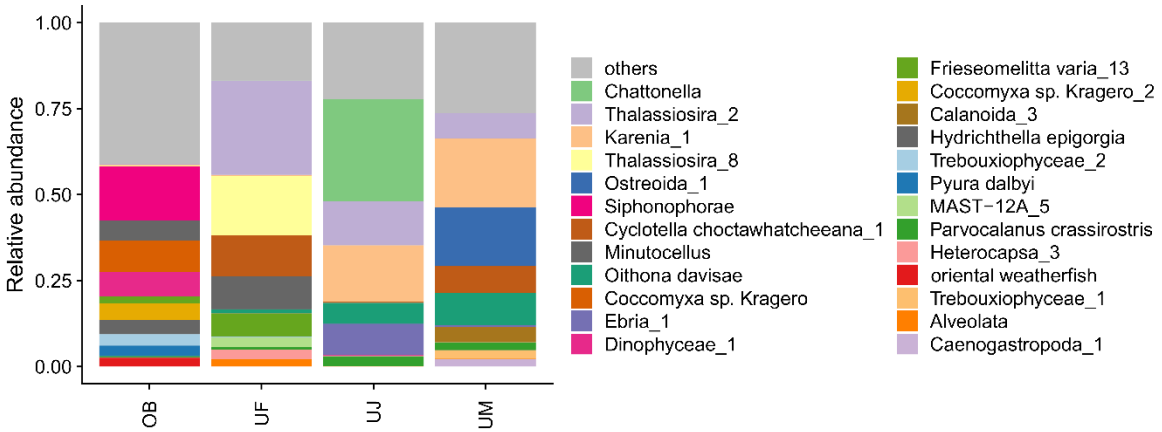

Figure 2-9: OTU composition of the eukaryotic communities of four samples. The eukaryotic OTUs with relative read frequency was >2% in any sample are plotted in color. OTUs with fewer reads were merged to others. Figure taken from (Prodinger et al., 2020) in accordance with CC BY 4.0.

2.5 Discussion

The goal of this study was to streamline a MEGAPRIMER analysis protocol using primer pair cocktails instead of individual primers and to evaluate a possible primer bias, which was hypothesized in the original MEGAPRIMER publication (Li et al., 2018). Optimization of the new method was achieved by testing different primer cocktail methods (i.e., MP5, MP10.v1, MP10.v2, MP20), various amplification PCR conditions (Table 2-5 and Table 2-6), and sequencing depths (Table 2-7 and Figure 2-2) on the sample that was previously analyzed by Li et al. (Li et al., 2018). The new methods of this study reproduced the previously published *Imitervirales* community composition (Li et al., 2018) (Figure 2-5B), while requiring less resources (Figure 2-6). Additional analysis of three samples taken in the Uranouchi Inlet further showed distinct OTU profiles and the analysis of the UJ sample with both MP10 and MP1 showed similar OTU profiles.

Methods based on the utilization of “primer–pair–cocktails” were introduce in this study. Most of these methods (i.e., MP1, MP10, and MP20) showed specific merits and disadvantages. MP1 produced the highest ratio of *Imitervirales* reads to raw reads, but required the most sample DNA (>80 ng) and preparation time. In contrast, MP20 required the smallest amount of sample DNA (<4 ng) and sample preparation, but produced many unusable reads. MP10 produced the

highest numbers of *Imitervirales* reads (Figure 2-6 and Table 2-7), but required more sample DNA than MP20 (roughly 8ng for MP10), and had a lower *Imitervirales* read to raw read ratio than MP1. Despite this MP10 was judged as the most suitable method, for its balance between preparation complexity, sample DNA usage, and overall performance (Figure 2-6). In order to compare MP10 to the MP1 method, both were applied on two seawater samples (i.e., the Osaka Bay and the Uranouchi station J sample). In both cases, the primer cocktail method yielded similar results to the single primer analysis protocols (Figure 2-3 and Figure 2-5). Furthermore, I applied the new MP10–MEGAPRIMER–method on another three coastal seawater samples, revealing the unique OTU profiles of each sample.

The primer pair cocktail strategy improved the MEGAPRIMER analysis workflow, but it was not the only critical factor of successful amplicon generation. Even though the Osaka Bay sample could be analyzed with different amplification protocols (i.e., protocol 1 to 5 of Table 2-6), the MEGAPRIMER amplicons of Uranouchi station J were repeatedly found to be unfit for sequencing. These failed experiments revealed the cleanup step to be critical for sequencing success. The sequencing failure was caused by high concentration of >100 bp oligonucleotides, an unwanted byproduct of MEGAPRIMER based *polB* amplification (Figure 2-1). The oligonucleotides were not removable with magnetic bead cleanup, since magnetic beads were specifically designed to remove oligonucleotides of <100 bp (www.beckmancoulter.com, accessed 23.3.2021). The influence of different factors on the oligonucleotides was tested by analyzing three coastal seawater samples of the Uranouchi Inlet alongside the Osaka Bay sweater sample with MP10, but with variating PCR conditions, cocktail mixtures, and cleanup protocols. Gel extraction was found to be an alternative to magnetic beads clean up, because gel extraction was able to remove unspecific amplification products of all sizes. This lead to a more consistent sequencing success. MP10.v2 in combination with protocol 3 (the only protocol with a gel extraction cleanup step) was

the only tested method that could generate amplicon sequencing data for all four seawater samples. In conclusion MP10v2 with protocol 3 facilitated *Imitervirales* community analysis for all tested samples, while being less laborious and using less sample DNA than the previously published methods.

In the original MEGAPRIMER study Li et al. already provided a solution for reducing unspecific short amplicons by checking each amplification separately by gel electrophoresis and removing the amplicons of primer pairs that did not show a clear amplicon band (Li et al., 2018). This approach was successful and resulted in a high *Imitervirales* to raw read ratio (51%) and a high richness (5,595 OTUs, Table 2-7). I repeated this experiment without removing unsuccessful primer pairs. Not only did I generate fewer *Imitervirales* reads and lower richness (2,607 OTUs), the read *Imitervirales* to raw read ratio was lower as well (23%, Table 2-7). Despite these differences in sample preparation and sequencing success, the community composition was similar in both experiments (Figure 2-5).

I quantified 6 OTUs with high relative read frequencies of the MEGAPRIMER barcoding data of this study in each of the four seawater samples by qPCR. The observed *Imitervirales polB* concentration ranged from 0.7×10^3 to 2.6×10^3 molecules mL^{-1}. A previous study found between 4×10^3 and 1.7×10^5 *Nucleocytoviricota* genomes for 1 mL of the >0.2µm size fraction of photic zone samples (Hingamp et al., 2013). Roughly 36% of these *Nucleocytoviricota* were assigned to *Imitervirales* (Hingamp et al., 2013), hence Hingamp et al. found between 1.4×10^3 and 6.1×10^4 *Imitervirales* genomes in 1 mL. The quantification results of my study are therefore consistent with previous *Imitervirales* abundance estimates (Hingamp et al., 2013).

The relative read percentages of *Mimviridae* OTUs based on MEGAPRIMER analysis was compared to the concentration of OTUs (i.e., qPCR analysis of *polB* genes per mL). As hypothesized in the original MEGAPRIMER study (Li et al., 2018), a sequencing bias was observed in the

MEGAPRIMER data (Figure 2-7). The reason for this bias may be that some OTUs can be amplified by up to 38 primers (Li et al., 2018). Therefore, comparing the relative read frequency of OTUs cannot be used to estimate which OTUs shows a higher absolute abundance.

However, the observed relative frequency profiles of OTUs was correlated to the marker gene concentrations, despite the previously described amplification bias. Therefore, the relative frequency of *Imitervirales* OTUs are not random and contain information about the relative abundance of an OTU in the *Imitervirales* community. Furthermore, the relative read frequency profile of OTUs that were found in at least two samples was similar to the qPCR inferred abundance profile (Figure 2-7). The comparison of relative read percentages of a single OTU in multiple samples (i.e., not between different OTUs in one sample) may still be used to estimate the relative abundance of this OTU in the different *Imitervirales* communities.

The *Imitervirales* richness of the Osaka Bay sample was higher than the richness of any Uranouchi Inlet sample (Figure 2-2). A previous study also found that the coastal seawater of the Osaka Bay had a higher *Imitervirales* richness than other environments like brackish water (mangrove), freshwater (of a hot spring) or the open ocean (Japanese sea) (Li et al., 2019). The hosts of oceanic *Imitervirales* are mostly eukaryotic algae. Therefore, I hypothesized that a high richness of the *Imitervirales* community might be explained by a high richness of eukaryotic algae. The eukaryotic community of the Osaka Bay showed a higher richness than two of the Uranouchi Inlet samples, but was less rich than the Uranouchi station M sample (Figure 2-8A). Most of the eukaryotic reads were assigned to six lineages, out of which only chlorophytes contain *Imitervirales* hosts (Figure 2-8C), other *Imitervirales* hosts were not abundantly present in any sample. The slightly higher relative abundance of chlorophytes in the Osaka Bay and the Uranouchi station M might contribute to a higher richness in *Imitervirales*. Aside from this

observation I was not able to explain the high richness of *Imitervirales* in the Osaka Bay sample with the eukaryotic community.

In terms of Jaccard similarity of the different communities, both the *Imitervirales* and eukaryotic communities showed a high similarity among the Uranouchi Inlet samples and a low similarity between any of the Uranouchi Inlet samples and the Osaka Bay sample (Figure 2-8B). The similarity of the eukaryotic communities was also reflected in the overlap of abundant eukaryotic OTUs (i.e., OTUs with high relative read frequencies). Three of the five most abundant OTUs found in the Uranouchi Inlet samples overlapped (i.e., OTUs assigned to *Thalassiosira*, *Karenia* sp., and *Cyclotella choctawhatcheeana*, Figure 2-9), while the most abundant eukaryotic OTUs of the Osaka Bay were not abundantly present in any Uranouchi Inlet sample (except for *Minutocellus* in Uranouchi station F sample, Figure 2-9).

This study provides a reasonable and robust MEGAPRIMER cocktail method by showing the merits and disadvantages of PCR amplification protocols and cleanup methods, and by evaluating the primer bias of MEGAPRIMER. In conclusion, the new method (i.e., MP10.v1 or v2 with protocol 3) necessitated less sample DNA, was less time consuming and reproduced two different analysis that used single primer pair (MP1) amplification. Thereby, the necessary amount of DNA and analysis time was sharply reduced. This evaluation of MEGAPRIMER and the new cocktail based amplicon preparation method provides a basis for future ecological analysis of *Imitervirales*. By employing these new methods the MEGAPRIMER analysis can be performed on a larger sample set, enabling more in depth studies *Imitervirales* ecology.

3 Chapter 3: Year-round amplicon sequence variant community dynamics of eukaryotes, *Imitervirales*, and prokaryotes and their differences

3.1 Abstract

Diverse microbial communities inhabit coastal seawater. These communities show repeating seasonal patterns, which are caused by abiotic factors like temperature and nutrient availability and by biotic factors like competition for nutrients and other interactions between communities. Here, I studied the community dynamics of eukaryotes, a monophyletic clade of large DNA viruses in the order *Imitervirales*, and prokaryotes by metabarcoding analysis of 43 samples taken during 20 months in the Uranouchi Inlet, Kochi, Japan. I analyzed the different microbial communities at the amplicon sequence variant (ASV) level to better understand the seasonal dynamics of coastal microbial communities. I found that the ASV communities of eukaryotes, prokaryotes, and *Imitervirales* changed in a similar seasonal cycle. Further analysis revealed that certain features of the community dynamics differed. For instance, the proportion of ASVs that were recovered after a year differed for each microbial community. These different community dynamics were explained in part by differences in the recurrence and persistence of individual community members of the cellular (i.e., prokaryotic, eukaryotic) and the viral community. Members of the prokaryotic community showed the highest persistency, while eukaryotic and *Imitervirales* ASVs were less persistent. In this study I show that the basis of this differences in community dynamics among eukaryotes, their viruses, and prokaryotes lies in the niche breadth of their respective community members as well as the specificity of their interactions with one another (virus–eukaryote–interactions vs. prokaryote–eukaryote–interactions).

Contributions

Sampling was performed by Yoshihito Takano, Keizo Nagasaki, and Etsunori Taniguchi. DNA sequencing support was provided by Yasuhiro Gotoh, Tetsuya Hayashi, Tatsuhiro Isozaki, Kento Tominaga, and Takashi Yoshida. I performed both wet experiments and dry analyses, and wrote the initial manuscript.

3.2 Introduction

The microbial communities of the ocean can be separated into three main groups: eukaryotes, prokaryotes, and viruses. These microbes contribute substantially to biogeochemistry by cycling different elements (Brum and Sullivan, 2015; Falkowski et al., 2008; Not et al., 2012). The composition of these microbial communities changes due to factors that are either abiotic (e.g., temperature or nutrient concentrations) (Gilbert et al., 2012; Sunagawa et al., 2015) or biotic factors (e.g., predation or infection) (Lima-Mendez et al., 2015; Nagasaki et al., 1994). Marine ecosystems are complex because the members of the aforementioned microbial communities are constantly interacting with each other and thereby change their environment. This means that in order to understand socioeconomically important processes like carbon export and elemental cycling we need to study the change in microbial communities.

High throughput sequencing based studies have previously described the dynamics of eukaryotic (Chen et al., 2017; Giner et al., 2019; Gran-Stadniczeñko et al., 2019b), prokaryotic (Chafee et al., 2018; Fuhrman et al., 2015; Gilbert et al., 2012; Milici et al., 2016; Sakami et al., 2016; Teeling et al., 2016; Ward et al., 2017), both eukaryotic as well as prokaryotic communities (Bock et al., 2018), and viral communities (Ignacio-Espinoza et al., 2020). Other studies have investigated the relationships between cellular communities (Martin-Platero et al., 2018; Needham and Fuhrman, 2016; Santi et al., 2019) or cellular and viral communities (Arkhipova et al., 2018; Chow and Fuhrman, 2012; Gran-Stadniczeñko et al., 2019b; Johannessen et al., 2017; Needham et al., 2013; Pagarete et al., 2013; Sandaa et al., 2018). The seasonality of these different communities was also explored using time series sampling over several years for prokaryotes (Fuhrman et al., 2015; Gilbert et al., 2012), eukaryotes (Giner et al., 2019), and small viruses (Ignacio-Espinoza et al., 2020).

The least well understood of the three aforementioned microbe groups are viruses. Large double–stranded DNA viruses are considered important top down regulators of eukaryotic communities in the ocean (Endo et al., 2020). In coastal areas these viruses were associated with algal bloom events with metagenomics (Moniruzzaman et al., 2016) or microscopy (Nagasaki et al., 2003; Tarutani et al., 2000; Tomaru et al., 2004). This suggests that *Imitervirales* and other *Nucleocytoviricota* infect bloom forming unicellular algae and contribute or even cause bloom collapse (Lehahn et al., 2014; Moniruzzaman et al., 2016; Tomaru et al., 2004). Another metabarcoding study linked members of the *Imitervirales* community and the eukaryotic community of South Norway (Gran-Stadniczeñko et al., 2019b). A recent study showed horizontal genes transfer between eukaryotes and *Imitervirales* (Schulz et al., 2020). A transcriptomic analysis of a high frequency sampling spanning three days (Ha et al., 2021) and a metatranscriptomics analysis of a global set of samples (Carradec et al., 2018) also showed that *Imitervirales* are ubiquitously present and active in the ocean.

To contribute to the research of oceanic microbes I am the first to study the microbial communities of the Uranouchi Inlet, which is located at the south eastern side of Shikoku Island. The climate of this area is subtropical with distinct seasons, typically hot summers and mild winters. During spring, summer, and fall algal blooms or red tides of *Heterosigma akashiwo* (Raphidophyceae), *Karenia mikimotoi* (Dinophyceae), and *Chattonella* spp. (Raphidophyceae) occur in Uranouchi Inlet. These blooms can damage local fisheries and aquaculture and are therefore considered harmful algae blooms (HABs).

The *Imitervirales* and eukaryotic community, as well as the prokaryotic community of the Uranouchi Inlet were analyzed using previously established highly sensitive metabarcoding methods. I analyzed seawater from Uranouchi Inlet that was sampled 43 times during 20 months from January 2017 to September 2018. The communities of these seawater samples were

characterized at an exact amplicon sequence level which distinguishes sequences with even a single nucleotide difference. *Imitervirales* and cellular microbial community dynamics were then compared to better understand their interactions. The host community of *Imitervirales* (i.e., eukaryotes) was analyzed, as well as the prokaryotic community, which can only be indirectly influenced by *Imitervirales*. I found striking similarities in the seasonality of the communities, but also differences in their specific community dynamics.

3.3 Methods

3.3.1 Seawater sampling and DNA extraction

Seawater sampling was performed by my collaborators at 4 distinct sampling locations in the Uranouchi Inlet, Kochi Prefecture, Japan during 20 months starting from 05th January 2017 until 25th September 2018 with an average sampling interval of 15 days. The sampling stations were station “j”: 33°25'43.2"N 133°22'49.5"E, station “m”: 33°25'60.0"N 133°24'38.3"E, station “f”: 33°26'33.6"N 133°24'41.8"E, and station “u”: 33°25'49.7"N 133°24'01.4"E. My collaborators sampled in total 43 samples. The samples were named in a “yymmdd-station” format, for instance the first sample taken at station “j” on 5th January 2017 was called 170105-j.

Sampling was performed by pumping 10L seawater from a depth of 5m into a bottle (polycarbonate). The seawater was transported to the laboratory for sequential filtering. Samples were filtered through a 3.0µm membrane filter (diameter 142 mm, polycarbonate, Merck), the filtrate was collected and then sequentially filtered through a 0.8µm membrane filter (diameter 142 mm, polycarbonate, Merck). The filtrate (1L) of the 0.8µm membrane filter step was then filtered through a 0.22µm filtration unit (Sterivex, polycarbonate, Merck). Each seawater sample yielded 3 filters, of which the residue was intended for metabarcoding analysis. The filters were stored at −80°C. DNA extraction was performed with either Proteinase–K method (Frias-Lopez et

al., 2008) for the 0.22µm filtration units or the xanthogenate–sodium dodecyl sulfate method (Yoshida et al., 2003) for the larger size fractions (3µm and 0.8µm).

3.3.2 PCR amplification and sequencing

Eukaryotic community diversity analysis was performed on the DNA extract of the 0.8µm – 3µm and >3µm size fractions with 18S rRNA gene primers targeting the V8/V9 region. The primers were called "V8f" (forward primer, ATAACAGGTCTGTGATGCCCT) and "1510r" (reverse primer, CCTTCYGCAGGTTCACCTAC) (Bradley et al., 2016). Five µL of both primers (1 µmol·L^{-1}), 2.5 µL template DNA (0.25 ng·μL^{-1}), and 12.5 µL 2× KAPA HiFi HotStart ReadyMix were mixed and the reaction was carried out with a thermal cycler (Thermal Cycler Dice Touch, Takara Bio Inc., Shiga, Japan) programmed to heat to 98°C (3 min), followed by 25 cycles of 98°C (20s), 65°C (15s), and 72°C (15s), the cycles were concluded with a final elongation step at 72°C (10 min). Amplicon cleanup was performed using magnetic beads according to the protocol of the manufacturer.

Prokaryotic community diversity was analyzed with a 16S rRNA gene amplification protocol performed on the DNA extract of the 0.22µm – 0.8µm size fraction. The primer pairs were called "Pro-16S NGS" (forward primer, CCTACGGGNBGCASCAG) and "Pro-16S NGS" (reverse primer, GACTACNVGGGTATCTAATCC) (Takahashi et al., 2014). I mixed 5 µL of both primers (1 µmol L^{-1}) with 2 µl of the DNA template (1 ng·μL^{-1}) and 12 µL 2× KAPA HiFi HotStart ReadyMix. The thermal cycler was programmed as follows, 95°C (3 min), then 25 cycles of a 3 step (30s each) cycle of 95°C, 55°C, and 72°C, a final extension with 72°C (5 min). Amplicon cleanup was performed using magnetic beads according to the manufacturer's protocol.

Imitervirales community composition was analyzed with MP10.v2 (Table 2-3) in combination with protocol 3 (Table 2-6) detailed in chapter two. MEGAPRIMER were mixed in cocktails of 10 or 11 primers (Table 2-3) and the DNA extracted from the 0.22µm – 0.8µm size fraction was used as template. I performed 8 PCR amplifications per sample using 7.5 µL primer

cocktail (20 µmol L^{-1}), 5 µL template DNA (1 ng $µL^{-1}$) and 12.5 µL KAPA HiFi HotStart ReadyMix. The thermal program started with 95°C (3 min), followed by 32 cycles of a 3 step (30s each) cycle of 95°C, 54°C, and 72°C, a final extension with 72°C (5 min). The 8 generated amplicons of each sample were merged, resulting in 1 amplicon mixture per sample with a volume of 200 µL. Ethanol precipitation was performed on each sample's merged amplicon and the resulting DNA pellet was resuspended in 25 µL of ultrapure water. Agarose gel (2%) electrophoresis was performed. The gel was stained (Gel Green, Takara Bio Inc., Shiga, Japan), the amplicon band between 400 bp and 600 bp was cut out and DNA was extracted using a spin down column extraction kit (Promega Wizard SV, Promega, Madison, WI).

Henceforth the different microbial amplicons (i.e., 16S, 18S and MEGAPRIMER amplification products) were processed nearly identically. Indexing PCR was performed according to Illumina's library preparation protocol. The indexed cellular amplicons were purified using magnetic beads, also according to Illumina's sample preparation guide. The indexed *Imitervirales polB* amplicons were purified using agarose gel extraction. The final library concentration was 10 pM (30% PhiX spike) and sequencing was performed on the MiSeq platform (2 × 300 nucleotides, paired–end reads).

3.3.3 Eukaryote and prokaryote data analysis

Both the 18S amplicon as well as the 16S amplicon reads were analyzed using a QIIME 2 (version 2020.2.0) (Bolyen et al., 2018) based pipeline, if not specified otherwise default settings were used. An analysis pipeline for each 18S and 16S amplicon data were written, however they were very similar in function. First, the raw reads were imported to QIIME 2 and further processed with the dada2 extension. Low quality reads were removed, primers were cut according to their length. High quality reads were truncated (at 240 bp for eukaryotic reads and 250 bp for prokaryotic reads), corrected, merged, and Chimeras were removed. All these steps were

automatically performed by dada2 (Callahan et al., 2016). Taxonomic annotation was performed using the SILVA 132 majority database and Qiime 2's "feature-classifier classify-consensus-vsearch" command with a 90% identity threshold accepting only the top hit. Singleton and unassigned ASVs were removed and an ASV table was exported for further analysis.

3.3.4 *Imitervirales* amplicon processing system based on dada2 (MAPS2)

An R script (R 3.6.3) (www.R-project.org, accessed 23.3.2021) was used to load dada2 and import, filter, trim, correct, dereplicated and merge raw sequences to ASV (Callahan et al., 2016). Unless further specified default settings were used. The script loaded the sequences using the "filterAndTrim" command, removing primers according to maximum length of MEGAPRIMER and truncating reads at 240 bp. Low quality regions were truncated as well and up to 2 reading errors were accepted. The sequencing error rates were learned using "learnErrors" on both forward and reverse reads. Reads were dereplicated using "derepFastq" and "dada" was called on the dereplicated reads, using the learned error rates to correct reads. Reads were merged with the "mergePairs" command with a minimum overlap of 12 and accepting no errors. Chimeras were removed with "removeBimeraDenovo". An ASV table and fasta file containing the amplicon sequences was exported. BLASTx (2.9.0)(Altschul et al., 1990) was used to search the amplicon sequences against a custom PolB amino acid sequence database (Li et al., 2018) with an e-value of 10^{-5} and ASVs with none viral hits were removed. A custom output format of the BLASTx search was used to save the translated nucleotide sequences and create a fasta file with amplicon PolB amino acid sequences. ASVs that matched viral sequences were processed further by adding their PolB amino acid sequence to a previously published PolB reference tree (Endo et al., 2020) using pplacer (1.1.alpha19) (Matsen et al., 2010). ASVs placed on the monophyletic *Imitervirales* branch were split into 14 clades, ASVs that were not inside the *Imitervirales* branch were removed. In the last step all ASV were trimmed in a common region and a final ASV table was exported.

3.3.5 Ecological analysis: Diversity and Dissimilarity metrics

The ASV tables of either eukaryotes (the 18S amplicon sequencing data), prokaryotes (the 16S amplicon sequencing data), or *Imitervirales* (the MEGAPRIMER sequencing data) were parsed with R. The "vegan" (2.5-6) (Oksanen et al., 2012) and "ggplot2" (3.2.1) (https://ggplot2.tidyverse.org, accessed 7.4.2021) packages were used for analysis and visualization. Datasets with fewer than 8,000 reads were removed, all other data sets were normalized (for community composition analysis) or subsampled at 8,000 reads (for dissimilarity and diversity analysis). Whether or not an ASV was present in a dataset was determined by it yielding more than 0.1% of reads after subsampling (i.e., 8 reads in the subsampled dataset). The dissimilarity of communities was calculated with the Sørensen-Dice method (i.e., vegan's "vegdist" function with the default Bray Curtis method and binary set to true). Thereby the dissimilarity of 2 communities is defined as follows:

$$dissimilarity_{SD} = \frac{2|community_A \cap community_B|}{|community_A| + |community_B|}$$

"cmdscale" and "sammon" (MASS package) were used to create non–metric dot plots to visualize dissimilarities. The similarities of communities among the same and opposite seasons was calculated by binning months into seasons: spring (March to May), summer (June to August), fall (September to November), and winter (December to February). I calculated the *p* value with either R's "t.test", or if more than 100 samples were compared and Benjamini–Hochberg correction was necessary with "pairwise.t.test".

3.3.6 Cell counts and environmental data (biotic and abiotic)

I received data about the temperature, salinity, phytoplankton cell counts, and concentrations of nutrients, organic and inorganic forms of nitrogen and phosphorus, dissolved oxygen, and chlorophyll *a* from my collaborator Etsunori Taniguchi (Kochi Prefectural Fisheries

Research Institute, https://www.pref.kochi.lg.jp/soshiki/040409, accessed 30.3.2021) with appropriate ethics approval. This data is henceforth called environmental data. The environmental data was investigated from 2013 to 2019 to monitor the severity of harmful algal blooms (HAB) in the Uranouchi Inlet. Therefore only HAB species' cell count data was available. In this study I used the maximum daily cell count of each station, day and sampling depth (mostly 0m, 2m, and 5m). Different HAB species are considered blooming at different cellular concentrations (https://www.pref.kochi.lg.jp/soshiki/040409, accessed 30.3.2021), depending on fish mortality rates. Chattonella spp. cellular concentrations as low as 10 to 100 cells mL^{-1} can cause fish to die. *K. mikimotoi* (an unarmored, solitary dinoflagellate, 24µm – 40µm long, 20µm –32µm wide (Omura et al., 2012)) can kill fish with cellular concentrations of several 100 to 1,000 cells·mL^{-1}. *H. akashiwo* (kidney shaped, motile dinoflagellate, 8µm – 25µm long (Omura et al., 2012)) concentrations over 10,000 cells·mL^{-1} can be deadly for fish.

3.3.7 Co-occurrence analysis and correspondence analysis

The co-occurrence of all microbial ASVs was analyzed using the "FlashWeave" package in Julia (julialang.org, accessed 28 October 2021). FlashWeave is an machine–learning based algorithm that can detect microbial associations (Tackmann et al., 2018). Co-occurrence networks of major eukaryotic and *Imitervirales* ASVs were generated using only ASVs with a relative read abundance ≥ 1% in at least 1 dataset. The networks analysis of major ASVs of viruses and eukaryotes used either the PicoNano dataset or the NanoPlus dataset with the *Imitervirales* dataset, resulting in two separate networks. The number of edges of inter and intra–community networks were analyzed with subsampled (8,000 reads) ASVs tables of all communities. ASVs that contributed less than 0.1% in any dataset were removed. After normalizing ASV tables, they were merged and a pseudo read count was added. The network was generated using Julia (v1.0.3) (Bezanson et al., 2017) with FlashWeave's "learn_network" command. I used sensitive and

heterogeneous modes, set make_sparse and normalize to false, set alpha to 0.01, and removed ASVs that were observed fewer than 3 times (Tackmann et al., 2018). R's "mantel" command (5 million permutations) was used to calculate the spearman correlation of dissimilarity matrixes (i.e., Mantel test). NMDS analysis was done by using the "cmdscale" command on a dissimilarity matrix. Stress of the NMDS was calculated with the "sammon" command (MASS package) (Venables and Ripley, 2002). An RDA (redundancy analysis or correspondence analysis) was performed on the community compositions of the different microbes using the abiotic data (i.e., temperature, salinity, and concentrations of dissolved oxygen, phosphor (total), and nitrogen (total)) as constraining matrix. The RDA was done by calling the "rda" command with default settings. The "forward.sel" command (adespatial package) was used to determine the predictive power of the abiotic factors on the microbial; communities. The dissimilarity of the microbial communities in different seasons was calculated by binning samples by month according to their season. Winter includes samples taken in December, January, and February, spring includes March, April, and May, summer includes June, July, and August, and fall includes the remaining months (September, October, and November). The statistical significance of analysis was confirmed by calculating the *p* value with the "t.test" command, if applicable *p* values were Benjamini–Hochberg corrected. The "niche.width" (spaa package) command was called on the ASV tables to calculate Levins' niche breadth (Levins, 1968) of single ASVs.

3.4 Results

3.4.1 Seasonality of abiotic factors and harmful algal bloom forming species

The environmental data recorded during six years shows an expected seasonal variation of temperature and salinity. The other biotic and abiotic factors show a less well pronounced seasonality, but the differences between seasons are still visible (Figure 3-1 and Figure 3-2). The HABs were monitored over cell count, which revealed that each of the HAB forming species has a

preferred month in which to bloom in. *K. mikimotoi* bloomed in June or the early summer, while *Chattonella* spp. bloomed later in summer around July. *H. akashiwo* bloomed mostly around April, however in three years a second bloom occurred later in summer (Figure 3-1 and Figure 3-2).

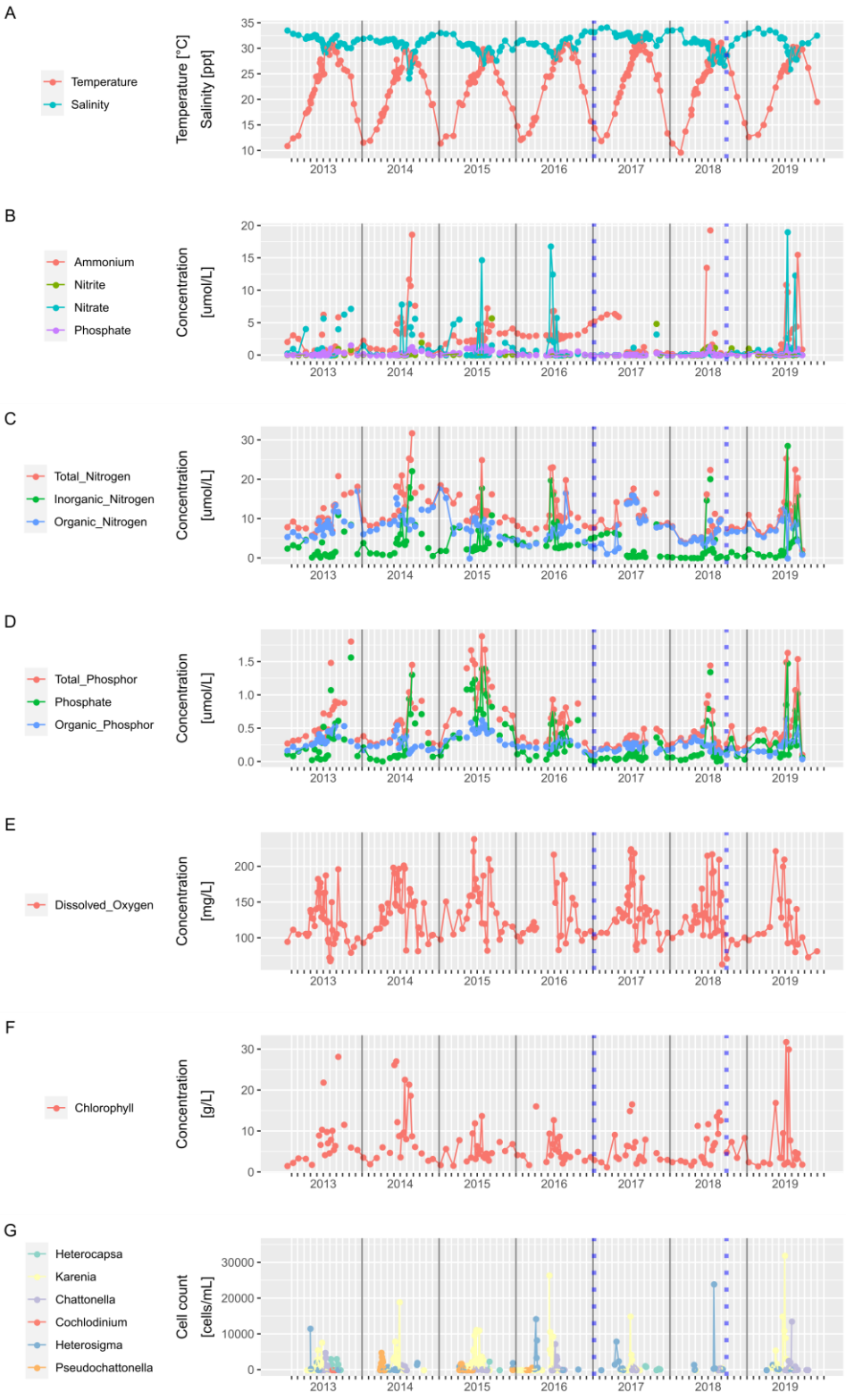

Figure 3-1: Environmental data was collected during six years (2013 to 2019) . Blue dotted lines indicate sampling period during which amplicon sequencing was performed. Figure taken from (Prodinger et al., 2021).

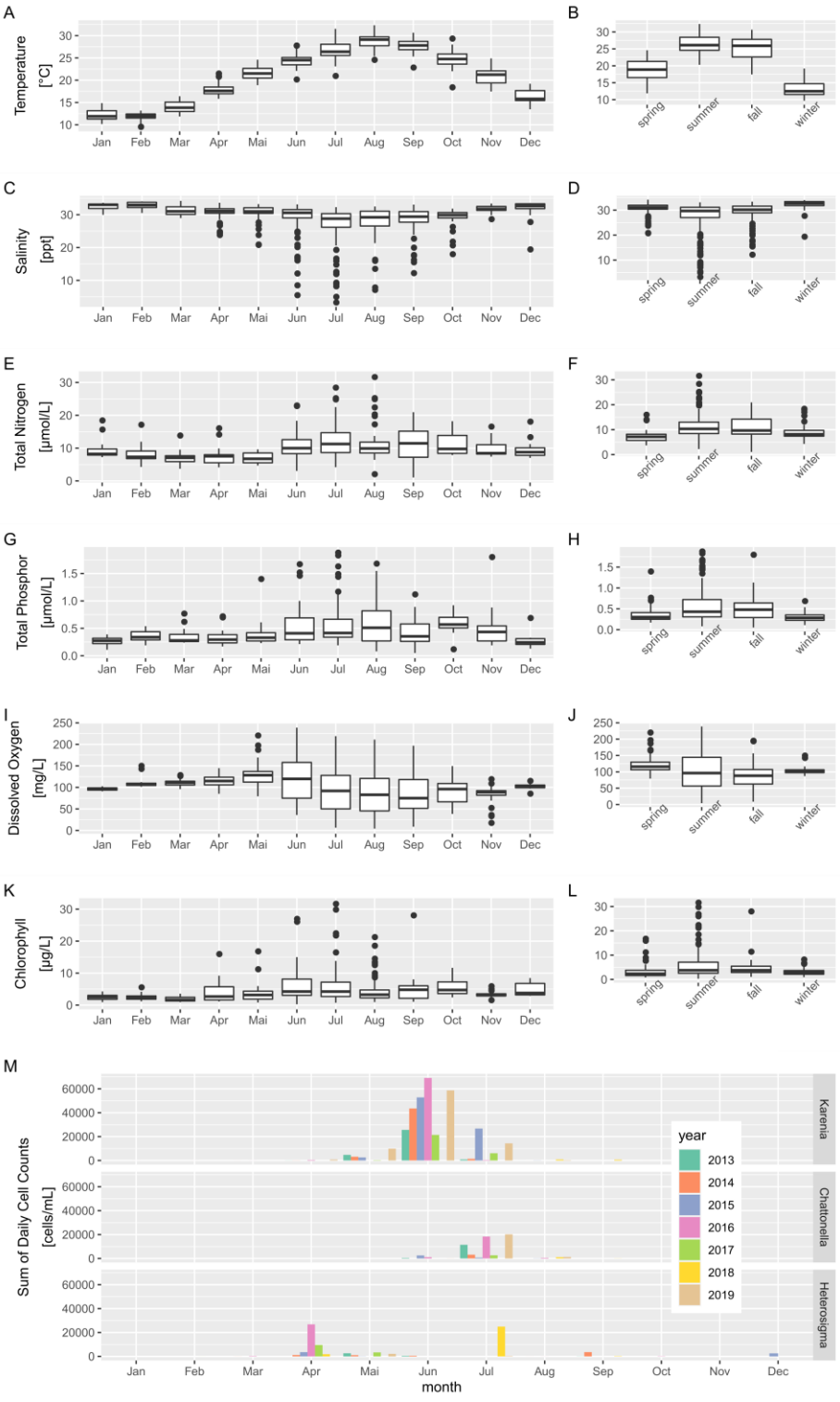

Figure 3-2: Monthly and seasonally bins of environmental data. (A–L) Boxplots show the monthly and seasonally summarized environmental data. (M) The sum of daily cell counts show a typical season for each bloom to occur. Figure taken from (Prodinger et al., 2021).

3.4.2 Generated sequencing data

The amplicons of each microbial group (prokaryotes, small and large eukaryotes, and *Imitervirales*) of each of the 43 different seawater samples were sequenced. This resulted in 172 data sets, some of which yielded no meaningful data or insufficient numbers of reads for analysis. The 4 different microbial groups are NanoPlus (large eukaryotes), PicoNano (small eukaryotes), prokaryotes and *Imitervirales*. NanoPlus (41 data sets) and PicoNano (38 data sets) refer to the eukaryotic communities of two different size fraction, respectively (i.e., >3µm and 0.8µm – 3µm). Prokaryotes (36 data sets) describes the prokaryotic community of the 0.22µm – 0.8µm fraction. *Imitervirales* (41 data sets) is the *Imitervirales* community also of the 0.22µm – 0.8µm fraction. The sum of all communities' reads before read depth normalization contained several thousand ASV (Table 3-1), after read depth normalization NanoPlus had the lowest richness (1,486 ASVs), PicoNano (2,184 ASVs) and prokaryotes (2,383 ASVs) were similar in richness and *Imitervirales* showed the highest richness (3,693 ASVs).

Table 3-1: Overview of the generated datasets for each microbial group. Read numbers and richness of all data sets with more than 8,000 reads are shown. Table taken from (Prodinger et al., 2021).

	Raw reads	Reads	Richness all reads	Richness per samples subsampled	Average richness subsampled
NanoPlus	3,613,399	3,001,267	3,356	52 – 253	120 ± 51
PicoNano	2,692,119	2,180,138	4,559	115 – 410	278 ± 68
Prokaryotes	4,044,427	3,427,158	4,478	141 – 415	275 ± 57
Imitervirales	45,106,746	5,712,580	6,261	108 – 394	246 ± 72

3.4.3 Community composition of different microbes

Most eukaryotic ASV belonged to one of 12 groups excluding "others", the taxonomic annotation levels of these groups were between phylum level and genus level. In the NanoPlus dataset these 12 groups contributed to 98.7% of total reads and the most abundant groups are Metazoa (28.6%), Diatomea (25.1%) and Dinoflagellata (including Dinophyceae: 22.5%). All other

groups contributed less than 7% of total reads. In the PicoNano these 12 groups contributed 88.8% of total reads and the most abundant groups were “other Alveolata” (Alveolata other than Dinoflagellata; 17.5%), Metazoa (17.2%), and “other Stramenopiles” (Stramenopiles excluding both Diatomea and Raphidophyceae, 12.4%). The other 9 groups produced less than 8% of total reads each.

ASV assigned to HAB species were grouped in their family or genus as Kareniaceae, *Chattonella*, and *Heterosigma* in Figure 3-3 and were detected in both size fractions. Even though several ASVs per HAB species were present (i.e., *Karenia* sp.: 24 ASVs, *Chattonella* sp.: 29 ASVs, and *H. akashiwo:* 4 ASVs), one ASV dominated each species with at least 88% of reads in both the NanoPlus and the PicoNano datasets. In total the HAB species contributed less than 6% of the NanoPlus and less than 8% of the PicoNano datasets, however during blooms the HAB species produced up to 80.2% (Kareniaceae), 30.6% (*Chattonella*) and 16.6% (*Heterosigma*) in either size fraction.

Nearly all prokaryotic reads (98.4%) belonged to one of six phyla, however the three most dominant phyla contributed 89.3% of total reads. These three phyla were Proteobacteria, Bacteroidetes, and Cyanobacteria (excluding chloroplasts). Proteobacteria was the most abundant phylum (53.6% of total reads), with Alphaproteobacteria being the most abundant class (72.1 % of Proteobacteria reads). Bacteroidetes contributed 20.4% of total reads most of which belonged to the classes Bacteroidia (91.4%). Cyanobacteria (15.3% of total reads) were nearly exclusively Oxyphotobacteria (99.9% of Cyanobacteria reads). The three less abundant phyla (i.e., Verrucomicrobiae, Actinobacteria, and Planctomycetes) contributed less than 5% of total reads each.

Most *Imitervirales* ASV (91.1% of ASVs, 96.0% of *Imitervirales* reads) belonged to 1 of 14 clades. The clades with the most total reads assigned were clade 6 (24.0%), clade 3 (20.7%), and

clade 5 (17.2%). The other clades contributed less than 9% each. Half of all clades (7 of 14 clades) contained known *Imitervirales* that have been isolated or of which the whole genome was previously described. The most abundant of these clades were clade 3 (including the Organic lake phycodnavirus 1 and 2 (Zhang et al., 2015)), clade 1 (7.8% of total reads, including several *Imitervirales* infecting haptophytes (Gallot-Lavallée et al., 2015; Johannessen et al., 2015; Santini et al., 2013)), and clade 10 (2.4% of total reads, including the recently isolated ChoanoV1 (Needham et al., 2019)). The other clades including known *Imitervirales* (clade 8, clade 9, clade 13, and clade 14) contributed less than 2% of total *Imitervirales* reads.

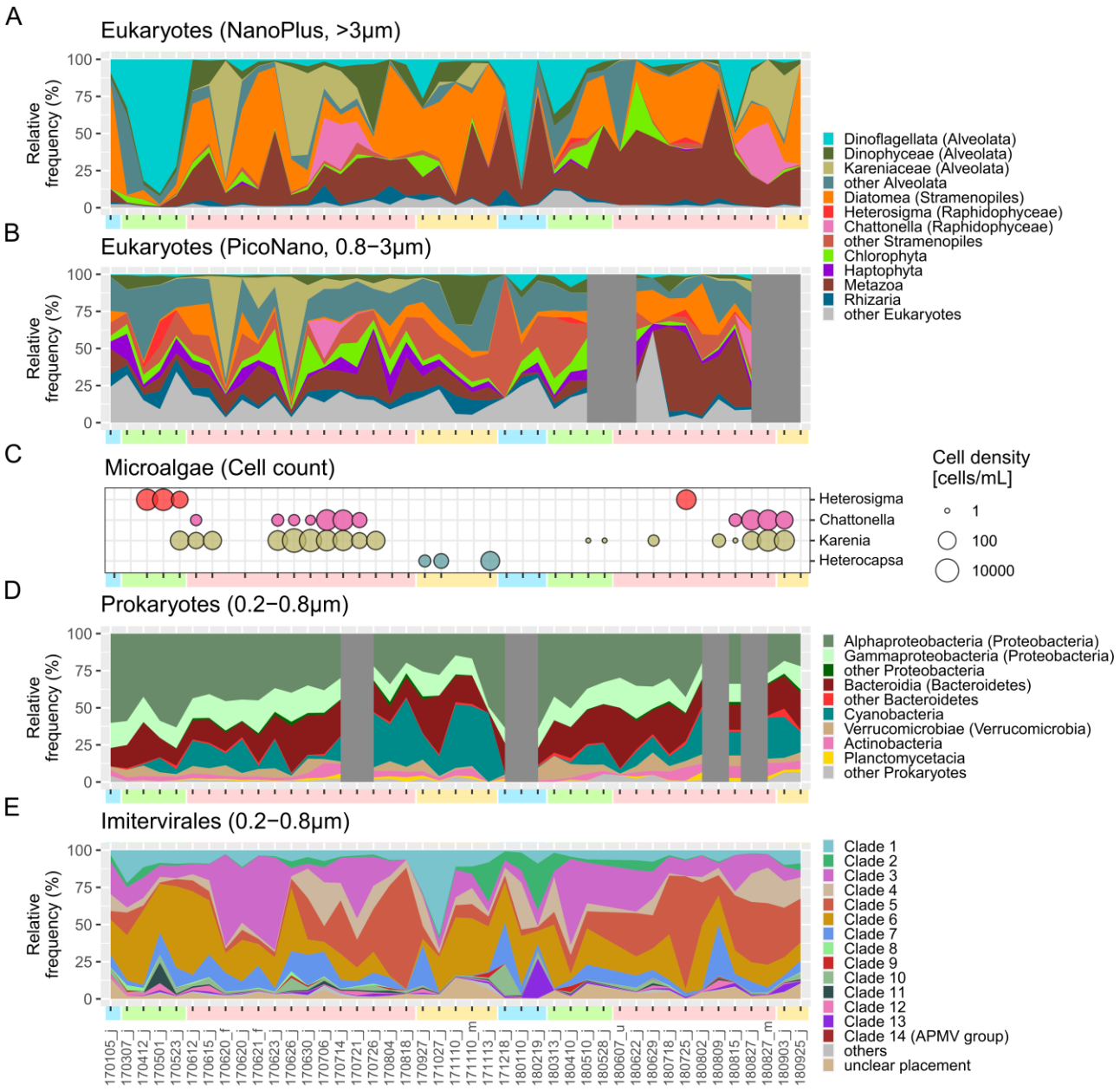

Figure 3-3: Community compositions of different microbes. A, B) The eukaryotic community of two size fractions is similar in terms of detected eukaryotic groups, but the most abundant eukaryotic groups are distinct. C) The cell count shows bloom periods. D) The prokaryotic community was dominated by the same phyla over the sampled period. E) Different abundant Imitervirales clades were dominant in the sampled period. Figure taken from (Prodinger et al., 2021).

3.4.4 The seasonal change of microbial ASV communities

The nonmetric multidimensional scaling (NMDS) of the dissimilarity between samples showed a clear seasonal cycle of the communities. Most samples of the same month clustered together with themselves or samples of the preceding or following month in every community (Figure 3-4). This seasonal change of community composition showed statistically significant correlation between any two microbe groups (Mantel test > 0.74, $p < 10^{-4}$).

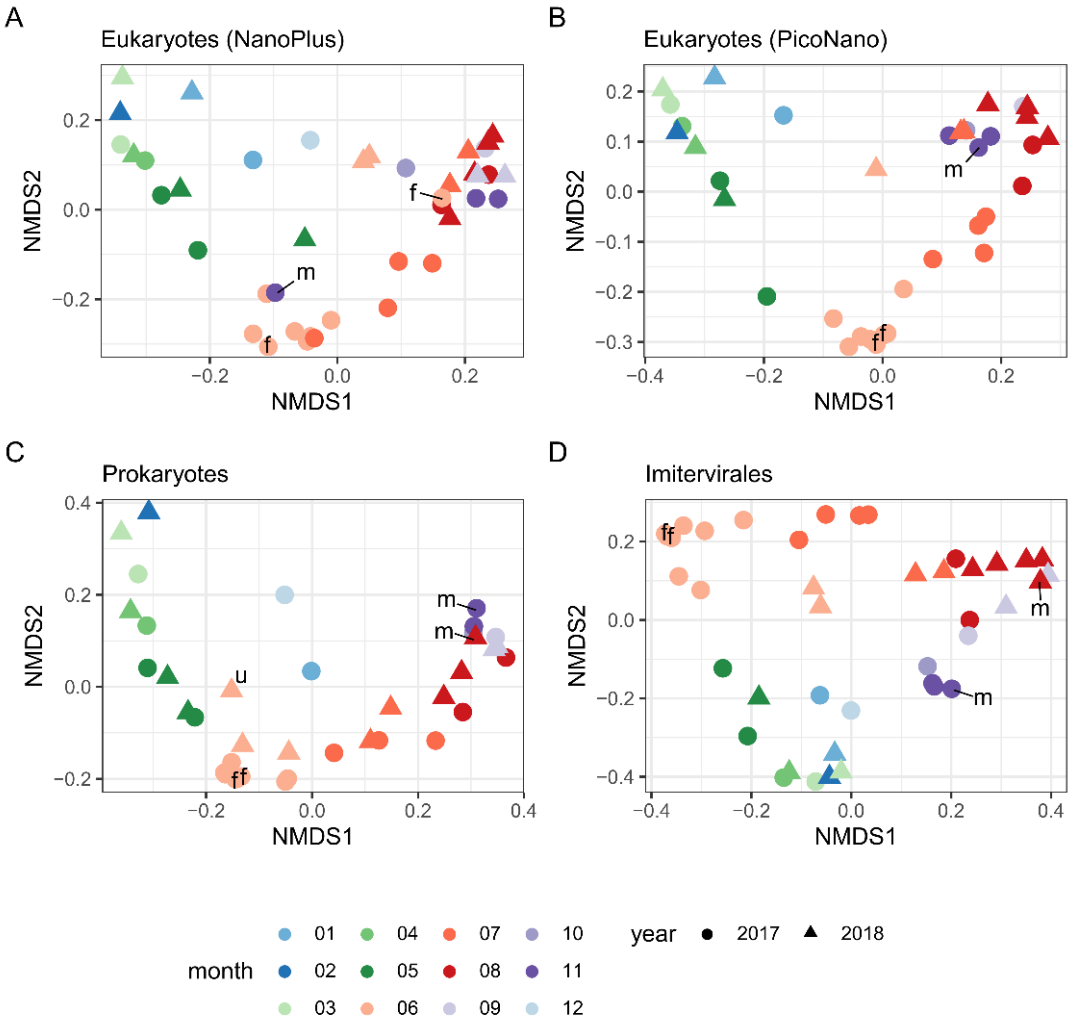

Figure 3-4: Two dimensional visualization of community dissimilarities. Each community showed a seasonal cycle. The stress of the dissimilarity nonmetric multidimensional scaling (NMDS) was below 0.12 for each plot. Figure taken from (Prodinger et al., 2021).

The average dissimilarity of the different microbe communities of the same season ranged from 0.58 to 0.78 and was lower (p <10^{-16}) than for opposite seasons, which ranged from 0.77 to 0.94. Opposite seasons are either spring and fall or summer and winter. The dissimilarity between communities of summer and winter was not higher than dissimilarities between spring and fall communities (Figure 3-5), in fact the prokaryotic spring and fall communities were more dissimilar than the summer and winter communities (p <10^{-6}).

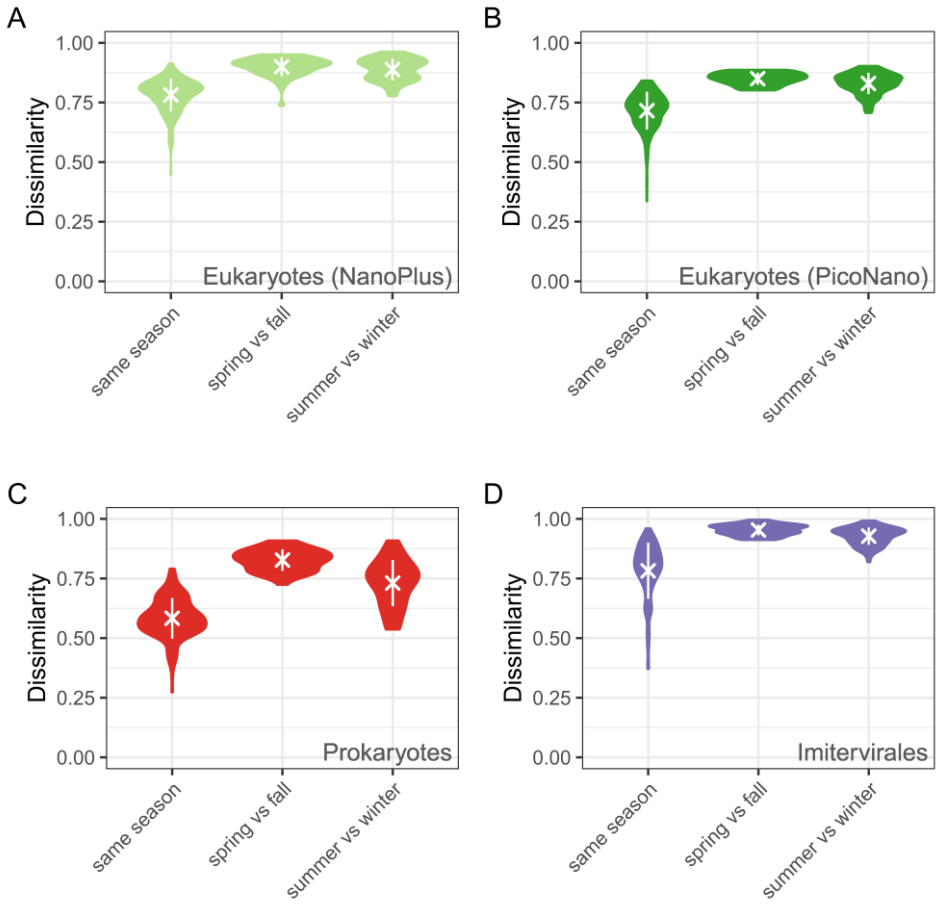

Figure 3-5: Dissimilarities of microbial communities of the same and opposite seasons. Figure taken from (Prodinger et al., 2021).

3.4.5 Distinct seasonal patterns of microbial ASV communities

Aside from communities from opposite seasons being more dissimilar than communities of the same season, I found that communities that are 30 days apart (i.e., one month) or 365 days apart (i.e., one year) were less dissimilar than communities that were 182 days (i.e., six months) apart. The dissimilarity at these time ranges (30, 182, and 365 days) will henceforth be called D30, D182 and D365, respectively. This yearly dissimilarity curve was similar for each community (Figure 3-6 A-D), however the level of dissimilarity was different among the four microbial communities (Figure 3-6 E-G, $p < 10^{-6}$), only D30 and D365 of the *Imitervirales* and NanoPlus communities were not statistically significantly different ($p > 0.2$). The D30, D182 and D365 of the prokaryotic community was lower than for the other communities. The D30, D182 and D365 of both the *Imitervirales* and NanoPlus community was higher compared to the prokaryotic community and the PicoNano community.

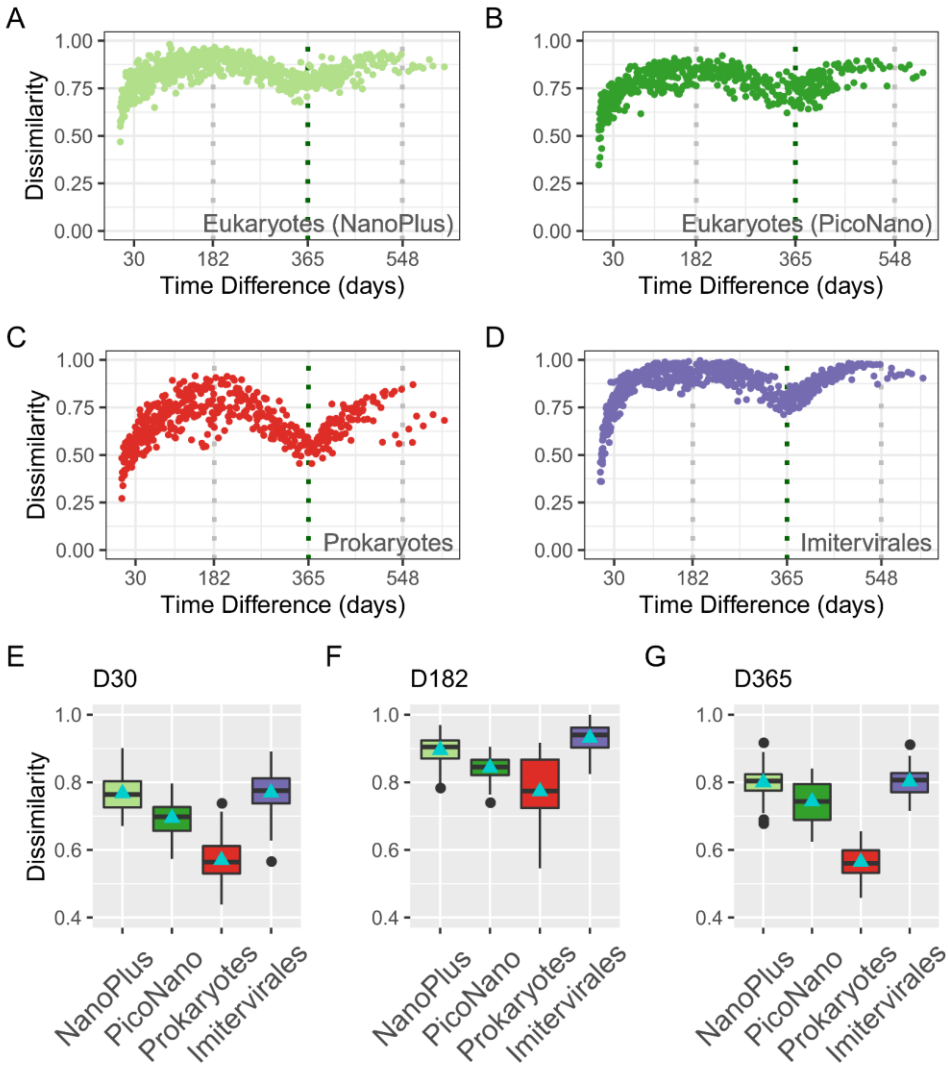

Figure 3-6: The dissimilarity between different microbial communities is highest for communities that are six months apart. The dissimilarity of communitis sampled 6 months apart is different for eukryotes, prokaryotes and viruses. Figure taken from (Prodinger et al., 2021).

3.4.6 Differences in ASV recurrence and persistence

To better understand the changes in community dissimilarity, the temporal patterns of abundant ASVs were also analyzed. The most abundant ASV of each microbe group showed different maximum relative abundances (Figure 3-7). The prokaryotic ASVs reached a comparatively low maximum relative abundance of 31%, while the eukaryotic and *Imitervirales* ASV reached up to 70%. In short, the different microbial communities' most abundant ASVs' relative abundance profiles were different. *Imitervirales* ASVs reached high relative read

frequencies, but they were found in few samples. The most abundant eukaryotic ASVs also reached high relative abundances similar to *Imitervirales* ASVs. The most abundant prokaryotic ASVs were present in many samples but did not reach high relative abundances (Figure 3-7).

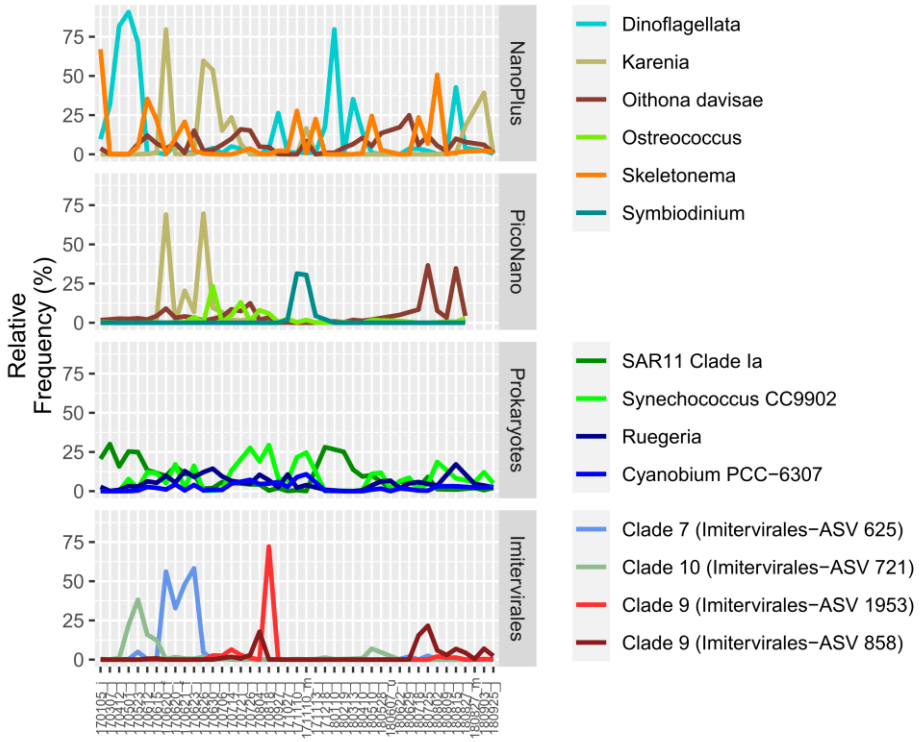

Figure 3-7: The four most abundant ASVs of each microbe group. The taxonomic asignment of the different ASVs is shown. Eukaryotic ASVs that were one of the most abundant ASVs in both size fractions were a Karenia*–ASV and an* O. Davisae*–ASV. Imitervirales ASV tended to be not present in many samples, but if found they had high relative read counts. Figure taken from (Prodinger et al., 2021).*

Another difference between the eukaryotic, *Imitervirales* and prokaryotic communities is the ASV recurrence (i.e., how frequently an ASV was found). Even though *Imitervirales* had more ASVs with high relative abundance, most *Imitervirales* ASVs (94.6%) were found in fewer than 20 samples. The recurrence was different between the microbe groups (Figure 3-8A and B). *Imitervirales* showed low recurrence. Eukaryotic ASV recurred more often than *Imitervirales* and had a similar recurrence in both size fractions. The prokaryotic ASVs generally reoccurred frequently. This higher recurrence and persistence was also reflected in the Levins' niche breadth of ASVs in the communities. In this study, ASVs with a high Levins' niche breadth (>6) are

considered “generalists”, while ASVs with a low Levins’ niche breadth (<2) are considered “specialists”. Each community contained generalists and specialists, however the ratio of specialist ASV to generalist ASV was different. Nearly half (45%) of the prokaryotic community were ASVs with a Levins’ niche breadth >6, the other communities contained >20% generalists (PicoNano eukaryotes: 18%, NanoPlus eukaryotes: 16%, and *Imitervirales*: 5%) (Figure 3-8C). The *Imitervirales* community consisted mostly of specialists (51%), while cellular communities contained >40% generalist ASVs (NanoPlus eukaryotes: 38%, PicoNano eukaryotes: 32%, and prokaryotes: 18%) (Figure 3-8C).

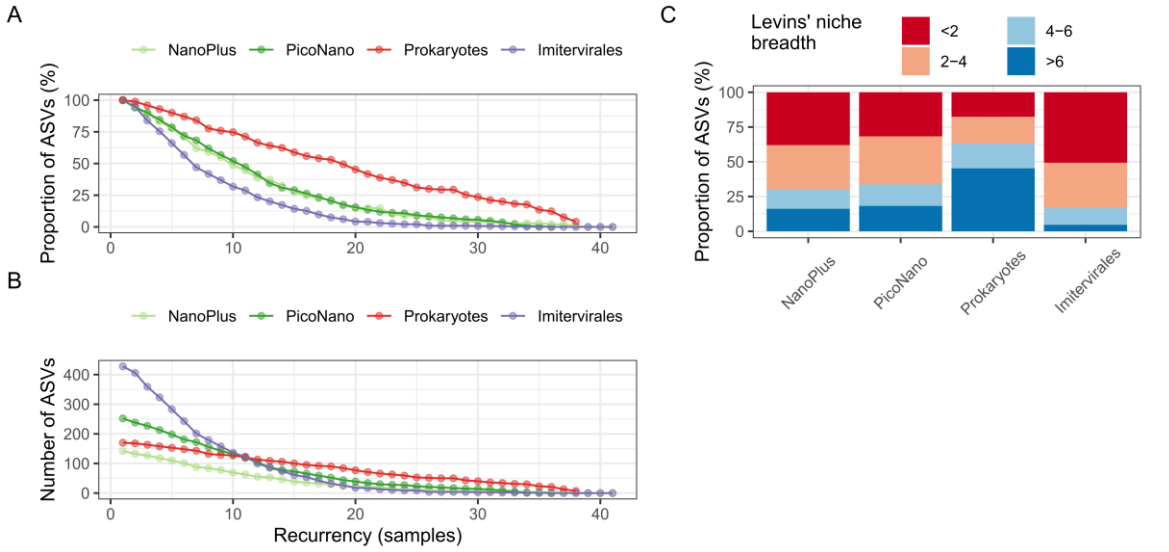

Figure 3-8: The recurrence of the most abundant ASVs of each microbe group and Levins’ niche breadth analysis results. (A) The percentage of abundant ASV that reoccurred was highest for Prokaryotes and lowest for Imitervirales. (B)The plot of absolute numbers of recurring ASVs shows that Imitervirales ASV tend to be present in only few samples. (C) The prokaryotic community included many ASVs with larger Levin’s niche breadth values, while eukaryotic and viral ASVs tended to produce lower Levin’s niche breadth values. All ASV producing over 1% relative abundance in at least one sample were used for this analysis. Figure taken from (Prodinger et al., 2021).

3.4.7 ASV co-occurrence among the different communities and correspondence analysis

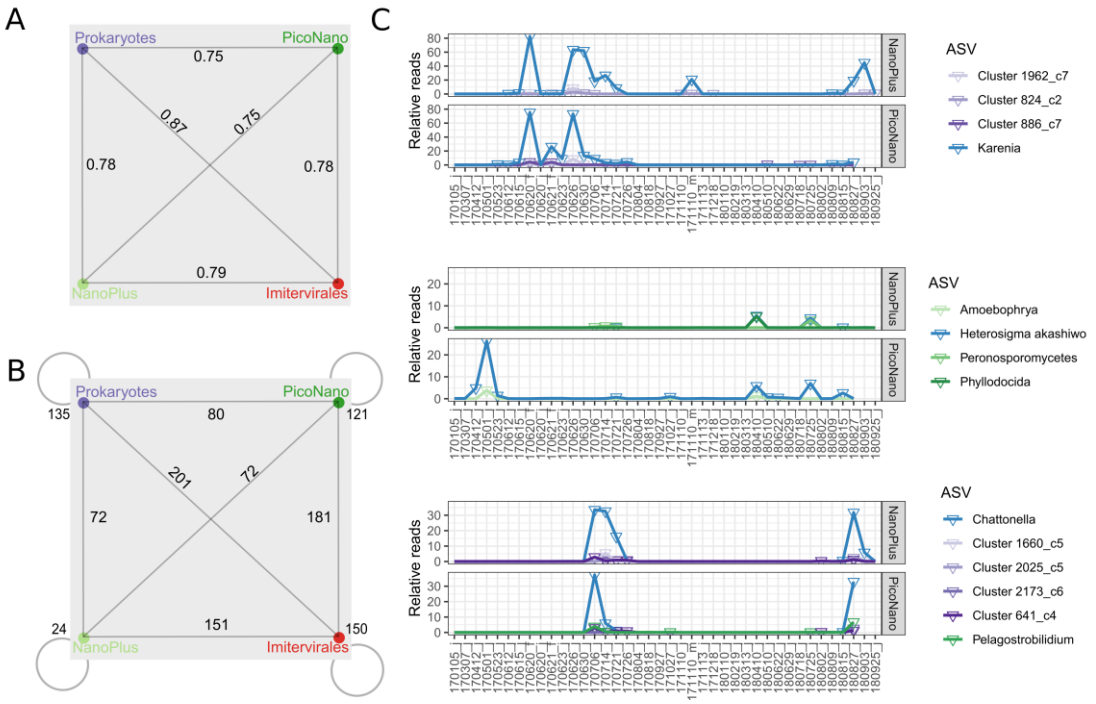

Figure 3-9: Strong connectivity of the microbial communities. (A) The Mantel test showed that the changes within the different communities are strongly correlated with one another. (B) The network analysis also showed that edges between communities were as frequent as edges inside a community. (C) Several major Imitervirales *and major eukaryotic ASV co-occurred with HAB species. Triangles show relative abundance over 0.1% (i.e.: whether the ASV was considered present or not). HAB forming ASVs are shown in blue, Imitervirales ASVs in purple and eukaryotic ASVs in green. Figure taken from (Prodinger et al., 2021).*

The Mantel test found a, strong correlation between the different microbial communities of at least 0.75 (p <10^{-4}, Figure 3-9A). A co-occurrence network that included ASVs of both of the eukaryotic communities (NanoPlus: 699 ASVs, PicoNano: 1,285 ASVs), the prokaryotic community (958 ASVs) and the *Imitervirales* community (1,771 ASVs) showed high intra–community (edges inside a given community) and inter–community (edges between members of different communities) connectivity (Figure 3-9B). Additionally, a co-occurrence network was created using the ASV tables of abundant eukaryotic ASVs (159 ASVs from the PicoNano data, or 110 ASVs from the NanoPlus data) and abundant *Imitervirales* ASVs (264 ASVs) to detect co-occurrence (Figure 3-9C). An ASV assigned to *Karenia* sp. had two neighbors in the NanoPlus network and two neighbors in the PicoNano network. All of the neighbors were *Imitervirales* ASV and one of the ASV

was the same in both networks. The *Chattonella* sp. ASV had three neighbors in the PicoNano network, they were two *Imitervirales* ASV and one eukaryotic ASV assigned to Pelagostrobilidium spp. (Alveolata, a genus containing nine species of mixotrophic Ciliates (Agatha et al., 2005)). The neighbors of the *Chattonella* sp. ASV in the NanoPlus network were three *Imitervirales* ASVs, one of which was also a neighbor in the PicoNano network. An ASV assigned to *H. akashiwo* had two neighbors in the NanoPlus network (an ASV assigned to the order Phyllodocida (Bilateria, a marine worm) and an ASV assigned to the class Peronosporomycetes (Stramenopiles, order: Peronosporales , fungi like plant parasites (Phillips et al., 2008)) and a single neighbor assigned to Amoebophrya (Synidiales, a parasite of dinoflagellates (Kim et al., 2008)) in the PicoNano network.

The correspondence analysis of the microbial communities as response variable and abiotic factors as explanatory variable showed that the abiotic factors had poor explanatory power over most communities (Table 3-2). Temperature (cellular communities) and total Nitrogen (viruses) were found to be the most important parameters. The abiotic factors explained <15% of the viral (4%) and both eukaryotic communities (NanoPlus: 12% and PicoNano: 9%). The prokaryotic community was explained better by abiotic factors (42%).

Table 3-2: The cumulative adjusted R^2 of abiotic parameters and the microbial communities. The table shows different abiotic parameters, ranked by their cumulative adjusted R^2 (left number) for each microbial community. The p-value is shown on the right. The highest possible cumulative adjusted R^2 is printed bold. When statistically insignificant abiotic parameters are used to predict the microbial community, the cumulative adjusted R^2 is decreased (lower rows of the table). Table taken from (Prodinger et al., 2021).

Rank by influence	Eukaryotes NanoPlus	Eukaryotes PicoNano	Imitervirales	Prokaryotes
1	Temperature (0.10; 0.004)	Temperature (0.05; 0.017)	Total Nitrogen (0.02; 0.092)	Temperature (0.35; 0.001)
2	Total Phosphor (**0.12**; 0.14)	Total Nitrogen (0.08; 0.048)	Temperature (0.035; 0.10)	Salinity (0.41; 0.004)
3	Dissolved Oxygen (0.12; 0.4)	Dissolved Oxygen (**0.09**; 0.3)	Salinity (**0.04**; 0.3)	Total Nitrogen (**0.42**; 0.4)
4	Salinity (0.10; 0.9)	Total Phosphor (0.07; 0.6)	Total Phosphor (0.04; 0.5)	Total Phosphor (0.41; 0.6)
5	Total Nitrogen (0.07; 0.9)	Salinity (0.06; 0.7)	Dissolved Oxygen (0.02; 0.6)	Dissolved Oxygen (0.40; 0.7)

3.5 Discussion

The eukaryotic communities consisted mostly of Metazoa, Alveolata, and Stramenopiles, which are all typically associated with coastal areas (Chen et al., 2017; Gran-Stadniczeñko et al., 2019b; Martin-Platero et al., 2018). However, The microbial eukaryote communities of Uranouchi Inlet were different at lower taxonomic ranks from those in other coastal areas of Northeast America (Martin-Platero et al., 2018), Southeast China (Chen et al., 2017), and Southern Norway (Gran-Stadniczeñko et al., 2019a). A unique aspect of the Uranouchi Inlet are the frequent and consistently occurring algal blooms of different species that have been observed through cell counting as well as metabarcoding (Figure 3-3). These HAB species are known to form blooms in different coastal areas across Asia (Sakamoto et al., 2020). The prokaryotic community consisted of phyla that are also found in other coastal areas such as the Massachusetts Bay and the Mediterranean (Martin-Platero et al., 2018; Sakami et al., 2016; Santi et al., 2019). Especially the most abundant prokaryotic clade which belonged to the Alphaproteobacteria group “SAR 11 Ia” is a typical coastal prokaryote. SAR 11 itself is known for high relative abundance in the eutrophic

ocean (Chow et al., 2013; Eiler et al., 2009) and the "Ia" subclade is a coastal ecotype of the SAR 11 clade (Stingl et al., 2007). The *Imitervirales* community showed comparable richness (>6000 ASVs) to previously published *Imitervirales* community studies of *Imitervirales* found in different coastal, freshwater, and brackish water environments (Li et al., 2019) and in the open ocean (Mihara et al., 2018).

All of the studied microbial communities showed a pronounced seasonal cycle, including the *Imitervirales* community (Figure 3-4). Similar observations have been reported for eukaryotes (Giner et al., 2019; Gran-Stadniczeñko et al., 2019b) and bacteria (Chafee et al., 2018; Ward et al., 2017). Reports about the seasonality of *Imitervirales* are fewer. A bimonthly sampling study reported approximate seasonal clustering of large DNA virus communities in the Arctic Ocean (Sandaa et al., 2018). However, a monthly sampling study reported a lack of seasonality of the *Imitervirales* community in a coastal region of Norway (Gran-Stadniczeñko et al., 2019b). The pronounced seasonality of *Imitervirales* reported here is however not unexpected, given that smaller viruses also show seasonality (Chow and Fuhrman, 2012; Ignacio-Espinoza et al., 2020; Pagarete et al., 2013).

I found that spring *vs.* fall communities were as different as summer *vs.* winter communities when I compared the different communities formed during their seasonal cycle. Previous studies also conducted a comparison of microbial communities sampled several months apart. They showed that eukaryotes (Giner et al., 2019), prokaryotes (Fuhrman et al., 2015), and small viruses (Ignacio-Espinoza et al., 2020) were most dissimilar if communities that were sampled six months apart were compared. Community dissimilarities between spring *vs.* fall and summer *vs.* winter were not reported in these studies. A previous study reported the importance of water temperature on community composition (Sunagawa et al., 2015). The forward selection analysis conducted here also showed that temperature had the most influence out of all abiotic

parameters on the different cellular communities and that temperature was the second most influential abiotic parameter for the *Imitervirales* community. Water temperature was comparable between fall and spring in the Uranouchi Inlet. Furthermore, samples taken in May and November showed similar environmental conditions like water temperature and nutrient concentrations. However, the communities in these two months were as dissimilar as communities sampled during winter and summer. This high dissimilarity between spring and fall communities should not be surprising considering that spring is preceded by a winter (the coldest season) and fall is preceded by summer (the warmest season).

These observations underscore the importance of the order of the changes in environmental conditions on the sequential changes of microbial communities. The data collected in Uranouchi Inlet suggests that the dynamics of microbial communities are not determined exclusively by abiotic factors at a given period. Instead, they are driven by a Markovian process whereby the current biotic state in combination with the current abiotic conditions substantially influence future biotic states. This pattern is clearly seen in macro-ecosystems; for example, a forest looks different in spring and fall even on days with similar weather conditions. Microbial communities with a much higher turnover rate (one to several days) (Martin-Platero et al., 2018) are no exception to this rule. A previous study on oceanic microbial communities suggested that biotic factors (such as subsets of communities) are better predictors of microbial community composition than abiotic factors (Lima-Mendez et al., 2015). This is reminiscent of the classical notion that the sum of biotic and abiotic factors forms the environment that affects the dynamics of individual microbial populations. I suggest that the microbial communities at a given time point strongly influence the formation of the next generation of communities in the Uranouchi Inlet as well as many other environments. A Similar idea has been previously proposed (Needham and Fuhrman, 2016; Nelson et al., 2008). Such a Markovian process suggests that microbial

communities leave a historical trace, like a “memory” of the original community structure. This “memory” influences the structures of successive communities and its influence may last for a week up to months.

The dissimilarity of samples that were taken approximate one year apart was on average higher than 0.5, which meant that around half of the community that was present one year prior, was not detected. Previous studies that observed eukaryotic communities (Giner et al., 2019), prokaryotic communities (Chow et al., 2013; Cram et al., 2015; Fuhrman et al., 2015) and viral (Ignacio-Espinoza et al., 2020) communities also described a similar trend that continued for several years. These previous studies’ time series sampling continued for a long time (5-10 years) and some found that fewer and fewer OTUs continuously returned each year for 4-5 years (Chow et al., 2013; Fuhrman et al., 2015; Ignacio-Espinoza et al., 2020). According to these observations, it seems that the marine microbial community structure never perfectly recovers over time. I visualized this concept schematically in Figure 3-10A. This irreversible progression of microbial communities implies a significant role of the “invisible” rare biosphere in the formation of a new community structure through their capacity to produce new abundant microbes over time (Giner et al., 2019; Ignacio-Espinoza et al., 2020). The rare biosphere is undetectable, even with the state-of-the-art metabarcoding approach that was used in this study. This limitation comes from sequencing depth, as well as the heterogeneity in time and space of community compositions around sampling locations. Size fractionation can also affect the detection of microbes, especially if their sizes change depending on their life stages or growth conditions. These technical limitations in uncovering the rare biosphere combined with the capacity of microbes to replace abundant populations over time may explain why microbes cannot completely recover their community structures. In addition to the rare biosphere, migration (Yoshida et al., 2018) and evolution may also have an effect on the emergence of new ASVs. Interestingly, previous studies

on bacterial communities found that their dissimilarity increased during the first four years of analysis, after which its basal level remained constant (overlaid with a seasonal sinusoidal pattern) (Fuhrman et al., 2015). This observation suggest that the influence of the "memory" may last for over a year (Figure 3-10A).

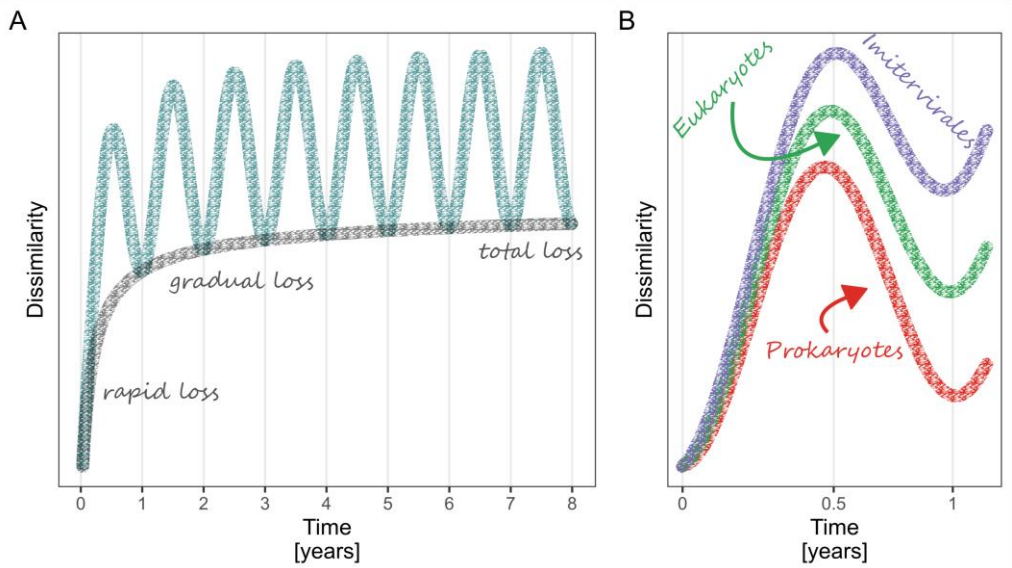

Figure 3-10: Visual representation of the community memory model. (A) Several studies reported a rapid loss of community members in the first year, followed by a gradual loss, and finally plateauing in constant sinusoidal community dissimilarity curve. (B) A schematic representation of the differences in ASV turnover, maximum dissimilarity, and ASV recurrence of the different communities observed in this study. Figure taken from (Prodinger et al., 2021).

In this study I showed that the degree and speed of changes in the community structure are different depending on which microbial community is observed. The D30, D182, and D365 values were lowest for prokaryotes, which have been observed to show a drift in dissimilarity for several years in previous studies (Fuhrman et al., 2015). In contrast *Imitervirales* showed the highest D182. The D30 and D365 were also highest for *Imitervirales* and NanoPlus eukaryotes. This suggests that *Imitervirales* and eukaryotes (PicoNano and NanoPlus) have a higher ASV turnover and shorter community memory than prokaryotes. A high resolution sampling series study also reported a higher OTU and species turnover for eukaryotes than for prokaryotes (Martin-Platero et al., 2018).

It is important to address the known drawbacks of my sampling and amplicon sequencing method (e.g., primer bias) and the possible influence of these drawbacks on ASV community analysis and the comparison between data generated by different primers. The influence of primer bias and duplicate genes was reduced by using a binary matrix (instead of relative abundance weighted matrix) for community dissimilarity analysis. Additionally, the commonly used nucleotide identity threshold based OTU clustering (e.g., at 97% sequence similarity) was replaced with exact amplicon sequence variants, since ASVs were found to be more sensitive than similarity threshold set by the user (Callahan et al., 2017). Another possible criticism is that the long persistency and lower community dissimilarity of prokaryotes might be a result of a lower resolution. However, this is inconsistent with the low relative abundances of even the most abundant prokaryotic ASVs. If the prokaryotic marker gene region had a lower resolution, higher relative frequencies and a lower richness compared to eukaryotes should be observed. Another possible source of bias is the sampling strategy. I specifically targeted certain size fractions with certain primers. This was meant to concentrate the DNA of differently sized microbes of several liters of seawater. For instance, I analyzed the prokaryotic community of the 0.22µm – 0.8µm fraction, however by separating microbes by size I missed out on particle attached and large prokaryotes. The same is true for the *Imitervirales* community. *Imitervirales* that are attached to or inside cells cannot be detected by analyzing the 0.22µm – 0.8µm fraction. However, the results of the analyzed data agrees with previous studies on many respects, therefore it is likely that despite these experimental biases, the data of this study is able to show differences in the dynamics of the targeted microbial groups.

The dynamics of dominant ASVs of each of the different microbial communities (eukaryotes, prokaryotes, and viruses) were found to differ in recurrence, persistence, and relative abundance. For instance, several of the most abundant eukaryotic ASVs that were taxonomically assigned to HAB species (i.e., *Chattonella* spp., *K. mikimotoi*, and *H. akashiwo*) showed typical

boom-and-bust dynamics. A large proportion of the *Imitervirales* ASVs showed HAB species-like dynamics. This means that many ASVs were not persistently present and if they emerged they formed peaks in their relative abundance. Viral ASVs and HAB species ASVs both showed a boom-and-bust strategy. They became abundant quickly but only if certain conditions were met. These ASVs are therefore specialists (or opportunists). This high level of specialization of the viral community can be explained by their dependence on the eukaryotic community for survival. The viral community has to adapt at least as quickly as the eukaryotic community to not lose the evolutionary arms race. Additionally, viruses have to compete among themselves, since it has been shown that several *Imitervirales* compete for the same host (Baudoux and Brussaard, 2005). Few *Imitervirales* ASVs that showed long persistence and reoccurred were found. A recent study suggested that *Imitervirales* infecting certain haptophytes have a long latent period (Blanc-Mathieu et al., 2021). However, as mentioned above, the sample strategy of this study limited the detection of the metabarcoding analysis to free *Imitervirales* virions. Viruses with long latent periods would not be abundantly present in the size fraction I used to study the *Imitervirales* community.

A co-occurrence analysis was used to investigate whether *Imitervirales* were associated with bloom forming algae (Figure 3-9). The *Chattonella* sp. ASV co-occurred with several *Imitervirales* ASV of neighboring clades and a single mixotrophic eukaryote ASV (genus *Pelagostrobilidium*) and the *Karenia* sp. ASV co-occurred exclusively with *Imitervirales* ASVs (Figure 3-9). *H. akashiwo* is known to be predated by Phycodnaviridae (Nagasaki and Yamaguchi, 1997; Nagasaki et al., 1994). The reason why the *H. akashiwo* ASV co-occurred exclusively with other mixotrophic or parasitic eukaryotes might be that it simply is not possible to detect the large DNA viruses infecting *H. akashiwo* with MEGAPRIMER. While co-occurrence is not strong evidence for a

virus–host pair (Edwards et al., 2016), this co-occurrence analysis may be a starting point for further virus–host studies.

Here I showed the pronounced seasonality of the different microbial communities. While this has been observed for cellular microbes, a pronounced seasonality of *Imitervirales* has not been described before. I argued that the seasonality of the different microbial communities do not exclusively depend on the current environmental conditions. I found that it is more likely that community changes originate from a Markovian like process involving the composition of previous communities and also abiotic parameters. This was highlighted the interaction network between communities and the strong correlation of changes between the different microbial communities, while abiotic factors were poor predictors of the microbial communities' composition and dynamics. Despite striking similarity of the seasonal cycle of all four microbe groups, the dissimilarities of the four microbial groups showed significant differences when compared at different time intervals (i.e., D30, D182, and D365). These differences were further explored with the recurrence of abundant ASVs and their abundance profiles. I also argued that this study and previously published findings support a figurative microbe "community memory" that differs in length depending on microbe group (Figure 3-10B). The relative abundance profiles were further used to assign an opportunistic behavior to *Imitervirales* ASV and to match co–occurring eukaryotes and *Imitervirales* ASVs. The results of this study provide a first step to understanding the complex interactions of different microbe groups in the Uranouchi Inlet and a better understanding of the survival strategies of *Imitervirales*.

4 Chapter 4: Conclusions and perspective

4.1 Summary

In the first step of this work, I optimized a metabarcoding method for *Imitervirales* by building on the previous work of Li et al. (Li et al., 2018). I found that mixing primer pair cocktails in batches of 10 heavily reduced the necessary analysis time and that gel extraction was a more reproducible amplicon clean up method than the previously used clean up kit. The *Imitervirales* community analysis method presented in this study offers a robust and inexpensive analysis workflow to survey the present *Imitervirales* in any aquatic sample. Thereby I improved the workflow and made MEGAPRIMER a suitable option for community analysis of *Imitervirales*.

In the next step of this work, I utilized the new MEGAPRIMER metabarcoding workflow to analyze 43 coastal seawater samples, which were taken during 20 months. I employed metabarcoding for different microbial communities of different size fractions. The analysis of the time series data revealed that *Imitervirales* like the other microbial communities possess a pronounced seasonal cycle. Previous studies already tried to show the seasonality of large viruses before (Sandaa et al., 2018) but could not confirm it most likely because the primer set only targeted a small fraction of the algal virus community (Gran-Stadniczeñko et al., 2019b). The analysis of other microbial communities (i.e., eukaryotes and prokaryotes) showed that all communities shared this seasonality and hence their community changes correlated. However while this seasonal cycle was similar, the turnover rate, persistence, and recurrence of individual community members differed from one community to another. The viral community especially showed low persistence and recurrence, indicating a rapid turnover of its members.

To describe these changes of the different communities I introduced the “community memory” hypothesis. It has been shown that community compositions can predict the abundances of community members (Lima-Mendez et al., 2015) and even influence the formation

of future communities depending on the current composition of a microbial community (Teeling et al., 2016). The community memory hypothesis suggests that this influence fades over time (Fuhrman et al., 2015) and that the "length" of the community memory varies among different microbial communities (i.e., prokaryotes, eukaryotes and large viruses). The community memory hypothesis was defined in chapter three and the described differences of the ASV community memory is the most important outcome of this work. This hypothesis could also provide an explanation for certain traits of the *Imitervirales* community, like why *Imitervirales* are more diverse than other communities (Mihara et al., 2018) as discussed below.

4.2 Discussion

This work found that *Imitervirales* were more diverse than other microbes. This has been reported previously using metagenomics (Hingamp et al., 2013; Mihara et al., 2018), as well as deep sequencing of MEGAPRIMER amplicons (Li et al., 2018). Yet, why *Imitervirales* are so diverse is still unknown. The simplest possible answer is that mutations occur frequently during replication, in part as a response to the high evolutionary pressure on viruses to adapt to their hosts (Sanjuán and Domingo-Calap, 2016). However this may not apply to every viral community or population, as viral communities and their genomic repertoire can vary considerably. In contrast to other smaller viruses, *Imitervirales* were shown to encode many functional DNA repair genes (Blanc-Mathieu and Ogata, 2016), they are therefore likely less prone to mutations. The high turnover rate of *Imitervirales* is also not shared with small dsDNA viruses, as the 97% OTU based community composition of small dsDNA viruses was shown to be stable with high persistence and recurrence of its members (Ignacio-Espinoza et al., 2020).

In the third chapter, I also analyzed two eukaryotic communities namely the PicoNano size fraction and the NanoPlus size fraction aside from the *Imitervirales* communities. The richness and turnover rate of the *Imitervirales* community was either about the same or slightly higher than

that of either eukaryotic communities. As hypothesized in a preceding study (Mihara et al., 2018) the high richness of *Imitervirales* may be an adaptation to the already high richness of eukaryotic microbes in the ocean, which is currently estimated based on ribosomal sequences to be around 150,000 OTUs (Vargas et al., 2015). Since several *Imitervirales* were shown to infect the same host the "genetic arms race" is not only carried out between eukaryotes and *Imitervirales*, but also between *Imitervirales* themselves (Johannessen et al., 2015). The hypothesis that *Imitervirales* diversity and turnover is a response to host diversity and turnover would also explain the comparatively low diversity and turnover of bacteriophages (Ignacio-Espinoza et al., 2020), since total prokaryotic richness was estimated to be lower than eukaryotic richness at around 40,000 species (Sunagawa et al., 2015).

Imitervirales were found to be rich and diverse in coastal and oligotrophic oceanic areas (Hingamp et al., 2013; Li et al., 2018; Mihara et al., 2018). Certain *Imitervirales* OTUs are furthermore considered specific to certain regions (Endo et al., 2020). In this study I also found that few *Imitervirales* OTUs were shared between both the Osaka Bay and the Uranouchi Inlet.

Aside from time (third chapter) and space (second chapter and a previous study (Li et al., 2019)), *Imitervirales* communities might also be different depending on which size fraction is analyzed. It was previously shown that the communities of eukaryotes of different size fraction vary considerably (Vargas et al., 2015). In my studies, I focused on the *Imitervirales* community of the 0.22µm – 0.8µm size fraction.

It is however likely that a MEGAPRIMER analysis of a larger size fraction (e.g., 0.8µm – 3µm) may also show *Imitervirales* ASVs to be present. The persistence of these *Imitervirales* in larger size fraction (<0.8µm) may be higher than that of the studied *Imitervirales* of the Uranouchi Inlet. A recent study showed the persistent infection of certain *Imitervirales* infecting haptophytes (Blanc-Mathieu et al., 2021). It was also suggested that *Imitervirales* that infect haptophytes

evolve to match their hosts' strategy of a persistent occurrence (Sandaa et al., 2021). This hypothesis can be tested with a MEGAPRIMER based metabarcoding study of the *Imitervirales* community of a size fraction that excludes free virions (e.g., 0.8µm – 3µm) of the oligotrophic ocean, where haptophytes were observed to be abundantly present (Liu et al., 2009).

4.3 Outlook

Our understanding of viral communities and their interactions with host communities is still limited. Viruses are able to rapidly change a host community by killing the winner (Winter et al., 2010), an example of this are algal bloom termination events (Tarutani et al., 2000). This showed how lytic viruses top–down–control their community to ensure the diversity of their hosts. However recent studies hinted at more benefits to the eukaryotic community through viral infection. For instance one study found that the virus of a non–bloom forming algae showed lower virulence and encoded many metabolism regulating genes (Blanc-Mathieu et al., 2021) that are available to the host through horizontal gene transfer (Needham et al., 2019). These viral traits seem to support the survivability of the host (for self–perseverance) and are not aimed at diminishing the host population as rapidly as possible. Such a co–evolution to a more persistent infection was previously described for phages (Correa et al., 2021) and might also be a successful strategy for large dsDNA viruses (Sandaa et al., 2021).

The aim of this study was to better understand the dynamics of *Imitervirales* and eukaryotic communities. I found a difference in the ASV community memory of viral and cellular communities and described them in this work. Yet, arguing that the hypothesis was proven with a sample set taken during 20 months and only analyzing one core gene for each community is hardly sufficient to prove the hypothesis. It should rather be considered a starting point for studies that further investigate the community memory of *Imitervirales* and other microbes.

The simplest approach, following this work, may simply be a prolonged sampling series in the Uranouchi Inlet, similar to Cram et al.'s and Fuhrman et al.'s work in which sampling was conducted during 10 years (Cram et al., 2015; Fuhrman et al., 2015). Thereby, it can be estimated whether the fading of the community memory continues, as suggested in Figure 3-10A and similar previous studies (Fuhrman et al., 2015; Ignacio-Espinoza et al., 2020) and if the community memory is actually lost more rapidly for *Imitervirales* than for cellular microbes and if the same is true for phages and bacteria.

This study exclusively used metabarcoding, but further analysis of the 43 seawater samples of this study with untargeted metagenomics are also incentivized to better understand major players of the large dsDNA virus community, as suggested in Figure 1-1. Untargeted metagenomics include fewer sample preparation steps (that can affect the outcome of the analysis) and do not suffer from downsides, like overestimating certain species that have duplicate amplicon target genes (Gong and Marchetti, 2019). Therefore verifying the community memory hypothesis should involve untargeted metagenomics of the Uranouchi Inlet samples of this study. The community dissimilarity analysis of this study can be conducted using different microbial core genes to further test the community memory hypothesis. If the hypothesis is true, plotting the microbial–community–dissimilarity against time should show a similar curve as observed here and a more rapid and higher turnover for the *Imitervirales* community compared to cellular communities.

If the genomes of the major players of the viral communities are sequenced through untargeted metagenomics, virus–host pairs can also be more reliably connected by checking horizontal gene transfer (Schulz et al., 2020) instead of co-occurrence like it was performed here. Another advantage of untargeted metagenomics is that, unlike with MEGAPRIMER metabarcoding, *Phycodnaviridae* can be detected. This includes *Heterosigma akashiwo virus* (HaV), which may be a

major player in the Uranouchi Inlet since *H. akashiwo* blooms occur yearly and HaV has previously been associated with bloom termination (Tomaru et al., 2004). Studying how the distribution of ASVs of the post bloom HaV community changes from one year to another may shed additional light on viral community dynamics and their inter–species competition (i.e., Red Queen dynamics) which is also an interesting topic (Ignacio-Espinoza et al., 2020).

While the *Imitervirales* persistence over time was found to be short here, a previous metabarcoding based study also used MEGAPRIMER and found that the *Imitervirales* communities around the world shared many OTUs (Li et al., 2019). Li et al. sampled different aquatic environments several hundred kilometers apart, these samples contained many shared OTUs and the identified *polB* gene fragments were also present in metagenes sampled on the opposite site of the globe (Li et al., 2019). Such an analysis of the ASVs found in Uranouchi Inlet may be conducted to examine whether *Imitervirales* that are displaced from a local community are preserved on a global scale.

Aside from viruses, other lowly abundant microbes have recently been recognized as attractive targets for community analysis. Fungi for instance, have recently been found to be important members of marine environments but are lowly abundant (Amend et al., 2019). Li et al.'s approach to primer design may be a suitable strategy to study their communities as well. MEGAPRIMER has proven to be more sensitive than previous methods and after the optimization it was also barely more inconveniencing than any other targeted metagenomics approach. This was mostly due to Li et al.'s innovative approach to primer design. Li et al.'s concept (Li et al., 2018) of searching gene catalogues using hidden–Markov–models to assemble environmental genes for degenerate primer design may be employed to study other lowly abundant microbes like oceanic fungi.

As of yet, the diversity and ecology of *Imitervirales* and other oceanic microbes are not well enough understood. *Imitervirales* communities were just recently found to be part of the ocean's vast assembly of microbes (Hingamp et al., 2013) and their influence on geochemical cycling (Kaneko et al., 2020) and cellular communities (Endo et al., 2020; Schulz et al., 2020) has just been estimated to be substantial. I hope that this work and the community memory hypothesis will be the basis of future studies and help to deepen the understanding of oceanic communities.

5 Data Availability

All the sequencing data of chapter two was deposited to DDBJ (Megaprimer amplicon data: DRA009129; 18S rRNA amplicon data: DRA009128). The sequencing data of chapter three was also deposited to the DDBJ (DRA010976). The MAPS2 analysis pipelines and analysis pipelines for cellular communities can be downloaded from a public repository (github.com/FlorianProdinger/pipeline_18S, github.com/FlorianProdinger/pipeline_16S, github.com/FlorianProdinger/MAPS2, accessed 17 February 2021).

Acknowledgements

I would like to express my sincere thanks and gratitude to Hiroyuki Ogata, who offered me an opportunity to study in his laboratory. He supported me even before I was a student in his laboratory and did not hesitate to offer help me when I needed it. I have learnt a lot from him, not only about the ecology of giant viruses. I would also like to thank Hisashi Endo, Romain Blanc-Mathieu, Yusuke Okazaki, Takashi Yoshida, and Keizo Nagasaki for their helpful comments and advice. I want to thank Yasuhiro Gotoh and Hayashi Tetsuya for supporting me with data generation and Etsunori Taniguchi for sharing the cell count and abiotic data of the Uranouchi Inlet as well as Yoshihito Takano for his sampling efforts. I want to thank the Kyoto University super computer systems and Hideya Uehara with computational support. I further want to thank Daichi Morimoto, Kimiho Omae, Yanze Li, Kento Tominaga, Tatsuhiro Isozaki, Xia Jun, and Lingjie Meng, who I am proud to have studied and conducted experiments with, as well as all other Ogata laboratory and Yoshida laboratory members. I want to thank the Japanese Ministry of Education, Culture, Sports, Science and Technology for the opportunity to study under Professor Ogata. Finally, I want to thank my parents for their unconditional support.

References

Abergel, C., Rudinger-Thirion, J., Giegé, R., and Claverie, J.-M. (2007). Virus-Encoded Aminoacyl-tRNA Synthetases: Structural and Functional Characterization of Mimivirus TyrRS and MetRS. J. Virol. *81*, 12406–12417.

Abrahão, J., Silva, L., Silva, L.S., Khalil, J.Y.B., Rodrigues, R., Arantes, T., Assis, F., Boratto, P., Andrade, M., Kroon, E.G., et al. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat. Commun. *9*, 749.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The Revised Classification of Eukaryotes. J. Eukaryot. Microbiol. *59*, 429–514.

Agatha, S., Strüder-Kypke, M.C., Beran, A., and Lynn, D.H. (2005). Pelagostrobilidium neptuni (Montagnes and Taylor, 1994) and Strombidium biarmatum nov. spec. (Ciliophora, Oligotrichea): phylogenetic position inferred from morphology, ontogenesis, and gene sequence data. Eur. J. Protistol. *41*, 65–83.

Aherfi, S., Colson, P., La Scola, B., and Raoult, D. (2016). Giant Viruses of Amoebas: An Update. Front. Microbiol. *7*.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amend, A., Burgaud, G., Cunliffe, M., Edgcomb, V.P., Ettinger, C.L., Gutiérrez, M.H., Heitman, J., Hom, E.F.Y., Ianiri, G., Jones, A.C., et al. (2019). Fungi in the Marine Environment: Open Questions and Unsolved Problems. MBio *10*, e01189-18.

Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic Acids Res. *9*, 3015–3027.

Arkhipova, K., Skvortsov, T., Quinn, J.P., McGrath, J.W., Allen, C.C., Dutilh, B.E., McElarney, Y., and Kulakov, L.A. (2018). Temporal dynamics of uncultured viruses: a new dimension in viral diversity. ISME J. *12*, 199.

Arnold, B.J., Huang, I.-T., and Hanage, W.P. (2021). Horizontal gene transfer and adaptive evolution in bacteria. Nat. Rev. Microbiol. 1–13.

Bar-On, Y.M., and Milo, R. (2019). The Biomass Composition of the Oceans: A Blueprint of Our Blue Planet. Cell *179*, 1451–1454.

Bar-On, Y.M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. Proc. Natl. Acad. Sci. *115*, 6506–6511.

Baudoux, A.-C., and Brussaard, C.P.D. (2005). Characterization of different viruses infecting the marine harmful algal bloom species Phaeocystis globosa. Virology *341*, 80–90.

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

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V.B. (2017). Julia: A Fresh Approach to Numerical Computing. SIAM Rev. *59*, 65–98.

Birtles, R., Rowbotham, T., Storey, C., Marrie, T., and Raoult, D. (1997). Chlamydia-like obligate parasite of free-living amoebae. The Lancet *349*, 925–926.

Blanc-Mathieu, R., and Ogata, H. (2016). DNA repair genes in the Megavirales pangenome. Curr. Opin. Microbiol. *31*, 94–100.

Blanc-Mathieu, R., Dahle, H., Hofgaard, A., Brandt, D., Ban, H., Kalinowski, J., Ogata, H., and Sandaa, R.-A. (2021). A persistent giant algal virus, with a unique morphology, encodes an unprecedented number of genes involved in energy metabolism. BioRxiv 2020.07.30.228163.

Blankenship, R.E. (2010). Early Evolution of Photosynthesis1. Plant Physiol. *154*, 434–438.

Bock, C., Salcher, M., Jensen, M., Pandey, R.V., and Boenigk, J. (2018). Synchrony of Eukaryotic and Prokaryotic Planktonic Communities in Three Seasonally Sampled Austrian Lakes. Front. Microbiol. *9*.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2018). QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science (PeerJ Inc.).

Boyer, M., Madoui, M.-A., Gimenez, G., Scola, B.L., and Raoult, D. (2010). Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. PLOS ONE *5*, e15530.

Bradley, I.M., Pinto, A.J., and Guest, J.S. (2016). Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. Appl. Environ. Microbiol. *82*, 5878–5891.

Brum, J.R., and Sullivan, M.B. (2015). Rising to the challenge: accelerated pace of discovery transforms marine virology. Nat. Rev. Microbiol. *13*, 147–159.

Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A., Chaffron, S., Cruaud, C., Vargas, C. de, Gasol, J.M., et al. (2015). Patterns and ecological drivers of ocean viral communities. Science *348*.

Brussaard, C.P.D., Wilhelm, S.W., Thingstad, F., Weinbauer, M.G., Bratbak, G., Heldal, M., Kimmance, S.A., Middelboe, M., Nagasaki, K., Paul, J.H., et al. (2008). Global-scale processes with a nanoscale drive: the role of marine viruses. ISME J. *2*, 575–578.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583.

Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. *11*, 2639–2643.

Caron, D.A., Countway, P.D., Jones, A.C., Kim, D.Y., and Schnetzer, A. (2012). Marine protistan diversity. Annu. Rev. Mar. Sci. *4*, 467–493.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., et al. (2018). A global ocean atlas of eukaryotic genes. Nat. Commun. *9*.

Case, R.J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W.F., and Kjelleberg, S. (2007). Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies. Appl. Environ. Microbiol. *73*, 278–288.

Castillo, Y.M., Forn, I., Yau, S., Morán, X.A.G., Alonso-Sáez, L., Arandia-Gorostidi, N., Vaqué, D., and Sebastián, M. (2021). Seasonal dynamics of natural Ostreococcus viral infection at the single cell level using VirusFISH. Environ. Microbiol. *n/a*.

Chafee, M., Fernàndez-Guerra, A., Buttigieg, P.L., Gerdts, G., Eren, A.M., Teeling, H., and Amann, R.I. (2018). Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. *12*, 237–252.

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

Chen, F., Suttle, C.A., and Short, S.M. (1996). Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. *62*, 2869–2874.

Chen, W., Pan, Y., Yu, L., Yang, J., and Zhang, W. (2017). Patterns and Processes in Marine Microeukaryotic Community Biogeography from Xiamen Coastal Waters and Intertidal Sediments, Southeast China. Front. Microbiol. *8*.

Choi, C.J., Jimenez, V., Needham, D.M., Poirier, C., Bachy, C., Alexander, H., Wilken, S., Chavez, F.P., Sudek, S., Giovannoni, S.J., et al. (2020). Seasonal and Geographical Transitions in Eukaryotic Phytoplankton Community Structure in the Atlantic and Pacific Oceans. Front. Microbiol. *11*, 2187.

Chow, C.-E.T., and Fuhrman, J.A. (2012). Seasonality and monthly dynamics of marine myovirus communities. Environ. Microbiol. *14*, 2171–2183.

Chow, C.-E.T., Sachdeva, R., Cram, J.A., Steele, J.A., Needham, D.M., Patel, A., Parada, A.E., and Fuhrman, J.A. (2013). Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. ISME J. *7*, 2259–2273.

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

Colson, P., Levasseur, A., La Scola, B., Sharma, V., Nasir, A., Pontarotti, P., Caetano-Anollés, G., and Raoult, D. (2018). Ancestrality and Mosaicism of Giant Viruses Supporting the Definition of the Fourth TRUC of Microbes. Front. Microbiol. *9*.

Correa, A.M.S., Howard-Varona, C., Coy, S.R., Buchan, A., Sullivan, M.B., and Weitz, J.S. (2021). Revisiting the rules of life for viruses of microorganisms. Nat. Rev. Microbiol. 1–13.

Cram, J.A., Chow, C.-E.T., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A., and Fuhrman, J.A. (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J. *9*, 563–580.

Cunha, V.D., Gaia, M., Ogata, H., Jaillon, O., Delmont, T.O., and Forterre, P. (2020). Giant viruses encode novel types of actins possibly related to the origin of eukaryotic actin: the viractins. BioRxiv 2020.06.16.150565.

Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W., and Frankenberg-Dinkel, N. (2008). Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. Curr. Biol. CB *18*, 442–448.

Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016). Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol. Rev. *40*, 258–272.

Eiler, A., Hayakawa, D.H., Church, M.J., Karl, D.M., and Rappé, M.S. (2009). Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic North Pacific subtropical gyre. Environ. Microbiol. *11*, 2291–2300.

Endo, H., Blanc-Mathieu, R., Li, Y., Salazar, G., Henry, N., Labadie, K., de Vargas, C., Sullivan, M.B., Bowler, C., Wincker, P., et al. (2020). Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. Nat. Ecol. Evol. 1–11.

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ *3*, e1319.

Falkowski, P.G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth. Res. *39*, 235–258.

Falkowski, P.G. (2001). Biogeochemical Cycles. In Encyclopedia of Biodiversity, S.A. Levin, ed. (New York: Elsevier), pp. 437–453.

Falkowski, P., and Knoll, A.H. (2007). Evolution of Primary Producers in the Sea (Amsterdam ; Boston: Academic Press).

Falkowski, P.G., and Raven, J.A. (2007). Aquatic Photosynthesis: (Second Edition) (Princeton University Press).

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science *320*, 1034–1039.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science *281*, 237–240.

Fischer, M.G., Allen, M.J., Wilson, W.H., and Suttle, C.A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. Proc. Natl. Acad. Sci. *107*, 19508–19513.

Forootan, A., Sjöback, R., Björkman, J., Sjögreen, B., Linz, L., and Kubista, M. (2017). Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). Biomol. Detect. Quantif. *12*, 1–6.

Forterre, P. (2012). Virocell Concept, The. In ELS, (John Wiley & Sons, Ltd), p.

Forterre, P. (2013). The virocell concept and environmental microbiology. ISME J. *7*, 233–236.

Forterre, P., Krupovic, M., and Prangishvili, D. (2014). Cellular domains and viral lineages. Trends Microbiol. *22*, 554–558.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008). Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. U. S. A. *105*, 3805–3810.

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

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015). Marine microbial community dynamics and their ecological interpretation. Nat. Rev. Microbiol. *13*, 133–146.

Gallot-Lavallée, L., Pagarete, A., Legendre, M., Santini, S., Sandaa, R.-A., Himmelbauer, H., Ogata, H., Bratbak, G., and Claverie, J.-M. (2015). The 474-Kilobase-Pair Complete Genome Sequence of CeV-01B, a Virus Infecting Haptolina (Chrysochromulina) ericina (Prymnesiophyceae). Genome Announc. *3*.

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

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P.J., Huse, S., and Joint, I. (2009). The seasonal structure of microbial communities in the Western English Channel. Environ. Microbiol. *11*, 3132–3139.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I., et al. (2012). Defining seasonal marine microbial community dynamics. ISME J. *6*, 298–308.

Giner, C.R., Balagué, V., Krabberød, A.K., Ferrera, I., Reñé, A., Garcés, E., Gasol, J.M., Logares, R., and Massana, R. (2019). Quantifying long-term recurrence in planktonic microbial eukaryotes. Mol. Ecol. *28*, 923–935.

Giovannoni, S.J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. Annu. Rev. Mar. Sci. *9*, 231–255.

Gong, W., and Marchetti, A. (2019). Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. Front. Mar. Sci. *6*.

Gran-Stadniczeñko, S., Egge, E., Hostyeva, V., Logares, R., Eikrem, W., and Edvardsen, B. (2019a). Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities as Revealed by Metabarcoding and Microscopy. J. Eukaryot. Microbiol. *66*, 494–513.

Gran-Stadniczeñko, S., Krabberød, A.K., Sandaa, R.-A., Yau, S., Egge, E., and Edvardsen, B. (2019b). Seasonal Dynamics of Algae-Infecting Viruses and Their Inferred Interactions with Protists. Viruses *11*.

Gregory, A.C., Solonenko, S.A., Ignacio-Espinoza, J.C., LaButti, K., Copeland, A., Sudek, S., Maitland, A., Chittick, L., dos Santos, F., Weitz, J.S., et al. (2016). Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. BMC Genomics *17*.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. Nature *532*, 465–470.

Ha, A.D., Moniruzzaman, M., and Aylward, F.O. (2021). High Transcriptional Activity and Diverse Functional Repertoires of Hundreds of Giant Viruses in a Coastal Marine System. BioRxiv 2021.03.08.434518.

Hain, M.P., Sigman, D.M., and Haug, G.H. (2014). 8.18 - The Biological Pump in the Past. In Treatise on Geochemistry (Second Edition), H.D. Holland, and K.K. Turekian, eds. (Oxford: Elsevier), pp. 485–517.

Hansen, P.J., Nielsen, T.G., and Kaas, H. (1995). Distribution and growth of protists and mesozooplankton during a bloom of Chrysochromulina spp. (Prymnesiophyceae, Prymnesiales). Phycologia *34*, 409–416.

Hardin, G. (1960). The Competitive Exclusion Principle. Science *131*, 1292–1297.

Henley, S.F., Cavan, E.L., Fawcett, S.E., Kerr, R., Monteiro, T., Sherrell, R.M., Bowie, A.R., Boyd, P.W., Barnes, D.K.A., Schloss, I.R., et al. (2020). Changing Biogeochemistry of the Southern Ocean and Its Ecosystem Implications. Front. Mar. Sci. *7*, 581.

Hingamp, P., Grimsley, N., Acinas, S.G., Clerissi, C., Subirana, L., Poulain, J., Ferrera, I., Sarmento, H., Villar, E., Lima-Mendez, G., et al. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. ISME J. *7*, 1678–1695.

Hirata, T., Hardman-Mountford, N.J., Brewin, R.J.W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., et al. (2011). Synoptic relationships between surface Chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types. Biogeosciences *8*, 311–327.

Hsieh, T.C., H. Ma, K., and Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol.

Hutchinson (1961). The Paradox of the Plankton | The American Naturalist: Vol 95, No 882.

Ignacio-Espinoza, J.C., Ahlgren, N.A., and Fuhrman, J.A. (2020). Long-term stability and Red Queen-like strain dynamics in marine viruses. Nat. Microbiol. *5*, 265–271.

Iyer, L.M., Aravind, L., and Koonin, E.V. (2001). Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. J. Virol. *75*, 11720–11734.

Jacob, F., and Wollman, E.L. (1961). Viruses and genes. Sci. Am. *204*, 93–107.

Johannessen, T.V., Bratbak, G., Larsen, A., Ogata, H., Egge, E.S., Edvardsen, B., Eikrem, W., and Sandaa, R.-A. (2015). Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). Virology *476*, 180–188.

Johannessen, T.V., Larsen, A., Bratbak, G., Pagarete, A., Edvardsen, B., Egge, E.D., and Sandaa, R.-A. (2017). Seasonal Dynamics of Haptophytes and dsDNA Algal Viruses Suggest Complex Virus-Host Relationship. Viruses *9*.

Kaneko, H., Blanc-Mathieu, R., Endo, H., Chaffron, S., Delmont, T.O., Gaia, M., Henry, N., Hernández-Velázquez, R., Nguyen, C.H., Mamitsuka, H., et al. (2020). Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. IScience 102002.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kijima, S., Delmont, T.O., Miyazaki, U., Gaia, M., Endo, H., and Ogata, H. (2021). Discovery of Viral Myosin Genes With Complex Evolutionary History Within Plankton. Front. Microbiol. *12*, 1450.

Kim, S., Park, M.G., Kim, K.-Y., Kim, C.-H., Yih, W., Park, J.S., and Coats, D.W. (2008). Genetic diversity of parasitic dinoflagellates in the genus amoebophrya and its relationship to parasite biology and biogeography. J. Eukaryot. Microbiol. *55*, 1–8.

Koonin, E., Dolja, V., Krupovic, M., Varsani, A., Wolf, Y., Yutin, N., Zerbini, F., and Kuhn, J. (2019). Create a megataxonomic framework, filling all principal taxonomic ranks, for DNA viruses encoding vertical jelly roll-type major capsid proteins.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinforma. Oxf. Engl. *23*, 1289–1291.

Koskella, B., and Brockhurst, M.A. (2014). Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. Fems Microbiol. Rev. *38*, 916–931.

Lane, N. (2015). The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals.' Philos. Trans. R. Soc. B Biol. Sci. *370*.

Larsen, J.B., Larsen, A., Bratbak, G., and Sandaa, R.-A. (2008). Phylogenetic analysis of members of the Phycodnaviridae virus family, using amplified fragments of the major capsid protein gene. Appl. Environ. Microbiol. *74*, 3048–3057.

Legendre, M., Arslan, D., Abergel, C., and Claverie, J.-M. (2012). Genomics of Megavirus and the elusive fourth domain of Life. Commun. Integr. Biol. *5*, 102–106.

Lehahn, Y., Koren, I., Schatz, D., Frada, M., Sheyn, U., Boss, E., Efrati, S., Rudich, Y., Trainic, M., Sharoni, S., et al. (2014). Decoupling physical from biological processes to assess the impact of viruses on a mesoscale algal bloom. Curr. Biol. CB *24*, 2041–2046.

Levins, R. (1968). Evolution in Changing Environments: Some Theoretical Explorations. (MPB-2) (Princeton University Press).

Li, W. (1995). Composition of ultraphytoplankton in the central North Atlantic. Mar. Ecol. Prog. Ser. *122*, 1–8.

Li, J., and Convertino, M. (2021). Inferring ecosystem networks as information flows. Sci. Rep. *11*, 7094.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinforma. Oxf. Engl. *22*, 1658–1659.

Li, Y., Hingamp, P., Watai, H., Endo, H., Yoshida, T., Ogata, H., Li, Y., Hingamp, P., Watai, H., Endo, H., et al. (2018). Degenerate PCR Primers to Reveal the Diversity of Giant Viruses in Coastal Waters. Viruses *10*, 496.

Li, Y., Endo, H., Gotoh, Y., Watai, H., Ogawa, N., Blanc-Mathieu, R., Yoshida, T., and Ogata, H. (2019). The Earth Is Small for "Leviathans": Long Distance Dispersal of Giant Viruses across Aquatic Environments. Microbes Environ. *advpub*.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., et al. (2015). Determinants of community structure in the global plankton interactome. Science *348*.

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., and Vargas, C. de (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. Proc. Natl. Acad. Sci. *106*, 12803–12808.

Lynch, M.D.J., and Neufeld, J.D. (2015). Ecology and exploration of the rare biosphere. Nat. Rev. Microbiol. *13*, 217–229.

Lyons, T.W., Reinhard, C.T., and Planavsky, N.J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. Nature *506*, 307–315.

Martin-Platero, A.M., Cleary, B., Kauffman, K., Preheim, S.P., McGillicuddy, D.J., Alm, E.J., and Polz, M.F. (2018). High resolution time series reveals cohesive but short-lived communities in coastal plankton. Nat. Commun. *9*, 266.

Maruyama, F., and Ueki, S. (2016). Evolution and Phylogeny of Large DNA Viruses, Mimiviridae and Phycodnaviridae Including Newly Characterized Heterosigma akashiwo Virus. Front. Microbiol. *7*.

Matsen, F.A., Kodner, R.B., and Armbrust, E.V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics *11*, 538.

Menge, B.A. (2000). Top-down and bottom-up community regulation in marine rocky intertidal habitats. J. Exp. Mar. Biol. Ecol. *250*, 257–289.

Mihara, T., Koyano, H., Hingamp, P., Grimsley, N., Goto, S., and Ogata, H. (2018). Taxon Richness of "Megaviridae" Exceeds those of Bacteria and Archaea in the Ocean. Microbes Environ. *33*, 162–171.

Milici, M., Deng, Z.-L., Tomasch, J., Decelle, J., Wos-Oxley, M.L., Wang, H., Jáuregui, R., Plumeier, I., Giebel, H.-A., Badewien, T.H., et al. (2016). Co-occurrence Analysis of Microbial Taxa in the Atlantic Ocean Reveals High Connectivity in the Free-Living Bacterioplankton. Front. Microbiol. *7*.

Monier, A., Chambouvet, A., Milner, D.S., Attah, V., Terrado, R., Lovejoy, C., Moreau, H., Santoro, A.E., Derelle, É., and Richards, T.A. (2017). Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. Proc. Natl. Acad. Sci. *114*, E7489–E7498.

Moniruzzaman, M., LeCleir, G.R., Brown, C.M., Gobler, C.J., Bidle, K.D., Wilson, W.H., and Wilhelm, S.W. (2014). Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. Virology *466–467*, 60–70.

Moniruzzaman, M., Gann, E.R., LeCleir, G.R., Kang, Y., Gobler, C.J., and Wilhelm, S.W. (2016). Diversity and dynamics of algal Megaviridae members during a harmful brown tide caused by the pelagophyte, Aureococcus anophagefferens. FEMS Microbiol. Ecol. *92*, fiw058.

Moniruzzaman, M., Martinez-Gutierrez, C.A., Weinheimer, A.R., and Aylward, F.O. (2020). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat. Commun. *11*, 1710.

Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature *409*, 607–610.

Moreira, D., and López-García, P. (2015). Evolution of viruses and cells: do we need a fourth domain of life to explain the origin of eukaryotes? Philos. Trans. R. Soc. B Biol. Sci. *370*.

Nagasaki, K., and Yamaguchi, M. (1997). Isolation of a virus infectious to the harmful bloom causing microalga Heterosigma akashiwo (Raphidophyceae). Aquat. Microb. Ecol. - AQUAT MICROB ECOL *13*, 135–140.

Nagasaki, K., Ando, M., Imai, I., Itakura, S., and Ishida, Y. (1994). Virus-like particles in Heterosigma-Akashiwo (Raphidophyceae) – a possible red tide disintegration mechanism. Mar. Biol. *119*, 307–312.

Nagasaki, K., Tomaru, Y., Tarutani, K., Katanozaka, N., Yamanaka, S., Tanabe, H., and Yamaguchi, M. (2003). Growth Characteristics and Intraspecies Host Specificity of a Large Virus Infecting the Dinoflagellate Heterocapsa circularisquama. Appl. Environ. Microbiol. *69*, 2580–2586.

Needham, D.M., and Fuhrman, J.A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat. Microbiol. *1*, 1–7.

Needham, D.M., Chow, C.-E.T., Cram, J.A., Sachdeva, R., Parada, A., and Fuhrman, J.A. (2013). Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. ISME J. *7*, 1274–1285.

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., et al. (2019). A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. Proc. Natl. Acad. Sci. 201907517.

Nelson, J.D., Boehme, S.E., Reimers, C.E., Sherrell, R.M., and Kerkhof, L.J. (2008). Temporal patterns of microbial community structure in the Mid-Atlantic Bight. FEMS Microbiol. Ecol. *65*, 484–493.

Not, F., Siano, R., Kooistra, W.H.C.F., Simon, N., Vaulot, D., and Probert, I. (2012). Chapter One - Diversity and Ecology of Eukaryotic Marine Phytoplankton. In Advances in Botanical Research, G. Piganeau, ed. (Academic Press), pp. 1–53.

Ogawa, H., Amagai, Y., Koike, I., Kaiser, K., and Benner, R. (2001). Production of Refractory Dissolved Organic Matter by Bacteria. Science *292*, 917–920.

Okazaki, Y., Fujinaga, S., Salcher, M.M., Callieri, C., Tanaka, A., Kohzu, A., Oyagi, H., Tamaki, H., and Nakano, S. (2021). Microdiversity and phylogeographic diversification of bacterioplankton in pelagic freshwater systems revealed through long-read amplicon sequencing. Microbiome *9*, 24.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Sólymos, P., Stevens, M.H.H., and Wagner, H. (2012). vegan: Community Ecology Package.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., and Stahl, D.A. (1986). Microbial Ecology and Evolution: A Ribosomal RNA Approach. Annu. Rev. Microbiol. *40*, 337–365.

Omura, T., Iwataki, M., Borja, V.M., Takayama, H., and Fukuyo, Y. (2012). Marine Phytoplankton of theWestern Pacific.

Pace, N.R., Stahl, D.A., Lane, D.J., and Olsen, G.J. (1986). The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. In Advances in Microbial Ecology, K.C. Marshall, ed. (Boston, MA: Springer US), pp. 1–55.

Pagarete, A., Chow, C.-E.T., Johannessen, T., Fuhrman, J.A., Thingstad, T.F., and Sandaa, R.A. (2013). Strong Seasonality and Interannual Recurrence in Marine Myovirus Communities. Appl. Environ. Microbiol. *79*, 6253–6259.

Phillips, A.J., Anderson, V.L., Robertson, E.J., Secombes, C.J., and van West, P. (2008). New insights into animal pathogenic oomycetes. Trends Microbiol. *16*, 13–19.

Poranen, M.M., Ravantti, J.J., Grahn, A.M., Gupta, R., Auvinen, P., and Bamford, D.H. (2006). Global Changes in Cellular Gene Expression during Bacteriophage PRD1 Infection. J. Virol. *80*, 8081–8088.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PloS One *5*, e9490.

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

Prodinger, F., Endo, H., Gotoh, Y., Li, Y., Morimoto, D., Omae, K., Tominaga, K., Blanc-Mathieu, R., Takano, Y., Hayashi, T., et al. (2020). An Optimized Metabarcoding Method for Mimiviridae. Microorganisms *8*, 506.

Prodinger, F., Endo, H., Takano, Y., Li, Y., Tominaga, K., Isozaki, T., Blanc-Mathieu, R., Gotoh, Y., Hayashi, T., Taniguchi, E., et al. (2021). Year-round dynamics of amplicon sequence variant communities differ among eukaryotes, Imitervirales, and prokaryotes in a coastal ecosystem. FEMS Microbiol. Ecol. fiab167.

Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. Nat. Rev. Microbiol. *6*, 315–319.

Raoult, D., Scola, B.L., and Birtles, R. (2007). The Discovery and Characterization of Mimivirus, the Largest Known Virus and Putative Pneumonia Agent. Clin. Infect. Dis. *45*, 95–102.

Riemann, L., Steward, G.F., and Azam, F. (2000). Dynamics of Bacterial Community Composition and Activity during a Mesocosm Diatom Bloom. Appl. Environ. Microbiol. *66*, 578–587.

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., Solonenko, N., Lara, E., Poulain, J., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature *537*, 689–693.

Sakami, T., Watanabe, T., Kakehi, S., Taniuchi, Y., and Kuwata, A. (2016). Spatial variation of bacterial community composition at the expiry of spring phytoplankton bloom in Sendai Bay, Japan. Gene *576*, 610–617.

Sakamoto, S., Lim, W.A., Lu, D., Dai, X., Orlova, T., and Iwataki, M. (2020). Harmful algal blooms and associated fisheries damage in East Asia: Current status and trends in China, Japan, Korea and Russia. Harmful Algae 101787.

Sandaa, R.-A., Heldal, M., Castberg, T., Thyrhaug, R., and Bratbak, G. (2001). Isolation and Characterization of Two Viruses with Large Genome Size Infecting Chrysochromulina ericina (Prymnesiophyceae) and Pyramimonas orientalis (Prasinophyceae). Virology *290*, 272–280.

Sandaa, R.-A., E. Storesund, J., Olesin, E., Lund Paulsen, M., Larsen, A., Bratbak, G., and Ray, J.L. (2018). Seasonality Drives Microbial Community Structure, Shaping both Eukaryotic and Prokaryotic Host–Viral Relationships in an Arctic Marine Ecosystem. Viruses *10*.

Sandaa, R.-A., Saltvedt, M.R., Dahle, H., Wang, H., Våge, S., Blanc-Mathieu, R., Steen, I.H., Grimsley, N., Edvardsen, B., Ogata, H., et al. (2021). Adaptive evolution of viruses infecting marine microalgae (haptophytes), from acute infections to stable coexistence. Biol. Rev. *n/a*.

Sanjuán, R., and Domingo-Calap, P. (2016). Mechanisms of viral mutation. Cell. Mol. Life Sci. *73*, 4433–4448.

Santi, I., Tsiola, A., Dimitriou, P.D., Fodelianakis, S., Kasapidis, P., Papageorgiou, N., Daffonchio, D., Pitta, P., and Karakassis, I. (2019). Prokaryotic and eukaryotic microbial community responses to N and P nutrient addition in oligotrophic Mediterranean coastal waters: Novel insights from DNA metabarcoding and network analysis. Mar. Environ. Res. *150*, 104752.

Santini, S., Jeudy, S., Bartoli, J., Poirot, O., Lescot, M., Abergel, C., Barbe, V., Wommack, K.E., Noordeloos, A.A.M., Brussaard, C.P.D., et al. (2013). Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc. Natl. Acad. Sci. U. S. A. *110*, 10800–10805.

Schroeder, D.C., Oke, J., Hall, M., Malin, G., and Wilson, W.H. (2003). Virus Succession Observed during an Emiliania huxleyi Bloom. Appl. Environ. Microbiol. *69*, 2484–2490.

Schulz, F., Yutin, N., Ivanova, N.N., Ortega, D.R., Lee, T.K., Vierheilig, J., Daims, H., Horn, M., Wagner, M., Jensen, G.J., et al. (2017). Giant viruses with an expanded complement of translation system components. Science *356*, 82–85.

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D.A., Denef, V.J., McMahon, K.D., Konstantinidis, K.T., Eloe-Fadrosh, E.A., Kyrpides, N.C., et al. (2020). Giant virus diversity and host interactions through global metagenomics. Nature *578*, 432–436.

Schvarcz, C.R., and Steward, G.F. (2018). A giant virus infecting green algae encodes key fermentation genes. Virology *518*, 423–433.

Scola, B.L., Audic, S., Robert, C., Jungang, L., Lamballerie, X. de, Drancourt, M., Birtles, R., Claverie, J.-M., and Raoult, D. (2003). A Giant Virus in Amoebae. Science *299*, 2033–2033.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J.A. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio *5*, e01371-01314.

Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G., and Cameron Thrash, J. (2019). High proportions of bacteria and archaea across most biomes remain uncultured. ISME J. *13*, 3126–3130.

Stingl, U., Tripp, H.J., and Giovannoni, S.J. (2007). Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. ISME J. *1*, 361–371.

Sukenik, A., Zohary, T., and Padisák, J. (2009). Cyanoprokaryota and Other Prokaryotic Algae. In Encyclopedia of Inland Waters, G.E. Likens, ed. (Oxford: Academic Press), pp. 138–148.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., et al. (2015). Structure and function of the global ocean microbiome. Science *348*, 1261359.

Suttle, C.A. (1994). The significance of viruses to mortality in aquatic microbial communities. Microb. Ecol. *28*, 237–243.

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

Tackmann, J., Rodrigues, J.F.M., and Mering, C. von (2018). Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. BioRxiv 390195.

Tada, Y., Taniguchi, A., Nagao, I., Miki, T., Uematsu, M., Tsuda, A., and Hamasaki, K. (2011). Differing Growth Responses of Major Phylogenetic Groups of Marine Bacteria to Natural Phytoplankton Blooms in the Western North Pacific Ocean. Appl. Environ. Microbiol. *77*, 4055–4065.

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., and Nishijima, M. (2014). Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. PLOS ONE *9*, e105592.

Tarutani, K., Nagasaki, K., and Yamaguchi, M. (2000). Viral Impacts on Total Abundance and Clonal Composition of the Harmful Bloom-Forming Phytoplankton Heterosigma akashiwo. Appl. Environ. Microbiol. *66*, 4916–4920.

Teeling, H., Fuchs, B.M., Bennke, C.M., Krüger, K., Chafee, M., Kappelmann, L., Reintjes, G., Waldmann, J., Quast, C., Glöckner, F.O., et al. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. ELife *5*.

Thomsen, H.A., Buck, K.R., and Chavez, F.P. (1994). Haptophytes as components of marine phytoplankton. Haptophyte Algae 187–208.

Tomaru, Y., Tarutani, K., Yamaguchi, M., and Nagasaki, K. (2004). Quantitative and qualitative impacts of viral infection on a Heterosigma akashiwo (Raphidophyceae) bloom in Hiroshima Bay, Japan. Aquat. Microb. Ecol. *34*, 227–238.

Tominaga, K., Ogawa-Haruki, N., Nishimura, Y., Watai, H., Yamamoto, K., Ogata, H., and Yoshida, T. (2021). Prevalence of viral frequency-dependent infection in coastal marine prokaryotes revealed using monthly time series virome analysis.

Torrella, F., and Morita, R.Y. (1979). Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, oregon: ecological and taxonomical implications. Appl. Environ. Microbiol. *37*, 774–778.

Trombetta, T., Vidussi, F., Roques, C., Scotti, M., and Mostajir, B. (2020). Marine Microbial Food Web Networks During Phytoplankton Bloom and Non-bloom Periods: Warming Favors Smaller Organism Interactions and Intensifies Trophic Cascade. Front. Microbiol. *11*, 2657.

Tynan, C.T. (1998). Ecological importance of the Southern Boundary of the Antarctic Circumpolar Current. Nature *392*, 708–710.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

Vargas, C. de, Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N.L., Probert, I., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science *348*.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S (New York: Springer-Verlag).

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science *304*, 66–74.

Wagstaff, B.A., Vladu, I.C., Barclay, J.E., Schroeder, D.C., Malin, G., and Field, R.A. (2017). Isolation and Characterization of a Double Stranded DNA Megavirus Infecting the Toxin-Producing Haptophyte Prymnesium parvum. Viruses *9*.

Ward, C.S., Yung, C.-M., Davis, K.M., Blinebry, S.K., Williams, T.C., Johnson, Z.I., and Hunt, D.E. (2017). Annual community patterns are driven by seasonal switching between closely related marine bacteria. ISME J. *11*, 1412–1422.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U. S. A. *95*, 6578–6583.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).

Wilhelm, S.W., Bird, J.T., Bonifer, K.S., Calfee, B.C., Chen, T., Coy, S.R., Gainer, P.J., Gann, E.R., Heatherly, H.T., Lee, J., et al. (2017). A Student's Guide to Giant Viruses Infecting Small Eukaryotes: From Acanthamoeba to Zooxanthellae. Viruses *9*.

Wilson, W.H., Gilg, I.C., Duarte, A., and Ogata, H. (2014). Development of DNA mismatch repair gene, MutS, as a diagnostic marker for detection and phylogenetic analysis of algal Megaviruses. Virology *466–467*, 123–128.

Winter, C., Bouvier, T., Weinbauer, M.G., and Thingstad, T.F. (2010). Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited. Microbiol. Mol. Biol. Rev. *74*, 42–57.

Woese, C.R., Stackebrandt, E., Weisburg, W.G., Paster, B.J., Madigan, M.T., Fowler, V.J., Hahn, C.M., Blanz, P., Gupta, R., Nealson, K.H., et al. (1984). The phylogeny of purple bacteria: The alpha subdivision. Syst. Appl. Microbiol. *5*, 315–326.

Worden, A.Z., Nolan, J.K., and Palenik, B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. Limnol. Oceanogr. *49*, 168–179.

Yoshida, T., Yuki, Y., Lei, S., Chinen, H., Yoshida, M., Kondo, R., and Hiroishi, S. (2003). Quantitative Detection of Toxic Strains of the Cyanobacterial Genus Microcystis by Competitive PCR. Microbes Environ. *18*, 16–23.

Yoshida, T., Nishimura, Y., Watai, H., Haruki, N., Morimoto, D., Kaneko, H., Honda, T., Yamamoto, K., Hingamp, P., Sako, Y., et al. (2018). Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. ISME J. 1.

Zhang, W., Zhou, J., Liu, T., Yu, Y., Pan, Y., Yan, S., and Wang, Y. (2015). Four novel algal virus genomes discovered from Yellowstone Lake metagenomes. Sci. Rep. *5*, 15131.