Virus-microbe interactions in marine environments determined by DNA proximity linkages

Christina Renee Rathwell

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Gabrielle Rocap, Chair

Jody Deming

Rick Keil

Program Authorized to Offer Degree:

Oceanography

University of Washington

Abstract

Virus-host interactions in marine-environments determined by DNA proximity linkages

Christina Renee Rathwell

Chair of the Supervisory Committee:
Gabrielle Rocap
Oceanography

Microbial processes in the ocean drive global biogeochemical cycles and within microbial communities the most abundant entities are viruses. Viruses in the marine environment can alter microbial community composition and genomic potential when they interact with host and non-host microbes. However, because many host microbes are not yet in culture, traditional methods of culturing viruses cannot yet be undertaken for most viruses. A key component of interpreting data collected about virus presence and abundance across the world's oceans is determining with which microbes they can interact and under what conditions. In this thesis, I apply the DNA linkage method known as Hi-C sequencing to map out virus-host interactions in three different environments.

In Chapter 1, I used this method to determine the impact and interaction of viruses with hosts in both particle-associated and free-living microbial communities at a depth below the euphotic zone in a shallow coastal fjord. I inferred that infected cells on particles were likely delivered to greater depths by particle flux and that viruses moved into new environments where host cells are sparse may interact with non-host organisms.

In Chapter 2, I used Hi-C sequencing to analyze viral activity within the deep chlorophyll maximum of the Eastern Tropical North Pacific. Viruses were linked across multiple phytoplankton populations. I obtained evidence of viral auxiliary metabolic genes (AMGs) within *Synechococcus* cells at the time of sampling. Hi-C links also provided evidence of a known *Phycodnaviridae*, as well as a *Mimivirus* infecting a *Bathycoccus prasinos* at the time of sampling. Furthermore, a CrAss-like phage was shown to be interacting with a Planctomycetes of the Pirellulales order, a novel host in the marine environment for the *Crassvirales* family.

In Chapter 3, I analyzed virus-microbe interactions in an oxygen deficient zone (ODZ), in the Eastern Tropical North Pacific, from which few microbes have been cultivated. The majority of organisms at this depth are prokaryotic and live amongst a diverse and unique community of viruses that have remained difficult to characterize. Hi-C links enhanced microbial binning and resulted in 268 medium-to-high quality bins with genomic potential for a range of methane, sulfur and nitrogen cycling metabolisms in the sample examined. We associated 75 viral genomes with microbial metagenomes in one sample and identified 19 virus-host interactions with high virus per host cell values. Though some virus-host interactions were predicted with computational software, the majority of viruses linking with hosts by Hi-C were not predicted computationally and these were statistically more abundant within the metagenome. Summarily, Hi-C method applied in this thesis revealed virus-microbe activity that other tools at our disposal could not, increasing our understanding of the complex path viral DNA may travel in marine environments, suggesting unique influences on biogeochemical cycling and potential genetic diversity.

TABLE OF CONTENTS

Page

Introduction 1

Chapter 1: Hi-C assembled genomes of estuarine populations reveal virus-microbe associations and a broad interaction range of a cyanophage 8

1.1 Abstract 8
1.2 Significance 9
1.3 Introduction 9
1.4 Results 12
1.5 Discussion 19
1.6 Materials and Methods. 23
1.7 Acknowlegements 29
1.8 Figures 30
1.9 Supplementary Figures and Tables 35

Chapter 2: Hi-C links viruses to eukaryotes and prokaryotes within a deep chlorophyll maximum of the Pacific Ocean 51

2.1 Abstract 51
2.2 Introduction 52
2.3 Materials and Methods 53
2.4 Results and Discussion 59
2.5 Acknowledgements 64
2.6 Figures 65
2.7 Tables 71

Chapter 3: The genomic potential and viral interactions of a prokaryotic community within the secondary chlorophyll maximum of an oxygen deficient zone 83

3.1 Abstract 83
3.2 Introduction 84
3.3 Materials and Methods 88
3.4 Results 94
3.5 Discussion 101
3.6 Acknowledgements 109
3.7 Figures 110
3.8 Tables 121

Conclusion 135
Appendix A 140
References 146

ACKNOWLEDGEMENTS

This work would not have been possible without a wide range of support professionally and personally. I must of course thank the members of my committee. To the chair, Gabrielle Rocap, I owe you so much. You have given me grace when I needed it—and fire too. Your ability to jump into a wide range of scientific discourse with insight and enthusiasm is remarkable. Thank you for helping me see this through. Thank you to Rick Keil, Jody Deming and Stanley Fields—you've held me up personally and scientifically. Jody, you have pushed my thinking about microbial communities and environments in such a delightful way and I am in awe of your passion and curiosity. Rick, you have been a cheerful support at sea and on land. Thank you for helping me make connections, both scientifically and in community. Stan, you are such a critical pillar of the formation of my identity as a scientist and I'm so grateful to have been supported by you even past my time in your lab, but through my graduate student journey as well. A special thank you to an honorary member of my committee, Arthur Nowell. Your kindness, steadfast belief, and thorough edits held my hand right across that finish line. Thank you for helping me with the grit to finish and with the dreams of Scotland on the other side of it.

I am indebted to the many forms of research support and funding that came along the way. I would like to thank members of the Ocean Remote Chemical Analyzer Buoy group, namely Al Devol, Wendi Ruef, Chris Williams and Zoe Archer. I am also grateful to employees of Phase Genomics, both past and present, Ivan Liachko, Maximilian Press, Shawn Sullivan, Anna Shultz and so many more for holding my hand through the Hi-C technique and answering my many questions gracefully. This work was also made possible with funding from the NSF Graduate Research Fellowship Program, the Benjamin Hall Conservation Genetics Award, and the Booth Award, and the School of Oceanography.

Thank you to all the graduate students who have helped me cope and grow and learn. To the Rocap Lab specifically for helping me curate my many talks and scientific stories. I am so thankful for you all! Michael Carlson, Jaci Saunders, Natalie Kellogg, Jordan Winter. Thank you to my cohort who have helped me psychologically as much as scientifically. Though every graduate student who has worked in both the School of Oceanography and those in the Center for Environmental Genomics have truly all played a role. Thank you to my cohort, Isaiah, Hilary, Caroline, Dylan, Andrew, Meghan, Katie, Charles, Rosalind Echols, and Sasha Seroy. Rosalind and Sasha, I've got to say, you are my people! Thanks for having a similar tone, similar interests and dislikes, and for being there for either a four hour "coffee" or some beer and fiber arts. You were such an important part of my sanity.

I am so grateful for the gracious and generous support of staff and administration—from the bioinformatics help (Chris Bee, Cedar McKay) to the birthday celebrations and office functioning (Rita Anderson, Grace Wong) and the help with billing, keys and registration (Romeo Balagot, Chanthavy Manikham, Michelle Townsend, Su Tipple, Lien Lai). I'd like to extend a special thank you to our janitorial staff at CEG, Bill and Chong, who cleaned our spaces even though they'd rather be spending time with their grandbaby. Congratulations on your retirement! Thanks to the new wave of administrators and janitorial staff and all your patience with us demanding graduate students as you learn the role.

So much of my time here has been spent growing as a teacher as well. I am very grateful for the community of educators I worked among. Namely, a huge thank you to Mikelle Newer and Sasha Seroy, Susanna Michaels and Anitra Ingalls for all your support and allowing me to try things on my own. I owe so much of my growth to you. Furthermore, I'd like to thank the students and interns who have shown up with enthusiasm and curiosity. I'd specifically like to thank the

interns who I oversaw in the lab, I hope you learned as much as I did while you were here: Emily Frink, Yuliana Maritza-Romo, Matthew Hays, Andrea Odell, Zehara Kedir and Jessica Petrochuck.

Lastly but with the most emphatic heart, I'd like to thank my families. Firstly, to my family of origin, who didn't know exactly what getting a PhD entailed, but stubbornly believed that I could do it. I felt your unwavering support and I couldn't have done it without you. Thank you to my husband, who should be earning an honorary degree simultaneously. Erik has listened so deeply to my struggles scientifically and mentally throughout this process, that he now knows more about Hi-C sequencing and viral metagenomes than any other UX designer before him. Every hour spent at work outside of childcare hours is absolutely made possible by him. Thank you to my daughters for putting so much faith in me, for your endless optimism and courage, and for showing me that we are truly all scientists at birth.

DEDICATION

To my daughters, Matilda and Ellory. May the work I have completed here lead to a greater belief in yourselves and a clearer route to your dreams. Your love, patience and joy have motivated me more than you could know.

To my love and best friend, Erik, for being a singular support through all of this. Here's to many more bike rides on the other side.

INTRODUCTION

Viruses occupy the distance between living and chemistry—as the bearer of the largest genetic diversity on the planet[1,2], they oscillate between actively reproducing entities and passive inert clusters of chemical compounds. Viruses, specifically their ancestral state, are believed to be critical to the origin of life and the expansion of the tree of life[3,4]. Viral evolutionary studies suggest that without viruses on this planet, complex multicellular beings would not exist[5]. Viruses are found in every livable medium yet studied on Earth, from ice cores in the Arctic to deep-sea sediment, oligotrophic ocean water columns to compacted soil, rivers to acidic hydrothermal plumes[2,6–10].

The first phage to be isolated from marine water was on a bacterial culture from a fish microbiome in 1955[11], and took repeated attempts across 60 liters of water. Intensive sampling efforts produced isolation of viral populations capable of lysing marine bacterial cultures, but prevalence was presumed low in marine waters, due to the limited rate of success with methodologies applied[12–14]. Microscopy of coastal environments showed high incidence of bacteriophage in 1978[15], but did not really take hold in the literature until a decade later when again microscopic measurements suggested high abundance of viruses in marine environments[16].

Despite the slow start, the discovery of abundant viruses in marine waters and consequent viral infection in marine organisms eventually led to an explosion of research to describe the ecological impact of viral infection on marine organisms[17–19]. Studies determining the infection rate in marine microbial communities concluded that between 3% and 50% of populations were killed daily[20]. Individual strains in laboratories around the world and microcosms on the decks of dozens of cruises were subjected to viruses and the impacts observed, suggesting that there are

viral populations targeting every marine microbe and that viral infection within a community may support growth of the remaining community[20–24]. Together this research, provided the view that viruses impact the environment by lysing microbial cells, delivering cell material to the dissolved pool and feeding the microbial community, thus preventing the transfer of microbial organic material from transferring to higher trophic levels[25].

With a stronger grasp of the significance of these tiny entities and novel tools at our disposal, massive metagenomic sequencing efforts were undertaken to genomically describe entire viral communities across the global ocean[1,26–28]. Each environmental sample sequenced revealed a nearly unique population, distinct from databases holding viral sequences from non-marine environments, but also unrelated to recently acquired ocean viromes[29,30]. Ocean virome collection and processing techniques were standardized by important breakthroughs, rapid iron chloride precipitation of viruses and identification software to search for viral sequence within metagenomes[31–33]. These methods have been applied continuously since, leading to the genomic description of nearly 200,000 unique viral populations across the world's oceans[34,35].

Continued innovations within culturing, sequencing, and imaging have led to an expanded understanding of the impact of viral populations. The greatest advances have come from studies where many technologies overlap, such as in well cultured organisms and their phage. One such study applied a multi-faceted approach to determine that the viral role is not limited to the viral shunt but can sometimes contribute to the biological pump, in what is now called the viral shuttle[36,37]. The biological pump is an essential biogeochemical feature of the carbon cycle where organic material produced in the surface waters is sequestered in the deep ocean and therefore removed from the ocean's interaction with the atmosphere. The mechanism driving carbon flux to the deep ocean was previously believed to be the sinking of large heavily ballasted cells, such as

diatoms[38,39] and coccolithophores[40], or due to zooplankton behavior and interaction with organic material[41]. Broadscale sequencing efforts from the surface and mid-depths, paired with an underwater video profiler (UVP), were able to determine that *Synechococcus* phage, which infect tiny bacterial photoautotrophs, were one of the highest correlating entities across the world's oceans with increased carbon export at depth (150 m)[36]. These results shifted the viral significance paradigm and our understanding of the ocean's potential response to, and impact on, the world's changing climate. These conclusions made from this work were dependent on understanding which host a *Synechococcus* phage could infect and classifying the database sequences by aligning to characterized cultures.

Many rich ecological conclusions of high-throughput sequencing efforts of viral communities come from data for viruses where the host is known. An abundant catalog of marine *Prochlorococcus* and *Synechococcus* cultures has allowed for the accrual of an equally abundant number of cultured marine cyanophage[22,23,42]. Laboratory experiments paired with isolated cyanophage genomes reveal that some cyanophage have a more virulent, highly specific infection pattern, while others are less virulent and more promiscuous within the Cyanobacteriaceae classes[43,44]. High-throughput environmental analyses, inherently dependent on knowledge of host and virus combined, have been deployed, revealing dramatic increases in abundance across every season, year after year, of the more promiscuous less virulent clade[45].

If not for *Prochlorococcus* cultures and their isolated phage *Prochlorococcus phage P-SSP7*, it would have been much more difficult to lay claim to the fact that viruses carry unique copies of host auxiliary metabolic genes (AMGs), which are expressed and translated during lytic viral infections, and supplement host growth critical for viral replication[46]. Furthermore, the continued characterization of AMGs in isolated viruses of clonal cultures has assisted in

understanding AMG sequences found in viral genomes in the environment[22,47,48]. The knowledge gained from well cultured genera from the marine environment has led to the discovery of AMGs across viral populations of unknown hosts as well, including AMGs for denitrification, sulfur-cycling, revising key carbohydrate metabolism pathways, cellular defenses, motility, stress responses and methane cycling[49–51]. The presence, abundance and expression of viral AMGs of known hosts suggests that virally infected microbial cells, or virocells, have a unique biogeochemical impact on marine systems and the cycling of nutrients[52,53].

If some of the greatest advancements in ecological theory of viruses is predicated on the previous assignment of virus-host relationships, how do we approach the vast majority of environmental viral sequences for which a host cannot be predicted[34,54–56]? The sequencing of microbial communities over the last two decades has led to databases of genomes from mostly uncultured classes or even phyla[57], making the hope to culture unknown marine viruses mostly untenable at present. In the face of a great need to develop cultivation-free techniques to ascertain the hosts of uncultured viruses in the environment several tools have proven their potential.

Genomic signatures and similarities can be mined within the growing amount of sequencing data already collected, allowing for virus-host predictions on a massive scale, that utilizes previously completed sequencing efforts and machine learning advances[56,58,59]. These analyses include searching for CRISPR homologous sequences[60] to understand which viruses that bacteria and archaea may have mounted a defense against in the past. Microbial and viral genomes can also be categorized by certain genomic similarities such as GC content, and tetranucleotide frequency, and presence of tRNA sequences. Because a virus and host must have compatible genomes for efficient metabolism during infection, these similarities can often be used to predict hosts from the same environment as the virus[61–63]. However, these analyses provide evidence of

efficient and sometimes even just historical infections rather than a direct link to current activity[64]. For example, for horizontal gene transfer to occur between a virus and host, it must happen through a non-lytic interaction, suggesting that all interactions, not simply the most dramatic, have a potential impact and should be monitored[65].

Two high-throughput techniques have emerged in the last decade for associating viruses with their hosts *de novo* at the time of sampling. Both single-cell amplified genomes (SAGs) and high throughput chromosomal conformation capture (Hi-C) sequencing allows DNA from viruses and hosts to be paired through sequencing. In SAG sequencing any viral genome on or in the cell may be amplified at the time of the cellular genome amplification as well with some signals within that data, suggesting probable delineation between an adsorbed virus and an infecting virus[66,67]. Hi-C is a sequencing technique that was first utilized to map the 3-D conformation and structure of chromosomes in human and model eukaryote cells to determine the relationship between folding and cell function[68,69]. This mapping is made possible by physically trapping DNA in place with formaldehyde prior to extraction from the cell and subsequently tagging, purifying and ligating those links into chimeric read pairs. These read pairs are used to associate individual contigs from the same assembly that were actually within the same cell at the time of sequencing. Because viral genomes and other mobile genetic elements are often most relevant when within microbial cells, their genomes may be linked with host genomes in a Hi-C library[70–73]. Both methods offer environmental virologists an opportunity to assess the immediate activity and interaction of viruses with microbial cells in their environment.

Understanding viral activity and infection dynamics *in situ*, in addition to *in vitro* with cultured isolates, may be critical to understanding the important effects of viruses on microbial populations and genomic potential. Bacteriophages isolated on clonal cultures in the laboratory

have been shown to have a high host-specificity often not even infecting across genera[74–76], shaping our belief that viruses in the marine environment are highly host-specific. However, do we anticipate that viruses in a complex ocean environment, riddled with substrate, nutrient and host heterogeneity[77], will behave identically to a viral strain continually grown on a monoculture under ideal and controlled conditions for host growth? Contrary to this anticipated conclusion, laboratory experiments show that the specificity between a virus and host can change by growing viruses on mixed host populations or alternating between different clonal hosts[78–80]. Furthermore, analyses of environmental viral communities with a limited or a broad host population vary in host specificity accordingly, supporting that host specificity is a somewhat plastic trait that adjusts in response to a diversity of hosts[81].

SAGs and Hi-C sequencing have been used minimally on environmental samples, but frequently the results have suggested that viruses may behave quite differently *in situ* than we can predict from host cultures and genomic similarities. Moreover, the viral impact may not be limited to a full lytic infection of susceptible hosts, but rather to entering any cell permissive to it. SAG sequencing from a marine sample found sequences similar to previously isolated cyanophage present and possibly replicating in an alphaproteobacterial cell[66]. Additionally, Hi-C has been used to associate viruses with microbes in a range of environments, showing that viral interactions may not be as straightforward as sequence predictions assume, specifically that single viruses may interact with more than one taxonomic class[72,82–85].

This dissertation is founded on the expansion of technological capabilities to answer a basic and fundamental question in viral ecology—with which microbial cells are viruses interacting? The answer to this question can help us determine how viruses contribute to the biological pump, how viruses may contribute to genetic transfer within microbial communities and

how viruses may be affecting the ultimate biogeochemical impact of those communities in marine environments. Applications of this method across many marine environments to further refine virus-host interactions will help us make sense of the sizable and mounting genomic and datasets we are collecting simultaneously.

In my first chapter Hi-C sequencing techniques are applied to both particle-attached and free-living samples of microbial communities in a seasonally hypoxic estuary below the euphotic zone to understand the interplay of virus-host interactions and particles distributed to depth from the surface. In Chapter Two, the Hi-C method is applied to a sample from a deep chlorophyll maximum in the Pacific Ocean. In this sample Hi-C links confirm previously reported virus-host interactions, but more importantly provide evidence of that infection occurring at the time of sampling. In Chapter Three, this method is used to understand virus-host interactions in an oxygen deficient zone where almost none of the microbes have been cultured, and both viral and microbial populations are uniquely adapted to anoxic water. Together these chapters elucidate the nature of virus-host interactions *in situ* in a way that metagenomic sequencing alone cannot, pushing the field towards a new wave of research in marine viral ecology.

Chapter 1.

HI-C ASSEMBLED GENOMES OF ESTUARINE POPULATIONS REVEAL VIRUS-MICROBE ASSOCIATIONS AND A BROAD INTERACTION RANGE OF A CYANOPHAGE

Rathwell, CR, McKay, C, Rocap, G

1.1 Abstract

Aquatic microbes play key roles in global biogeochemical cycles and their viral-induced mortality influences the flow of carbon and nutrients between the dissolved and particulate pools. However, many microbes remain uncultivated, hindering understanding of their metabolic capabilities and preventing isolation of viruses that infect them. Here we augment metagenomic sequencing with Hi-C, a proximity-linkage method whereby DNA within a cell is physically bound and then sequenced to link contigs within a metagenome that originated from the same cell. In a size-fractioned water sample from beneath the euphotic zone in a hypoxic estuarine fjord in Puget Sound, WA we resolved 49 proximity-linked bins above 50% complete, including 21 Hi-C Assembled Genomes (HAGs) over 90% complete and a nearly complete genome of the eukaryotic green alga *Picochlorum*. Viral and microbial sequence within the same HAG identified 18 virus-microbe interactions. A myovirus and a siphovirus were associated with 2 different genera within the Saltatorellus clade of Planctomycetes, a phylum for which no virus has been identified. A partial *Phycodnaviridae* genome linked to Haptophyte sequence is consistent with contemporaneous observations of a dissipating coccolithophore bloom. A cyanophage S-CAM7-like sequence had a broad interaction range. It was associated with a partial *Synechococcus* genome

in the >3.0 µm size fraction and with a Gammaproteobacteria related to *Alcanivorax* in the 0.2µm-3.0µm fraction. We suggest that viruses produced in surface waters that are shuttled to depth on sinking aggregates may interact with different hosts in deeper waters, providing an important avenue for gene transfer across broad taxonomic ranges.

1.2 Significance

Aquatic microbes are important in global elemental cycling. Knowing which viruses infect them in the environment remains a challenge. Using Hi-C, a molecular technique to physically link DNA within a cell, we assembled nearly complete genomes of both prokaryotes and eukaryotes from a hypoxic estuary. Hi-C links captured virus-host interactions for known virus-host pairs and for hosts with no previously known viruses. The same virus was linked to two distinct microbes in different size fractions of water, suggesting it has a broad host range. Viral lysis in surface waters generates sinking particles that deliver newly produced viruses to deeper waters where they interact with different potential hosts, providing an opportunity for gene exchange between unrelated microbes.

1.3 Introduction

Marine microbes influence global biogeochemical cycles through carbon fixation, oxygen production and the sequestration of carbon to the deep ocean via the biological pump[86,87]. Within microbial communities, viruses play a large role in the distribution and composition of organic material[88]. Both heterotrophic and autotrophic microbial populations experience infection and undergo lysis, redistributing nitrogen- and phosphate-depleted organic material into dissolved pools in the viral shunt[19,20,89]. However, viral infection and cell lysis can also cause increases in particle aggregation and sinking, strengthening the biological pump in the viral shuttle[36,37,90]. Viruses alter the genomic potential of microbial communities as agents of density-dependent

mortality of specific hosts[18,20,28,91] and as vehicles for the exchange of genetic material[92,93]. Furthermore, a virocell expresses both viral and microbial genes, resulting in altered functions and contributions to biogeochemical cycling[53,94,95].

Unfortunately, the majority of environmental viruses remain uncultivated and uncharacterized. The fraction of viruses that have been characterized through cultured isolates can be identified in the environment for presence, adsorption, infection rates, and transcription, which can lead to insights into viral ecology[96–99]. In the absence of an isolated reference, growing amounts of metaviromic sequence data[34] have demonstrated the vast genetic diversity of viruses but are less well able to connect them to specific biogeochemical functioning. A critical remaining need is to determine which viruses are interacting with which hosts in the environment, in order to fully define the lifestyles and behaviors of viruses within different environmental contexts.

Isolation of novel viruses relies on having a cultivated host, yet approximately 81% of all microbial cells on Earth are from an uncultivated class or genus[57]. In the marine environment, although some of the most abundant organisms are in culture, single amplified genomes (SAGs) and metagenome assembled genomes (MAGs), make up over 40% of the genomes available for microbial species in the Genome Taxonomy Database, and these techniques are now commonly used to characterize uncultivated communities[100–111]. SAGs sample a very small volume, can identify microscale variation within a bacterial population, and can elucidate virus-host interactions for individual cells[66]. MAGs are derived from bulk environmental samples and provide population level characteristics of communities[112]. However, MAGs are unable to link viral genomes with host genomes as sequences are binned through nucleotide signatures such as tetranucleotide frequency and sequence coverage across samples, both of which may differ more between viruses and their hosts than among microbial populations[112].

High throughput chromosome conformation capture (Hi-C)[68] is a promising technique for binning sequences within a metagenomic assembly that can link mobile genetic elements, such as viruses, with host metagenomes[71,113,114]. The Hi-C method involves crosslinking DNA in a live population using formaldehyde to hold DNA shape in place. Next, DNA is sheared and ligated in a dilute reaction, allowing DNA that was proximal in a cell but perhaps distant along the linear genome sequence, or even from a separate DNA molecule, to be joined together. These junctions are labeled, concentrated, and sequenced. This approach results in short DNA sequence reads that are composed of two different regions of DNA which were physically close to each other in the live cell. Hi-C has been used extensively to determine the scaffolding of chromosomes[69,115], demonstrating how inter-chromosomal contacts might affect cell function[116]. This method has also been applied to cluster contiguous sequences (contigs) of a mixed culture metagenome into individual genomes[71,113,114]and Hi-C has been extended to more complex communities to cluster over 900 MAGs from a rumen microbiome metagenome[117] and 428 MAGs from sheep gut microbiomes[118]. Hi-C has also been used to investigate virus-host linkages within a cattle rumen microbiome[72], to capture actively replicating viruses in the human gut microbiome[70] and has detected broader host ranges of viral elements in a sheep gut microbiome than previously seen through cultured isolates[118]. Here we explore the potential of Hi-C in the aquatic environment, home to complex microbial communities that are much more dilute than mammalian gut microbiomes. We applied Hi-C linkages to two size-fractionated metagenomes in Hood Canal, Washington, and used a conservative Hi-C clustering approach to improve binning of microbial metagenomic sequences and identify specific viruses and the potential hosts they are interacting with in the environment.

1.4 Results

1.4.1 Sampling and Proximity Linkage-Enabled Bin Generation

Hood Canal is a long narrow estuary connected by a shallow sill to Puget Sound, Washington at its northern entrance (Figure 1A), resulting in relatively slow flushing times and seasonal hypoxia in the southern end. The water in southern Hood Canal (bottom depth 35 m) on August 15, 2016, was sharply stratified with a shallow (<10 m) mixed layer and chlorophyll maximum at 5 m; water at the sampling depth of 25 meters was hypoxic (0.5 mg/L, or 15.5 μmol/kg^3; Figure 1B-C). Continuous data collected by the Twanoh Bay ORCA mooring in Hood Canal indicated that the bottom waters had been hypoxic for over a month (Figure 1B). Additionally, earlier in July a coccolithophore bloom, visible from space, had established in Hood Canal (Figures 1A and S1), and was visually confirmed to be present at time of sampling. Virus like particle concentration at 25 m was 8.76 x 10^8 ml^{-1}, which was 18.4 times that of microbial cells (4.77 x 10^7 ml^{-1}) (Figure 1D). Independently size-fractionated samples from 25 m were used to generate both Hi-C sequence libraries and standard short read metagenomic sequence libraries (Table S1).

We used two different established methods to cluster assembled contigs based on proximity linkages as well as a linkage-independent method. MetaBat attempted to cluster all contigs greater than 1500 bp and generated 165 and 74 bins from 107,590 and 54,957 contigs in the 0.2–3 μm and >3 μm size fractions respectively (Figure 2A-B, Table S2). Proximity linkage-enabled binning employed only the subset of contigs that were part of a Hi-C read pair (37.5% of contigs in the 0.2–3 μm library, and 18.9% in the >3 μm library; Figure 2A-B Table S2). HiCBin generated 65 and 38 bins from 56,179 and 38,116 contigs in the 0.2–3 μm and >3 μm size fractions respectively. In contrast, the ProxiMeta algorithm which also utilized the Hi-C read pairs incorporated more contigs than HiCBin and resulted in more bins; 301 and 261 bins from 85,156 and 50,925 contigs

in the 0.2–3 μm and >3 μm size fractions respectively. None of the ProxiMeta bins were more than 10% contaminated as determined by checkM, while HiCBin resulted in 17 bins that were greater than 10% contaminated and MetaBat generated 51 bins greater than 10% contaminated (Figure 2E-F).

Because a central goal of this study was to connect viral sequences with their potential hosts, we used a conservative approach to generate the final bins, prioritizing stringency of proximity linkages over maximizing bin size. First, the ProxiMeta bins were further refined using a Giant Component analysis to retain only contigs that were part of the most connected web in each bin (Figure S2). Like the original bins, these ProxiMeta Giants were not contaminated above 10% (Figure 2 E-F). Because HiCBins were generally larger than ProxiMeta bins but more likely to be contaminated (Figure 2 E-F), we combined the more aggressive clustering of HiCBin with the strict filtration of the ProxiMeta clustering to create intersection bins (Figure S3) and thus enhanced completeness without having contaminated bins. This approach resulted in 170 final bins (98 in 0.2–3 μm and 72 in >3 μm), none of which were contaminated above 10%. Of those 170 bins, 21 are high quality draft genomes (>90% complete, <5% contaminated), 28 are medium quality (>50% complete, <10% contaminated) and the remainder can be considered low quality draft genomes (<50% complete, <10% contaminated; Figure 2 E-F, Table S2).

The overall representation of each taxonomic class among assembled contigs, linked contigs and contigs binned using the various methods was generally similar, although there were some notable exceptions (Figure 2 C-D). In both size fractions Planctomycetes represent a much larger fraction of contigs with Hi-C links than of total contigs: in the >3 μm library Planctomycetes represent 10% of total contigs assembled, 21% of contigs with links, and 41% of contigs in final bins (Figure 2D). In contrast, in that same >3 μm library, Bacillariophyta contigs made up 25% of

assembled contigs, but only 7% of linked contigs and only 1-5% of any subsequent linkage-dependent binned contigs (Figure 2D). The 170 final intersection bins were assigned to 19 different phyla with the majority of bins identified as Proteobacteria (70 bins) or Planctomycetes (34 bins) (Figure 2 G-H). Bins greater than 50% complete spanned 13 different phyla assignments.

1.4.2 High Quality Hi-C Assembled Genomes

The 21 Hi-C assembled genomes (HAGs) that were >90% complete and <5% contaminated, included 7 Planctomycetes, 6 Proteobacteria, 3 Verrucomicrobia, 3 Actinobacteria, and 2 Bacteroidetes (Table S3). Notably, the 6 bins classified as Proteobacteria represented 3 pairs from different orders (*Oligoflexales*, *Cellvibrionales* and *Rhodobacterales*) in which a highly similar genome (>99% ANI within each pair) was independently assembled in each size fraction (Figure S4). The *Oligoflexales* (IB_0.2_006 and IB_3.0_013), whose class *Oligoflexia* has recently been reclassified as a new phylum *Bdellovibrionota*[119], each had a best blast hit to the predatory bacterium *Pseudobacteriovorax antillogorciicola*. The *Cellvibrionales* bins (IB_0.2_005 and IB_3.0_006,) were members of the genus *Halioglobus* and the *Rhodobacterales* bins (IB_0.2_003 and IB_3.0_004) were classified as *Rhodobacteraceae.*

High representation of Planctomycetes in contigs with Hi-C links enabled assembly of 7 high quality draft genomes and 7 medium quality draft genomes that represent at least 8 distinct Planctomycetes populations from five major clades including *Pirellula*, *Bythopirellula*, *Gimesia*, *Phycisphaerae,* and the newly cultured *Saltatorellus* clade[120] (Figure 3). Within the Planctomycetes there were also cases of independent assembly of closely related genomes in each size fraction, including one pair most closely related to *Mariniblastus fucicola* (IB_0.2_001, IB_3.0_005, 99.4% ANI), as well as a pair in the *Bythopirellula* clade (IB_0.2_007, IB_3.0_010, 99.2% ANI; Figure 3, Table S3). Proximity linkages also enabled lineage resolution of closely

related taxa in the same size fraction in the *Saltatorellus* clade (IB_3.0_008, IB_3.0_014). Phylogenetic placement of concatenated core Planctomycetes genes from the HAGs frequently places them closest to MAGs from metagenomic sequencing in the Mediterranean Sea[100] (named TMEDX in Figure 3). This placement is not surprising, as many cultured Planctomycetes isolates come from populations present in freshwater, soil, peat moss and even plumbed water, and may not be the best representatives of marine populations[121]. There were no bins of any quality identified as members of the *Candidatus* Brocadiaceae clade, which contains organisms capable of performing anammox.

In addition to the 21 high quality draft genomes of prokaryotes, two ~12.4 Mb bins, one from each size fraction (IB_3.0_017, IB_0.2_010), each contained a nearly complete genome of *Picochlorum,* a unicellular green alga from the phylum Chlorophyta (Figure 2G). Phylogenomic analysis indicates that they are closely related to each other (99.86 % ANI) and to *Picochlorum* NBRC102739 (Figure S5). Although these two bins were judged as 47–77% contaminated using checkM, which employs a marker set intended for bacterial and archaeal genomes, analysis using BUSCO[122] with a set of 1519 Chlorophyta specific gene markers suggests that they were approximately 86% complete with less than 1% of the single-copy orthologs duplicated. This completion statistic is in the range of complete published *Picochlorum* genomes from cultured isolates, 68–96% (Table S4) and may be an underestimate, as the gene prediction on these contigs using prodigal likely failed to fully capture accurate eukaryotic gene models. Thus, as with the three pairs of Proteobacteria and two pairs of Planctomycetes, proximity linkages were able to independently assemble nearly complete and nearly identical eukaryotic genomes from the two size fractions.

1.4.3 Proximity Linkages Identify Viral Host Associations

Virus-host associations were examined by searching for viral sequences within the final proximity linkage bins, which were generated with a conservative binning approach to minimize contamination (Figure S3). We identified 18 bins that contained >10 kb of viral sequence and that could be classified with DemoVir (Table S5, Figure 4). Viruses identified included *Myoviridae* (7)*, Siphoviridae* (3), *Podoviridae* (2)*,* unclassified Caudovirales (2) and *Phycodnaviridae* (4). CheckV determined two of these viral sequences to be potential prophage (Figure 4). We used the taxonomy of the non-viral sequence in each bin to determine the microbial populations with which each viral genome interacted. The HAGs containing viral sequence varied in size and completeness of both host and viral genomes and included five of the high-quality draft genome bins as well as six bins that had minimal host genomes (3% complete or less; Figure 4, Table S3). The majority of the viral contigs within HAGs were not placed in bins by the Hi-C independent binning method MetaBat2 (Figure S6)

Although viruses infecting marine or estuarine Planctomycetes have not been previously identified, 4 Planctomycetes bins ranging in size from 2.5 Mb to 7.8 Mb contained viral genomes (Figures 3 and 4). In the 0.2–3 μm size fraction two *Myoviridae* viruses with many viral genes were found in IB_0.2_018 (*Gimesia* clade) and IB_0.2_014 (*Saltatorellus* clade) (Figures 3 and 4). IB_0.2_018 contained genes encoding RNA polymerase, portal protein, several peptidases, nuclease, a major capsid protein, lysins and DNA polymerase (Figure S7). IB_0.2_014 had fewer canonical viral genes but included genes encoding the terminase large subunit, a major capsid protein, DNA recombination, helicases and an integrase (Figure S7). In the >3 μm size fraction IB_3.0_008 is a near-complete HAG related to *Saltatorellus ferox* containing *Siphoviridae* including genes for many structural and assembly proteins, such as major capsid protein, portal protein, 6 tail proteins, tape measure protein, terminase and head assembly protein (Figure S7,

Table S6). Finally, also in the >3 μm size fraction, a potential prophage was identified in IB_3.0_003, a near-complete HAG closely related to *Mariniblastus fucicola*. This HAG contained a *Myoviridae* sequence including genes for Thy1, multiple terminase units, tail protein, RNA polymerase, portal protein, DNA polA, and more (Figures 4 and S7). Although Pfam searches conclusively identified canonical viral genes in all four of these bins, Blastn searches against viral RefSeq resulted in very low identity matches, generally to non-marine viruses (i.e., *Bacillus phage, Pseudomonas phage*, *Caulobacter phage*), suggesting that these are novel viruses with no reference genomes sequenced yet.

Four of the 18 identified viruses, all from the >3 μm size fraction, were associated with photoautotrophic hosts, even though the sampling depth of 25 m was below the euphotic zone (Figure 1C, Figure 4). Specifically, the chlorophyll trace is absent as shallow as 6 m, so we don't anticipate a growing population of photoautotrophs at the sampled depth of 25 m. A *Phycodnaviridae* was identified in the *Picochlorum* bin IB_3.0_017, and a Caudovirales was identified in a very incomplete *Synechococcus* bin (IB_3.0_058) that contained nearly 10% viral sequence (Table S3, Figure 4). A third bin, IB_3.0_05, contained majority viral sequence, a *Phycodnaviridae* with 3 viral contigs totaling 92.8 kb. A smaller fraction of this bin sequence was identifiable as microbial in origin and included a plurality of Haptophyte contigs (6 contigs, 88.3 kb) classified in 3 different orders (*Isochrysidales, Prymnesiales* and unclassified Haptophyta) as well as 4 unclassified contigs (81.2 kb) and 4 others that were identified as some other Eukaryote, each a different one (51.4 kb). Genes on viral contigs in IB_3.0_055 most frequently had a blastx against nr best match to *Emiliana huxleyii* virus *EhV86*. However, percent identities ranged from 23% to 47%. Overall, these results suggest the presence of a haptophyte and its associated virus, both of which are only distantly related to sequences currently in the database. Their presence is

consistent with contemporaneous observations of a coccolithophore bloom occurring in the sampled waters (Figures 1A and S1).

The fourth photoautotroph and virus bin, IB_3.0_51, was also predominantly viral and contained a *Myoviridae* (98.9 kb, 69 contigs) most closely related to the myovirus, *Synechococcus* phage S-CAM7, which has a 214 kb genome and was isolated off the Southern California Pacific Coast on *Synechococcus* strain WH7803[47] (Figure 5). Also in this bin are 21 contigs (80.1 kb) most closely related to clade I *Synechococcus* strain ROS8604 (Figure S8). As observed generally with viral sequence with HAGs (Figure S6), both HiCBin and ProxiMeta were able to pair *Synechococcus* contigs with S-CAM7-like viral contigs, but MetaBat2 was unable to place most of the S-CAM7-like viral contigs into any bin. Though this bin was very incomplete, there were 5 other incomplete clade I *Synechococcus* bins present across both size fractions, ranging in size from 15 kb to 160 kb, with a checkM completeness of 0–6% (Table S3, Figure S8). Furthermore, the viral sequence in IB_3.0_51 was one of two of the viruses identified that had more than 50 genes for canonical viral proteins as determined by Pfam searches (Figures 4 and S7), and possessed genes for tail proteins, portal proteins, phage lysozyme, peptidases, neck proteins, major capsid proteins, terminase, helicases, head assembly proteins and baseplate proteins.

An additional viral genome closely related to *Synechococcus* phage S-CAM7 was identified in the other size fraction (0.2–3 μm); however, this *Myoviridae* genome (122.2 kb, 9 contigs) was present in a bin (IB_0.2_031) containing 1.65 Mb of a Gammaproteobacterium most closely related to *Alcanivorax* (Figures 4 and 5). The viral sequence within this bin also contained more than 50 canonical viral proteins, most with high percent identity (63–100%) to both the S-CAM7 0910SB42 genome and to the S-CAM7-like viral sequence in IB_3.0_51 (Figures 5 and S7). Both S-CAM7-like viral genomes contain auxiliary metabolic genes (AMGs) previously

described in cyanophage (Table S7). Notably, each virus contained the photosynthesis-related AMGs for *cp12*, a gene responsible for inhibiting the Calvin Cycle during infections in the light[123], and *cpeT*, a bilin lyase, important for pigment adaptation in the freshwater cyanobacteria *Fremyella diplosiphon*[124]. These two genes are commonly reported in cyanophage genomes but are not present in the S-CAM7 0910SB42 genome[47].

1.5 Discussion

1.5.1 HAGs and Virus-Host Interactions Made Possible Through Hi-C Sequencing

Despite their significant impact on biogeochemical cycles most members of aquatic microbial and viral communities remain uncultivated[34,57]. In order to characterize uncultivated microbial and viral populations and their intracellular interactions in the environment, we generated Hi-C proximity linkage libraries in tandem with bulk metagenome samples from two size fractions in the deeper, more saline layer of Hood Canal estuary in Puget Sound, WA. Using a stringent approach to linking contigs with Hi-C read pairs, we generated 49 bins greater than 50% complete and 21 bins qualifying as high-quality draft genomes[125]. Among the high quality HAGs were 5 pairs of closely related genomes (>99% ANI within each pair) independently assembled in each size fraction, provided an unintended replication of the method. In addition, two nearly complete genomes of the green alga *Picochlorum*, one in each size fraction, demonstrate the potential of Hi-C for assembling genomes of marine eukaryotes. Hi-C was particularly successful in assembling genomes from the phylum Planctomycetes, perhaps because these cells have cytoplasmic invaginations that order their genomes in condensed, membrane-bound, nucleoid structures[126]. The generation of over 20 high-quality draft genomes from one sampling event shows the power of using proximity-links in environmental metagenomics. Enhanced assemblies

will help elucidate the genomic potential of uncultivated microbes and can help bring more microbial taxa into culture, expanding available hosts for isolation of marine viruses[127].

We specifically use the term HAG for the proximity linkage bins, to differentiate them from MAGs, since the Hi-C links enable binning of non-contiguous sequence (such as distinct chromosomes or plasmids) and even unrelated genomes that co-occur within the same cell. Indeed, some proximity linkage bins contained both viral and microbial sequences, potentially indicating the presence of viruses within cells, which ultimately allowed for designation of virus-microbe associations *in situ* for 18 HAGs. The majority of the viral contigs within HAGs were not placed in bins by Hi-C independent binning methods, suggesting the unique ability of Hi-C to capture virus-host interactions from disparate sequences in a metagenome.

Although Hi-C sequencing is a promising window into virus-microbe interactions in the environment, some caution is warranted in interpreting sequences from HAGs. First, as with all methods aiming to detect viruses based on sequence, viral and microbial sequence can be hard to disentangle. Second, shorter contigs may not allow distinguishing between an infecting virus and a prophage integrated in the genome or present on a plasmid, as all will be linked with the microbial HAG using Hi-C reads, as observed in two cases here. Finally, as one HAG represents a community of microbial cells, if multiple viruses were infecting a given population at a time, one HAG could potentially contain multiple viral genomes. Closer inspection of viral sequences found within HAGs, including DemoVir assignment, Pfam assignment of genes, examination of single copy canonical genes and alignment of genomes, was necessary to identify the 18 virus-microbe associations here. These included four Planctomycetes HAGs ranging in size from 2.5 to 7.8 Mb containing viral genomes (3 *Myoviridae* and a *Siphoviridae*), the first potential identification of

viruses infecting marine Planctomycetes, demonstrating this method's ability to assemble viral genomes for greatly-under characterized viral populations.

1.5.2 Phytoplankton and their Viruses Delivered to Aphotic Depth on Particles

The viral shuttle hypothesis proposes that viral infection of surface primary producers contributes to particle flux to depth[128]. Coccolithophore blooms experiencing viral infection contribute heavily to particulate carbon flux to the deep ocean[37], in part due to the ballast of their disc-like structures on the cell surface but potentially also due to the aggregation of cell material post-lysis[90]. The abundance of picophytoplankton such as *Synechococcus* and their viruses are also associated with generation of particles at depth[36]. In Hood Canal, at a sampling depth of 25 m, below the stratified surface layer and just 10 m above the seafloor, 4 bins in the >3 μm size fraction were composed of photoautotrophs associated with viruses, supporting the viral shuttle hypothesis[88]. Here the two bins containing both *Synechococcus* and cyanophage were in the >3 μm size fraction, instead of the 0.2–3 μm fraction where free-living *Synechococcus* might be expected. The assembly characteristics of some of these bins also appear to suggest a progressed lytic infection. The Haptophyta and one of the *Synechococcus* HAGs were predominantly viral, with very little host sequence present in the final bin, which might occur in the later stages of infection as viral genomes are replicated and host genomes are degraded[129]. To further support this hypothesis, model virus-host systems could be used in the laboratory to characterize the precise Hi-C linkage signal across an infection cycle with a view towards linking Hi-C sequencing data to stage of infection or type of viral interaction *in situ.*

1.5.3 S-CAM7 Synechococcus Phage Interacts With a Broad Range of Hosts

The proximity links we obtained in this study identified previously characterized virus-microbe associations[47,130] *in situ* while also revealing alternate virus-microbe interactions. In the particle-

associated size fraction a cyanophage closely related to *Synechococcus* phage S-CAM7 paired with a clade I *Synechococcus*. S-CAM7 was isolated on clade V *Synechococcus* strain WH7803, but the ability to infect across clades of *Synechococcus* is not unexpected for a generalist cyanophage[131]. However, in the free-living fraction a virus even more closely related to S-CAM7 was linked to microbial sequence related to *Alcanivorax,* a genus of hydrocarbon-degrading Gammaproteobacteria[132]. Although this interaction may be surprising, there is some precedent for Proteobacteria to be interacting *in situ* with cyanophage, as SAG sequencing has identified cyanophages sequences related to S-CAM7 in *Roseobacter*[66]. Laboratory experiments have demonstrated that a generalist cyanophage may enter the cells of a resistant cyanobacterial strain and replicate the viral genome and assemble capsids, yet not complete a lytic infection[131]. The Hi-C linked sequences presented here cannot determine whether S-CAM7-like viruses in Puget Sound are producing a lytic infection in *Alcanivorax* populations but are evidence that the virus had entered these cells. Notably, S-CAM7 carries the fewest photosynthesis AMGs of cultured cyanomyoviruses[47], which might occur if the evolutionary landscape of this virus included a broad range of interactions[80]. Future work assessing the ability of cultured S-CAM7 to enter non-cyanobacterial hosts and replicate or generate capsids is required to understand the evolutionary constraints on this viral population. Whether this S-CAM7 interaction with *Alcanivorax* represents a truly lytic infection or an aborted lytic infection, it provides an avenue for gene transfer across diverse microbial populations.

The majority of aquatic microbial virus-host interactions to date have been witnessed *in vitro*, leading to the hypothesis that viruses in the environment have a narrow host range[74]. However, some laboratory evidence suggests viruses isolated on a single host population may increase their selectivity in future passages but that alternating hosts for the isolation increases

viral host range, even if at an apparent cost to virus fitness[80]. The marine environment differs from cultures in many ways, particularly in its complex community composition, dilute environment and physical transport processes[34]. Viruses produced in surface populations may contribute to particle aggregation and be shuttled to depth where the primary host population is no longer abundant. Here viruses may interact with alternate species, allowing for genetic transfer across a broad range of taxa and potentially selecting for broad host ranges in viral populations[128,133]. Expansion of the use of Hi-C linkages and SAG sequencing[66,67,134] to determine virus-host interactions *in situ*, and assessment of host ranges with cultured isolates will clarify the role of viral host selectivity in influencing both microbial population dynamics and carbon cycling and flux of carbon to depth in the ocean.

1.6 Materials and Methods

1.6.1 Collection Site and Environmental Conditions

Samples were collected from Twanoh Bay in Hood Canal in Puget Sound in Mason, Washington on August 15, 2016 from 4 pm to 6 pm. Sampling was done adjacent to an Ocean Remote Chemical Analyzer (ORCA) profiling mooring[135], at 47° 22.5'N, 123° 0.5'W. The bottom depth at this site is at 35 m. Water was collected from 25 m depth with a Niskin bottle deployed from the deck of R/V *Mackinaw*, a vessel used to service the ORCA moorings.

1.6.2 Epifluorescent Microscopy

Fifty milliliters of water were collected from 25 m and processed using established methods for enumeration via microscopy[136,137]. Water was treated with microscopy grade paraformaldehyde, immediately frozen on deck in liquid nitrogen, and stored back in the lab at –80°C until it could be processed further. The sample was thawed and filtered onto two 0.02 μm Anodisc filters. Dried

filters were incubated in 2.5x concentration SYBR Gold for 25 minutes in the dark. Excess dye was removed, and filters were mounted on a slide using 50:50 Glycerol/1X TE. Samples were viewed at 1000x under excitation with a X-Cite 120 bulb. Images were captured of fields of view using the Nikon Eclipse 80i. Fluorescing regions were classified as microbial or virus like particles based on size and counted across 20 to 200 fields of view. Pixel to micrometer conversion was calculated using a 10 μm stage micrometer ruler at equal magnification. Sample counts were extrapolated to entire surface area of filter.

1.6.3 Shotgun Metagenomes

Twenty liters of seawater were collected and kept cold for less than 4 hours during the commute back to the lab, where they were then immediately filtered. The sample was filtered sequentially through 3.0 μm and then 0.2 μm polycarbonate filters, using a peristaltic pump and 142 mm inline filter rigs. Filters were frozen at –80°C until further processing. DNA was extracted from the 0.2 μm and 3.0 μm filters with standard phenol chloroform methods. Extracted DNA was cleaned and purified with ethanol precipitation. Genomic DNA was processed with the Nextera XT Library Preparation kit. At least 3 replicate PCR amplification reactions were pooled with less than 10 cycles each to add Illumina adapter sequences and barcodes. Sequences were read on the Illumina NextSeq 500 platform with a NextSeq high output 300 cycle kit; 150 bases were read in each direction in addition to the 8 bp barcode used to differentiate reads from different size fractions. Metagenomic samples were sequenced at 207 and 156 million reads for the 0.2–3 μm, and >3 μm size fractions, respectively (Table S1).

1.6.4 Hi-C Libraries

Five liters of water were collected and immediately treated with 1% formaldehyde. After 30 minutes, cross-linking reactions were quenched with the addition 10 mg/L of glycine. Within hours

treated water was size-fractionated onto 3.0 µm and subsequent 0.2 µm polycarbonate filters. Material was then resuspended from the filters in 1x TE and pelleted by centrifugation. Pellets were frozen at –80°C and then delivered to Phase Genomics in Seattle, WA, for Hi-C library processing using established methods[71,138]. Briefly, restriction endonucleases with diverse AT content, Sau3AI and MluCI, were employed to digest formaldehyde fixed DNA in each sample. Digested ends were filled in, biotinylated, and blunt end ligated in a dilute environment to promote self-ligation. Biotinylated junctions were purified[39,52] and sequenced on an Illumina HiSeq 2500, with 2 x 150 paired end reads, resulting in 80.6 million and 56.2 million reads for the 0.2–3.0 µm and the >3.0 µm libraries respectively (Table S1).

1.6.5 Assembly and Hi-C independent Binning

Metagenomic reads were quality-filtered using Trimmomatic, targeting and extracting Illumina adapter sequences and reads with low phred scores[139]. Filtered reads were then assembled using MEGAHIT on default settings, with multiple kmers[140]. The total assembled sequences were 1,147 Mb and 360 Mb with N50s of 1286 and 1454 for the 0.2–3 µm and the >3 µm libraries respectively (Table S1). Quality of assemblies was analyzed using QUAST[141]. Protein coding sequences were predicted using Prodigal[142] (v2.6.3). For each library contig coverage was determined by recruiting reads to the assembly using Bowtie 2[143] (v1.2.2). Assembled contigs in each size fraction were binned independently using MetaBAT 2[112] (v2.15), with a minimum contig length of 1500 bp and default settings otherwise.

1.6.7 Binning with Hi-C Reads

Hi-C read pairs recruited to two distinct contigs were used as proximity linkages and bins were generated with two separate algorithms that employed these linkages. First, ProxiMeta, processed by Phase Genomics in Seattle, WA, utilized adapted MetaPhase methods as previously described[71].

Secondly, the same Hi-C read pairs were run through an open-source pipeline, HiCBin[144]. In both methods contigs greater than 1000 bp were clustered together into bins using a normalized link frequency. In ProxiMeta, link frequency between two contigs is determined by mapping reads to an assembly using Bowtie2[143] and then normalized between contigs based on the number of self-links for each contig, which considers contig size and frequency of enzyme digestion sites. With the HiCBin pipeline, HiC reads were mapped to assembled contigs using BWA-MEM with -5SP setting and HiCZin normalization was employed[145]. Because HiCBin requires coverage and taxonomy information for normalization, taxonomy assignments generated from a custom DIAMOND blastx (described below) were formatted with kaiju using the addTaxonNames script, with subsequent brief formatting modifications to meet HiCBin taxonomy requirements and include virus information[146]. Contig coverage information was determined using the MetaBAT 2 jgi_summarize_bam_contig_depths script. Normalized link frequencies were used by each method independently to cluster contigs into bins. ProxiMeta bins were further refined further to extract the largest connected component or giant component of its network graph (Figure S2) using CINNA and igraph libraries in R[147].

The majority of contigs in each ProxiMeta giant component bin were assigned to the same HiCBin, but HiCBin typically lumped several ProxiMeta giant components together. Therefore, intersection bins were created by filtering ProxiMeta giant components to include only those contigs that went to the same HiCBin cluster and then combining those filtered giant components when the HiCBin method had combined them (Figure S3). Contigs not binned together by both methods were excluded from the final intersection bins. These intersection bins were then assessed for contamination and quality using CheckM. Bins which were >10% contaminated were split back into their ProxiMeta giant components, still filtered for contigs that all were binned in the

same HiCBin cluster. Final bin names were IB_x.x_yyy, where IB means intersection bin, x.x is either 0.2 or 3.0 for size fraction library and yyy is the bin number.

1.6.8 Bin Completeness and Contamination

Bins from every method were assessed for completion and quality using CheckM[148] (v1.1.3) with the lineage_wf command. CheckM results were used to classify bins as high quality (HQ:>90% complete, <5% contaminated), medium quality (MQ:>50% complete, <10% contaminated), low quality (LQ:<50% complete, <10% contaminated) and contaminated (>10% contaminated) as previously described[149]. Two bins identified as *Picochlorum* were also assessed using BUSCO[122] (v5.3.0) using the chlorophyta_odb10 lineage dataset (creation date: 2020-08-05, number of genomes: 16, number of BUSCO-derived single-copy orthologs: 1519).

1.6.8 Taxonomic Identification of Assembled Contigs

A custom database was made by compiling three separate publicly available databases: MarDB, MMETSP, and Viral RefSeq. MarDB is a database of 8000 marine akaryote genomes[150]. The MarDB database was downloaded from the Marine Metagenomics Portal website in February of 2019[150]. The Viral RefSeq database was composed of all available viral genomes from RefSeq (release84). These databases were combined with the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP)[151] with custom additions of zooplankton genomes. Each contig in the size-fractionated assemblies was identified using a translated blast with the fast local aligner DIAMOND[152] (v0.9.24.125) using the Lowest Common Ancestor (LCA) parameter.

1.6.9 Bin Taxonomy Assignment

Best LCA assignments and best phylum assignments were determined for each bin, by normalizing contig taxonomy assignments by length. Each bin was composed of contigs of varying sizes and

best LCA taxonomic matches; therefore, the percentage of total DNA length of a bin identified as a specific taxon was used to determine the most likely taxonomy for every bin. Best LCA taxonomic identities were determined for each bin by the largest percentage of total DNA that was identified as that specific taxonomic ID. All LCA matches from the same Phylum were also summarized so that each bin had a best phylum, when possible, even if individual contigs could have been identified more specifically. If phylum assignment was not possible, highest order of taxonomic assignment possible was given.

1.6.10 Identifying Viruses

Virus sequences were identified within each metagenome using several methods: VirSorter, VirFinder and DIAMOND blastx results. VirSorter[68] (v1.0.5) was run within the CyVerse[69] discovery environment and was applied to all contigs greater than 5 kb. Sequences returned as category 1 (most confident) or category 2 (very likely) were accepted as viral. VirFinder[153] (v1.0.0) was run on Rstudio using default settings against all contigs greater than 5 kb. Contigs determined to be viral by either VirSorter, VirFinder or a virus best hit in the DIAMOND search against our custom database, as previously described, were considered potentially viral. All final intersection bins that contained more than 10 kb of potential viral sequence (27 bins) were investigated further using CheckV[154] (v0.8.1); those identified as viral genomes (from 18 bins) were classified using DemoVir (https://github.com/feargalr/Demovir). Genes on these viral contigs were assessed for protein family hits on the pfam.xfam.org website[155] and blasted against Viral RefSeq (release 211) using command line blastp[156]. Comparison of S-CAM7 virus genomes was generated in Easyfig[157].

1.6.11 Phylogenetic Analysis

Whole genome phylogenies for Planctomycetes, Chlorophyta and *Synechococcus* were constructed in PhyloPhlAn[158] (v3.0.60) using settings –diversity medium and the config_file

supermatrix_aa.cfg. Marker genes for each tree used the BUSCO-derived single-copy orthologs for each lineage (planctomycetes (571 genes), Chlorophyta (1519 genes) Synechococcales (788 genes)). For final phylogenetic trees the best protein model for the PhyloPhlAn generated concatenated alignment (LG+G4+F in all 3 cases) was found using ModelTest-NG[159] (v0.1.7) and then used for a maximum likelihood search using RAxML-NG[160] (v1.1) with 20 distinct starting trees. Trees were visualized and annotated using iTOL[161] (v6.7). Average nucleotide identity (ANI) values were calculated using FastANI[162] (v1.33).

1.6.12 Data Deposition

Sequences have been deposited in GenBank under BioProject PRJNA992579. Hi-C reads SRR25203940-SRR2520394, short reads SRR25203943-SRR25203944.

1.7 Acknowledgements

This work was supported by NSF DEB-1542240 to GR. CR was also supported by a National Science Foundation Graduate Research Fellowship. We thank I. Liachko and Phase Genomics for processing and sequencing of Hi-C libraries. Additionally, we would like to thank the ORCA team and crew of the R/V Mackinaw for assistance in collecting samples and sharing chemical data as well as J. Deming for helpful comments on the manuscript.

CR and GR conceived the study. CR, CM and GR performed analyses. CR and GR wrote the paper.

The authors declare no competing interests.

1.8 Figures

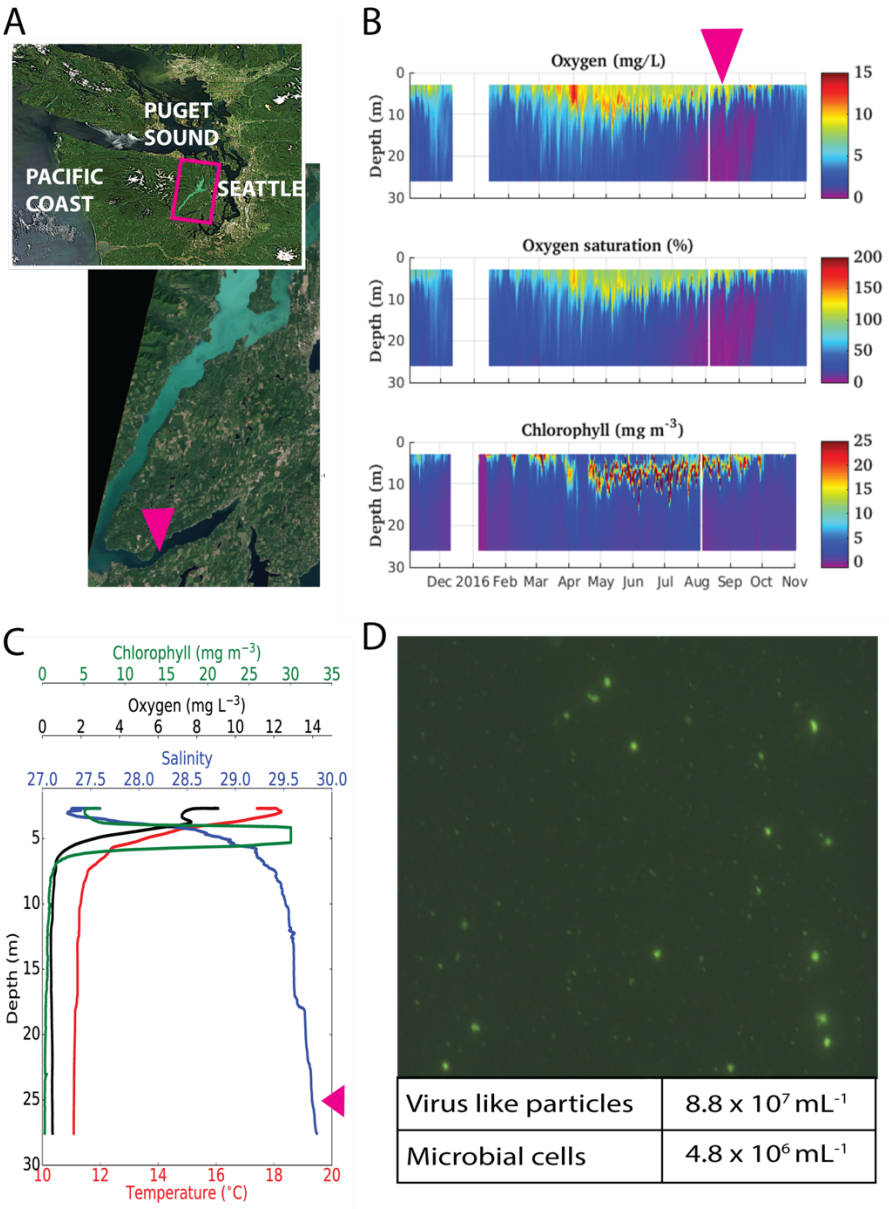


Figure 1. Environmental context. **A**) Satellite imagery from July 26th, 2016, downloaded from NASA MODIS archive, showing turquoise coccolithophore bloom spreading from the north of Hood Canal into the southern reaches of Hood Canal. Pink rectangle and inset map indicates location of Hood Canal. Pink arrow indicates mooring location and sampling site. **B**) Annual data from ORCA mooring located in Twanoh Bay, Hood Canal showing twice daily oxygen concentrations, oxygen saturation and chlorophyll concentrations with depth for 2016. Pink arrow indicates time of sampling. **C**) Depth profile of salinity, chlorophyll, oxygen and temperature from ORCA mooring on day of sampling, August 15, 2016. Pink arrow indicates depth sampled. **D)** Epifluorescent microscopy image of whole seawater at 25 m, with the average virus like particle and microbial concentrations at 25 m.

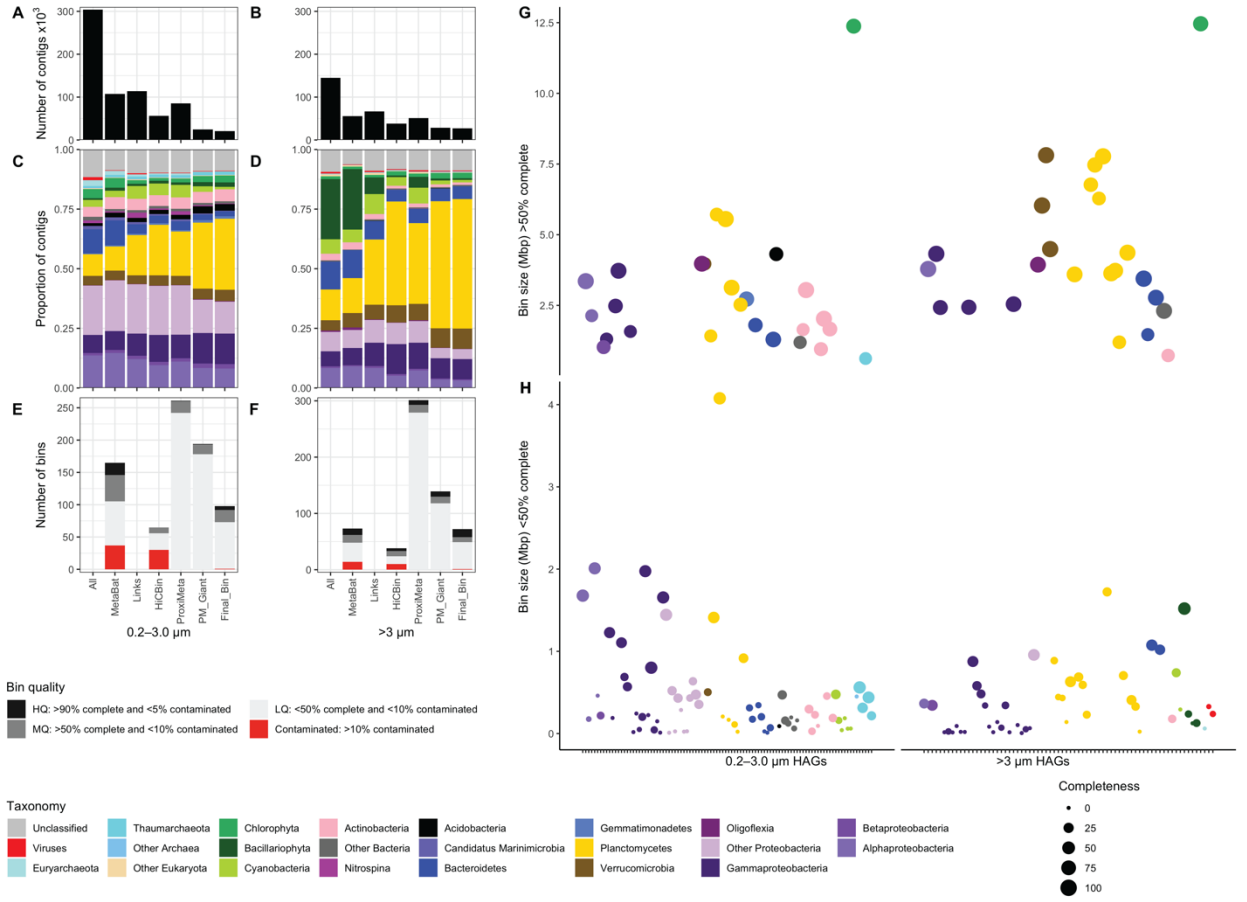


Figure 2. Overview of Hi-C linked bin generation and assessment. **A-B)** Number of contigs greater than 1 kbp for each step including: the initial assembly (All), bins generated without using Hi-C links (MetaBat), contigs containing a link to a Hi-C read (Links), bins generated by HiCBin (HiCBin), bins generated by ProxiMeta (ProxiMeta), ProxiMeta refined giant component clusters (PM_Giant), and the ultimate intersection of HiCBin bins with ProxiMeta refined giant component clusters (FinalBin). Values are shown for the 0.2–3.0 µm library (A) and the >3 µm library (B). **C-D)** Proportion of the contig pool included in each step above assigned to different taxonomy for the 0.2–3.0 µm library (C) and the >3 µm library (D). **E-F)** Summary of checkM data for each binning methods above for the 0.2–3.0 µm library (E) and the >3 µm library (F). Contaminated bins (>10%) are shown in red regardless of completeness and non-contaminated bins are shown in grey and black based on quality classification. **G-H)** Completeness, size and taxonomy of final proximity linkage bins separated by those >50% complete (G) and < 50% complete (H). Each bin is represented by a dot color-coded by DIAMOND blast taxonomy. Dot radius is relative to checkM completeness. Note different scales for y-axis (bin size).

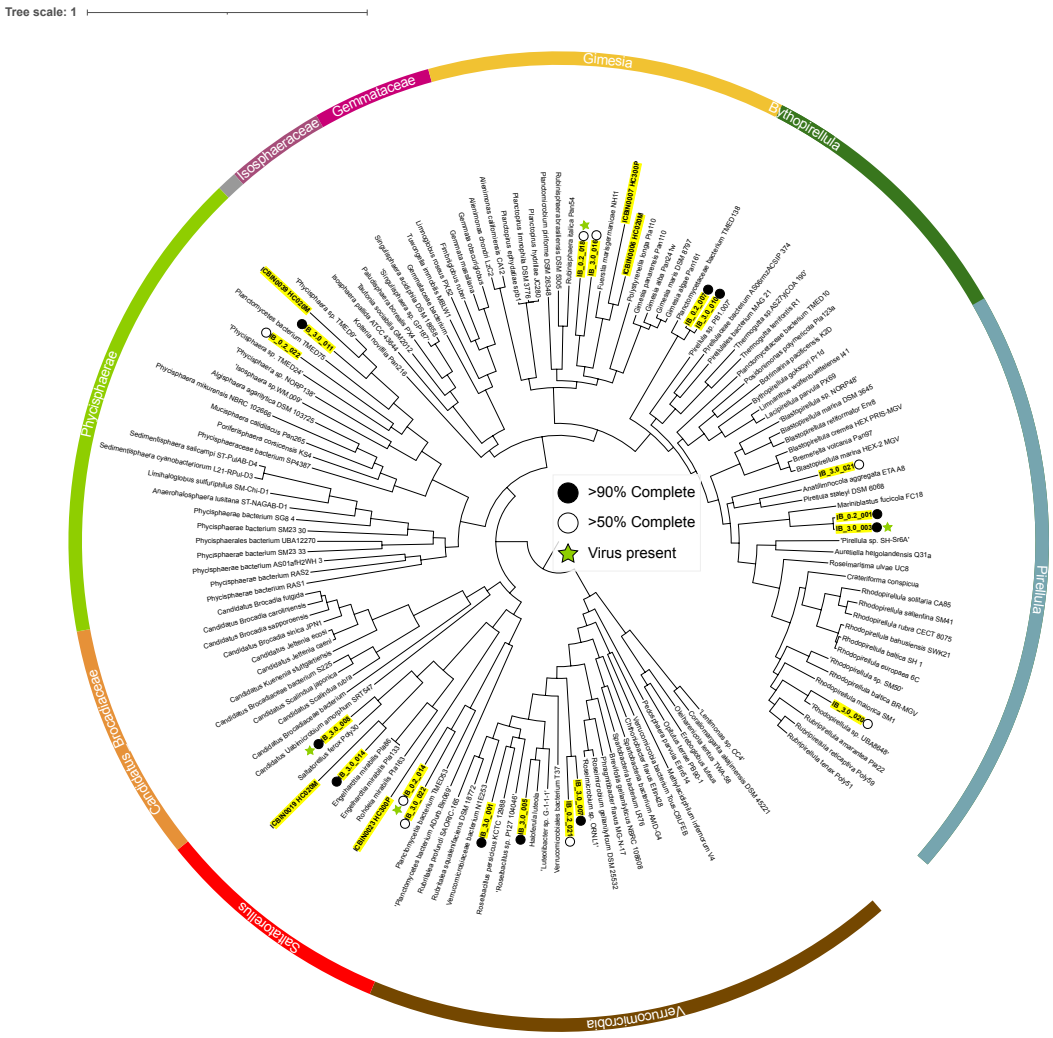


Figure 3. Maximum likelihood phylogenetic tree of Planctomycetes and Verrucomicrobia based on concatenated proteins of 571 single copy genes from the Planctomycetes lineage. The 18 Hi-C assembled genomes >50% complete are indicated as medium quality (MQ; 50-90% complete, 8 HAGs) or high quality (HQ; >90% complete, 10 HAGs). Viral sequence present within a HAG is indicated by a green star. The Verrucomicrobia were used as the outgroup. Numbers at nodes represent % of Felsenstein bootstrap values (100 bootstrap replicates were performed, at which point the replicates were assessed to have converged).

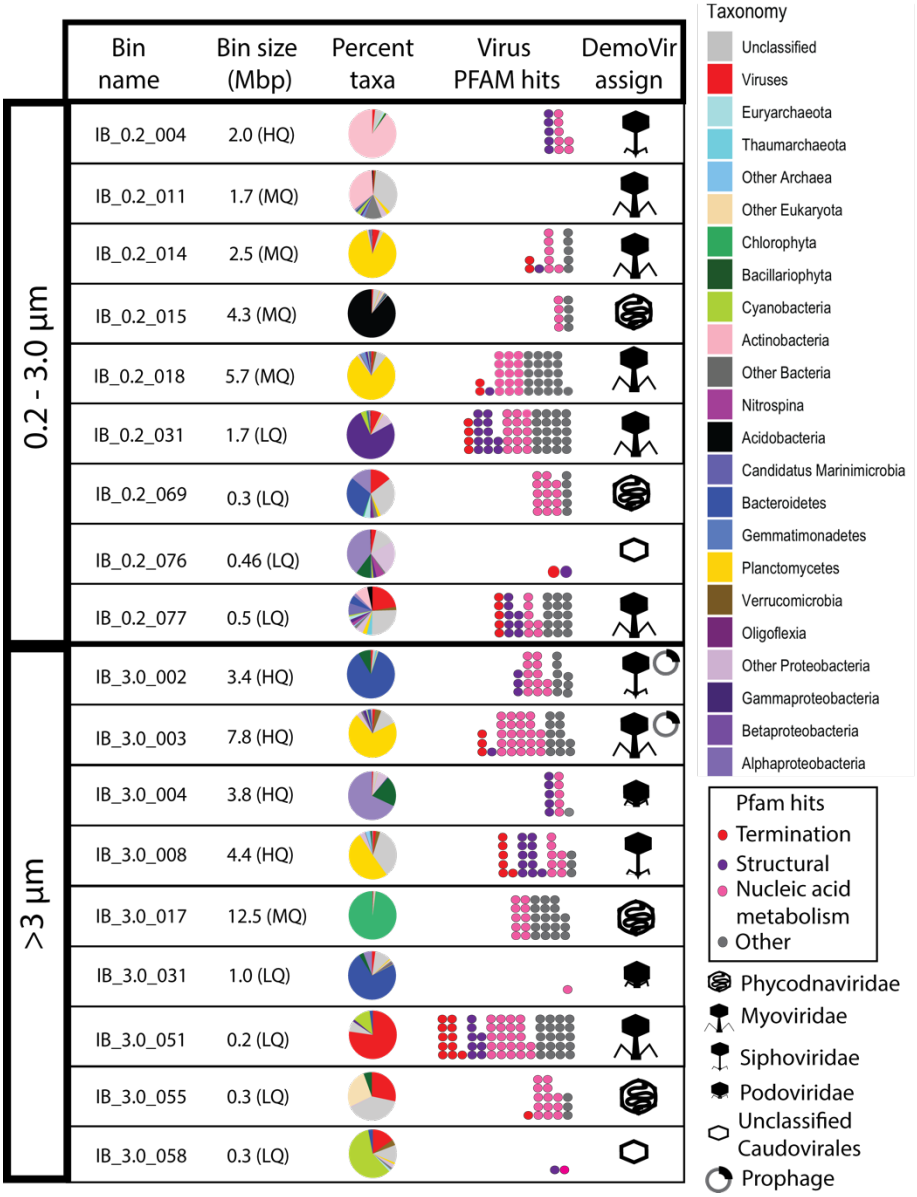


Figure 4. Virus-microbe interactions detected in HAGs. Each row contains data about the bin in which the virus sequence was clustered in and the virus sequence itself. Bin information includes name, size, and quality (where high, medium and low quality determination is indicated by HQ, MQ and LQ, respectively) and pie chart representing percentage of total sequence length assigned taxonomy by DIAMOND blast. Virus information includes dot plot of color-coded circles indicating viral proteins identified by Pfam search and icon with DemoVir viral taxonomy assignment.

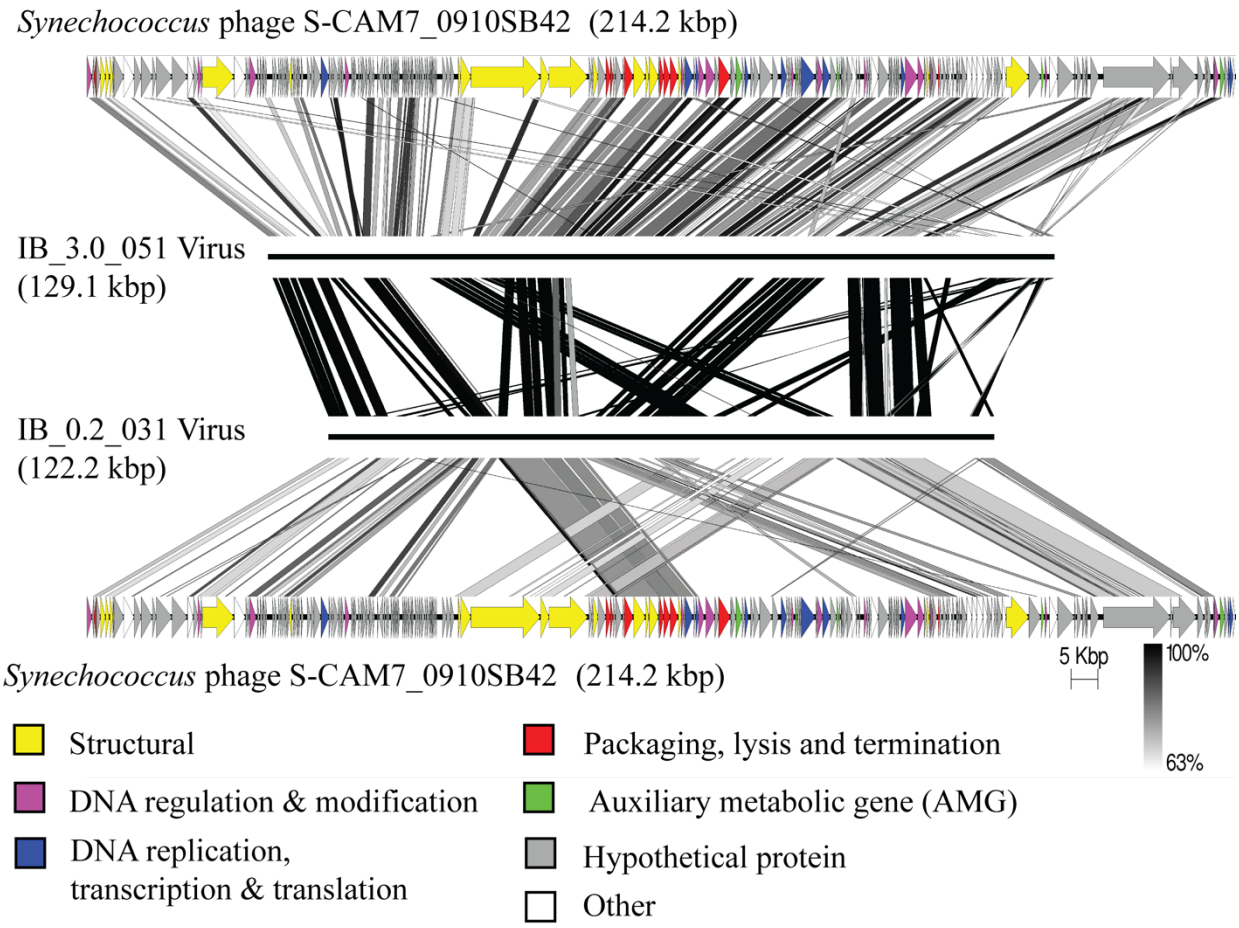
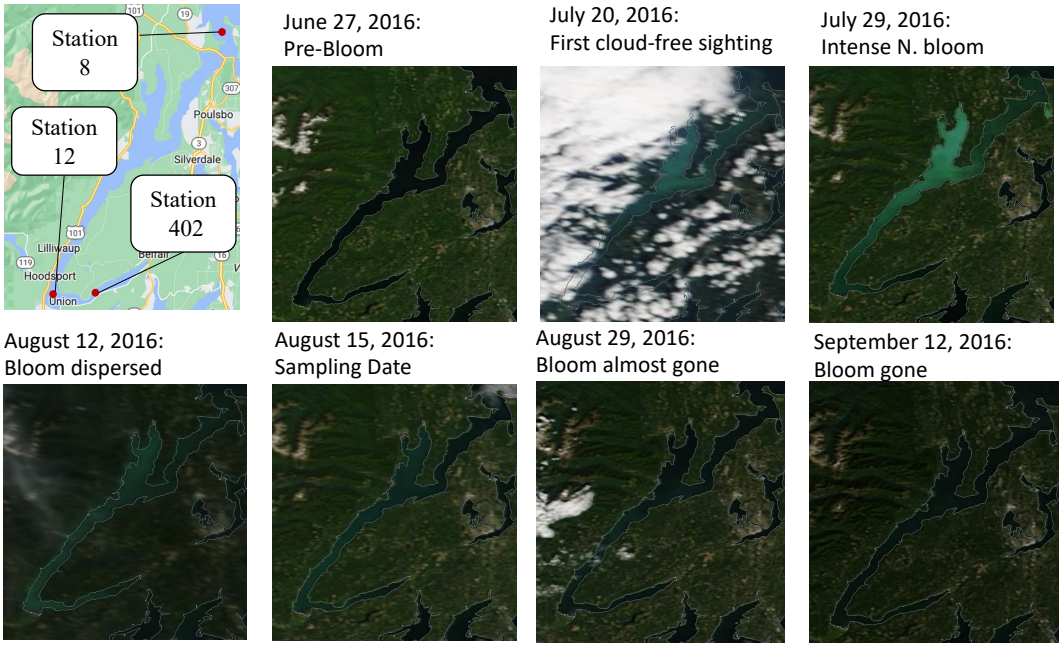
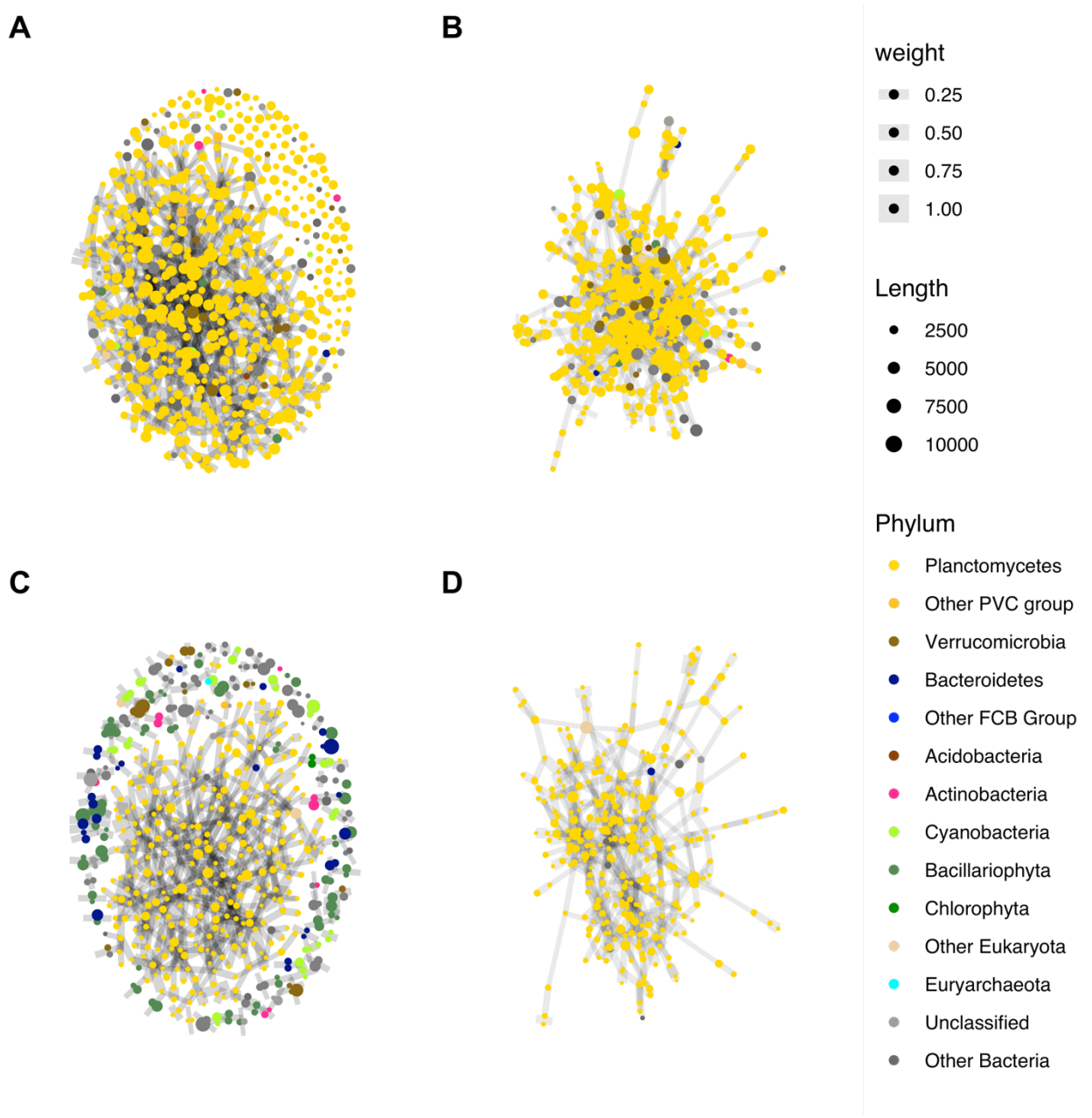


Figure 5. S-CAM7 virus detected in both size fractions interacting with different hosts. Genome similarity between the *Synechococcus* phage S-CAM7 0910SB42 published genome (top and bottom rows) and viral genomes from bins IB_3.0_051 (*Synechococcus*) and IB_0.2_031 (*Alcanivorax)*. Virus genes are annotated on the S-CAM7 genome and color coded by function. Scale bar represents percent nucleotide identity between each pair of genomes. See table S7 for list of AMGs in each genome.

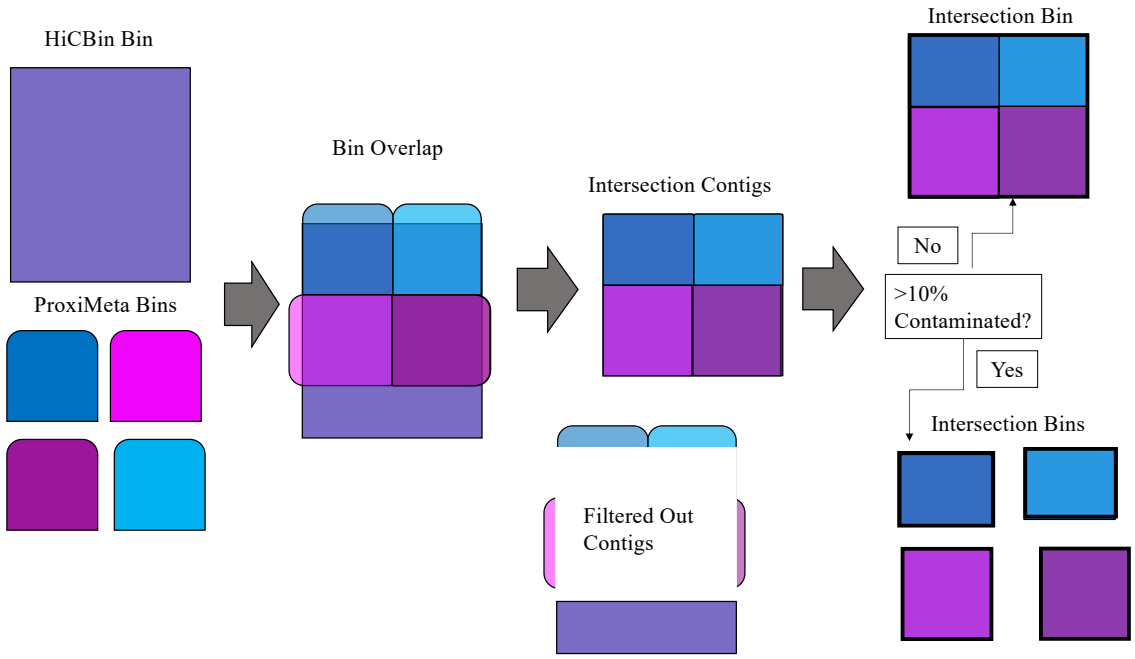
1.9 Supplemental Figures and Tables



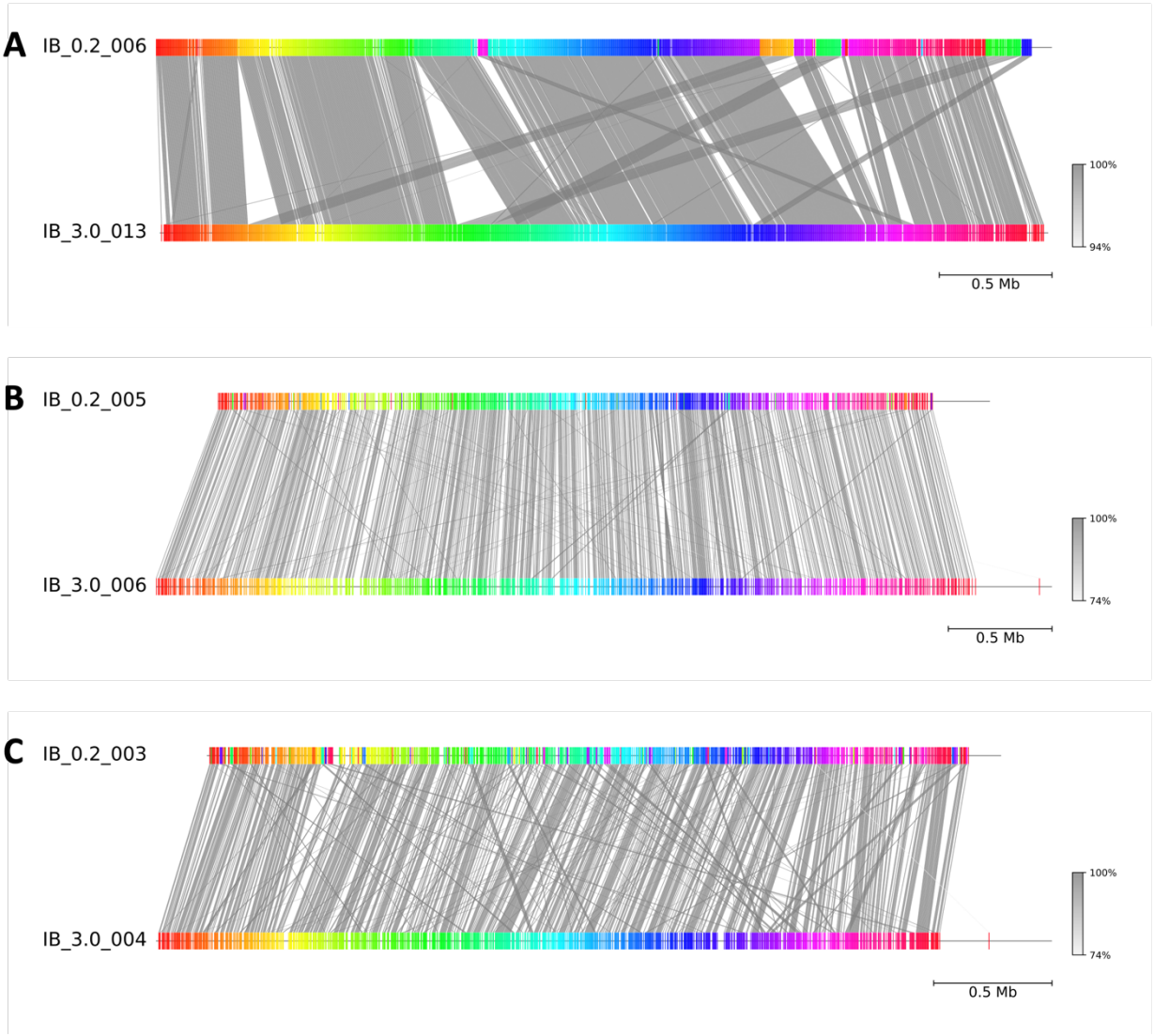
Supplementary Figure 1
MODIS satellite images tracking coccolithophore bloom progression from pre-bloom conditions in June 2016 through post-bloom conditions in September 2016. Bloom intensity in northern reach of Hood Canal peaks in late July, and bloom disperses throughout Hood Canal, with low levels of coccolithophores visible by sampling date on August 15, 2016 in Twanoh Bay (Station 402).



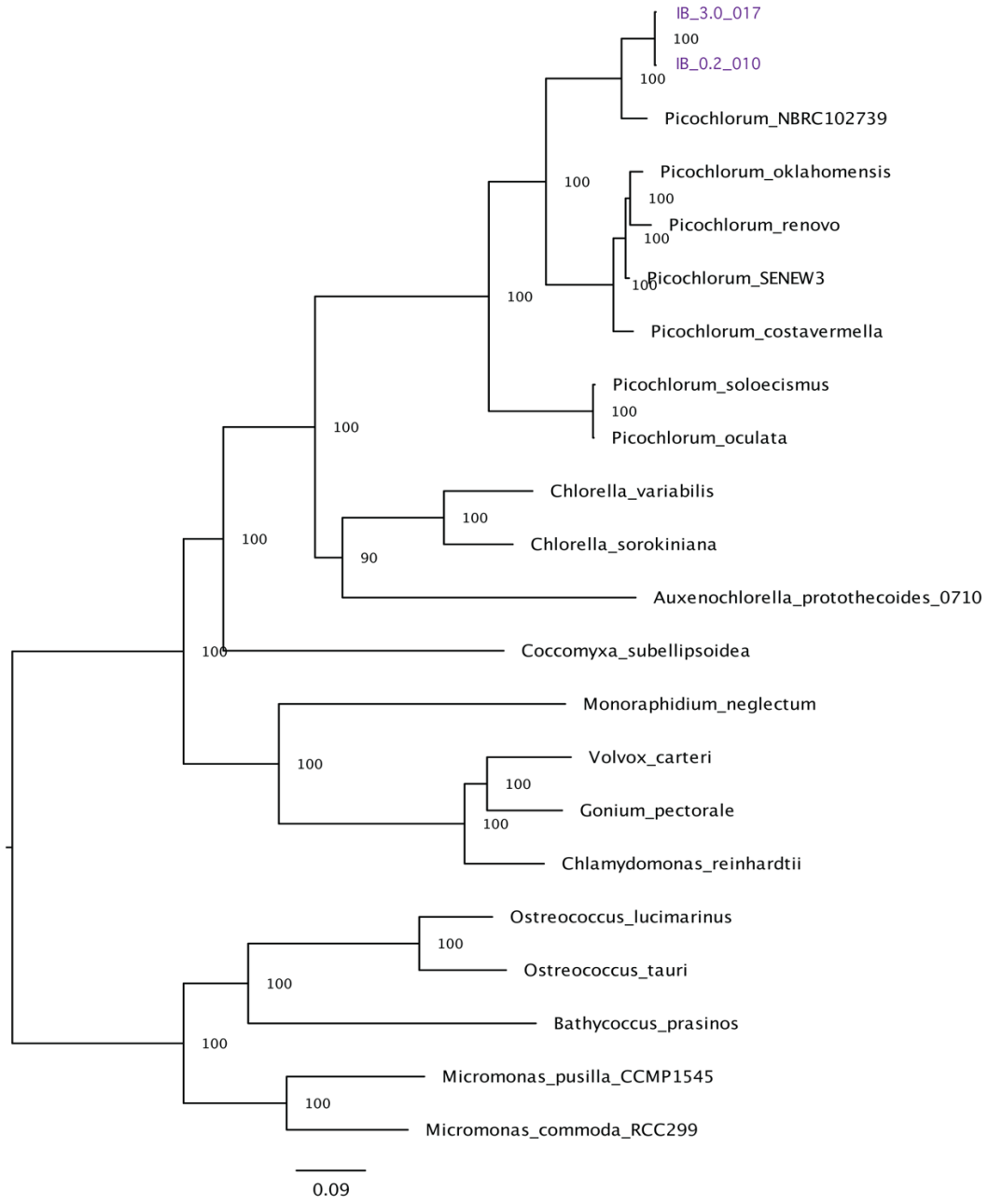
Supplementary Figure 2
Before and after giant component analysis for two different ProxiMeta bins. Each node is a contig found in the bin, color coded by taxonomy and relative width determined by length of contig. Edges represent number of Hi-C reads connecting the two nodes, and weight or relative frequency of links is represented by line width. **A** The ProxiMeta cluster.212 from the 0.2 µm size fraction. **B.** Giant component derived from cluster shown in A. **C.** The ProxiMeta cluster.300 from the 3.0 µm library. **D.** Giant component derived from cluster shown in C.



Supplementary Figure 3
Schematic describing intersection bin protocol. Bins were determined by two separate algorithms that used Hi-C linkages to bin contigs, the more aggressive HiCBin and the more conservative ProxiMeta giant component clusters (GCC). Overlap was calculated and intersection contigs retained. Contigs not included in same bin by both method were removed. Intersection bins could contain multiple ProxiMeta GCCs, but only one HiCBin bin. If final intersection bin was contaminated it was split back into its intersection-trimmed ProxiMeta GCC bins.

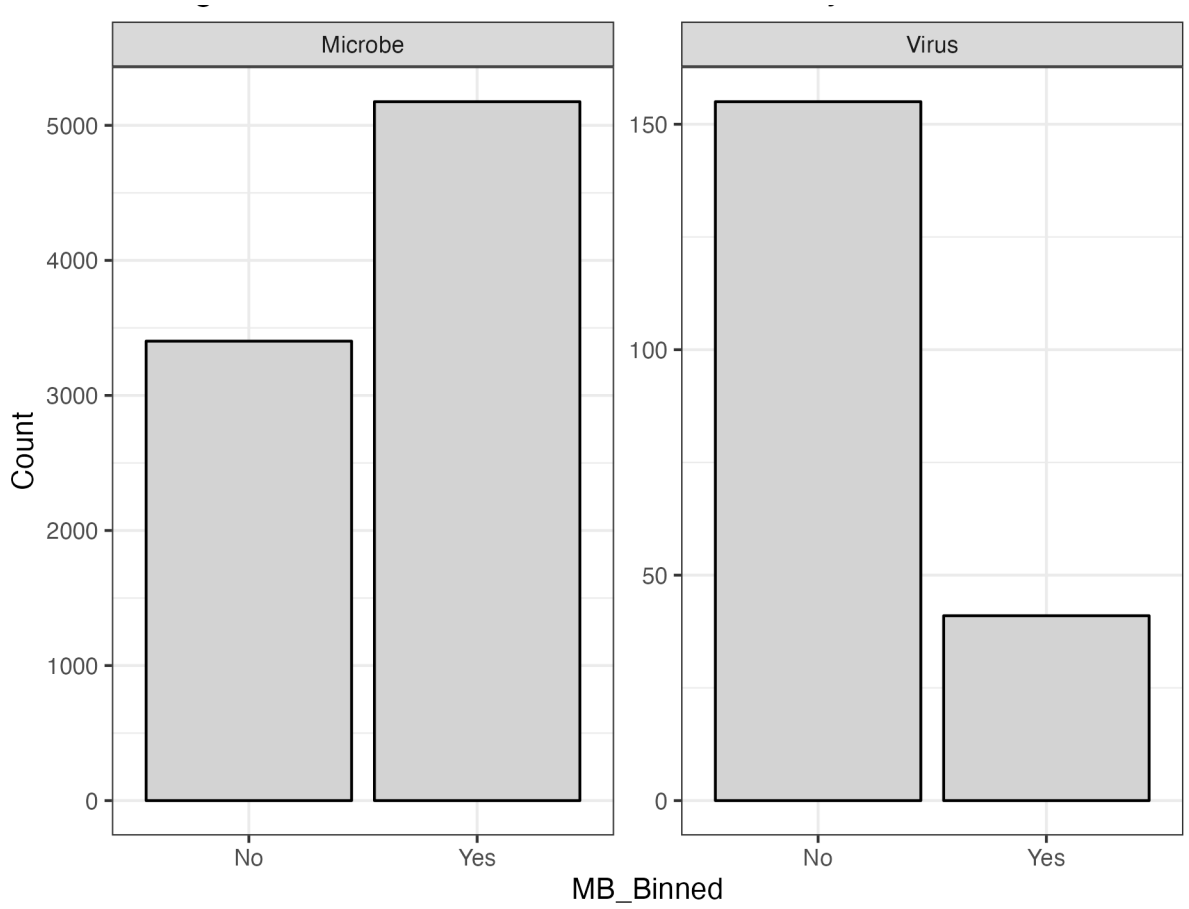


Supplementary Figure 4
Genome alignment of pairs of HAGs in which a highly similar genome (>99% ANI within each pair) was independently assembled in each size fraction. Colored segments represent orthologous genome segments within each pair, gray scale bar indicates % nucleotide identity. **A)** *Oligoflexales*: *Pseudobacteriovorax antillogorciicola* **B)** *Cellvibrionales*: *Halioglobus* and **C)** *Rhodobacterales*: *Rhodobacteraceae*



Supplementary Figure 5

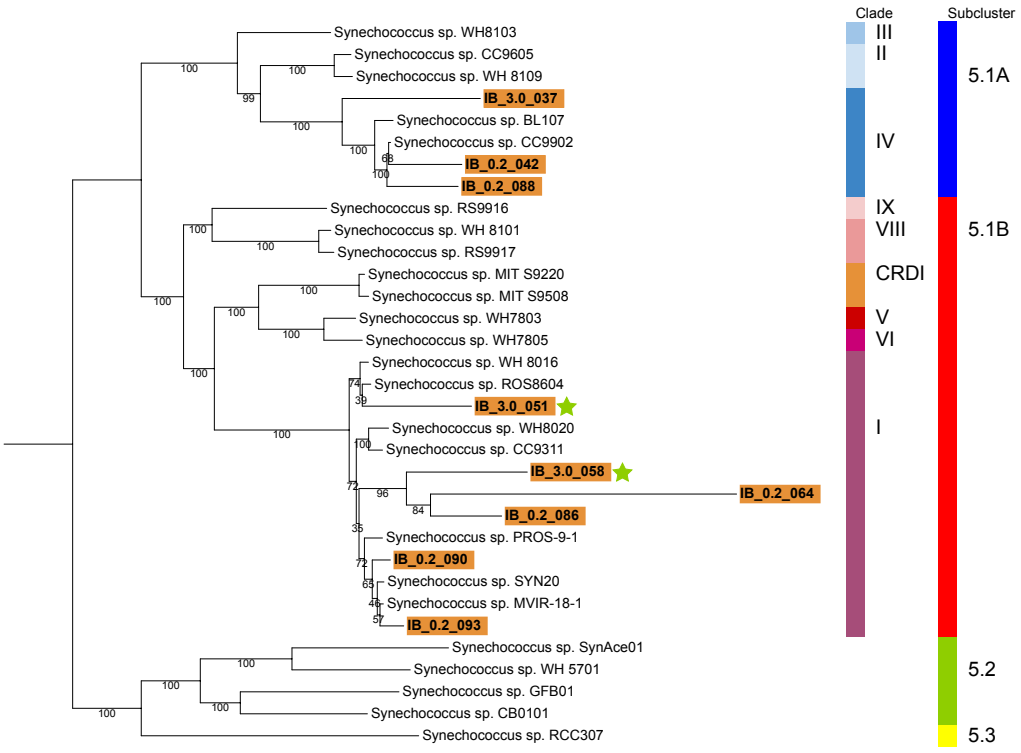
Maximum likelihood phylogenetic tree of green algae based on concatenated proteins of 1519 single copy genes from the Chlorophyta lineage. The two *Picochlorum* HAGs assembled from the 0.2–3 µm and >3 µm fractions cluster with other sequenced *Picochlorum* genomes. Numbers at nodes represent % of Felsenstein bootstrap values (50 bootstrap replicates were performed, at which point the replicates were assessed to have converged).



Supplementary Figure 6
Outcome of Metabat binning of contigs from proximity linked viruses and microbes. All contigs for each of the 18 HAGs that contained substantial virus sequence are included, classified as either microbe (left panel) or virus (right panel), and whether Metabat, which does not use proximity links to assign bins, was able to place these contigs in a bin (no or yes).

Supplementary Figure 7

Heatmap of presence of canonical viral genes in each final bin containing >10kbp viral sequence. Presence was determined through Pfam searches. For simplification some rows are specific genes (DNA polymerase A), and some rows are suites of genes (Tail proteins). See table S6 for list of Pfams in each row. Bins are clustered based on protein counts, which is not phylogenetic placement, but completion ranking.



Supplementary Figure 8

Maximum likelihood phylogenetic tree of marine *Synechococcus* based on concatenated proteins of 788 single copy genes from the Synechococcales lineage. The 9 *Synechococcus* bins in this study (highlighted in orange) are all < 50% complete and are members of clades I and IV. Bins containing viral sequence are indicated by green star. IB_3.0_051 contains the S-CAM7-like viral genome. *Synechococcus* from subclusters 5.2 and 5.3 were used as an outgroup. Numbers at nodes represent % of Felsenstein bootstrap values (164 bootstrap replicates were performed, at which point the replicates were assessed to have converged).

Supplementary Table 1. Sequencing and assembly statistics for each size fractionated library.

	Size Fractionated Library	
	0.2 µm – 3.µm	>3.0 µm
Shotgun Sequencing Reads (millions)	207	156
HiC Sequencing Reads (millions)	56	81
Total length (Mbp)	1,147	360
N50 (bp)	1,286	1,454
L50 (contigs)	198,415	79,521
Number of Contigs	2,307,155	1,049,922
Number of Contigs (>1000bp)	304,112	145,302
Longest Contig (bp)	355,649	350,327

Supplementary Table 2. Size and number of bins for each binning method:

	Size Fractionated Library			
	0.2 µm – 3.µm		>3.0 µm	
Binning Method	Contigs	Bins	Contigs	Bins
Total Hi-C Linked Contigs	114,129	NA	66,332	NA
MetaBat	107,590	165	54,957	73
ProxiMeta	85,156	301	50,925	261
HiCBin	56,179	65	38,116	38
ProxiMeta Giant	24,575	194	28,640	139
Intersection Bins	21,113	98	27,552	72

Supplementary Table 3. Classification and statistics for each final bin included in study.

Bin Name	Group	Best Phylum	Best Class	Best Lowest Common Ancestor Hit	Viral Sequence Length (bp)	Bin size (bp)	Completeness (%)	Contamination (%)	Longest contig	Number contigs	Mean contig length	Size Fraction
IB_0.2_015	Acidobacteria	Acidobacteria	unclassified Acidobacteria	Acidobacteria bacterium	46524	4314479	65.8749	9.7400	84985	617	6993	0.2 −µm
IB_0.2_081	Acidobacteria	Acidobacteria	unclassified Acidobacteria	Acidobacteria bacterium	0	90110	0.0000	0.0000	9495	29	3107	0.2 −µm
IB_0.2_030	Archaea	Thaumarchaeota	Nitrosopumilales	Nitrosopumilus	0	561210	44.1748	2.2654	31841	99	5669	0.2 −µm
IB_0.2_039	Archaea	Thaumarchaeota	unclassified Thaumarchaeota	Candidatus Nitrosopelagicus	30153	312925	26.6990	0.0000	30153	46	6803	0.2 −µm
IB_0.2_048	Archaea	Thaumarchaeota	unclassified Thaumarchaeota	Candidatus Nitrosopelagicus	9216	216247	16.8036	0.0000	11978	68	3180	0.2 −µm
IB_0.2_035	Archaea	Thaumarchaeota	unclassified Thaumarchaeota	Thaumarchaeota archaeon casp-thauma4	0	439732	40.1294	2.7508	11039	119	3695	0.2 −µm
IB_0.2_026	Archaea	Thaumarchaeota	unclassified Thaumarchaeota	unclassified Thaumarchaeota	0	620714	49.8382	2.9126	24222	92	6747	0.2 −µm
IB_3.0_057	Archaea	Euryarchaeota	unclassified root	Euryarchaeota archaeon	10670	61723	0.0000	0.0000	10670	27	2286	3.0 −µm
IB_0.2_010	Eukaryota	Chlorophyta	Trebouxiophyceae	Picochlorum sp. RCC944	74135	12376972	78.3549	47.6935	175797	659	18781	0.2 −µm
IB_3.0_059	Eukaryota	Bacillariophyta	Thalassiosirales	Skeletonema	12940	125028	0.0000	0.0000	20379	40	3126	3.0 −µm
IB_3.0_025	Eukaryota	Bacillariophyta	Thalassiosirales	Thalassiosira miniscula	0	1519999	47.3354	0.0000	173205	22	69091	3.0 −µm
IB_3.0_047	Eukaryota	Bacillariophyta	Thalassiosirales	Thalassiosira miniscula	0	237783	7.2595	0.0000	70217	8	29723	3.0 −µm
IB_3.0_044	Eukaryota	Bacillariophyta	Thalassiosirales	Thalassiosira oceanica CCMP1005	0	127410	9.5722	3.7167	11358	46	2770	3.0 −µm
IB_3.0_017	Eukaryota	Chlorophyta	Trebouxiophyceae	Picochlorum oklahomense	62957	12461553	82.6019	77.4242	189686	710	17551	3.0 −µm
IB_0.2_013	FCB	Bacteroidetes	Bacteroidales	Lentimicrobiaceae bacterium	0	1802456	70.3315	2.4780	106169	63	28610	0.2 −µm
IB_0.2_065	FCB	Bacteroidetes	Bacteroidales	Lentimicrobiaceae bacterium	0	312363	6.1755	0.0000	57473	48	6508	0.2 −µm
IB_0.2_066	FCB	Bacteroidetes	Bacteroidales	Lentimicrobiaceae bacterium	0	174200	5.9140	0.0000	48553	6	29033	0.2 −µm
IB_0.2_061	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriaceae bacterium	2297	203910	7.6803	0.0000	11622	69	2955	0.2 −µm
IB_0.2_069	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriaceae bacterium	49888	344782	3.0434	0.5032	11418	108	3192	0.2 −µm
IB_0.2_082	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriaceae bacterium	0	26195	0.0000	0.0000	6510	11	2381	0.2 −µm
IB_0.2_063	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriales	0	70881	6.4577	0.0000	26060	13	5452	0.2 −µm
IB_0.2_008	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriales bacterium UBA3442	0	1294301	87.0968	1.1828	59170	108	11984	0.2 −µm
IB_0.2_084	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriales bacterium UBA3442	0	11053	0.0000	0.0000	5824	5	2211	0.2 −µm
IB_0.2_009	FCB	Gemmatimonadetes	unclassified Gemmatimonadetes	Gemmatimonadetes	1119	2723233	82.9024	6.7227	54915	233	11688	0.2 −µm
IB_3.0_002	FCB	Bacteroidetes	Flavobacteriales	Crocinitomix algicola	44946	3444777	98.1081	0.5405	58885	266	12950	3.0 −µm
IB_3.0_024	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriaceae bacterium	0	1464529	52.7133	0.2145	88223	158	9269	3.0 −µm
IB_3.0_027	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriaceae bacterium	0	1074915	34.4828	1.7241	16443	218	4931	3.0 −µm
IB_3.0_012	FCB	Bacteroidetes	Flavobacteriales	Flavobacteriales	15985	2773257	90.9186	1.0750	43591	253	10961	3.0 −µm
IB_3.0_031	FCB	Bacteroidetes	Saprospirales	Saprospirales bacterium	22131	1019672	26.2278	0.0000	15386	563	1811	3.0 −µm
IB_0.2_003	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	0	3356138	95.1515	3.4052	71494	282	11901	0.2 −µm
IB_0.2_028	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Thalassiosira miniscula	0	1677294	45.3357	1.5898	19428	857	1957	0.2 −µm
IB_0.2_071	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Thalassiosira miniscula	0	176032	1.6928	0.0000	13493	41	4293	0.2 −µm
IB_0.2_024	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae bacterium	0	2134956	52.2124	3.9594	23841	718	2973	0.2 −µm
IB_0.2_032	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae bacterium	0	2009409	43.1034	0.0000	37261	212	9478	0.2 −µm
IB_0.2_076	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae bacterium	16287	460858	0.0000	0.0000	15694	112	4115	0.2 −µm
IB_0.2_057	Proteobacteria	Betaproteobacteria	Betaproteobacteria	unclassified Betaproteobacteria	0	218942	9.3699	0.0000	24001	36	6082	0.2 −µm
IB_0.2_017	Proteobacteria	Betaproteobacteria	Nitrosomonadales	unclassified Nitrosomonadaceae	0	1023231	63.6024	8.5190	12047	428	2391	0.2 −µm
IB_0.2_020	Proteobacteria	Gammaproteobacteria	Acidiferrobacterales	Acidiferrobacteraceae bacterium	0	1305048	54.7967	0.6098	27295	219	5959	0.2 −µm
IB_0.2_037	Proteobacteria	Gammaproteobacteria	Acidiferrobacterales	Acidiferrobacteraceae bacterium	0	1227545	32.7744	2.1341	19023	398	3084	0.2 −µm
IB_0.2_098	Proteobacteria	Gammaproteobacteria	Acidiferrobacterales	Acidiferrobacteraceae bacterium	0	189366	0.0000	0.0000	11012	66	2869	0.2 −µm
IB_0.2_025	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Gammaproteobacteria	0	1576126	50.7383	6.0037	22736	387	4073	0.2 −µm
IB_0.2_054	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Gammaproteobacteria	0	202377	10.1097	0.0000	7404	77	2628	0.2 −µm
IB_0.2_005	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Haieaceae	5585	3723877	92.2289	1.6417	68294	395	9428	0.2 −µm
IB_0.2_012	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	2478162	73.1975	6.8182	70539	293	8458	0.2 −µm
IB_0.2_046	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	5290	569637	17.5549	0.0000	39481	109	5226	0.2 −µm
IB_0.2_053	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	688236	11.9122	0.0000	19008	147	4682	0.2 −µm
IB_0.2_072	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	49843	0.9404	0.0000	5308	19	2623	0.2 −µm
IB_0.2_038	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	1105012	29.9373	0.0000	27616	245	4510	0.2 −µm
IB_0.2_096	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	243647	0.0000	0.0000	79110	12	20304	0.2 −µm
IB_0.2_091	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Porticoccaceae bacterium	0	17859	0.0000	0.0000	6882	7	2551	0.2 −µm
IB_0.2_062	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Eukaryota	8075	44867	7.1429	0.0000	11662	6	7478	0.2 −µm
IB_0.2_033	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria bacterium	10548	1974739	42.1884	5.4155	24633	792	2493	0.2 −µm
IB_0.2_097	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria bacterium SG8_47	3555	11413	0.0000	0.0000	4725	3	3804	0.2 −µm
IB_0.2_029	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Proteobacteria	0	800967	44.3998	0.0654	33773	80	10012	0.2 −µm
IB_0.2_075	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	unclassified Gammaproteobacteria	0	224396	0.0000	0.0000	19193	63	3562	0.2 −µm
IB_0.2_079	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	unclassified Gammaproteobacteria	0	149119	0.0000	0.0000	7655	56	2663	0.2 −µm
IB_0.2_031	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Alcanivorax	122164	1655312	43.2175	4.2534	30158	1014	1632	0.2 −µm
IB_0.2_006	Proteobacteria	Oligoflexia	Oligoflexales	Psuedobacteriovorax antillogorgiicola	0	3973528	90.7563	0.8403	355649	111	35798	0.2 −µm
IB_0.2_034	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria bacterium UBA7452	5003	1444374	42.1782	9.0447	23315	429	3367	0.2 −µm
IB_0.2_040	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	473597	24.3347	0.8403	23475	99	4784	0.2 −µm
IB_0.2_043	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	430851	18.6992	0.0000	10482	124	3475	0.2 −µm
IB_0.2_045	Proteobacteria	Unclassified Proteobacteria	Unclassified Proteobacteria	Proteobacteria	0	520878	18.4959	0.0000	22183	139	3747	0.2 −µm
IB_0.2_047	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	354701	17.5407	2.4390	27852	48	7390	0.2 −µm
IB_0.2_051	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	636036	12.4138	0.0000	15760	141	4511	0.2 −µm
IB_0.2_059	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	432285	8.3333	0.0000	21231	89	4857	0.2 −µm
IB_0.2_080	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	10043	0.0000	0.0000	5453	3	3348	0.2 −µm
IB_0.2_083	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	26538	0.0000	0.0000	8637	5	5308	0.2 −µm
IB_0.2_085	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	286678	0.0000	0.0000	7449	104	2757	0.2 −µm
IB_0.2_087	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	23209	0.0000	0.0000	4172	10	2321	0.2 −µm
IB_0.2_089	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	27736	0.0000	0.0000	8822	9	3082	0.2 −µm
IB_3.0_033	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Candidatus Pelagibacter ubique	0	364967	22.5312	3.2816	3578	237	1540	3.0 −µm
IB_3.0_004	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	29103	3786570	97.3636	4.8261	118995	581	6517	3.0 −µm
IB_3.0_030	Proteobacteria	Betaproteobacteria	Nitrosomonadales	unclassified Nitrosomonadales	0	343271	30.7143	1.3721	10474	150	2288	3.0 −µm
IB_3.0_018	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Gammaproteobacteria	0	2433871	81.7369	2.7424	38821	366	6650	3.0 −µm
IB_3.0_028	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Gammaproteobacteria	0	876171	31.6795	3.3293	23933	225	3894	3.0 −µm
IB_3.0_052	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Gammaproteobacteria	0	341016	3.4483	0.0000	8281	124	2750	3.0 −µm
IB_3.0_015	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	5714	2543488	88.5107	0.4934	49634	280	9084	3.0 −µm
IB_3.0_034	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	580890	18.9655	0.0000	29707	69	8419	3.0 −µm
IB_3.0_039	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	481980	15.6740	0.1567	11875	223	2161	3.0 −µm
IB_3.0_045	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	342729	8.6207	0.0000	22240	59	5809	3.0 −µm
IB_3.0_053	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	71112	3.4483	0.0000	4463	37	1922	3.0 −µm
IB_3.0_067	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	83810	0.0000	0.0000	5090	30	2794	3.0 −µm
IB_3.0_068	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	16503	0.0000	0.0000	4624	5	3301	3.0 −µm
IB_3.0_069	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	139815	0.0000	0.0000	30354	10	13982	3.0 −µm
IB_3.0_071	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae bacterium	0	15578	0.0000	0.0000	5426	5	3116	3.0 −µm
IB_3.0_019	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	2423528	77.4573	1.4614	62351	214	11325	3.0 −µm
IB_3.0_050	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	25775	4.1667	0.0000	15203	6	4296	3.0 −µm
IB_3.0_060	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	16062	0.0000	0.0000	4369	5	3212	3.0 −µm
IB_3.0_061	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	13790	0.0000	0.0000	4111	5	2758	3.0 −µm
IB_3.0_064	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	91666	0.0000	0.0000	25614	5	18333	3.0 −µm
IB_3.0_065	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	20785	0.0000	0.0000	10750	7	2969	3.0 −µm
IB_3.0_066	Proteobacteria	Gammaproteobacteria	Cellvibrionales	marine gamma proteobacterium HTCC2080	0	19273	0.0000	0.0000	6497	4	4818	3.0 −µm
IB_3.0_006	Proteobacteria	Gammaproteobacteria	Cellvibrionales	unclassified Halieaceae	0	4322876	96.4067	4.1893	38423	1057	4090	3.0 −µm
IB_3.0_054	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria	0	52462	3.4483	0.0000	7135	17	3086	3.0 −µm
IB_3.0_070	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria	0	6163	0.0000	0.0000	2583	4	1541	3.0 −µm
IB_3.0_072	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria	0	65298	0.0000	0.0000	8195	20	3265	3.0 −µm
IB_3.0_056	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria bacterium	0	103384	0.0000	0.0000	5523	58	1782	3.0 −µm
IB_3.0_013	Proteobacteria	Oligoflexia	Oligoflexales	Psueodobacteriovorax antillogorgiicola	0	3939947	90.7563	0.0000	350327	88	44772	3.0 −µm
IB_3.0_026	Proteobacteria	Unclassified Proteobacteria	unclassified Proteobacteria	Proteobacteria	0	956674	38.3808	1.2195	26248	246	3889	3.0 −µm
IB_0.2_036	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	1411303	36.9795	0.0668	19150	346	4079	0.2 −µm
IB_0.2_022	PVC	Planctomycetes	Phycisphaerae	Planctomycetaceae bacterium	0	1419349	52.7018	5.1805	10444	575	2468	0.2 −µm
IB_0.2_078	PVC	Planctomycetes	Planctomycetales	Gimesia sp.	0	22353	0.0000	0.0000	2515	13	1719	0.2 −µm
IB_0.2_001	PVC	Planctomycetes	Planctomycetales	Marinblastus fucicola	1617	5553504	96.2944	3.0104	50921	669	8301	0.2 −µm
IB_0.2_007	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	0	3127455	90.3713	0.5747	60468	187	16724	0.2 −µm
IB_0.2_018	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	97552	5712375	62.5759	7.8652	50865	1926	2966	0.2 −µm
IB_0.2_073	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	5217	207161	0.0000	0.0000	5217	95	2181	0.2 −µm
IB_0.2_074	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	0	167321	0.0000	0.0000	5527	108	1549	0.2 −µm
IB_0.2_027	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae;	13313	4076832	45.5221	0.5047	13518	1339	3045	0.2 −µm
IB_0.2_068	PVC	Planctomycetes	Planctomycetales	Rhodopirellula sp.	6268	110501	4.0584	0.0000	6268	55	2009	0.2 −µm
IB_0.2_014	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	121864	2524005	67.0002	1.0753	29958	1165	2167	0.2 −µm

Supplementary Table 2 cont. Classification and statistics for each final bin included in study.

Bin Name	Group	Best Phylum	Best Class	Best Lowest Common Ancestor Hit	Viral Sequence Length (bp)	Bin size (bp)	Completeness (%)	Contamination (%)	Longest contig	Number contigs	Mean contig length	Size Fraction
IB_0.2_041	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	0	916319	21.2194	0.0000	6760	450	2036	0.2 −µm
IB_0.2_021	PVC	Verrucomicrobia	Verrucomicrobiales	Verrucomicrobiales	22464	3970211	53.9530	1.4006	13078	1258	3156	0.2 −µm
IB_0.2_055	PVC	Verrucomicrobia	Verrucomicrobiales	Verrucomicrobiales bacterium	0	503433	9.6552	0.0000	6760	220	2288	0.2 −µm
IB_3.0_011	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	3596571	91.4773	2.8409	84753	261	13780	3.0 −µm
IB_3.0_029	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	630725	31.1999	0.0000	33889	44	14335	3.0 −µm
IB_3.0_035	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	1162	687544	18.4211	0.0000	30456	134	5131	3.0 −µm
IB_3.0_042	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	591833	12.0690	0.0000	15781	203	2915	3.0 −µm
IB_3.0_043	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	229553	10.5263	0.0000	27028	19	12082	3.0 −µm
IB_3.0_046	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	887289	8.3333	0.0000	43162	60	14788	3.0 −µm
IB_3.0_048	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	444207	4.1667	0.0000	18298	135	3290	3.0 −µm
IB_3.0_049	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	432655	4.1667	0.0000	25300	31	13957	3.0 −µm
IB_3.0_063	PVC	Planctomycetes	Phycisphaerae	Phycisphaerae bacterium	0	140142	0.0000	0.0000	12291	42	3337	3.0 −µm
IB_3.0_003	PVC	Planctomycetes	Planctomycetales	Mariniblastus fucicola	132393	7771796	97.4709	2.7807	77801	1256	6188	3.0 −µm
IB_3.0_010	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	0	3631431	91.8864	2.2989	42858	657	5527	3.0 −µm
IB_3.0_016	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	0	7467438	82.7735	4.2691	10114	3491	2139	3.0 −µm
IB_3.0_021	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	5378	6281475	61.3967	1.9409	10507	2998	2095	3.0 −µm
IB_3.0_036	PVC	Planctomycetes	Planctomycetales	Planctomycetaceae bacterium	0	1722795	17.1756	1.1111	6786	1295	1330	3.0 −µm
IB_3.0_020	PVC	Planctomycetes	Planctomycetales	Rhodopirellula sp.	0	6769195	71.5817	10.4662	10828	3278	2065	3.0 −µm
IB_3.0_008	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	95283	4364051	94.6237	1.0753	101610	241	18108	3.0 −µm
IB_3.0_014	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	5912	3727161	90.0696	0.6354	22016	1097	3398	3.0 −µm
IB_3.0_022	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	4713	1196359	59.7744	1.2397	7285	755	1585	3.0 −µm
IB_3.0_032	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	0	407514	23.5582	0.5376	4816	279	1461	3.0 −µm
IB_3.0_038	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	0	705377	16.2833	0.7130	3123	535	1318	3.0 −µm
IB_3.0_040	PVC	Planctomycetes	unclassified Planctomycetes	Planctomycetes bacterium	0	328563	12.6924	0.0000	3308	246	1336	3.0 −µm
IB_3.0_062	PVC	Planctomycetes	unclassified root	Deltaproteobacteria bacterium	0	23345	0.0000	0.0000	4477	11	2122	3.0 −µm
IB_3.0_001	PVC	Verrucomicrobia	unclassified Verrucomicrobia	Verrucomicrobia	6421	6033156	99.3197	2.3810	89895	413	14608	3.0 −µm
IB_3.0_005	PVC	Verrucomicrobia	Verrucomicrobiales	Verrucomicrobiales	1480	4490638	97.2789	2.2109	40739	799	5620	3.0 −µm
IB_3.0_007	PVC	Verrucomicrobia	Verrucomicrobiales	Verrucomicrobiales bacterium	23731	7811881	96.2585	3.0908	30263	2005	3896	3.0 −µm
IB_0.2_023	Terrabacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiaceae bacterium	0	1643271	52.6677	1.7094	26507	339	4847	0.2 −µm
IB_0.2_002	Terrabacteria	Actinobacteria	Acidimicrobiales	Acidimicrobiales	0	3042328	95.7265	4.2735	90164	202	15061	0.2 −µm
IB_0.2_049	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria	0	297307	15.5172	0.0000	43135	37	8035	0.2 −µm
IB_0.2_060	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria	0	27883	8.3333	0.0000	12680	6	4647	0.2 −µm
IB_0.2_094	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria	0	93421	0.0000	0.0000	14636	20	4671	0.2 −µm
IB_0.2_016	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria bacterium	5074	951575	64.9573	2.5641	45828	152	6260	0.2 −µm
IB_0.2_056	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria bacterium	0	228464	9.4828	0.0000	12752	43	5313	0.2 −µm
IB_0.2_004	Terrabacteria	Actinobacteria	unclassified Actinobacteria	Actinobacteria	36057	2027472	94.0171	5.5556	118424	147	13792	0.2 −µm
IB_0.2_050	Terrabacteria	Actinobacteria	Unclassified Bacteria	Bacteria	0	189503	12.8527	0.0000	25937	18	10528	0.2 −µm
IB_0.2_058	Terrabacteria	Actinobacteria	unclassified Bacteria	Bacteria	0	456022	8.6207	0.0000	24395	87	5242	0.2 −µm
IB_0.2_011	Terrabacteria	Actinobacteria	unclassified Bacteria;	Bacteria	12270	1662629	75.2730	1.6523	24615	330	5038	0.2 −µm
IB_0.2_086	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcales	0	41182	0.0000	0.0000	5060	17	2422	0.2 −µm
IB_0.2_090	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcales	0	36469	0.0000	0.0000	6544	22	2709	0.2 −µm
IB_0.2_093	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcus	0	15757	0.0000	0.0000	11689	19	3206	0.2 −µm
IB_0.2_042	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcus sp. CC9902	0	476054	20.6343	0.5435	27766	61	7804	0.2 −µm
IB_0.2_088	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcus sp. CC9902	0	186283	0.0000	0.0000	8174	59	3157	0.2 −µm
IB_0.2_064	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcus sp. WH 8016	0	159755	6.2500	0.0000	12617	52	3072	0.2 −µm
IB_3.0_023	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria bacterium	0	727922	57.9487	0.4274	65281	57	12771	3.0 −µm
IB_3.0_041	Terrabacteria	Actinobacteria	Actinobacteria	Actinobacteria bacterium	0	178708	12.5000	0.0000	9013	62	2882	3.0 −µm
IB_3.0_058	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcus	19557	292904	0.0000	0.0000	17138	102	2872	3.0 −µm
IB_3.0_037	Terrabacteria	Cyanobacteria	Synechococcales	Synechococcus sp. CC9902	9349	739389	16.4577	1.7241	27179	171	4324	3.0 −µm
IB_0.2_077	Unclassified	Unclassified	Unclassified	Candidatus Pacearchaeota archaeon	100951	450590	0.0000	0.0000	20434	176	2560	0.2 −µm
IB_0.2_019	Unclassified Bacteria	Unclassified Bacteria	Unclassified Bacteria	Bacteria	0	1190925	55.1724	0.0000	79037	43	27696	0.2 −µm
IB_0.2_044	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	0	470334	18.5526	0.0000	45884	22	21379	0.2 −µm
IB_0.2_052	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	0	158156	12.0690	0.0000	42808	32	4942	0.2 −µm
IB_0.2_067	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	0	123661	4.1667	0.0000	21230	10	12366	0.2 −µm
IB_0.2_070	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	0	62077	2.3511	0.0000	28574	5	12415	0.2 −µm
IB_0.2_092	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	0	195910	0.0000	0.0000	25455	16	12244	0.2 −µm
IB_0.2_095	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	0	160264	0.0000	0.0000	32206	14	11447	0.2 −µm
IB_3.0_009	Unclassified Bacteria	Unclassified Bacteria	unclassified Bacteria	Bacteria	6568	2306271	93.4942	4.7748	55262	300	7688	3.0 −µm
IB_3.0_055	Virus	Viruses	unclassified Viruses	Viruses	92776	328027	0.8267	0.0000	66272	19	17265	3.0 −µm
IB_3.0_051	Virus	Viruses	Viruses	Viruses	98893	238654	3.4483	0.0000	7262	108	2210	3.0 −µm

Supplementary Table 4. Comparison of BUSCO summaries for published *Picochlorum* genomes and the two *Picochlorum* HAGs from this study.

Name	Source	File Downloaded	# of Proteins	Busco Results
Picochlorum*SENEW3* v2.0	http://cyanophora.rutgers.edu/picochlorum/	PseV2.0*prot* nr99.faa.txt	6760	C:91.2%[S:90.7%,D:0.5%],F:0.3%,M:8.5%,n:1519
Picochlorum_oklahomensis	http://cyanophora.rutgers.edu/picochlorum/	Pok*prot* nr99.faa	6968	C:68.0%[S:67.1%,D:0.9%],F:7.5%,M:24.5%,n:1519
Picochlorum_oculata	http://cyanophora.rutgers.edu/picochlorum/	Poc*prot* nr99.faa.txt	6328	C:86.7%[S:86.3%,D:0.4%],F:1.6%,M:11.7%,n:1519
Picochlorum_NBRC102739	http://cyanophora.rutgers.edu/picochlorum/	Pnb*prot* nr99.faa.txt	11366	C:91.5%[S:32.3%,D:59.2%],F:0.5%,M:8.0%,n:1519
Picochlorum renovo	https://phycocosm.jgi.doe.gov/Picre1/Picre1.home.html	Picre1*GeneCatalog* proteins_20200805.aa.fasta.gz	9010	C:67.6%[S:66.1%,D:1.5%],F:11.0%,M:21.4%,n:1519
Picochlorum soloecismus DOE101	https://phycocosm.jgi.doe.gov/Picsp_1/Picsp_1.home.html	Picsp*1* GeneCatalog*proteins* 20170909.aa.fasta.gz	6861	C:87.0%[S:85.8%,D:1.2%],F:0.7%,M:12.3%,n:1519
Picochlorum costavermella RCC4223	https://bioinformatics.psb.ugent.be/orcae/overview/RCC4223	mRNA*pseudo* RCC4223_*pep* 20151207.tfa	10135	C:95.6%[S:94.9%,D:0.7%],F:0.5%,M:3.9%,n:1519
IB_0.2_010	this study	--	11815	C:86.6%[S:85.8%,D:0.8%],F:4.2%,M:9.2%,n:1519
IB_3.0_017	this study	--	12351	C:86.7%[S:86.0%,D:0.7%],F:4.0%,M:9.3%,n:1519

Supplementary Table 5. Results and comparison of viral identification methods.

Number of Contigs Identified as Viral by Each Method	Size Fractionated Library	
	0.2 - 3.0 µm	>3.0 µm
VirFinder	1460	101
VirSorter	3077	462
DIAMOND blastx	21910	6382
VirFinder but not VirSorter	781	108
VirSorter but not by VirFinder	282	38
VirSorter and VirFinder	681	64
Unique to DIAMOND blastx	21721	6362
Found by each of the 3 methods	189	19

Supplementary Table 6: Pfam categories from Supplementary Figure 5 names of Pfams included in each category.

Pfam Figure Label	Unique_ Pfams	Pfam Names
Acetyltransferase	2	Acetyltransf_1, Acetyltransf_10
AMG	7	2OG-FeII_Oxy_3, CP12, CsrA, HSP20, MazG, PhoH
Baseplate protein	6	Gp5_OB, GPW_gp25, Phage_gp53, Phage-Gp8, T4_baseplate, YHYH
Capsid assembly	1	Phage_Mu_F (minor head protein, head morphogenesis)
DNA ligase	1	DNA_Ligase_A_M
DNA Metabolism	10	D5_N, NTP_transferase, CRAL_TRIO, CRAL_TRIO_N, MRC1, ORC5_C, Pyr_redox_2, Carbam_trans_C, Cupin_2, gp32, Pyr_excise
DNA polymerase A	1	DNA_pol_A
DNA polymerase B	1	DNA_pol_B
DNA recombination	3	RecA, RuvB_N, ERF
DNA regulation	9	Elf1, TFIIS_C, Whib, Topoisom_bac, Response_reg, DNA_topoisoIV, HTH_3, HTH_26, HTH_36
DNA repair	2	UVR, UvsY
Endonucleaes	8	HNH, ResIII, GIY-YIG, HNH_5, SegE_GIY-YIG, Las1, Phage_endo_I, RusA
Epimerase	1	Epimerase
Exonuclease	1	5_3_exonuc_N
Glycosyltransferase	9	Glyco_trans_1_4, Glyco_transf_11, Glyco_transf_25, Glyco_transf_7N, Glyco_transf_8, Glyco_transf_9, Glyco_transf_90, Glyco_transf_92, Glycos_transf_1
Head Assembly	3	Tn7_TnsA-like_N, Head-tail_con, Phage_connect_1
Helicase	4	DnaB_C, Helicase_C, SNF2-rel_dom, SWI2_SNF2
Integrase	2	Phage_integrase, rve_3
Lysin	2	HemolysinCabind, PG_binding_1
Major Capsid Protein	4	Gp23, Phage_capsid_2, Phage_capsid, Phage_T7_Capsid
Methyltransferase	8	Methyltransf_21, Methyltransf_23, Methyltransf_31, MethyltransfD12, N6_N4_Mtase, PEMT, PrmA, TylF
Neck protein	1	T4_neck-protein
Nuclease	2	PARP, ParBc,VRR_NUC
Other Structural Protein	2	Collar, Laminin_G_3
Peptidase	6	Peptidase_C39_2, Peptidase_M28, Peptidase_S74, Peptidase_S77, Peptidase_S78, Peptidase_S78_2
Phage lysozyme	1	Phage_lysozyme2
Portal protein	3	Phage_portal, Phage_portal_2, Portal_Gp20
RNA ligase	3	2_5_RNA_ligase2, RNA_lig_T4_1, RNA_ligase
RNA metabolism	7	dUTPase, mRNA_cap_enzyme, NUDIX, NYN, NYN_YacP, Ribonuc_red_lgC, Ribonuc_red_sm
RNA polymerase	4	Sigma70_r2, RNA_pol_Rpb2_6, RNA_pol, RNA_POL_M_15KD
Tail protein	12	fn3, Pectate_lyase_3, Phage_sheath_1, Phage_sheath_1C, Phage_T4_gp19, Phage_T7_tail, Phage_T7_tail, Phage_tail_L, Phage_tube_2, Phage-tail_3, PhageMin_Tail, T4-gp15_tss
Tape measure protein	1	TMP_3
Terminase	5	Terminase_2, Terminase_6N, DNA_Packaging, Terminase_3, Terminase_6C
Thy1	1	Thy1
Translation regulation	2	Thg1, Translat_reg
Transposase	1	DDE_3

Supplementary Table 7. Auxiliary metabolic genes found in the two S-CAM7-like viral genomes from bins IB_0.2_031 and IB_3.0_051 and the published genome of S-CAM7 isolate 0910SB42.

Gene	Function	IB_0.2_031	IB_3.0_051	SCAM-7
Hli	AMG Photosynthesis	No	No	Yes
Hsp20	AMG Cell Protection	No	Yes	Yes
CobS	AMG Other	No	Yes	Yes
MazG	AMG Other	No	Yes	Yes
PhoH	AMG Cell Protection	Yes	Yes	Yes
CP12	AMG Carbon regulation	Yes	Yes	No
CpeT	AMG Photosynthesis	Yes	Yes	No
Glutaredoxin	AMG Other	Yes	No	No

Chapter 2.

HI-C LINKS VIRUSES TO EUKARYOTES AND PROKARYOTES WITHIN A DEEP CHLOROPHYLL MAXIMUM OF THE PACIFIC OCEAN

Rathwell, C, Rocap, G

2.1 Abstract

Phytoplankton are critical drivers of biogeochemical cycles, thereby impacting the Earth's atmosphere and ocean. Viral infection in the surface ocean can reroute biogeochemical cycling to increase the biological pump or redistribute organic material to dissolved pools. The ultimate impact of viral communities is constrained by the susceptibility and availability of microbial hosts in their environment. Hi-C sequencing pairs viral sequences with host sequences when each are housed in the same cell at the time of sampling. Here we applied Hi-C sequencing to metagenomes in the oligotrophic Eastern Tropical North Pacific within a deep chlorophyll maximum to determine viral activity. Viruses were linked to eukaryotic phytoplankton including *Ostreococcus*, *Bathycoccus prasinos* and unclassified Chlorophyta. The *Bathycoccus prasinos* linked with both a known *Prasinovirus* virus and a *Mimivirus*. Viruses were also linked to prokaryotes, including a *Kyanoviridae* with a *Synechococcus* population, several distinct Caudoviricetes sequences with Alphaproteobacteria and Gammaproteobacteria and a crAss-like phage with Planctomycetes. Hi-C applied to a microbial community of a deep chlorophyll maximum thus provided evidence for multiple active infections within both prokaryote and eukaryote populations, and suggests alternate routes for DNA proximity in unlikely virus-host pairings.

2.2 Introduction

Phytoplankton photosynthesis in the surface ocean contributes nearly half of the oxygen in the planet's atmosphere[163,164]. Viral infection of phytoplankton communities has been shown to decimate populations[40], and contribute to biogeochemical cycling through the expression of virally encoded host auxiliary metabolic genes (AMGs)[46,47]. Virus-to-microbe ratios in the surface ocean have higher values and variation than at other ocean depths, which cannot be explained by a linear relationship with increasing microbial abundance[128], suggesting the need to study these dynamics on the scale of individually interacting pairs of viruses and microbes. The majority of viruses in the ocean are tailed bacteriophage belonging to the class Caudoviricetes[19]. Another class of marine viruses is Megaviricetes, which contains both the alga-infecting *Phycodnaviridae*[165] and *Mimiviridae,* which was first described infecting Amoebozoa [166] and has since been found to infect a range of eukaryotes including algal cells in the surface ocean[167]. The presence of both small bacteriophage and large mimiviruses that infect phytoplankton in the surface correlates to shifts in carbon export [36,168]. However, the sign and amplitude of the effect varies by location, suggesting that viral presence alone cannot determine viral impact, but rather the interactions of viruses and their hosts within their environment determine the ultimate impact of viral populations[168]. Understanding the role of viruses at the time of sampling is critical to tying viral behavior and virus-host interactions to ecological and biogeochemical findings. Here we apply Hi-C to assess viral interactions within a deep chlorophyll maximum in the Eastern Tropical North Pacific (ETNP). By linking DNA found within the same cell at the time of sampling we aim to provide evidence for viral activity and potential impact of both known and unknown viral populations in an important marine environment..

2.3 Materials and Methods

2.3.1 Sample Collection

Seawater samples were taken at an offshore site within the ETNP (P2: 16.58°N, 107.05°W) from the research vessel *Kilo Moana*, during cruise KM1920 (cast 33, October 6, 2019, local time of 08:15). The water properties were determined using a CTD-Rosette equipped with conductivity, temperature, transmissivity, chlorophyll fluorescence and oxygen (SBE 043) sensors. Duplicate 10-liter water samples were collected at a depth targeting the peak of the primary chlorophyll maximum (65 meters). The samples were filtered on 142 mm sequential 3.0 μm and 0.2 μm polycarbonate filters using in-line filtration, with the 0.2–3.0 μm size fraction processed further for microbial community analysis.

2.3.2 Filter Processing Onboard

One of the duplicate filters was used for metagenome sequencing and assembly and one for Hi-C library generation. The metagenome filter was frozen immediately and stored at –80°C. The second filter was treated with formaldehyde to generate Hi-C links. Soaking solutions of 1% formaldehyde and 10% glycine were created by mixing with seawater, from the same depth, that had been filtered through an Anodisc syringe-tip filter with a pore size of 0.02 μm. Filters were folded in half eight times with the sample-side facing in to create a cone-like shape and placed tip-first in 15 mL Falcon tubes containing 13.5 mL of the ultra-filtered 1% formaldehyde solution and incubated for 30 minutes. After 30 minutes a serological pipette between the filter and the side of the Falcon tube was used to deliver and mix in 1.5 mL of the 10% <0.02 μm glycine seawater solution. The filter was then incubated for another 30 minutes. A fresh serological pipette was used to remove all liquid volume from the Falcon tube, and the filters were frozen and maintained in –80°C until further processing.

2.3.3 Metagenome Library:

One half of the metagenome filter (not treated with formaldehyde) was used to extract DNA for a short-read sequencing library. The half filter was cut into smaller pieces which were placed in 1.5 mL 1X TE and subjected to a freeze-thaw at extreme temperatures. DNA was extracted using QuickGene-Mini80 with the DNA Tissue kit (Autogen, Cat No. DTS) and concentration assessed using a Qubit (dsDNA HS assay kit, Invitrogen Q3285). DNA was sequenced at the Northwest Genomics Center using NovaSeq sequencing on an S4 flow cell with a 2x150 base pair (bp) reads kit.

2.3.4 Hi-C Library

The formaldehyde-treated filter from the same depth was thawed and material was resuspended and pelleted from the filter. Each filter was cut into 1/16th pieces and the pieces were placed in a 50 mL Falcon tube containing 20 mL of 0.02 µm-filtered autoclaved artificial seawater (ASW) base. Filters were rotated in the dark at 16°C for 36 hours. The soaking pre-ASW was then removed and distributed among sterile 2.0 mL centrifuge tubes for pelleting, while a second wash was initiated. For both the first and second filter wash volumes, tubes were spun at 21,000 x g for 1.5 hours at 4°C. All but 50–100 uL of supernatant was removed. The remaining volume was used to resuspend each pellet. Resuspensions were combined into a single Eppendorf tube. A repeat spin cycle of 21,000 x g and 4°C was performed for 1.5 hours, and the supernatant was removed. The pellet was sent to Phase Genomics (Seattle, WA) for Hi-C library generation and HiSeq sequencing.

2.3.5 Assembly

Short read shotgun sequences generated with NovaSeq chemistry were cleaned using Trimmomatic (v0.39)[139], with the following specifications: `ILLUMINACLIP:bin/Trimmomatic-`

`0.39/adapters/TruSeq3-PE-2.fa:2:30:10:1:true`. These short-read sequences were assembled with MEGAHIT (v1.2.9)[140] using default parameters. Library quality was assessed for using QUAST[141] (v5.2.0) with default parameters.

2.3.6 Identifying Viral Contigs:

Virus contigs were identified using VirSorter2 (v2.2.4)[31], VIBRANT (v1.2.1)[169] and geNomad (v1.7.4, database 1.7)[170]. VirSorter2 was trained to include *Autolykiviridae* as described on https://github.com/jiarong/VirSorter2 , then run with the following parameters: `--include-groups 'dsDNAphage, ssDNA, NCLDV, lavidaviridae, autolykiviridae'.` VIBRANT was run using default parameters. Analysis in geNomad used the `end-to-end` pipeline with default settings. Any contig that was identified as viral by two out of three of these software programs was considered viral in this study. This approach was taken to utilize the strengths of each of these algorithms while avoiding false positives.

2.3.7 Viral Quality Check and Annotation

Quality of viral contigs from each cluster were used as input for CheckV (v1.0.1, database v1.5)[154] using the `end_to_end` pipeline. Output from geNomad was used to determine viral contig taxonomy. AMGs were found using DRAM-v (v1.5.0) [171], with kofam databases and skipping UniRef. All databases were setup and downloaded using `DRAM-setup.py` script on February 21, 2024. AMG summary files, generated by DRAM-v only included AMGs with a score of 1-3 and a flag of M (microbial metabolic gene) and an absence of flag B (three microbial metabolic genes in a row). Flags of K (known AMG), E (experimentally verified AMG), and F (AMG within k5bp of end of contig) were permitted. AMGs were also identified via VIBRANT viral analysis described above. AMG summary files from VIBRANT were combined with DRAMv summary files, and an AMG found using either method on our final viral contigs was considered valid.

Taxonomy of each virus was generated by geNomad[170] pipeline (see above). All viruses considered for potential virus-host links had to include at least one viral gene and be classified to virus family or lower.

2.3.8 Microbial Contig Taxonomy

Genes on all non-viral contigs were predicted using Prodigal (v2.6.3) and amino acid sequences were DIAMOND (v.2.0.15)[152] blasted through CAT (v5.2.3)[172] against a custom-db with the following parameters: `--index_chunks 1 --top 11 --I_know_what_Im_doing`. The custom database included all protein sequences from MarDB_(v1.5), MMETSP, MarFun (v3), and Viral RefSeq (v208). To capture eukaryotic sequences in our sample, all genes on the contigs were also predicted with MetaEuk (v6-a539d0) using `easy-predict` pipeline against the co-released taxonomically aware protein database which includes a broad representation of marine eukaryotes (https://wwwuser.gwdg.de/~compbiol/metaeuk/2020_TAX_DB/). Taxonomic assignments by MetaEuk were given by running the lowest common ancestry analysis with the command `tax2contig –majority 0.5 –tax-lineage 1 –lca-mode 2.` All non-viral contigs were processed with EukRep (v0.6.6), using the `strict` command. Final designation of a contig as Eukaryotic was determined using an integration of the results from the MetaEuk, Prodigal DIAMOND blast and EukRep. Contigs that had either Prodigal genes or MetaEuk genes aligned to eukaryotic sequences in the database were analyzed for taxonomy. When both gene-calling and alignment methods determined a sequence to be eukaryotic, that taxonomical classification remained. If those two data points disagreed but EukRep determined the contig to be Eukaryotic, the eukaryotic classification remained. Any contig that was determined to be eukaryotic through EukRep, but whose genes did not match any eukaryotic sequences in either the MetaEuk pipeline

or the Prodigal to custom DIAMOND blast pipeline described above was not considered eukaryotic.

2.3.9 Metagenome Assembled Genomes (MAGs)

A combinatorial approach was used to maximize the benefits of both proximity-linked bins and computationally derived bins. Short reads were mapped back to the assembly using Bowtie2 (v.2.3.5.1)[143] using default settings and then coverage files were generated with the MetaBat2 script `jgi_summarize_bam_contig_depths`. MetaBat2 (v2.15) and MaxBin2 (v2.2.7) were each used to create linkage-independent bins using default methods. ProxiMeta (July 27, 2023) was used to generate linkage-dependent bins, as a service provided by Phase Genomics (Seattle, WA). Bins produced by these three methods were then used as input to DAS Tool[173] (v1.1.7).

2.3.10 Microbial MAG Taxonomy

Initial taxonomic assignments were made for all bins by average contig taxonomy, weighted by length. Any bin which was determined to be Eukaryotic in this method utilized the taxonomy of the contigs within it as determined by MetaEuk. All other bins were assigned using GTDB-tk (v2.3.2, database r214)[174] with the classify_wf command. Completeness and contamination of each prokaryotic MAG was assessed using checkM (v1.1.3) [148].

2.3.11 Linking Assembled Contigs with Hi-C Reads

Hi-C read pairs were used to link microbial and viral contigs. Hi-C reads were quality checked using Phase Genomics script `hic_qc.py` found on https://github.com/phasegenomics. Hi-C reads were cleaned using BBDuk from the BBTools package (guide and downloads available at https://jgi.doe.gov/data-and-tools/software-tools/bbtools). First a file of known Illumina adapter sequences and the following settings, `ktrim=r k=23 mink=11 hdist=1 minlen=50 tpe tbo`, were used to trim adapter sequences. Next reads were quality trimmed with the command

`trimq=10 qtrim=r frm=5 minlen=50`. Finally, the first ten base pairs were trimmed using the option `ftl=10`[10]. These cleaned Hi-C paired end reads were mapped to assembled sequences using BWA-MEM (v0.7.17-r1188)[175] with the `-5SP` setting. To remove non-informative reads such as duplicates or unpaired reads, Hi-C reads were analyzed with SAMBLASTER (v0.1.26)[176]. SAMtools (v1.3.1) [177] was used to convert the sam file to a bam file using the `-F 0x904 -bS` settings, and subsequently sorted by name. Bam files were then converted to link frequency tables with the following columns: contig_a, contig_b, number of links. To visualize relative strength of links, a custom script was then run to generate a normalized score for each paired contig interaction. To roughly account for a broad range of biases simultaneously, such as contig length, contig abundance, restriction enzyme site abundance, etc., the number of links were normalized by self-links for each contig using the following equation:

$$\frac{(\#\ a{:}\ b\ links)}{\left[\frac{(\#a{:}\ a\ links) + (\#\ b{:}\ b\ links)}{2}\right]}$$

If both contig_a and contig_b had no self-links then a 1 was inserted into the denominator rather than 0 to avoid non-real numbers.

2.3.12 Determining Virus-Host Links

All contig-contig links that included a virus contig linking with a non-virus contig were examined as a potential informative virus-host link. These potential virus-host links were further reduced by only including contigs of at least 4000 bp in length that had at least two unique Hi-C read pairs between them [83,178]. These pairs were filtered further by only including host contigs that were taxonomically classified and were part of an uncontaminated bin (checkM lower than 10%) that had the same overall classification.

2.3.13 Analysis of Crassvirales and Prasinovirus Contigs

The *Phycodnaviridae* contig was blasted against the nr/nt database (April, 2024) to determine which sequences it most closely matched. Closest matches from each category, strains of *Bathycoccus* virus BpV and *Bathycoccus* virus BII-V, as well as environmental sequences were selected for alignment, in Mauve (v2.4.0)[179]. Gene annotations for the *Kyanoviridae, Crassvirales* and BpV1 from DRAM-v were plotted with the gggenes package in R.

2.4 Results and Discussion

Hi-C sequencing was paired with a traditional metagenome to determine the virus-host activity within the phytoplankton community from the oligotrophic Eastern Tropical North Pacific Ocean. The deep chlorophyll maximum spanned roughly 65 m to 85 m depth (Figure 1). We sampled at 65 meters, below which a steep oxycline marked the transition to the oxygen deficient zone at 112 m. The 0.2–3.0 μm size fraction was used to create both a short read metagenomic and a Hi-C sequencing library (Table 1). Assembled contigs were taxonomically identified, and 65,225,537 Hi-C read pairs were aligned to 520,282 contigs in the assembly to link DNA occupying the same cell at the time of sampling (Figure 2). Interactions, determined by Hi-C read pairs, between viruses and taxonomically classified sequences found in a classified MAG were examined. There were 270 final MAGs that were classified and not contaminated, by checkM analysis(Figure 3, Table 3), and therefore potential viral hosts. Of those 270 MAGs, 81 were classified as eukaryotic. Within the prokaryotic bins, 14 were above 90% complete and 64 were at least 50% complete. Out of the 68,294 contigs that were greater than 4000 bp, 11,131 were viral, and only 491 were taxonomically identified (Table 2, Figures 4 and 5). Most classified viral sequences were determined to be Megaviricetes and Caudoviricetes types (Figure 4), with varying checkV

completeness. Of the 491 classified viral contigs greater than 4kb, 22 were linked to a host with our Hi-C methods (Figure 6), indicating they were present in host cells at the time of sampling and abundant enough to be assembled and Hi-C linked.

By targeting the deep chlorophyll maximum, the Hi-C method revealed linkages between viruses and photoautotrophic hosts. One viral contig just over 4000 bp, identified as *Kyanoviridae,* was linked with two *Synechococcus* contigs found in a single *Synechococcus* bin (Figure 8, Table 3). This *Kyanoviridae* virus contained *psbA* and *psbD*, known auxiliary metabolic genes for cyanophage[47] (Figures 6 and 8), which suggests that Hi-C could be used as a method to understand which viral AMGs are found within host cells, and therefore which AMGs may have a deeper impact on host genomic potential. Twelve of the 22 viruses that could be linked to hosts were linking with eukaryotic photoautotrophs. One unclassified Megavirus linked to the picoeukaryote *Ostreococcus* (Figure 6). Two bins classified only as Chlorophyta were linked with viral sequences as well (Table 3, Figure 6). One of the viral contigs found within the unclassified Chlorophyta bin was identified as Mimiviridae while the other two were determined to be Unclassified Caudoviricetes, showing an unexpected interaction between a eukaryote and a bacteriophage. Each of the three contigs interacting with the unclassified chlorophytes were less than 10 kb, which may contribute to the unexpected taxonomic assignments. Perhaps most notable though was a *Bathycoccus prasinos* bin which had 11 separate contigs that linked with various virus contigs, including 3 *Mimiviridae* contigs and a *Phycodnaviridae* previously isolated on this species (Figure 6). Two contigs less than 10 kb were also identified as bacteriophage and carried bacteriophage genes, which is not likely a misclassification and therefore is an unanticipated interaction. The multiplicity of viral interactions with the *Bathycoccus prasinos* bin may be explained by the fact that Hi-C linkages detect interactions across a population; if different cells were experiencing

infection from different viruses concurrently, we cannot distinguish the infection of individual cells by different viruses[178].

The Phycodnaviridae interacting with *Bathycoccus prasinos* was annotated and blasted against NCBI databases. We call the phycodnavirus sequence in this sample Prasinovirus P2_DCM_2019, for the station (P2), 65m (deep chlorophyll maximum) and year (2019). Prasinovirus P2_DCM_2019 had most gene calls aligned with the Viral RefSeq sequences of *Bathycoccus sp. RCC1105* virus BpV1 (BpV1), a lytic virus isolated off the Eastern coast of France in the Mediterranean Sea[180] (Figure 7, Table 4). The BpV1 genome is nearly 200 kbp, but this contig was only 50,000 bp (Figure 7), suggesting that it is an incomplete viral genome. The *Prasinovirus* in our sample also aligned to additional *Prasinovirus* sequences, including BpV2 and BpV3 also isolated on *Bathycoccus sp* RCC1105 (Table 4), an environmental sequence collected in the Western Pacific Ocean in 2019[181], and the more recently described *Bathycoccus sp.* RCC716 viruses 1,2 and 3, (BII-V1, BII-V2, BII-V3) which were isolated from surface water at the Bermuda Atlantic Time Studies station[182]. Interestingly, metaviromes from the Eastern Tropical Pacific showed mutually exclusive dominance of BVII viruses over BpV near our station, suggesting that our viral sequence could be a BVII type[182]. Analysis of the Prasinovirus_DCM_P2_2019 sequence in the absence of the Hi-C method would be sufficient to infer that the host was *Bathycoccus prasinos* in our sample, but the Hi-C method confirms that at the time of sampling this virus was found in this population of cells. Based on the laboratory analyses we can predict that the virus-host interaction here is likely a lytic infection.

Hi-C was also able to link bacteriophage with members of the heterotrophic prokaryotic community within the deep chlorophyll maximum (Figure 6). Of the 9 bacteriophage linked with prokaryotic members of the microbial community 7 were unclassified, suggesting these are novel

virus-host interactions. Both Alphaproteobacteria and a Gammaproteobacteria of the order Arenicellales were linked with viral sequences through Hi-C (Figure 6). One Caudoviricetes contig linked with three Arenicellales contigs all within the same Arenicellales MAG, and a second Caudoviricetes contig linked with a fourth Arenicellales contig of the same MAG. One 4,009 bp unclassified Caudoviricetes contig linked with a fairly incomplete Rhodobacterales MAG, but three unclassified Caudoviricetes viral contigs ≥ 10 kb linked with a complete Alphaproteobacteria MAG, with GTDB-tk classifications down to genus level. Hi-C sequencing within gut populations suggest active lytic reproducing viruses are targeted with Hi-C analyses[70]. Data here suggests Alphaproteobacteria populations were experiencing infection in the chlorophyll maximum at the time of sampling, in October of 2019.

A crAss-like phage was also linked to a nearly complete Planctomycetes bin of the order Pirellulales (Figures 6 and 8). CrAss-like phage from the human gut have been described through cross-assembly (crAss) sequencing efforts[183] and later isolated on Bacteroidetes cultures[184,185]. However, the Crassvirales appear to be polyphyletic with many different crass-like virus sequences found in a wide range of environments[186]. This study is the first, to our knowledge, to directly link *Pirellulales* with a crAss-like phage. Apart from the isolation of a lytic phage, Verrucophage P8625[187], infecting Verrucomicrobia and the previous Hi-C studies linking viral sequences to Planctomycetes MAGs[82], the viruses that infect the members of the Planctomycetes, Verrucomicrobia and Chlamydiae (PVC) supergroup remain uncharacterized. This 66 kb Crassvirales contig linking with a nearly complete Pirellulales bin indicates a novel virus-host relationship not detected before, and a novel environmental role for a Caudoviricetes predominantly characterized in animal microbiomes.

Unexpectedly, links between viruses and bacteria from the PVC group show a broad range of interactions. There is a contig identified as a Mimivirus linked strongly to both *Bathycoccus prasinos* and Verrucomicrobiales. Additionally, a *Pirellulales* bin from the Planctomycetes phylum is found linked with a megavirus sequence, and one Caudoviricetes sequence is linked with that same Pirellulales bin and an Alphaproteobacteria bin. The cause of this interaction cannot be determined from this study alone. There is evidence of heterotrophic protistan cells consuming large viruses, suggesting an alternate route to DNA proximity other than infection[134,188]. Futhermore, the recent analysis of the heterotrophic eukaryote *Symbiomonas scintillans* in culture led to the discovery of the endosymbiosis of multiple Megaviruses, including BpV and Mimivirus, with no detectable bacterial endosymbiont[189]. Chlorophyta have not been shown to prey on viruses, but as eukaryotic organisms may be capable of taking up material through endocytosis, forcing proximity for these viruses and microbes beyond viral infection. Alternately, another endosymbiotic route could explain the cross-kingdom interactions. Verrucomicrobia have been identified as symbionts of ciliates and even other endosymbionts[190,191]. The Verrucomicrobia bin interacting with Mimiviridae has a GC content of 38%, indicative of endosymbiosis[192,193] and markedly lower than the average GC content of the remaining Verrucomicrobiota bins (51.9%). Interestingly, a fellow member of the PVC supergroup *Parachlamydia acanthamoebae* has been shown to prevent infection by *Mimiviridae* when present as an intracellular symbiont to their *Acanthamoeba* host[194], suggesting that members of this supergroup may play an interactive role with *Mimiviridae*.

Ultimately, Hi-C linkages can detect virus-microbe interactions across diverse taxa. Hi-C links can confirm the occurrence of infection in two known host-virus populations, but also linked microbial hosts for viruses that were unclassified and hosts that were previously unknown.

Specifically, Hi-C has the ability to reveal which of a subset of viral sequences are within their susceptible host at the time of sampling. This ability is important in understanding viral interactions and how they may range across environmental gradients which might change over time. The work overall underscores the utility and significance of Hi-C as a tool for assessing viral ecology and impact in the marine environment.

2.5 Acknowledgements

We thank the captain and crew of the R/V *Kilo Moana*, Bailey Armos for her help collecting samples at sea, and Susan Burke for extracting the DNA. Thanks to Arthur Nowell for generous feedback on the manuscript. This work was funded by a National Science Foundation Graduate Research Fellowship and the UW Hall Conservation Genetics Research Award to CR and NSF grants DEB-1542240 and OCE-2022911 to GR.

2.6 Figures

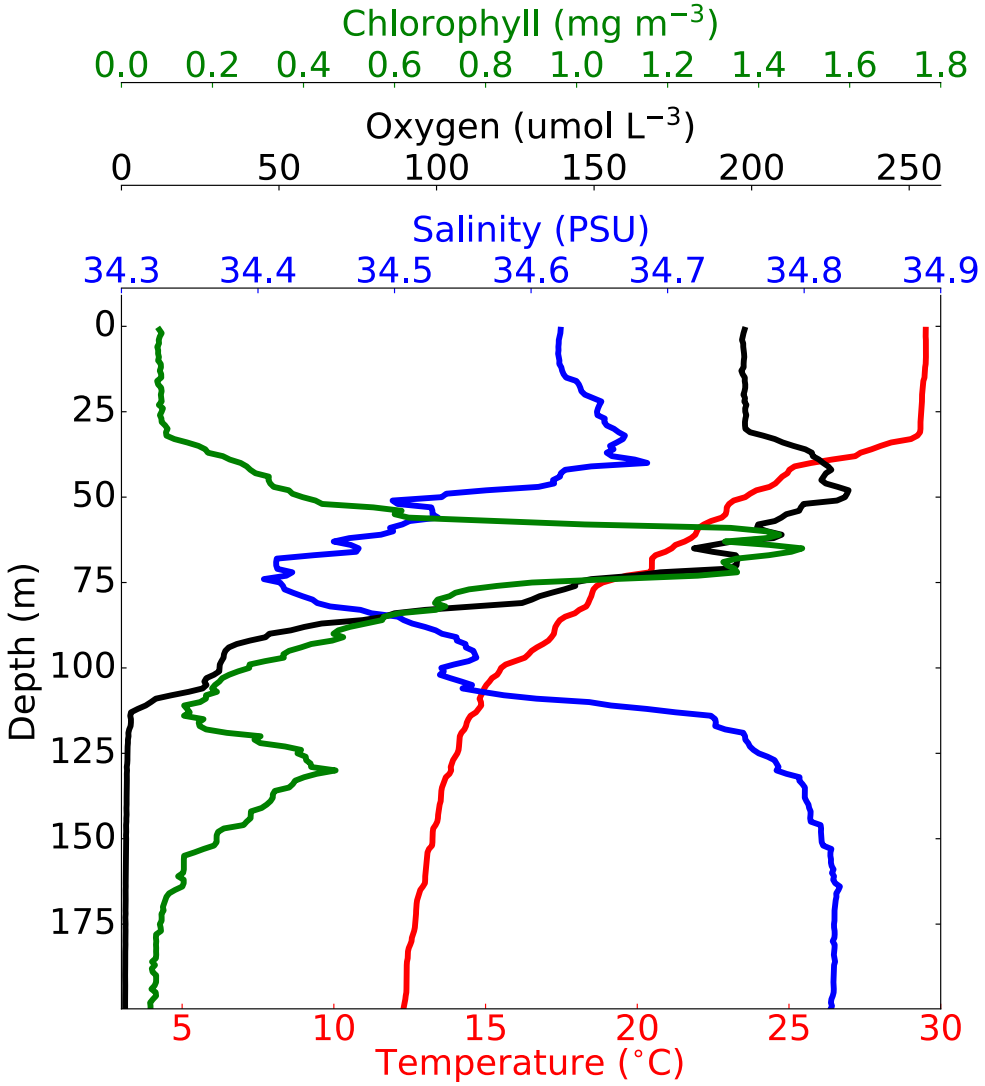


Figure 1. Depth profile of station from the same cast as sampling, as measured by CTD instrument package. Salinity (blue), temperature (red), chlorophyll fluorescence (green) and oxygen (black) from 0 to 200 meters depth are displayed.

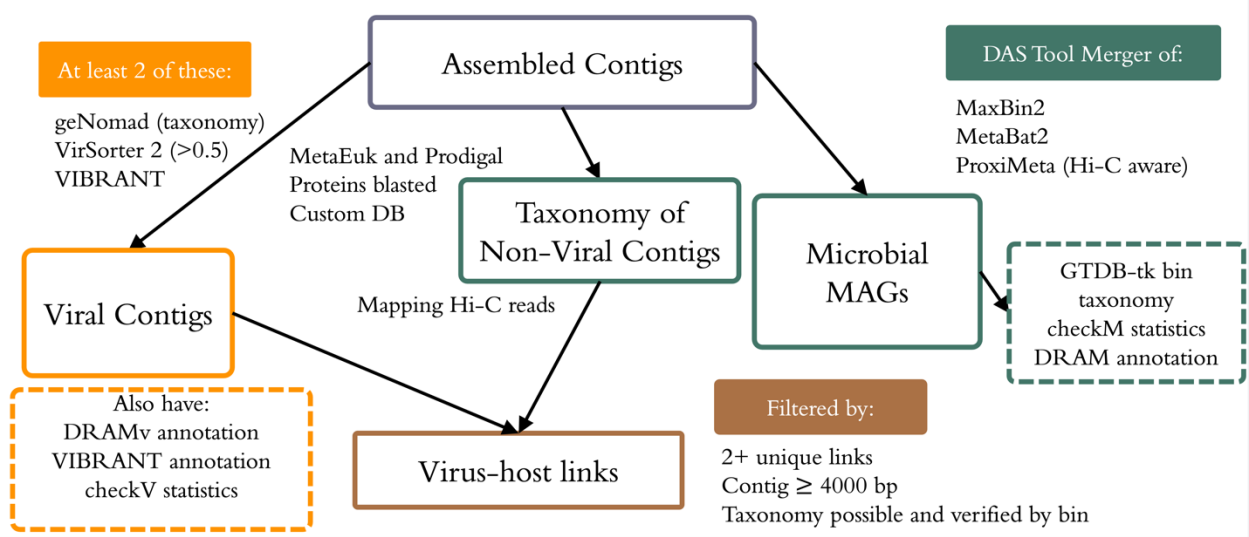


Figure 2. Computational workflow to generate virus host pairs. Assembled contigs were binned into metagenome assembled genomes using a DASTool merger of three binning methods. Bins were assessed and annoatated. Viral contigs were determined through 3 different algorithms and also assessed and annotated. Contigs from microbial bins that were linked with viral contigs through Hi-C reads were then evaluated further for virus-microbe interactions. Only contigs that were 4000 basepairs or more were included. Contigs also had to be classified to some degree and microbial contig needed to be classified and in a classified bin where those classifications matched.

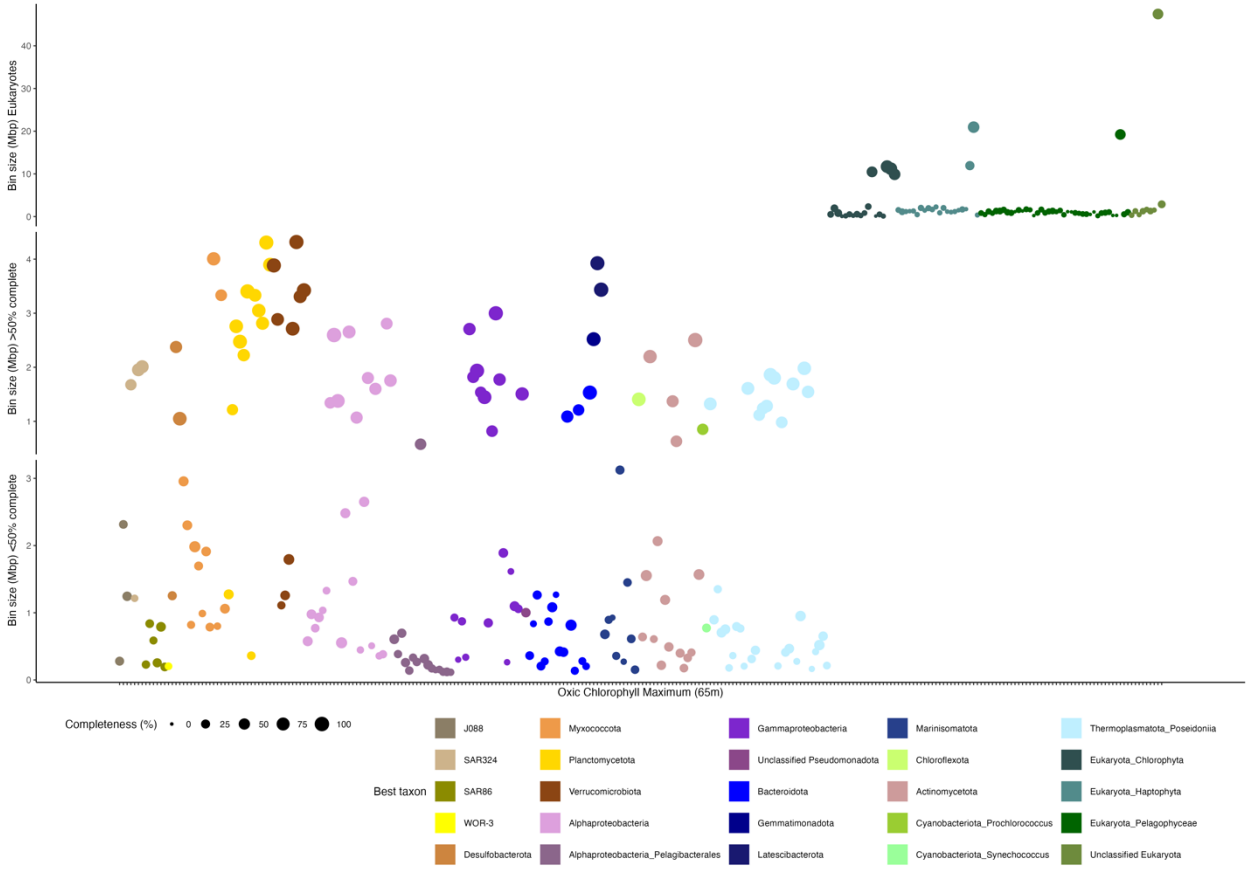


Figure 3. All metagenome assembled genomes (MAGs) that qualified as potential hosts. Each dot represents a MAG colored by taxonomy, which was determined by either GTDB-tk (prokaryote bins) or contig identification by a blast against a marine focused database. CheckM completion determines the width of each dot.

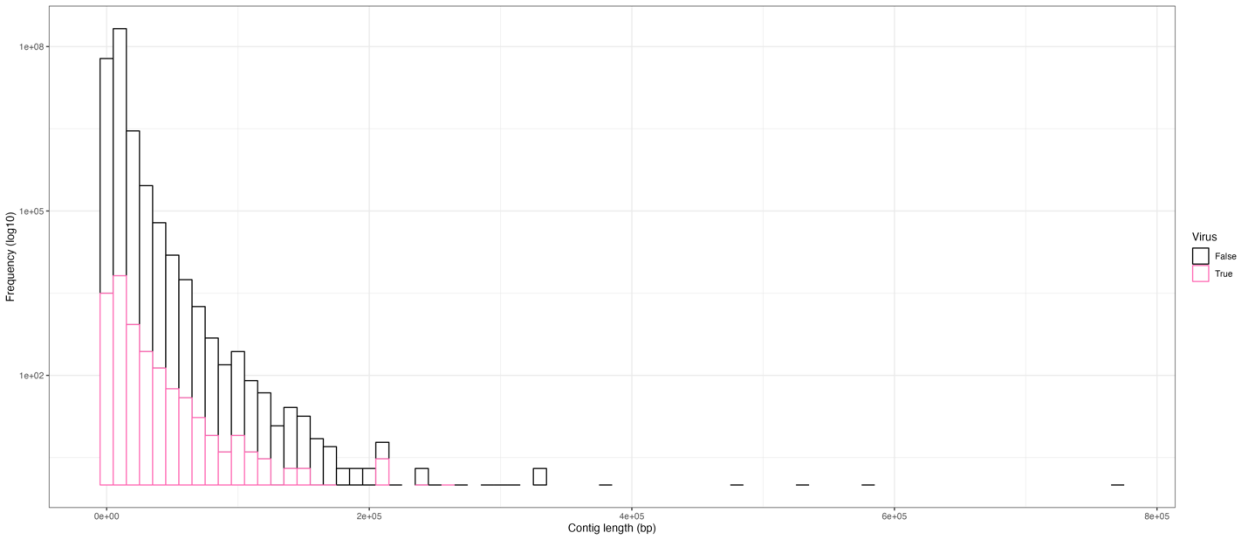


Figure 3. Histogram of the frequency of virus and non-virus contigs of different lengths. Frequency is given as the log-10 scale for clarity. Virus contigs are in pink, and non-virus contigs are shown as black bars.

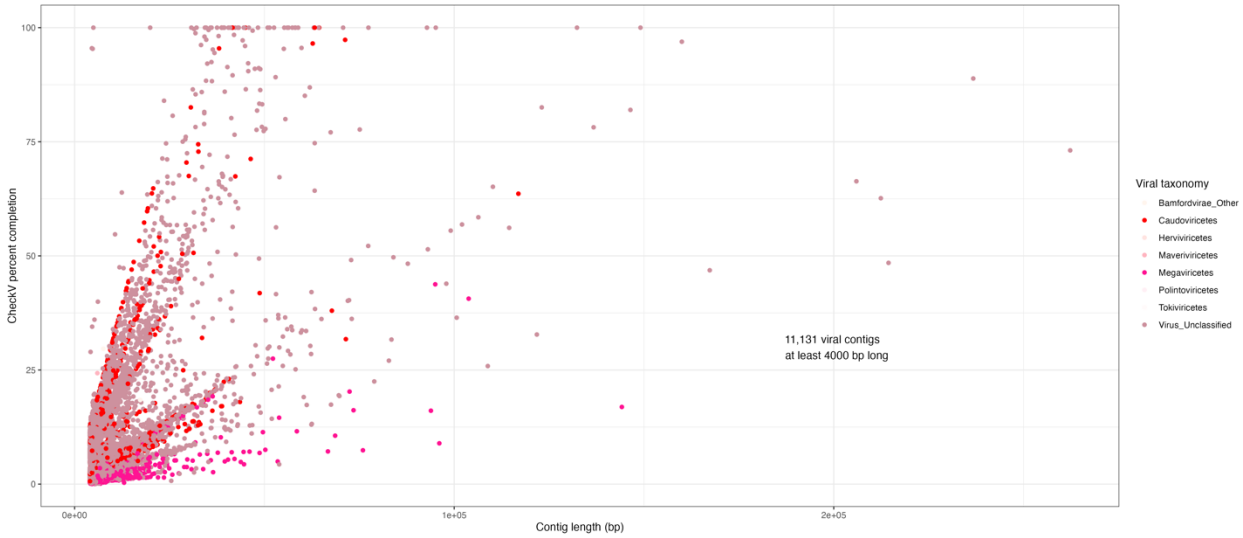


Figure 4. CheckV completion percentage versus virus contig length for the 11,131 viral contigs 4000 bp or longer. Each dot represents a viral contig and the dots are colored by viral taxonomy as determined by geNomad.

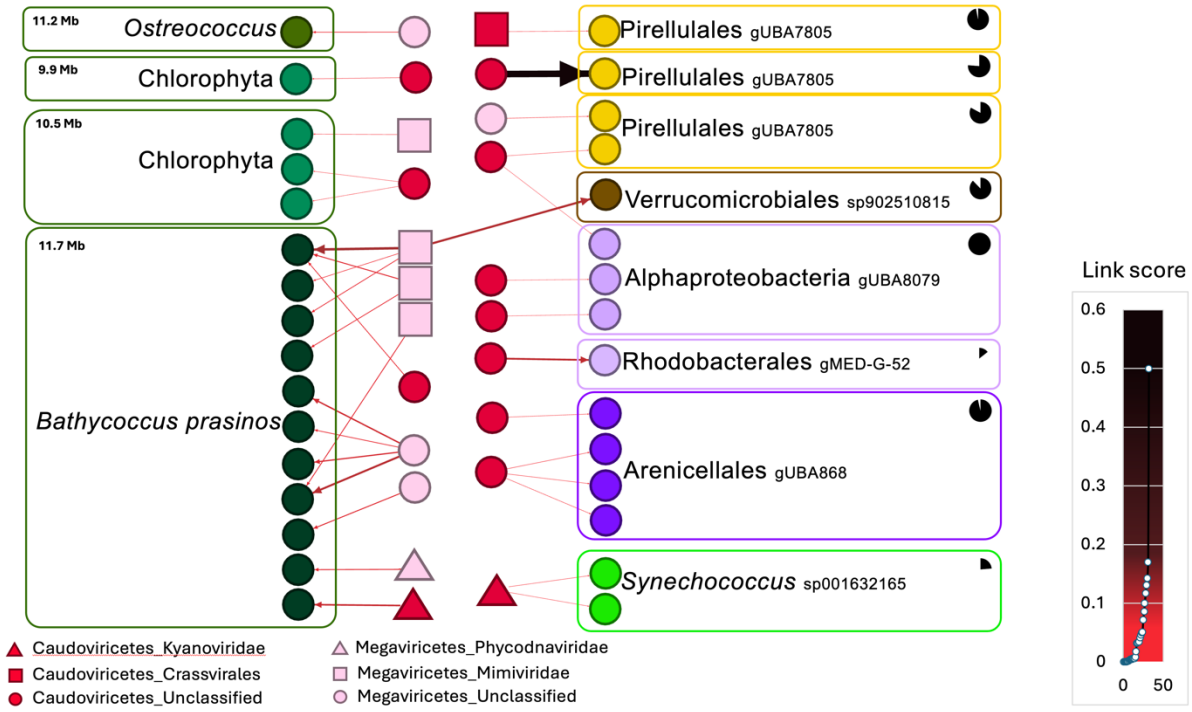


Figure 4: All virus-host links between classified contigs of at least 4000 bp in a Pacific Ocean deep chlorophyll maximum. Color of each circle represents taxonomic identification of each contig. The square around each microbial contig represents the bin that contig was found in. Taxonomy is written for each bin. Viral taxonomy generated by geNomad is coded by shape and color. Each bin's CheckM completeness is included as a pie chart in the upper-outer corner for prokaryotic bins. Bin size is displayed in the upper left corner of eukaryotic bins The links are colored based on scored weight of the frequency of links between the two contigs. A darker color represents a high frequency of links compared to the self-links of each contig from the pair.

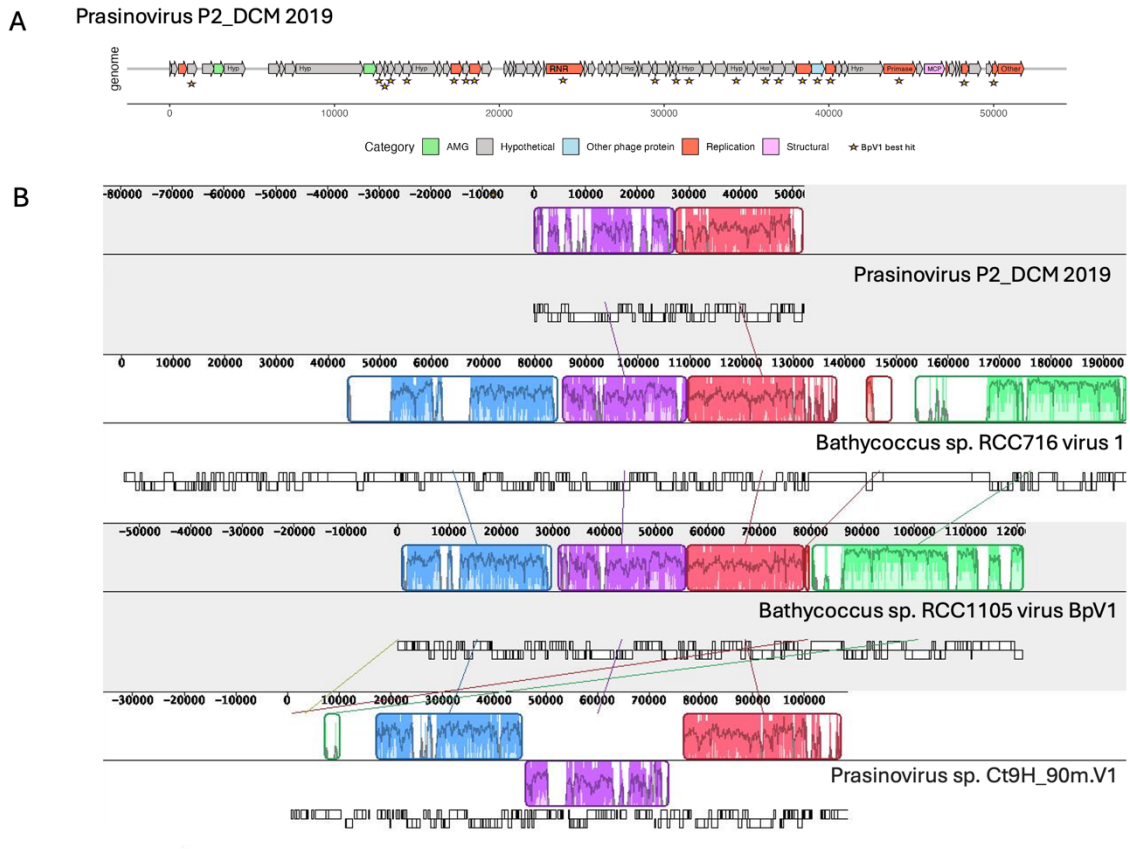


Figure 5:**A.** Gene content of *Prasinovirus P2_DCM_2019* as determined by DRAM-v annotations of Prodigal gene calls. Best hit to viral refseq proteins are shown and colors are described in the key. **B.** Alignment *Prasinovirus P2_DCM 2019* to reference sequences of cultured *Bathycoccus* virus genomes and environmental *Prasinovirus* Ct9H_90m.V1.

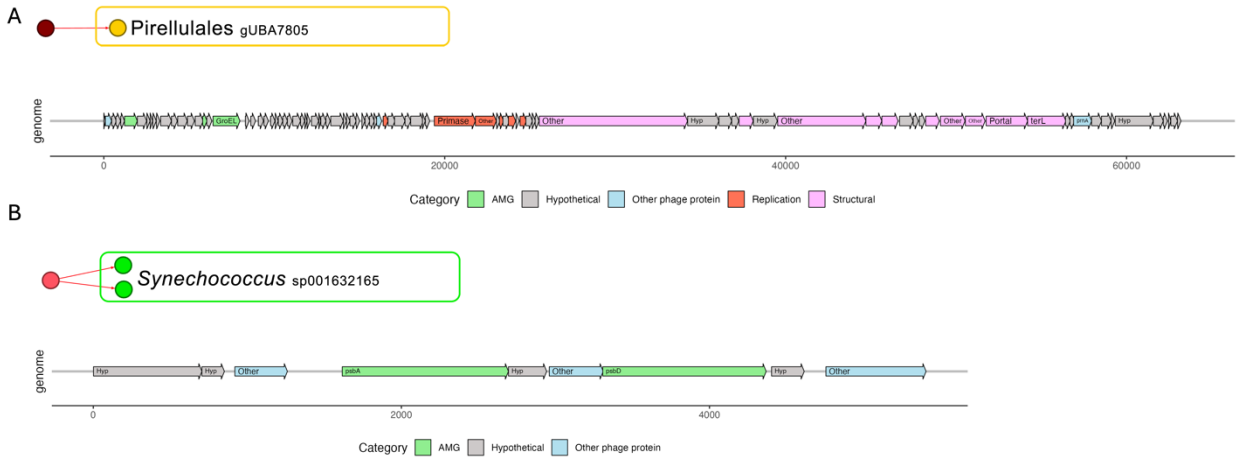


Figure 6. DRAM-v annotation for **A.**CrAss-like phage interacting with *Pirellulales* bin and **B**. *Kyanoviridae* that was linked with *Synechococcus* bin. Total length plotted on x axis and different scales for each.

2.7 Tables

Table 1 Sequencing and Assembly Statistics for All Contigs Greater Than 1000 bp

Metric	Value
NovaSeq read pairs (2x150 bp reads)	309,018,748
Assembly length (bp)	2,044,256,727
Number of contigs	967,198
Longest contig	772,665,893
N50	3,380
L50	1,690

Table 2 Number of Contigs Identified as Viral By Various Virus Identification Methods

Method	Number of Contigs (≥ 1000 bp)	Number of Contigs (≥ 4000 bp)	Number of Contigs (≥10000 bp)
Total contigs in assembly	967,198	68,294	14,000
geNomad	19,596	12,385	2,817
VirSorter2 (>0.5)	23,771	14,289	3,055
VIBRANT	11,421	8,038	2,173
Virus by at least 2 of 3 of the methods above	16,815	11,131	2,733
Virus by all three of the methods above	9,408	7,797	2,137

Table 3 270 Metagenome assembled genomes that were classified beyond kingdom. The 189 prokaryotic bins are less than 10% contaminated with single copy genes, assessed by CheckM.

Bin	CheckM complete	CheckM contam.	N50	GC	Genome size	Number contigs	GTDB Taxonomy
bin.89_sub	39.37	6.4	4329	34.9	519139	129	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__
2030_maxbin.232_sub	24.58	4.65	2786	35.1	794890	315	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-I;s__
bin.101_sub	10.12	1.89	3775	43.6	205459	54	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-K1;s__
bin.217	83.73	3.2	49762	42.2	1861758	71	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-K2;s__MGIIa-K2 sp002699425
bin.55	83.73	0	84334	45.6	1980981	40	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-L1;s__
bin.7	66.6	0.8	9075	61.6	1545066	229	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-L3;s__MGIIa-L3 sp002496845
bin.221	83.33	0	50026	54	1803370	60	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__Poseidonia;s__
bin.184	72.52	0.8	10194	48.3	1609147	203	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__Poseidonia;s__Poseidonia sp002726495
2030_maxbin.075_sub	27.95	3.76	2801	48.6	894946	351	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__;s__
bin.70_sub	6.07	0	5829	47.9	163862	27	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-N1;s__MGIIb-N1 sp002504845
2030_maxbin.142_sub	18.57	4.7	3581	54.2	1350325	490	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-N2;s__
bin.36	74.93	1.6	39877	40.7	1691549	85	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-N2;s__MGIIb-N2 sp002503045
bin.91	28.58	0.93	4746	41.8	651957	146	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-N2;s__MGIIb-N2 sp002718215
2030_maxbin.002	71.6	8.84	11410	45.3	1322551	194	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O1;s__MGIIb-O1 sp002497025
bin.293_sub	23.32	5.34	5797	37.5	411119	83	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O2;s__MGIIb-O2 sp002497245
2030_maxbin.182_sub	33.87	5.26	6982	37.3	753327	189	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O2;s__MGIIb-O2 sp002497245
bin.210	67.85	2.56	30288	37.4	1283286	79	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
2030_maxbin.205_sub	11.6	1.2	8949	52.8	362250	73	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
2030_maxbin.147_sub	33.93	6.86	4641	47.2	705075	223	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
2030_maxbin.323_sub	18.22	8.88	1624	51.1	766223	453	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.74	10.28	0	18532	55.2	420355	40	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.298	31.6	2.8	24773	47.8	466036	31	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.49	40.11	8	19946	49	951247	80	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.98	14.8	0	14565	55.2	214158	21	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.187_sub	20.67	3.2	28500	48.6	312402	23	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.268_sub	12	4	10705	54.3	208382	23	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
2030_maxbin.204_sub	13.31	0.29	3210	51.2	179318	71	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O5;s__
bin.193	54.36	8.06	24819	46.3	1115784	61	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O5;s__MGIIb-O5 sp002499865
bin.190	27.26	0.93	15179	46.8	441314	41	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O5;s__MGIIb-O5 sp002501805
bin.272	55.87	1.6	47096	47	980805	27	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O5;s__MGIIb-O5 sp002730095
bin.43_sub	9.35	0	41107	48.6	276320	11	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O5;s__MGIIb-O5 sp002730095
bin.201_sub	61.08	2.72	9384	48.1	1239180	162	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-P;s__
bin.244	17.95	0	49854	42.1	408600	16	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__MedAcidi-G1;s__MedAcidi-G1 sp002699555
2030_maxbin.004_sub	47.89	4.56	8658	45.3	1553362	291	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__MedAcidi-G2B;s__
2030_maxbin.093_sub	35.25	1.88	3954	43.7	1191621	364	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__MedAcidi-G2B;s__
2030_maxbin.046_sub	34.84	6.44	4346	45.4	2065567	608	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__MedAcidi-G2B;s__

bin_66	21.37	0	39347	44.7	176221	4	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__MedAcidi-G2B;s__MedAcidi-G2B sp003215175
bin.4_sub	43.79	1.72	6147	51.7	1569054	272	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__
2030_maxbin.013	81.2	2.14	84213	51.9	2197474	101	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__MedAcidi-G3 sp002722565
bin.145_sub	21.55	0	92881	52.1	328760	8	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__MedAcidi-G3 sp002722565
bin.295	97.75	6.84	61552	51.9	2503325	73	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__TK06;g__UBA7388;s__UBA7388 sp002470695
2030_maxbin.444_sub	28.57	8.79	1600	44.8	491675	306	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__UBA8592;g__UBA8592;s__
bin_33	55.21	0.85	27653	45.5	630850	29	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__UBA8592;g__UBA8592;s__
bin_39_sub	25.86	0	51036	44.6	398726	11	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__UBA8592;g__UBA8592;s__
2030_maxbin.036_sub	15.33	5.17	3127	36.2	608273	219	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Actinomarinales;f__Actinomarinaceae;g__;s__
2030_maxbin.072_sub	29.27	7.89	1981	30.5	218677	116	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Actinomarinales;f__Actinomarinaceae;g__;s__
2030_maxbin.003_sub	23.24	2.56	5342	33.1	640032	176	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Actinomarinales;f__Actinomarinaceae;g__Actinomarina;s__
bin_21_sub	61.46	3.11	3540	62.2	1371275	412	d__Bacteria;p__Actinomycetota;c__Actinomycetia;o__Nanopelagicales;f__S36-B12;g__S36-B12;s__S36-B12 sp002728915
bin.105_sub	34.48	1.72	12000	26.5	425187	51	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__MED-G16;g__GCA-2715645;s__
bin.283_sub	15.52	0	13444	24.9	280809	36	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__MED-G16;g__GCA-2715645;s__
2030_maxbin.412	35.97	7.55	2194	28.2	1081971	538	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__BACL11;g__CAJWCP01;s__
2030_maxbin.515	7.23	2.41	1467	28.9	1268472	827	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__BACL11;g__CAJWCP01;s__
2030_maxbin.296	19.91	6.74	2937	30.7	868361	355	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__BACL11;g__GCA-2716345;s__
bin.73	93.8	7.71	17042	34	1531801	138	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__BACL11;g__TMED123;s__
bin.280_sub	52.59	3.45	4984	27.2	1209801	285	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__GCA-2719315;s__
2030_maxbin.089_sub	9.47	0.86	2601	30.3	835832	365	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__GCA-2719315;s__
bin.255_sub	48.75	5.16	4102	29.9	815415	229	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Marisimplicoccus;s__
bin.266_sub	17.55	0	8330	29.7	136857	20	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Marisimplicoccus;s__
2030_maxbin.203	16.17	0.93	2282	34	276709	126	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Marivariicella;s__
bin.62	13.79	0	3638	39.6	203514	60	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA8316;s__UBA8316 sp002687665
2030_maxbin.038	22.56	4.2	9018	32.7	362187	83	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__SAT41;g__;s__
2030_maxbin.124_sub	25.86	8.62	3614	30.7	1263974	456	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__SAT41;g__;s__
bin.166_sub	63.49	7.96	5447	30.5	1087793	242	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__TMED113;g__SP256;s__SP256 sp002722215
bin.124_sub	30.17	0	4459	38.4	413770	107	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__UA16;g__UBA974;s__
2030_maxbin.190_sub	23.69	4.08	1793	31.8	206314	108	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__NORP286;f__NORP286;g__;s__
bin.297	86.8	1.19	51184	35.3	1406976	56	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA826;g__UBA826;s__
bin.121	51.65	4.65	4690	38.1	852372	208	d__Bacteria;p__Cyanobacteriota;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Prochlorococcus;s__Prochlorococcus sp003214695
bin_26_sub	23.67	1.62	2321	54.2	773168	319	d__Bacteria;p__Cyanobacteriota;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Synechococcus_C;s__Synechococcus_C sp001632165
bin_14	63.2	1.29	4141	60.9	2374942	615	d__Bacteria;p__Desulfobacterota_B;c__Binatia;o__UBA12015;f__UBA12015;g__UBA12015;s__
2030_maxbin.574_sub	25.92	3.51	1648	61.1	1253191	780	d__Bacteria;p__Desulfobacterota_B;c__Binatia;o__UBA12015;f__UBA12015;g__UBA12015;s__
bin_25	86.04	0.61	182027	31.2	1048528	15	d__Bacteria;p__Desulfobacterota_D;c__UBA1144;o__UBA1144;f__UBA1144;g__TMED58;s__
2030_maxbin.084	91.14	8.16	13108	55.1	2519896	405	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Longimicrobiales;f__UBA6960;g__UBA2589;s__UBA2589 sp002500925
2030_maxbin.650_sub	22.57	4.86	1651	41.1	2314718	1402	d__Bacteria;p__J088;c__;o__;f__;g__;s__
2030_maxbin.674_sub	29.3	1.91	1495	55	1245124	809	d__Bacteria;p__J088;c__;o__;f__;g__;s__
2030_maxbin.154_sub	25.01	3.69	1974	35.1	279889	140	d__Bacteria;p__J088;c__;o__;f__;g__;s__
bin.13_sub	93.41	9.89	18777	42.5	3925013	366	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA8231;f__UBA8231;g__GCA-2693325;s__
bin.3	97.25	4.4	79794	40.9	3435164	102	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA8231;f__UBA8231;g__GCA-2693325;s__
2030_maxbin.032_sub	17.24	5.17	4739	31.5	894558	277	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__Marinisomataceae;g__;s__

2030_maxbin.007_sub	30.63	7.05	4344	32.6	679164	222	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__Marinisomataceae;g__;s__
2030_maxbin.207_sub	21	0	3140	35.8	1450524	566	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__Marinisomataceae;g__;s__
2030_maxbin.060_sub	5.95	0.95	7160	35.9	927171	244	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__Marinisomataceae;g__Marinisoma;s__
bin.87_sub	20.69	0	22889	31.5	152674	11	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__Marinisomataceae;g__Marinisoma;s__
2030_maxbin.165_sub	25.8	5.17	3990	36.2	3123767	1034	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA1611;g__CAJWHU01;s__
2030_maxbin.202_sub	6.78	0.02	2209	34.3	272176	130	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA1611;g__GCA-002708715;s__
bin.273_sub	22.41	1.72	381896	33.5	611848	20	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__SCGC-AAA003-L08;f__;g__;s__
2030_maxbin.079_sub	17.54	1.75	2578	28.7	357960	163	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__SCGC-AAA003-L08;f__TMED6;g__;s__
2030_maxbin.704_sub	13.08	0	1244	67.1	988809	780	d__Bacteria;p__Myxococcota_A;c__UBA9160;o__UBA9160;f__UBA4427;g__UBA4427;s__UBA4427 sp002709835
bin.178_sub	12.93	0	1977	67.5	800892	391	d__Bacteria;p__Myxococcota_A;c__UBA9160;o__UBA9160;f__UBA4427;g__UBA4427;s__UBA4427 sp002709835
2030_maxbin.723	22.4	0.74	1341	54.4	786438	569	d__Bacteria;p__Myxococcota;c__UBA4151;o__UBA4151;f__UBA4151;g__CACGMS01;s__
bin.171_sub	80.25	1.68	11605	58.5	4006652	447	d__Bacteria;p__Myxococcota;c__UBA796;o__UBA796;f__;g__;s__
2030_maxbin.707_sub	30.98	0.69	1628	62.3	1911831	1180	d__Bacteria;p__Myxococcota;c__UBA796;o__UBA796;f__UBA796;g__;s__
bin.18	56.3	1.75	8215	55.5	3332419	481	d__Bacteria;p__Myxococcota;c__UBA796;o__UBA796;f__UBA796;g__UBA796;s__
2030_maxbin.503_sub	32.46	0	2783	55.8	2301750	919	d__Bacteria;p__Myxococcota;c__UBA9042;o__GCA-2863065;f__;g__;s__
bin.250_sub	32.22	0	5154	53.3	1060229	223	d__Bacteria;p__Myxococcota;c__UBA9042;o__GCA-2863065;f__GCA-2863065;g__;s__
2030_maxbin.696	48.43	3.23	1665	62.8	1982421	1194	d__Bacteria;p__Myxococcota;c__UBA9042;o__PHBI01;f__;g__;s__
2030_maxbin.693_sub	17.41	7.42	1437	59.7	819884	550	d__Bacteria;p__Myxococcota;c__UBA9042;o__UBA3505;f__;g__;s__
2030_maxbin.701_sub	24	3.51	1561	54.9	1695817	1081	d__Bacteria;p__Myxococcota;c__UBA9042;o__UBA3505;f__UBA3505;g__;s__
2030_maxbin.356	35.53	6.9	2042	50.3	2955798	1548	d__Bacteria;p__Myxococcota;c__XYA12-FULL-58-9;o__XYA12-FULL-58-9;f__JABJBS01;g__JABJBS01;s__
bin_42_sub	21.58	0.88	5092	42.8	361004	81	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__GCA-002718515;s__GCA-002718515 sp913046275
2030_maxbin.653	49.76	1.75	1913	65.6	1215170	657	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__JABAAC01;s__
bin_13	90.43	4.04	6007	56.2	2473984	505	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__UBA12567;s__
bin_6_sub	71.48	3.21	5108	49.2	3331581	755	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__Mariniblastus;s__
bin.209	89.88	1.25	13473	49.1	3897525	368	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__UBA721;s__UBA721 sp002348315
bin.107	94.05	1.18	100350	48.2	4307246	76	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__UBA721;s__UBA721 sp002705165
bin_10_sub	85.71	6.32	3248	48.9	2757739	928	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__UBA1268;g__UBA1268;s__
2030_maxbin.593_sub	34.81	6.29	1461	48.7	1273556	826	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__UBA1268;g__UBA1268;s__
bin_9	76.23	1.36	3898	54.1	2813790	795	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__UBA7805;g__UBA7805;s__
bin_4	97.01	3.55	6627	48.7	3403021	640	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__UBA7805;g__UBA7805;s__
bin_7	82.85	5.66	3991	51.7	3049352	863	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__UBA7805;g__UBA7805;s__
bin_15_sub	66.05	0	3744	52.4	2224547	637	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Planctomycetales;f__Planctomycetaceae;g__S139-18;s__
bin.231_sub	29.39	6.9	5808	38.6	1001817	211	d__Bacteria;p__Pseudomonadota;c__;o__;f__;g__;s__
2030_maxbin.099_sub	34.41	7.02	1971	33.5	576235	282	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Bin98;f__;g__;s__
bin_22	88.16	0.11	25262	28	1377413	67	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__HIMB59;f__GCA-002718135;g__;s__
bin_34_sub	48.08	1.81	8655	31.1	553436	84	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__HIMB59;f__HIMB59;g__HIMB59;s__
bin.172	62.64	3.3	72707	32	1069102	64	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MED-G09;f__MED-G09;g__;s__
bin.66_sub	18.61	1.74	7295	33.4	382747	70	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MED-G09;f__MED-G09;g__;s__
bin.63_sub	12.5	0	10170	31.8	358071	46	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MED-G09;f__MED-G09;g__AG-430-B22;s__AG-430-B22 sp002728335
bin.164_sub	24.14	5.17	7949	54.5	1467393	212	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MPNO01;f__UBA830;g__UBA830;s__
2030_maxbin.540_sub	31.49	1.72	1732	60.4	930962	551	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MPNO01;f__UBA830;g__UBA830;s__UBA830 sp002471575
2030_maxbin.078_sub	31.9	5.17	1409	26.5	606010	398	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__;g__;s__

2030_maxbin.698_sub	20.52	6.04	1240	35.5	267930	205	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__;g__;s__
bin_84_sub	16.98	0	6281	31.3	120058	23	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__;g__;s__
2030_maxbin.523	27.03	9.86	1420	27.4	259094	173	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__;s__
2030_maxbin.278_sub	27.44	7.12	2096	28	695853	347	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__;s__
2030_maxbin.542	20.09	9.18	1317	30.2	139243	104	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__;s__
2030_maxbin.577	18.61	7.29	1257	26.1	332616	253	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__;s__
bin_47	30.19	0	30174	29.1	316845	13	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__;s__
bin_74_sub	13.89	0	18041	28.8	156029	12	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__AG-414-E02;s__
bin_35	53.61	0	45840	26.6	576307	15	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__MED-G107;s__
bin_72_sub	8.33	0	34361	29.6	148577	5	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__MED-G40;s__
bin_56	30.87	0	35283	29.7	224498	8	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__MED-G40;s__
bin_92	12.5	0	36715	26.5	114796	5	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__
2030_maxbin.206	18.78	3.77	1664	29.6	385114	227	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__
bin_88	22.96	0	68353	29	120824	4	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__
bin_68	18.87	0	13107	27.2	171734	14	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__
2030_maxbin.695_sub	54.1	7.6	2018	62.7	1342146	709	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__UBA8309;s__UBA8309 sp001627655
bin.90	67.72	8.75	20699	37.1	1754144	152	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__CACKLF01;s__CACKLF01 sp002725175
bin.290_sub	9.48	0	5235	39	508622	111	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__MED-G52;s__
2030_maxbin.209_sub	32.21	6.9	1802	38.4	977490	528	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__MED-G52;s__
2030_maxbin.586_sub	13.43	1.1	1518	46.7	1035351	670	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__;g__;s__
bin.56_sub	63.79	9.78	42920	39.4	1600653	116	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__Casp-alpha2;g__Casp-alpha2;s__
bin.154_sub	73.72	8.68	6951	56.1	2653612	450	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__Casp-alpha2;g__UBA3470;s__
2030_maxbin.687_sub	14.66	0	1533	50.5	1328903	851	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__UBA2165;g__;s__
bin.176_sub	12.02	1.72	2202	38.5	446581	198	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__UBA2165;g__GCA-002721295;s__
bin.2_sub	37.88	6.03	7195	40	2650555	471	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__UBA2165;g__TMED23;s__
2030_maxbin.380	21.38	7.82	1613	31.6	770619	458	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__TMED109_A;f__;g__;s__
bin.257_sub	61.63	8.06	7837	29.8	1801374	320	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__TMED109;f__TMED109;g__TMED109;s__TMED109 sp002720535
bin_12	100	2.87	579560	55.9	2596324	22	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA6615;f__UBA6615;g__UBA8079;s__
bin.82_sub	57.82	9.59	7695	49.8	2807630	407	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA8366;f__UBA8366;g__GCA-002724395;s__
bin.102_sub	35.09	1.75	13807	51.2	2481167	256	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA8366;f__UBA8366;g__UBA6544;s__
bin_8	96.34	6.91	17966	53.7	2998934	272	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA868;s__
bin_27	27.5	0.61	2886	55	849225	313	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA868;s__
bin_16	95.73	6.54	14533	38.5	1935318	206	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA868;s__
2030_maxbin.664_sub	65.87	6.2	2010	45.4	2706778	1420	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Enterobacterales_A;f__Alteromonadaceae;g__Alteromonas;s__Alteromonas macleodii
bin_29	54.81	3.1	5328	31.6	818445	176	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__GCA-002705445;f__GCA-002705445;g__GCA-2705445;s__GCA-2705445 sp902595225
bin_18_sub	54.47	1.02	4549	56.6	1536333	348	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__
2030_maxbin.145_sub	18.53	2.88	2675	43.2	872195	374	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA4421;s__
2030_maxbin.673_sub	60.78	7.43	2228	53.5	1819992	895	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA4421;s__UBA4421 sp002725095
bin.284_sub	8.33	0	6610	48.1	1614715	262	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9659;s__UBA9659 sp002694945
bin.78_sub	20.9	0.86	4905	45	1058506	230	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Porticoccaceae;g__TMED48;s__TMED48 sp002591625
bin.179_sub	31.11	4.39	9175	42.5	1890163	261	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__
2030_maxbin.171_sub	16.74	9.76	1498	38.6	587535	371	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__;g__;s__

2030_maxbin.065_sub	19.59	8.62	2040	34.1	229211	113	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__;g__;s__
bin.84_sub	22.41	0	4356	34.8	193814	49	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__AG-339-G14;g__CACNYO01;s__CACNYO01 sp902596665
2030_maxbin.110_sub	22.18	1.72	1856	36.8	837798	459	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__D2472;g__;s__
bin.226_sub	31.9	6.9	6090	36	790646	170	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__D2472;g__D2472;s__
bin.116	26.36	1.92	4877	35.5	255092	55	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__GCA-2730855;s__
2030_maxbin.074_sub	17.19	0.95	2691	37.5	928383	382	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA11654;f__;g__;s__
2030_maxbin.169_sub	11.15	3.1	2210	38.6	337892	162	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__UBA11654;s__
2030_maxbin.118_sub	6.12	0.76	1732	38.5	302218	182	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__UBA11654;s__
bin.41_sub	34.09	6.03	4763	38.9	1097738	269	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__UBA11654;s__
bin_19	87.81	4.2	10084	34.9	1446198	213	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA4486;f__UBA4486;g__UBA11869;s__UBA11869 sp002698845
bin.95	80.24	2.34	111838	33.5	1506734	44	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA6522;f__UBA6522;g__TMED78;s__
bin.180_sub	6.9	0	5116	36.8	263795	58	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Woeseiales;f__Woeseiaceae;g__SP4260;s__SP4260 sp002728725
bin.113	65.52	1.72	18831	35.5	1772943	202	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Woeseiales;f__Woeseiaceae;g__SP4260;s__SP4260 sp003331585
bin.5	75.44	6.04	22258	31.4	2010502	144	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__
bin.211	72.59	0.84	9694	39.3	1953061	264	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002704555
2030_maxbin.596_sub	12.86	1.72	1733	44.6	1215178	715	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__JCVI-SCAAA005;s__JCVI-SCAAA005 sp000224765
2030_maxbin.086	51.44	6.96	1809	47.5	1676555	964	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA1014;s__UBA1014 sp001469005
2030_maxbin.221_sub	70.91	6.1	4618	52.5	2885470	875	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__MB11C04;g__MB11C04;s__
2030_maxbin.668_sub	43.35	2.59	1818	50.4	1793374	1011	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__Opitutaceae;g__UBA5691;s__
2030_maxbin.403_sub	31.19	3.45	2563	57.2	1258535	562	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__UBA1096 sp002299785
2030_maxbin.267_sub	18.45	3.45	2352	60.7	1110611	533	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__UBA1096 sp002299785
bin.165	95.5	6.32	30404	41.8	3423441	203	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__UBA1096 sp002684915
bin_5_sub	76.82	4.26	6154	46.8	3307752	657	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__UBA1100;s__UBA1100 sp002719595
2030_maxbin.135	95.24	8.17	10423	54.9	3883906	637	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__Roseibacillus_B;s__Roseibacillus_B sp002380675
bin_3	95.58	3.37	11219	51.1	4316877	573	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__DEV007;g__CACIUL01;s__
bin_11_sub	87.8	2.72	28129	38.8	2713130	122	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__DEV007;g__EC70;s__EC70 sp902510815
2030_maxbin.090_sub	14.35	3.51	2583	29.9	203808	90	d__Bacteria;p__WOR-3;c__32-111;o__;f__;g__;s__
bin.161	70.02	30.39	23271	48.5	11711430	720	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Bathycoccus_Bathycoccus prasinos
2030_maxbin.053_sub	14.38	2.34	26306	45.5	1946351	191	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Bathycoccus_Bathycoccus prasinos
2030_maxbin.069_sub	15.62	4.67	10898	38.2	831335	212	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Bathycoccus_Bathycoccus prasinos
2030_maxbin.052_sub	8.26	0	3189	50.1	512394	177	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Bathycoccus_Bathycoccus prasinos
bin.256	70.02	46.02	21228	61.1	11213469	761	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Ostreococcus_Ostreococcus sp. RCC789
2030_maxbin.077_sub	0.16	0	2900	61.1	83592	31	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Ostreococcus_Ostreococcus sp. RCC789
bin.110	4.69	0.52	43397	54.6	516233	21	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Ostreococcus_Ostreococcus sp. RCC789
bin.158_sub	2.23	0	28731	44.8	107129	6	Eukaryota_Chlorophyta_Mamiellophyceae_Mamiellales_Ostreococcus_Ostreococcus sp. RCC789
2030_maxbin.310	4.03	0	2025	54.7	829187	418	Eukaryota_Chlorophyta_Mamiellophyceae_nan_nan_nan
bin_2_sub	44.48	9.92	3102	67.4	10477949	3561	Eukaryota_Chlorophyta_no support_no support_no support_no support
bin.38_sub	54.06	31.21	2841	71.1	9899905	3639	Eukaryota_Chlorophyta_no support_no support_no support_no support
bin_32_sub	0	0	4536	62.8	120519	32	Eukaryota_Chlorophyta_no support_no support_no support_no support
2030_maxbin.144_sub	7.94	0	1238	74.3	561554	455	Eukaryota_Chlorophyta_no support_no support_no support_no support
bin_1_sub	7.88	2.16	3192	63.7	2344660	780	Eukaryota_Chlorophyta_no support_no support_no support_no support
2030_maxbin.132_sub	5.42	0	1277	71.1	540568	424	Eukaryota_Chlorophyta_no support_no support_no support_no support

2030_maxbin.098_sub	2.72	0	1279	69.1	167553	131	Eukaryota_Chlorophyta_no support_no support_no support_no support
2030_maxbin.141_sub	1.18	0	1362	67.3	231000	166	Eukaryota_Chlorophyta_no support_no support_no support_no support
2030_maxbin.150_sub	3.04	0	1488	63.1	243380	142	Eukaryota_Chlorophyta_no support_no support_no support_no support
2030_maxbin.473_sub	2.34	0	1475	61.5	1173918	756	Eukaryota_Haptophyta_nan_nan_nan_nan
2030_maxbin.457_sub	1.39	0	1586	60	1040788	643	Eukaryota_Haptophyta_nan_nan_nan_nan
2030_maxbin.448_sub	1.02	0	1350	66.9	1164514	824	Eukaryota_Haptophyta_nan_nan_nan_nan
bin.237	53.71	20.32	2194	70.6	20954132	9278	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.709_sub	0.52	0.52	1300	64.7	1738130	1288	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.260_sub	2.64	0	1222	71.1	1306724	1066	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.390_sub	2.79	0	1437	69.1	884232	604	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.708_sub	6.19	0	1311	66.3	1669171	1247	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.295_sub	5.33	0	1198	74.8	1972744	1642	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.253_sub	0.86	0	1232	72	1270174	1038	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.282_sub	6.15	0	1220	72.7	2027515	1676	Eukaryota_Haptophyta_nan_no support_no support_no support
2030_maxbin.546_sub	0.93	0	1492	62.4	1489010	991	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_no support
2030_maxbin.399_sub	4.59	0.47	1615	67.3	2023171	1258	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_no support
2030_maxbin.316_sub	3.74	0	1963	63.2	1658807	888	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_Phaeocystis antarctica
2030_maxbin.273_sub	2.8	0	1936	64.9	462067	245	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_Phaeocystis antarctica
2030_maxbin.286_sub	1.39	1.39	1836	67.3	1448537	811	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_Phaeocystis antarctica
bin.153_sub	27.28	4.05	3044	67.2	11933531	4188	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_Phaeocystis antarctica
2030_maxbin.320_sub	2.16	0	1722	65.6	2256248	1334	Eukaryota_Haptophyta_nan_Phaeocystales_Phaeocystis_Phaeocystis antarctica
bin.252_sub	2.01	0	6755	65.1	365889	73	Eukaryota_Haptophyta_no support_no support_no support_no support
2030_maxbin.246_sub	8.79	2.34	1305	67.8	1132864	837	Eukaryota_Haptophyta_no support_no support_no support_no support
2030_maxbin.219_sub	4.49	0	6804	67.3	1541207	393	Eukaryota_Haptophyta_no support_no support_no support_no support
2030_maxbin.248_sub	2.34	0	1242	69.9	1172224	949	Eukaryota_Haptophyta_no support_no support_no support_no support
2030_maxbin.281_sub	0	0	1274	68.6	1142272	850	Eukaryota_nan_Pelagophyceae_no support_no support_no support
2030_maxbin.404_sub	6.15	1.25	1222	68.8	1207928	958	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Aureococcus_Aureococcus anophagefferens
2030_maxbin.349_sub	4.75	0	1204	74.5	1704398	1412	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Aureococcus_Aureococcus anophagefferens
2030_maxbin.342_sub	5.2	0	1233	72.4	1523971	1234	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Aureococcus_Aureococcus anophagefferens
2030_maxbin.419_sub	8.26	0.31	1268	70.3	1306669	1006	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Aureococcus_Aureococcus anophagefferens
2030_maxbin.347_sub	0	0	1216	71.7	1301013	1060	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Aureococcus_Aureococcus anophagefferens
bin.174_sub	5.33	0	1797	70.1	539408	286	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Aureococcus_Aureococcus anophagefferens
2030_maxbin.183	6.42	0.65	7394	50	1579465	385	Eukaryota_nan_Pelagophyceae_Pelagomonadales_no support_no support
bin.122_sub	42.33	34.71	2093	67.4	19231311	8926	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.703	4.39	0.63	1256	56.9	817258	632	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.081_sub	2.73	0	2298	61.5	785152	367	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.153_sub	8.09	0.86	1852	63.3	1294215	725	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.200_sub	7.3	2.8	1790	59.1	1001913	565	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.648_sub	4.02	0	1210	59	611291	483	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.460_sub	3.56	0	1193	66.5	1428783	1167	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.392_sub	3.97	0	1201	64.3	875983	700	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.028_sub	2.34	0	2035	65.2	462140	231	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata

2030_maxbin.223_sub	6.23	0	2126	64.2	870383	395	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.534_sub	2.88	0	1205	62.4	422184	341	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.647_sub	2.3	1.72	1178	65.5	765985	632	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.016_sub	4.17	0	3422	70.7	824144	281	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.051_sub	6.03	0.25	4754	50.3	1193233	341	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.369_sub	0	0	1245	63.9	235588	184	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.563_sub	0	0	1211	64.5	1149136	909	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.379_sub	4.89	0	1169	67.9	1649478	1387	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.714	4.82	1.87	1187	59.3	1039877	827	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.354_sub	2.18	0.93	1161	69.5	1575514	1331	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.690_sub	0	0	1249	60	1080898	834	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
bin.195_sub	6.9	0	3090	68.8	1001347	349	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.712_sub	5.8	0	1281	55.4	915819	682	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.578_sub	0	0	1221	63	1091254	854	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.589_sub	4.17	8.33	1150	61.5	869112	728	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.700	1.72	0	1163	58.2	172433	143	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.436_sub	1.88	0	1214	65.5	1122517	908	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.376_sub	3.45	0	1302	65.1	844207	609	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.176_sub	9.28	0	1388	67.3	1299303	879	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.722	0	0	1213	57.4	268227	211	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.658	4.17	4.17	1253	60.1	575148	440	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.724	0	0	1225	57	276822	216	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.666_sub	0	0	1216	60.8	505235	401	Eukaryota_nan_Pelagophyceae_Pelagomonadales_Pelagomonas_Pelagomonas calceolata
2030_maxbin.697_sub	2.17	0	1327	63.7	1522613	1102	Eukaryota_no support_no support_no support_no support_no support
bin.112_sub	42.97	15.5	2572	67.1	47457619	18736	Eukaryota_no support_no support_no support_no support_no support
2030_maxbin.611_sub	7.24	0.93	1222	68.2	1256272	1001	Eukaryota_no support_no support_no support_no support_no support
bin.34	15.37	0.49	1931	68.2	2878164	1448	Eukaryota_no support_no support_no support_no support_no support
2030_maxbin.378_sub	4.64	1.29	1429	64.9	1282339	831	Eukaryota_no support_no support_no support_no support_no support
2030_maxbin.151_sub	3.58	0	1304	65.9	326656	244	Eukaryota_no support_no support_no support_no support_no support
2030_maxbin.531_sub	2.05	0	1479	64.2	427192	276	Eukaryota_no support_no support_no support_no support_no support
2030_maxbin.654_sub	8.23	0.22	1244	67.1	1294099	1028	Eukaryota_no support_no support_no support_no support_no support
2030_maxbin.623_sub	3.56	0	1186	70.9	1652044	1380	Eukaryota_no support_no support_no support_no support_no support

Table 2.4 Best blast hits to Prasinovirus_P2_DCM_2019

Description	Scientific name	Max Score	Total Score	Query Cover	E value	Percent ID	Accesssion Length	Accession Number
Prasinovirus sp. Ct9H_90m.V1 DNA, partial genome sequence	Prasinovirus sp.	3112	13354	0.38	0	0.7635	108466	LC629605.1
Bathycoccus sp. RCC716 virus 2 isolate BII-V2, sequence	Bathycoccus sp. RCC716 virus 2	2985	12699	0.33	0	0.7595	207870	MK522038.1
Bathycoccus sp. RCC716 virus 3 isolate BII-V3, complete genome	Bathycoccus sp. RCC716 virus 3	2924	12344	0.32	0	0.7579	211597	MK522039.1
Bathycoccus sp. RCC716 virus 1 isolate BII-V1 clone scaffold 1, sequence	Bathycoccus sp. RCC716 virus 1	2560	10494	0.28	0	0.748	121467	MK522034.1
Bathycoccus sp. RCC1105 virus BpV1, complete genome	Bathycoccus sp. RCC1105 virus BpV1	2326	7138	0.24	0	0.8108	198519	NC_014765.1
Bathycoccus sp. RCC1105 virus BpV2, complete genome	Bathycoccus sp. RCC1105 virus BpV2	2281	7001	0.26	0	0.7396	187069	HM004430.1

Chapter 3.

THE GENOMIC POTENTIAL AND VIRAL INTERACTIONS OF A PROKARYOTIC COMMUNITY WITHIN THE SECONDARY CHLOROPHYLL MAXIMUM OF AN OXYGEN DEFICIENT ZONE

Rathwell, C, Rocap, G

3.1 Abstract

Marine oxygen deficient zones (ODZs) are important regions for the loss of bioavailable nitrogen from the oceans and house largely uncultivated microbial and viral populations. Here we paired high fidelity long read sequencing and Hi-C sequencing technology with a traditional metagenomic approach to yield 268 medium quality and 106 high quality metagenome assembled genomes (MAGs) from a single depth within the secondary chlorophyll maximum. The genomic potential of this community suggests a modular denitrification approach, with enzymes responsible for the respiration of oxidated nitrogen species within populations from almost all phyla and no single MAG containing enzymes for complete denitrification. Many MAGs also had genes capable of the demethylation of trimethylamine, with two MAGs having the potential to create the methane precursor Coenzyme M, one of which was an autotrophic desulfobacterium. Furthermore, 861 virus contigs greater than 10 kb were identified and 75 of these had a significant Hi-C link to a classified MAG, with 19 having virus per host cell counts of 1 or greater, indicative of active viral reproduction in cells. The terL phylogeny of these linked viruses suggests that a large proportion of active phage in this sample were related most closely to sequences obtained from cultured prophage. Inducible temperate lifestyles may be an important feature of ODZ viral communities.

This study shows the utility of a multifaceted metagenomics approach, by generating a more complete view of the microbial community and allowing a better understanding of the activity of viruses found within.

3.2 Introduction

Oxygen deficient zones (ODZs) are regions in the open ocean where large swaths of the water column are permanently anoxic[195]. These ODZs occur in the Eastern Tropical North Pacific (ETNP), the Eastern Tropical South Pacific (ETSP), and the Arabian Sea. Though they comprise a small fraction of the total volume of the marine environment, these regions house microbial communities potentially responsible approximately one third of the total loss of bioavailable nitrogen from the world's oceans[195–197]. Models and recent measurements suggest that as temperatures rise in the ocean, these regions are expanding vertically and horizontally, making them critical sites for understanding one of the impacts of climate change[198,199].

In all ODZs there is a reduction in bioavailable nitrogen by two microbially driven processes, denitrification and anaerobic ammonium oxidation (anammox)[5]. The process of reducing oxidized nitrogen species to nitrogen gas can be modular, with many microbial populations possessing incomplete denitrification pathways[200–206]. Incomplete denitrification can produce N_2O, one of the largest contributors to ozone depletion and non-CO_2 greenhouse gas[207]. Unlike denitrification, every step in the anammox pathway is present in the same organism[208], all of which are members of the genus *Candidatus Scalindua* from the *Brocadiales* order of the *Planctomycetes* phylum[209]. The rates of denitrification and anammox have been measured with differing results, suggesting that each process can be a major contributor to bioavailable nitrogen removal given the right environmental context[197,209–212].

Methane and sulfur cycles are also impacted by microbial metabolisms within anoxic water of the ODZs[213,214]. Though methanogenic archaea are largely studied in anoxic marine sediments[215], the largest pool of the open ocean methane can be found in the Eastern Tropical North Pacific[216]. Though the origin and fate of methane in the ETNP is not clearly constrained, some evidence suggests methane generated in sediments can travel off shelf within oxygen deficient water[108,216–218], and methane concentrations as high as105 nM have been detected within the ETNP far from the shelf[214]. Anaerobic methane oxidation is suggested to be an important sink of methane within the ODZ[217]. Though ODZ waters are relatively sulfide-poor compared to the active sulfur cycling sites of sediments[219], there is some evidence of a cryptic sulfur cycle within the water column[202,220], and that sulfur cycling in ODZs may even increase carbon sequestration[221].

Key components in the microbially driven biogeochemical cycling in ODZ waters are the viruses and their interactions with microbial populations. Though molecular methods, including single gene approaches and metagenomic sequencing have identified genes for major elemental cycling pathways and some of their more dominant microbial sources, nearly all the microbial populations in the ODZ remain uncultivated[110,208,222], greatly inhibiting the study of potential virus-host interactions in the ODZ water column. Within microbial communities, viruses affect the flow of nutrients[19,20,88,89]. Viral infection can also lead to an increase in the biological pump when infection and lysis cause particle aggregation and subsequent sinking in what is called the viral shuttle[36,37,90]. In an ODZ, where heterotrophic denitrification is dependent on sources of organic matter, viruses have great potential to impact microbial processes[223]. Furthermore, viruses are also vehicles for the exchange of genetic material, which is critical for enhancing diversity in asexually reproducing host populations[92,93] and potentially for providing adaptive genes to unique niches, such as in a stratified redox environment. Lastly, during infection the microbial cell begins to

function in a completely altered way, expressing both viral and microbial genes prior to lysis, suggesting that the virally infected cell (virocell) itself is a unique participant in biogeochemical cycling[53,94,95]. Because viral activity is predicated on the presence of a suitable host, the ultimate impact of each virus can only be evaluated appropriately once its host range has been identified.

In marine ODZs viral genomes have been described through sequencing collected metaviromes. Viruses in ODZs are distinct from viral populations elsewhere in the global ocean, suggesting specific adaptations to the microbial populations present[54,55,224]. Viral metagenomes across depths and stations in the ETSP showed that variation in oxygen is the most significant driver of changes in the viral community[54] and that some viral populations carry important genes for denitrification as auxiliary metabolic genes (AMGs), suggesting unique adaptations to anoxic microbial communities[51]. Viral metagenomics in the ETNP did not reveal denitrification pathways in viral sequences, but rather AMGs for stress response, sporulation and motility at anoxic depths[55]. Additionally, the frequency of infected cells as determined through microscopy showed a significant rate of cells experiencing infection in a permanently anoxic water column[55]. Though viral sequences can be gathered in bulk and analyzed, assigning a host in the ODZ where the majority of hosts are uncultivated is much more difficult. In the ETSP study, no attempt was made to assign hosts to the viral sequences gathered[55] and, in the ETNP study, hosts were estimated for no more than 1% of the viral communities analyzed[54]. Furthermore, predictions based on sequence similarities do little to describe activity of viruses *in situ*. Importantly, research that can describe activity of environmental viruses, such as single-cell amplified genomes (SAGs)[134] and Hi-C sequencing of environmental samples, have shown either unexpected interactions or interactions that span taxa and even kingdoms[72,83,85,225]. These results indicate that viruses in environmental

samples such as the ODZ may be interacting with a broader range of microbial cells than can be predicted from cultures and sequences alone.

Hi-C involves wet-lab techniques that mark DNA that is physically proximal to other DNA, tagging it for sequencing in order to enable clustering of assembled sequences based on those tags. This method was first utilized as a tool to examine the three-dimensional structure of DNA within a clonal population of cells[68]. However, the ability to understand which sequences are physically proximal was soon realized to be an important asset in understanding short-read sequencing of environmental microbial populations[71,113,114]. This method has expanded to describe virus-host interactions within mixed population samples from gut microbiomes[72], wastewater[73], groundwater[226], soil[83], and estuaries[82]. Historical culturing shows that lysis of clonal cultures by cultivated viruses is very host-specific but measurable interactions in sampled environments may be much more wide-ranging taxonomically[74]. Hi-C proximity links that can demonstrate virus-host interactions *in situ* allow us to pair viruses with the microbial populations they are actively infecting, as well as any populations where viral genetic material may have entered the hosts in the environment.

In this study, we paired the Hi-C technique with short and long-read sequencing libraries to describe the microbes and their viruses from the secondary chlorophyll maximum within the anoxic water column of the ETNP. Hybrid assemblies and Hi-C sequencing are used to generate 268 medium-to-high quality MAGs, which allowed the assignment of modular denitrification roles to different microbial communities and the detection of a novel methanogenic sulfur-reducing Desulfobacteriota. Furthermore, 75 viral contigs present in the metagenome were assigned to hosts across a range of microbial taxa based on Hi-C links. Most virus-microbe interactions occurred with less than 1 virus per host cell in the population, although 19 appeared to be experiencing lytic

infection as determined by Hi-C. Hi-C and hybrid sequencing techniques combined allowed for a more thorough examination of microbial genomic potential and virus activity at depth in the secondary chlorophyll maximum.

3.3 Materials and Methods

3.3.1 Sample Collection

Seawater samples were taken at an offshore site within the ETNP (P2: 16.58°N, 107.05°W) from the research vessel *Kilo Moana*, during cruise KM1920 (cast 33, October 6, 2019, local time of 08:15). The water properties were determined using a CTD-Rosette equipped with conductivity, temperature, pressure, transmissivity, chlorophyll fluorescence and oxygen (SBE 043) sensors. Duplicate 20-liter water samples were collected in Niskin bottles from the same cast at a depth targeting the peak of the secondary chlorophyll maximum (130 meters). Twenty-liter samples were filtered on 142 mm sequential 3.0 μm and 0.2 μm polycarbonate filters using in-line filtration rigs until volume was consumed. The 0.2–3.0 μm size fraction, presumed to contain the free-living microbial community, was processed further for microbial community analysis.

3.3.2 Filter Processing Onboard

One filter was used for metagenome assembly and one filter was used for Hi-C library generation. The metagenome filter was frozen immediately and maintained at –80°C. The second filter was treated with formaldehyde to generate Hi-C links. Soaking solutions of 1% formaldehyde and 10% glycine were created by mixing chemicals with seawater from the same depth and filtering through an Anodisc syringe-tip filter with a pore size of 0.02 μm. Filters were folded in half 8 times with the sample-side facing in, creating a cone, and placed tip-first in 15

mL Falcon tubes containing 13.5 mL of the <0.02 μm 1% formaldehyde 99% seawater solution and incubated for 30 minutes. After 30 minutes a serological pipette between the filter and the side of the Falcon tube was used to deliver 1.5 mL of the 10% <0.02 μm glycine seawater solution. The filter was then incubated for 30 minutes. A fresh serological pipette was used to remove all volume from the Falcon tube and the filters were frozen and maintained at –80°C until further processing.

3.3.3 Metagenome Libraries:

The metagenome filter (not treated with formaldehyde) was cut in half and used to extract DNA for both short- and long-read metagenomic libraries. For the short-read sequencing library the half filter was cut into smaller pieces that were placed in 1.5 mL 1X TE and experienced a freeze-thaw at extreme temperatures of 95°C and –80°C. DNA was then extracted using QuickGene Mini80 with the DNA Tissue kits (Autogen, Cat No. DTS) and concentrations assessed using a Qubit (dsDNA HS assay kit, Invitrogen Q3285). DNA was sent to Northwest Genomics Center for NovaSeq sequencing on S4 flow cell with a 2 x 150 basepair reads kit.

The same extraction protocol was modified to generate high molecular weight genomic DNA for HiFi sequencing from the other half of the filter. Samples were mixed by swirling or inversions (a vortex was not used). Freeze-thaw cycles were limited to the first thaw from storage, and wide-bore pipette tips were used for each step. This high molecular weight library was sent to University of Washington's PacBio for HiFi SMRT read sequencing.

3.3.4 Hi-C Libraries

The formaldehyde-treated filter was thawed, and organic material resuspended and pelleted from the filter. The 142 mm filter was cut into 16 roughly equivalent pieces and placed in a 50 mL Falcon tube containing 20 mL <0.02 μm filtered sterile base for artificial seawater

(pre-ASW)[227]. Filters were rotated in the dark at 16°C for 36 hours. The soaking pre-ASW was then removed and distributed among sterile 2.0 mL centrifuge tubes for pelleting, while a second wash was initiated. For both first and second filter wash volumes, tubes were spun at 21,000 x g for 1.5 hours at 4°C. All but 50–100 uL of supernatant was removed. The remaining volume was used to resuspend each pellet and then combine them all into a single Eppendorf tube. A repeat spin cycle of 21,000 g at 4°C was performed for 1.5 hours, and the supernatant was removed. Removed supernatant from each spin was stored in the dark at 4°C for an additional week and pelleted with the same spin protocol as the first pellets, which were used for library generation when first pellets did not produce high yields. All supernatant was centrifuged in the spin protocol described above and generated sufficient library for Hi-C sequencing.

3.3.5 Assembly

Short-read shotgun sequences generated with NovaSeq chemistry were cleaned using Trimmomatic(v0.39)[139], with the following specifications: `ILLUMINACLIP:bin/Trimmomatic-0.39/adapters/TruSeq3-PE-2.fa:2:30:10:1:true`. The high-fidelity long reads were paired with the short-read sequences to generate a hybrid assembly. Specifically, NovaSeq forward and reverse reads were used as inputs with Spades (v3.15.2)[228] with `–meta` and `–only-assembler` parameters. PacBio HiFi reads were added using the `–pacbio` command to generate hybridized assemblies. Library quality was assessed for using QUAST[141] (v5.2.0) with default parameters. All sequences were named and filtered for a minimum length of 1000 bases.

3.3.6 Identifying Viral Contigs

Virus contigs were the hybrid library using VirSorter2 (v2.2.4)[31], VIBRANT (v1.2.1)[169] and geNomad (v1.7.4, database 1.7)[170]. VirSorter2 was trained to include *Autolykiviridae* as described on https://github.com/jiarong/VirSorter2, then run with the following parameters: `--min-length 1000 --include-groups 'dsDNAphage, ssDNA, NCLDV, lavidaviridae,`

`autolykiviridae'`. VIBRANT was run using default parameters and a minimum length of 1000 bp. Analysis in geNomad used the `end-to-end` pipeline with default settings. Any contig that was identified as viral by two of the three of these software programs, was considered further, thus utilizing specific strengths of each of these algorithms and simultaneously avoiding false positives. Each viral contig above 10,000 bases was considered a potential virus operational taxonomic unity (vOTU) and assessed further for quality.

3.3.7 Viral Quality Check and Annotation

Quality of vOTUs was assessed with CheckV (v1.0.1, database v1.5)[154] using the `end_to_end` pipeline. Output from geNomad was used to determine viral cluster taxonomy. Any virus with the checkV warning of "no viral genes detected" was excluded from our final analysis. AMGs were found using DRAM-v (v1.5.0), using kofam databases and skipping UniRef. All databases were setup and downloaded using `DRAM-setup.py` script on February 21, 2024. AMG summary files, generated by DRAM-v only included AMGs with a score of 1-3 and a flag of M (microbial metabolic gene) and an absence of flag B (three microbial metabolic genes in a row). Flags of K (known AMG), E (experimentally verified AMG), and F (AMG within 5 kb of end of contig) were permitted. AMGs were also identified via VIBRANT viral analysis described above. AMG summary files from VIBRANT were combined with DRAMv summary files, and an AMG found by either method was considered valid. Genes determined to be viral AMGs were filtered manually to exclude methyltransferases, peptidases and other enzymes that likely have a primarily viral function rather than supplementing host metabolism. Viral AMGs each had a KEGG ID, and many KEGG IDs have multiple possible pathways. Pathway names that made sense, given the sample, were selected; for example, when possible pathways included human disease, an alternate pathway name was chosen.

3.3.8 Metagenome Assembled Genomes

A combinatorial approach to generating MAGs was used to maximize the benefits of both proximity-linked bins and computationally derived bins. Short reads were mapped to the assembly using Bowtie2(v.2.3.5.1)[143] with default settings, and coverage files were generated with the MetaBat2 script `jgi_summarize_bam_contig_depths`. MetaBat2 (v2.15) and MaxBin2 (v2.2.7) were each used to create linkage-independent bins using default methods. ProxiMeta (July 27, 2023) was used to generate linkage-dependent bins, serviced by Phase Genomics (Seattle, WA). Bins produced by these three methods were then used as input to DAS Tool[173] (v1.1.7). Bins less than 10% contaminated and classified as bacterial or archaeal were used to assess virus-host pairs.

3.3.9 Microbial MAG Taxonomy and Annotation

Completeness and contamination of each MAG was assessed using checkM (v1.1.3)[148]. Taxonomy of each bin was determined using GTDB-tk (v2.3.2, database r214)[174]. Medium-to-high quality MAGs were annotated using the DRAM annotate pipeline to annotate all binned contigs 1500 bp or greater.

3.3.10 Computational Host Assignments

Host predictions were assigned independently of Hi-C links using iPHoP (v1.3.3)[59], which is a machine-learning algorithm that iteratively picks the best host from the results of six other host prediction algorithms and optimizes host identities for each virus[62,229–232]. Specifically, the tools include CRISPR sequence analyses, genome metric similarities and tRNA sequence similarity to determine likely host assignments. The latest database was downloaded on April 13, 2024.

3.3.11 Proximity-linked interaction assignments

Hi-C reads were cleaned using BBDuk from the BBTools package (guide and downloads available at https://jgi.doe.gov/data-and-tools/software-tools/bbtools). First, a file of known Illumina adapter sequences and the following settings, `ktrim=r k=23 mink=11 hdist=1 minlen=50 tpe tbo`, were used to trim adapter sequences. Next, reads were quality-trimmed with the command `trimq=10 qtrim=r frm=5 minlen=50`. Finally, the first 10 basepairs were trimmed using the option `ftl=10`[10]. These cleaned Hi-C paired end reads were mapped to assembled sequences using BWA-MEM (v0.7.17-r1188)[175] with the `-5SP` setting. In order to remove non-informative reads such as duplicates or unpaired reads, Hi-C reads were analyzed with SAMBLASTER (v0.1.26)[176]. SAMtools (v1.3.1)[177] was used to convert the sam file to a bam file using the `-F 0x904 -bS` settings, and subsequently to sort by name. Cleaned Hi-C reads were used to link contigs found in vOTUs with contigs found in microbial MAGs generated in the steps outlined above. Resultant link tables included the following columns: contig_a, contig_b, and number of links. Links were then filtered by several steps. First, virus per cell and phage connectivity were calculated as previously described[178]. Specifically, the average virus copy per cell in interacting "host" population (VPH) was calculated using Hi-C links between the virus contig and a specific population (L), all links between that virus and all possible host populations (L_v), host MAG abundance (H), and virus abundance (V) in Equation 1.

$$VPH = \frac{V}{H}\frac{L}{\sum L_v} \quad (1)$$

Next, density of virus-microbe interactions per kb^2 of sequence (D_{VH}), density of within-host MAG interactions per kb^2 of sequence (D_H), and VPH results from Eq. 1 were used to normalize phage host interactions and calculate phage-host connectivity (R'; Eq. 2).

$$R' = \frac{D_{VH}}{D_H}\frac{H\sum L_v}{VL} \quad (2)$$

These values were used to determine valid virus-microbe linkages within the sample in several rounds of filtration. Only virus-host interactions with at least two links were included in order to exclude false positives. In addition, only R' values greater than 0.1 were included. A minimum VPH of 0.1 was calculated using a receiver operating characteristic curve (ROC)[178], which maximized both the virus-host links included and the virus-host links lost.

3.3.12 TerL Phylogeny of Linked Viruses

Within all viral sequences that linked to classified microbial bins and passing filtration metrics, annotations (see Methods 3.3.7) were used to search for the phylogenetic marker gene, large terminase (*terL*). Gene calls made by Prodigal within VIBRANT (see above), and annotations made by DRAMv and VIBRANT (see above) were used to isolate *terL* amino acid sequences. Those amino acid sequences were aligned with AliView[233]. Trees were made using RaxML-ng, with automated bootstrapping (`--bootstrap`) and the `LG+G` model.

3.4 Results

The pelagic station of the ODZ in the ETNP had a steep oxycline between 75 meters and 125 meters (Figure 1). The primary chlorophyll maximum sat at the top of the oxycline and spanned a depth of roughly 65 m to 85 m. There was also a secondary chlorophyll maximum between 125 m and 150 m where oxygen was undetectable.

Seawater collected within the secondary chlorophyll maximum at 130 meters was processed with both high fidelity (HiFi) long-read sequencing from PacBio and short-read sequencing on the Illumina platform (Table 1). Over 309 million short reads were paired with over 100 thousand long reads with an N50 of almost 10,000 to create a hybrid assembly with a total length of 2 billion bases and an N50 of 4,375. There was a total of 662,351 contigs of at

least 1000 bases in this hybrid assembly, and the longest contig assembled was 755,536 basepairs.

Contigs were binned with various methods, both dependent upon and independent from Hi-C read alignment (Figure 2). MetaBat2 and MaxBin2 binned 207,734 and 382,422 contigs into 72 bins and 561 bins, respectively (Table 2). ProxiMeta, utilized more than 85 million Hi-C read pairs to generate 180 Hi-C linked bins from 23,590 contigs. All three of these binning results were used as input for DASTool, which combined them and limited contamination of single copy core genes, to create a final bin set of 819 bins from 361,952 contigs. Of those final bins 268 were determined to be medium quality (MQ) bins, with a single copy core gene completion score of 50% or higher and a contamination of single copy core genes of less than 10%. Of these, 106 of those were 90% complete or higher (Figure 3, Table 4).

Analysis of these medium-to-high quality MAGs shows a broad range of taxonomy present. These near-complete metagenomes spanned prokaryotic diversity with both archaeal and bacterial MAGs from 27 different phyla (Figure 4, Table 4). Bins from several archaeal phyla including Thaumarchaeota's Marine group II (named GTDB phylum is Thermoplasmatota, class is Poseidoniia) are represented. There are MAGs across Alphaproteobacteria and Gammaproteobacteria classes, as well as the Flavobacteria, Chlorobi and Bacteroidetes (FCB) and the Planctomycetes, Verrucomicrobia and Chlamydia (PVC) supergroups. Many key microbial members of this community were represented in these MAGs, including *Nitrospina*, Chloroflexota, *Desulfobacteria*, SAR324, *Pelagibacter* (SAR11), *Prochlorococcus* and Marinisomatota. Some phyla had many less abundant MAGs, such as Chloroflexota with 73 MAGs with some of the lowest coverage measurements of any of the high-quality MAGs. Conversely, there were only three *Prochlorococcus* MAGs, though those MAGs had high

coverage in our short-read library. Coverage could also vary greatly between MAGs from a specific phylum as well, such as with Marinisomatota (also known as Marinimicrobia), which had the most abundant MAGs but also two MAGs with some of the lowest abundance measurements.

Each medium- and high-quality MAG was assessed for presence of the genomic potential for denitrification. As there were no anammox bins in our medium-to-high quality MAGs, none of the MAGs in this set contained enzymes for the anammox pathway. Among the various combinations of denitrification genes present, MAGs that carried no genomic potential for denitrification (Figures 4 and 5) were the most numerous. The next largest group of MAGs contained just the *napAB* or *narGH* genes for the reduction of nitrate in anaerobic respiration. Many phyla (15) were capable of this first step of modular denitrification, with most only capable of the first step. The least common in our set of MAGs were genomes with the potential for reducing nitric oxide to nitrous oxide (*norB*), and the second least common were genomes with the gene coding for the reduction of nitrous oxide to nitrogen gas (*nosZ*). The most abundant MAGs from the Actinomycetota and Marinisomatota phyla carried either no denitrification genes or just the first step of nitrate reduction to nitrite. There were 20 MAGs that carried only genomic potential for the reduction of nitrite and another 20 that had genomic potential for the first two steps of denitrification. There were no MAGs that had the genomic potential for complete denitrification. The closest were two MAGs that had 3 of the 4 steps of denitrification. These genomes were some of the least abundant and had large genomes of 7.5 to 8.7 million bases. One was a Latescibacterota with *nap*, *nir* and *nos* genes, and the other was a Planctomycetota with *nap*, *nir* and *nor* genes.

Gene annotations performed by DRAM also suggest a broad range of metabolisms present in these populations of the secondary chlorophyll maximum. Genes for dissimilatory sulfate reduction, including *soxYDCYAXBZ*, were found in members of Acidobacterota and those from the FCB super group, specifically Bacteroidota and Gemmatimonadota phyla and one SAR324 MAG (Figure 4). This SAR324 also contained *dsrA* and *phsA* genes suggesting a more complete sulfur reduction pathway. Dissimilatory sulfur reduction genes were also present in MAGs of the Desulfobacterota phylum. However, most sulfur reduction genes were present in Alphaproteobacteria and Gammaproteobacteria MAGs. Methane metabolism genes were also notably present across a wide range of taxa. Many organisms contained the first methyltransferase enzyme in the methanogenic pathway that breaks down trimethylamine (*mttB*) to monomethylamine and Coenzyme M (Figures 4 and 6). The complete breakdown of trimethylamine generates a final methyl group and Coenzyme M and produces ammonia as a biproduct. One Acidobacteria and one Desulfobacteria MAG each carried the complete pathway of demethylated trimethylamine (*mttB*, *mtbB*, *mtmB*).

Viral sequences over 10,000 bases long were identified within the hybrid assembly using multiple algorithms and assessed for genome completeness, abundance and quality. Of the 861 viral contigs fewer than 70 could be classified further than Unclassified Caudoviricetes through geNomad (Figure 7). Most of these sequences contained at least one viral hallmark gene and almost half were considered 50% complete or higher by checkV, though fewer than 10 were considered 100% complete. There was a wide range in viral contig abundance in the sample, with most having an average depth lower than the average MAG (Figures 4 and 7). Within these viral sequences DRAM and VIBRANT annotations were used to determine presence of AMGs. A total of 87 unique KEGG IDs were annotated as AMGs and occurred 257 times in viral

genomes (Figure 8). Within those AMGs, genes for cofactors and vitamins, carbohydrate metabolism, and amino acid metabolism were the most frequent. Genes for the metabolism of sulfur (*cysH*) and methane (*cofF*) were also found. AMGs also included *queE*, *queC* and *queF* for transfer RNA biosynthesis, as well as genes for chemotaxis (*motA*) and flagellar synthesis (*fliC*). No viral AMGs for genes within the denitrification and anammox pathways were annotated.

Hi-C read pairs were aligned to the hybrid assembly, and links between viral contigs and microbial MAGs were screened to determine virus-host interactions. Predicted hosts for viral contigs were also generated computationally and compared with Hi-C linkages (Figure 9). Host predictions were possible with 208 virus contigs or approximately one quarter of all viral contigs present. Significant Hi-C links between viruses and microbial MAGs occurred for 75 of the virus contigs, less than 10% of the total virus contig pool. Notably, most viruses had no host predictions or Hi-C links (Figure 9D). Only 25 virus contigs had both a host prediction and a Hi-C linked MAG. Of the Hi-C linked viruses there was a wide range of phage-microbe connectivity values (R') and virus per cell values (VPH). Virus host predictions, based on a database of GTDB genomes, resulted in predictions of similar taxa to the microbial genomes recovered from this sample, with the exception of 18 virus contigs which were predicted to have genomes from the phylum Bacillota as microbial host.

Of the links that were between a virus with a predicted host, 4 of them were direct matches between predictions and links, 5 were more loosely associated taxonomically, and 16 were interacting with phyla markedly different than the predicted host (Figure 9A). Generally, the lowest ranking interactions between MAGs and viruses were either with unpredicted hosts or for interactions computational methods could not predict. An example of a more loosely

associated taxonomical link was between a virus predicted to infect SAR11 and a Gammaproteobacteria MAG. An unexpected interaction was between a virus predicted to infect Gammaproteobacteria and a MAG from the Nitrospinota phylum. Three viruses predicted to infect Alphaproteobacteria were instead found interacting with Alphaproteobacteria, one of which had a VPH of 3.6. Some virus host predictions were loosely similar to the Hi-C linked interactions including 3 viruses predicted to infect Bathyarchaeota linked with Nitrosphaeria and both archaeal classes of the Thermoproteota phylum. Also, a virus predicted to infect Bacillota was linked by Hi-C with Marinisomatota, which are both phyla from the supergroup FCB. A virus predicted to infect Verrucomicrobiota and was linked by Hi-C to a Planctomycetota, and both bacteria are part of the PVC superphylum. Interactions that more closely aligned with predictions were typically those with higher VPH values. One exception is the linkage between separate virus contigs predicted to infect archaeal phyla and linked to the bacteria phylum J088 as determined by the GTDB taxonomy toolkit.

The strongest Hi-C interactions with the highest VPH values were between viruses and microbial MAGs where the virus had no host prediction at all (Figure 9B). Additionally, Hi-C linked viruses were statistically more abundant than non-linked viruses in the short-read assembly (Figure 9E). The highest VPH values were between Planctomycetes and viruses. There were also Chloroflexota MAGs, and a Poseidoniia MAG that linked with a virus with high VPH values. Furthermore, 183 viruses with predicted hosts had no significant Hi-C links, which was a majority of the virus contigs with predictions. Viruses with VPH values greater than one were assumed to be actively reproducing viruses and were analyzed further for genomic content (Figure 10). There were 19 viruses included that had VPH values greater than one infecting 19 MAGs of only nine phyla. Of these viruses most had no host prediction and two had host

predictions that were aligned with taxonomy of the linked MAG (Figure 10A). These virus contigs ranged in size from 12 kb to almost 60 kb. Analysis of gene content confirmed the presence of many viral hallmark genes, including an abundance of structural and replication genes, as well as other phage protein such as those responsible for lysis and maturation (Figure 10B). Six of the 19 viruses contained the large terminase gene, three had a gene for the capsid protein, two had a portal gene, and four contained a viral polymerase gene.

The phylogenetic marker gene for *terL* was used to create maximum likelihood phylogenetic trees of linked viral genomes when they contained the *terL* gene. A total of six viruses with high VPH and likely undergoing lysis at the time of sampling, as well as five viruses that were linked but had a range of VPH values (0.24–0.98), were placed on a *terL* phylogenetic tree. Nine of these 11 viral *terL* genes placed near a characterized phage, and two could not be classified phylogenetically (Figure 11). Viral genes from a broad range of *terL* diversity were included to capture the unclassified Caudoviricetes viruses in our study. Three viruses in this study were grouped with T1-like viruses, two of which were closely related and each linked with Gammproteobacteria bins. One of the two Gammaproteobacteria-linking T1-like viruses had a VPH just under 1, suggesting perhaps that this virus was also lytic. A Planctomycete-linked virus was grouped with the D3112-like phages. A virus linking with a high VPH to an Alphaproteobacteria was phylogenetically most similar to *Clostridium phage C-st*, and three viruses from this study were most closely related to the Bacteriophage P1. A virus linking with Chloroflexota was most closely related to the *Bacillus* virus SP01. There were no viruses linking with microbial genomes with a *terL* gene present that were linked to either the T4-like or the T7-like families of Caudoviricetes.

3.5 Discussion

This study shows the power of combining long reads, short reads and Hi-C sequencing, resulting in over 250 medium-to-high quality MAGs within a single sample. Multiple efforts in the last decade have collected dozens to thousands of metagenomic samples using short read sequencing in order to bin many medium-to-high quality MAGs[100,102,205,233–236]. Within these seven cited studies, the average number of medium quality bins per sample assembled is 11.6, with only 1.6 high quality MAGs binned per sample collected. This study produced 106 high quality MAGs and 162 medium quality MAGs from a single sample. These numbers represent a larger rate of return by more than an order of magnitude. Increased sequencing depth of short reads, incorporation of HiFi long reads, and the use of Hi-C likely all have played a role in this successful recovery of so many high quality bins. For future oceanographic studies, when a thorough examination of specific locations is the research goal, then extending sample processing to include both high molecular weight and low molecular weight extractions with a subset of sample volume processed for Hi-C processing may be worthwhile.

Due to the strength of this method, we were able to determine the genomic potential of the microbial community at large, specifically within the secondary chlorophyll maximum of the ETNP. Marine ODZs have complex microbial communities that are adapted to anaerobic conditions and responsible for significant loss of marine bioavailable nitrogen[196]. Broadscale metagenomics have shown that denitrification is likely modular in ODZs, with many microbial populations only capable of one or two steps in the process[110,200,204,206]. Here we obtained no evidence for complete denitrification, but MAGs are not closed genomes, therefore we cannot confirm that a population capable of denitrification does not exist. In the Zhang *et al.* ODZ co-assembly and meta-analysis, one MAG from the phylum Alphaproteobacteria capable of complete

denitrification was recovered from within the ODZ core of the ETNP in 2016[206]. The genus determined by GTDB-tk was *GCA-2731375*, with an NCBI taxonomy of Rhodospirillaceae. In this study we recovered a MAG from the same taxonomy that was 97.76 percent complete and over 4.6 Mbp, but it did not contain either *nor* or *nos* genes. Though genes can certainly be omitted methodologically in metagenomic binning, the absence of two denitrification genes within such a complete genome suggests that the more likely explanation may be that varying strains for this species exist and their abundances differ across space or time.

Other differences exist between our analysis, which had high yields of MAGs for a single depth, and the Zhang *et. al.* meta-analysis [206], which co-assembled metagenomes from multiple depths, years and global ODZs to recover MAGs. For example, their MAGs could not confirm previous SAG findings that SAR11 populations in the ODZ have the capacity for heterotrophic respiration of nitrate[110], so they hypothesized that step in the pathway had higher contributions from SAR324 microbial populations than SAR11. In this study, we observed not only SAR324 but also multiple SAR11 bins with genomic potential to reduce NO_3^-. Furthermore, we recovered a Planctomycetes bin with 3 of the 4 necessary steps for denitrification which they were unable to recover with their methods. This Planctomycetes bin was not abundant in our sample and its recovery may be attributed to the disproportionate binning of Planctomycetes genomes in Hi-C data[82], perhaps due to their complex nucleoid cellular compartment[126] which increases genome density and potentially Hi-C linkage data. Significant differences between co-assembled metagenomes across many depths and multiple ODZ regions could also be due to methodological biases or a demarcation between the microbial community situated within the anoxic chlorophyll maximum in 2019 and those of other depths, regions, and years.

The microbial genomes recovered here give an insight to methane and sulfur cycling in the upper portion of marine ODZs. Trimethylamine n-oxide (TMAO) is an abundant molecule in marine environments[238], generated in the bodies of larger eukaryotic organisms and found in high concentrations in surface waters[239], where it has been shown to be an important nitrogen source for heterotrophic bacteria[238,240]. In our sample, just below the oxycline, 81 of the 268 medium-to-high quality MAGs contain the gene *mttB*, which codes for an enzyme used in metabolism of TMA to dimethylamine (DMA; Figure 4). Genes encoding subsequent enzymes to convert DMA to monomethylamine (MMA) (*mtbB*) and the final step which can release ammonia (*mtmB*) were less pronounced throughout the MAGs. Two MAGs had the genomic potential for complete catabolism of TMA, including an Acidobacterota and a Desulfobacterota. Trimethylamine metabolism with methyltransferases is studied predominantly in methylotrophic methanogenesis[241–245], as each of these methyltransferases (*mttB*, *mtbB*, *mtmB*) can generate a methylated Coenzyme M, which can then be converted into methane in archaeal methanogenesis (Figure 6A). The final critical step is catalyzed by the enzyme, McrA, but we found no MAGs with the gene for this enzyme (Figure 4), suggesting that TMA metabolism in the upper ODZ is driven by non-methanogenic organisms. However, the complete demethylation of TMA releases ammonia, which is scarce in ODZs and is an important fuel for anaerobic ammonia oxidation or anammox[209,246]. The transferred methyl group can be used to methylate important cofactors in the bacterial cell, such as tetrahydrofolate[247]. A previous study has shown *mttB* sequences were distributed widely in bacterial phyla from within anoxic sediment, suggesting a benefit beyond methanogenesis[242].

Interestingly, the Desulfobacterota MAG with all three methylamine methyltransferases carried many genomic similarities to a recently described genomic population of a Desulfobacterota found to be a dominant organism within the sediment of chemical dump sites off

the coast of California[247]. The archaeal methyltransferases *mttB* and *mtbB* each carry amber stop codons that require the production of the tRNAs necessary for the rare Pyl and Sec amino acids[248]. The additional genes required for Pyl synthesis were also found in the Desulfobacterota MAG from this study, as well as 7 of the 8 genes from the Wood-Ljungdahl pathway that were also found in the sediment Desulfobacterota. The Wood-Ljungdahl pathway is an ancient anaerobic carbon fixation pathway well-studied[249] in methanogenic archaea that combines two CO_x molecules into Acetyl-CoA [250–252] but can be used in the reverse direction for dissimilatory oxidation of acetate in sulfate reducers[253]. The presence of this unique genomic potential in Desulfobacterota in the upper portion of the water column suggests that conditions in the 0.2–3.0 µm size fraction of an anoxic water column, as well as the sediment, may be suitable for growth. Furthermore, measurements of TMA metabolism in sediments before modern sequencing techniques suggested that 20% of TMA metabolized in sediments could be due to sulfate-reducing bacteria, though the majority was due to methanogens[244]. The genomic potential of Desulfobacterota to metabolize TMA suggests that this organism could compete for TMA or with any potential methyl-based methanogenesis in the anoxic water column[241].

In addition to binned genomes, Hi-C reads in this study enabled us to capture virus-host interactions at the time of sampling, by showing the viral sequences within the genome that were most likely infecting the microbial community. Hi-C sequencing within gut populations has suggested that actively reproducing lytic viruses are targeted with Hi-C analyses[70]. Host predictions with the state-of-the-art computational methods we used were able to generate predictions of which microbes each virus identified in our metagenome could infect. These predictions work on the assumption that viruses with genetic similarity to hosts, such as tetranucleotide frequency, GC content and even tRNA sequence, will be optimized for host cell

machinery for efficient lytic infections[62]. Even the presence of CRISPR spacer sequences within a host genome is used to determine a virus host population, and yet these sequences indicate acquired immunity to a given virus rather than current susceptibility[63,75]. Therefore, though possibly evolutionarily relevant, such host predictions may fall short of describing viral ecology through current interactions with microbial hosts in the ocean.

Hi-C links occurred primarily with viruses that had no host predictions, suggesting the utility of multiple tools to understand viral ecology and impact. Furthermore, there were over a dozen associations that counteracted host prediction techniques, suggesting that either predictions are not very accurate or viruses are interacting with hosts they are not optimized to infect, or perhaps both. Recent Hi-C analyses in environmental samples, as well as previous Hi-C sequencing of gut microbiomes, have shown links between a single virus and microbial populations from multiple phyla[72,82,83,225]. Despite the advancement of statistical analyses and rigorous filtration of spurious links in Hi-C datasets[145,178], environmental viruses continue to defy predictions, when the predictions are based on awareness of viral behavior rooted in lytic viral isolations in laboratory settings rather than the complex communities of which they are a part *in situ*[128]. At possibly any given time within a sample of water, viruses may have a range of associations with microbial communities, from adsorption to non-infective entry and even suppressed infections at various stages before lysis[64,131,254,255].

A small fraction of all the viruses we documented had VPH values high enough to suggest a lytic infection rather than a non-infective interaction. Of course, a low VPH does not directly equate to laboratory measurements of burst size, but rather an average across the entire sample[178]. There may be several explanations of a low VPH with a lytic infection in a sample. Specifically, microbial populations may experience a low rate of infection at any given time[256]. Additionally,

an infection advanced enough that packaged viral genomes were not accessible via Hi-C links, or the host genome was sufficiently degraded[129], could explain a low VPH for a bulk sample analysis such as in this study. A moderately sized VPH could occur in lysogenized populations if the phage was integrated multiple times per genome of host population, but in well studied lambda phage, attachment sites are highly specific and unique to genomes[257]. Though there are reported cases of multiple prophage attaching to a single site, it is rare and often disabling[258]. There are secondary attachment sites for lambda phage, but they still maintain some homology and have lower frequency of integration than primary attachment sites[259]. Therefore, we maintain a VPH of 1 or greater is most likely the result of a lytic infection amongst cells of a microbial population.

The lytic infections that were captured through Hi-C links were novel viruses with canonical genes (Figure 10). The high representation of critical genes for viral reproduction and the absence of host genes confirm that these are viral sequences and not simply host sequences that had not been binned. Furthermore, only two of these interactions were *in silico* predictions which matched, i.e., MAGs from the Alphaproteobacteria and Verrucomicrobiota phyla. The rest of these lytic infections were not predicted accurately or at all by computational efforts. This predictive ability is critical in regions like the ODZ where the vast majority of microbes are uncultivated, leaving no option to characterize viruses but by sequencing and computational efforts[195,277,278].

The examined metagenome is a microbial metagenome rather than a viral metagenome[260], which has some limitations but allows us to observe viruses that have the largest potential for immediate interaction. One possible caveat is that viruses may have been present on the outside of cells in high numbers yet had lower internal counts; however, this scenario is unlikely to occur in the absence of recent lytic propagation of viruses in a sample. In this study, Hi-C links are more

likely to occur with more abundant virus contigs, suggesting that Hi-C link host assignment is more effective for the viruses that carry greater immediate significance or impact within a microbial metagenome, such as those actively infecting the microbial population sequenced (Figure 9D).

Linked bacteriophage in this study were phylogenetically similar to a diverse range of prophage genomes, including three that were similar to *Escherichia virus P1* and one Alphaproteobacteria phage similar to *Clostridium phage C-st*, each a cultured plasmid prophage[261,262] (Figure 11). A virus linking with Planctomycetes had a *terL* most similar to D3112, a prophage which prevents superinfection in *Pseudomonas* bacteria[263]. There were 3 viruses that placed phylogenetically near T1-like phage and, though all T1–T7 phage could lyse plaques on plated cultures of bacteria, T1 has since been shown to be capable of lysogeny[264]. There were no *terL* sequences within our linked viruses closely related to T7 or T4 like phage, which have been found to be abundant in surface water samples[43,45,98,265]. Lysogeny as a viral life strategy has been hypothesized to occur in extreme environments where host productivity or abundance may be low[266,267]. Many studies induce a lytic switch of environmental prophage by applying stressors to lysogenic cells, such as adding mitomycin C, exposing to high levels of UV light, or subjecting to extreme temperatures or harsh chemicals[268–273]. A study in the Baltic Sea, using mitomycin C to induce prophage, suggested increased lysogeny within the anoxic portion of the water column, and metagenomics of the Eastern Tropical South Pacific suggested a higher incidence of prophage genes within the free-living community of the anoxic water column and an increase in Siphoviridae, known for higher incidence as temperate phage[224,274]. The phylogenetic placement of linked viruses in our study supports the claim that lysogeny is an important lifestyle in ODZ environments.

The presence of potentially temperate phage with lytic signatures in Hi-C data suggests that some ODZ lysogenic phage were actively reproducing at time of sampling or by sampling. A prophage sequence can be deactivated within a host genome, but the ability to activate lysogens is not easy to infer from metagenomic data[275,276], so lytic activity of lysogens confirms their viability. Induction to the lytic cycle may occur through a number of mechanisms that could occur during sampling, such as increased host density, exposure to reactive oxygen species, and generalized cell stress[264,268,279,280](Appendix A). Future work should take great care to prevent oxygen exposure at the time of sampling to understand the role of oxygenation and possible induction or activity of ODZ prophage. If oxygen can induce prophage from anoxic depths in the ODZ, the repeated intrusion of oxygenated water by eddies[281] could create periodic episodes of mass mortality of bacteria. The viral induced lysis of members of the microbial community could contribute to enrichment of pools of DOM[18,25,27] and POM[36] in the ODZ. Because heterotrophic denitrification relies on organic material, increases in organic material can create biases towards denitrifaction over anammox[223] which in turn increases the potential for nitrous oxide production, a potent greenhouse gas[282].

Therefore, understanding viral activity not just genomic potential is critical in determining viral impact within ODZs. This study highlights the role that multifaceted sequencing approaches can play to elucidate viral activity and better describe viral lifestyle and ultimate effects on uncultivated populations *in situ* in marine environments.

3.6 Acknowledgements

We thank the captain and crew of the R/V Kilo Moana, Bailey Armos for her help collecting samples at sea, and Susan Burke for extracting the DNA. Thanks to Arthur Nowell for generous

feedback on the manuscript. This work was funded by a National Science Foundation Graduate Research Fellowship and the UW Hall Conservation Genetics Research Award to CR and NSF grants DEB-1542240 and OCE-2022911 to GR.

3.7 Figures

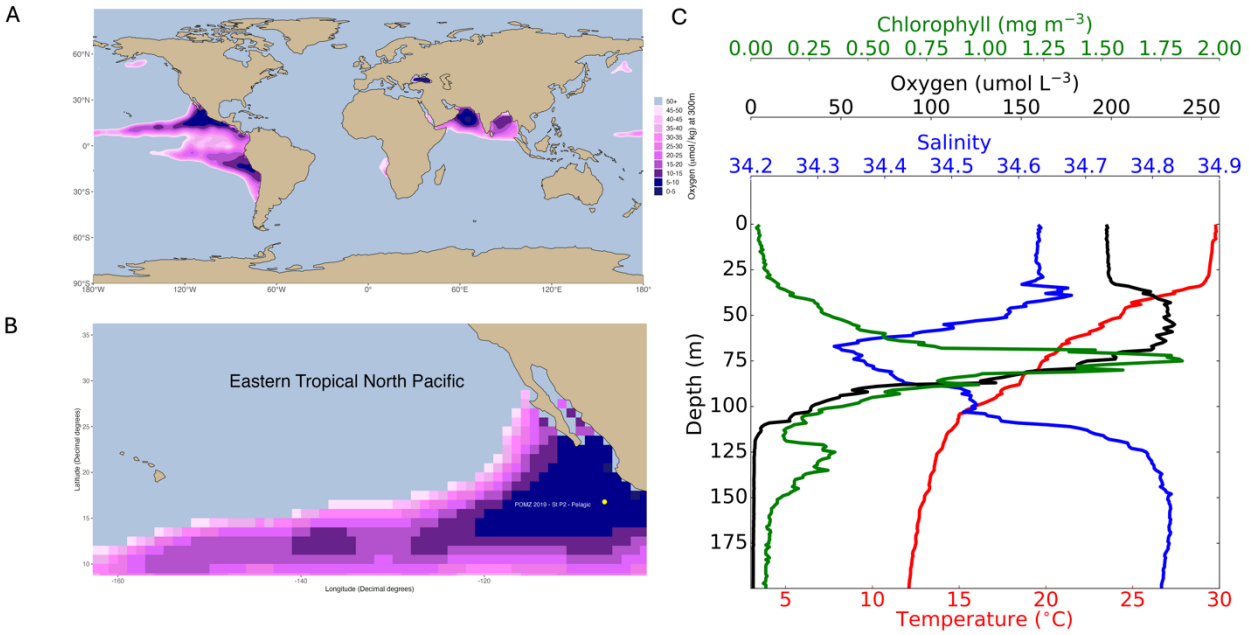


Figure 1.A. Global location of oxygen deficient zones (ODZ), including the Eastern Tropical North and South Pacific and the Arabian Sea. Oxygen concentrations are mapped from 2018 World Ocean Atlas data which preceded sample collection by one year. **B.** Zoomed in view of sample site location (yellow dot) in the Eastern Tropical Pacific Ocean ODZ. Oxygen concentration at 300m depth is depicted with color shown in legend on panel A. **C.** Measurements taken by the CTD instrument during the sampling cast, showing oxygen depletion to zero and a secondary chlorophyll maximum contained in anoxic depths.

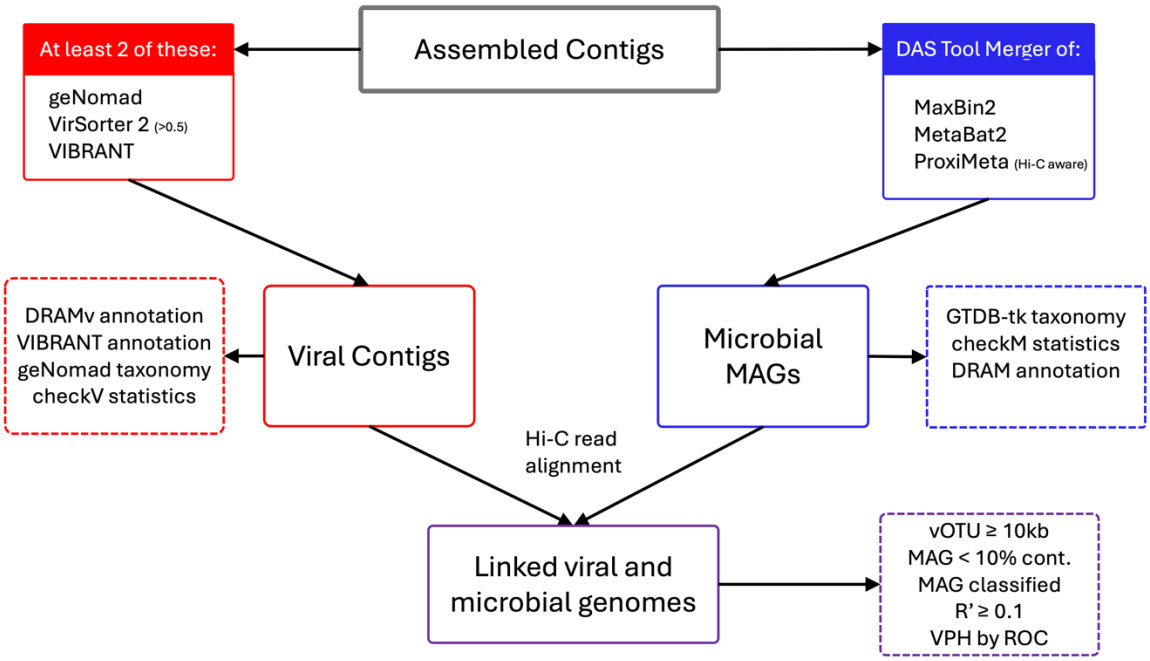


Figure 2. Computational workflow to determine virus-host pairs. Contigs from a hybrid assembly were identified as viral using 3 different computational methods. Contigs identified by at least two of these methods were pursued further as viruses. Potential viral contigs were annotated with DRAMv and VIBRANT, and taxonomy and statistics were derived from geNomad and checkV. Microbial contigs were binned with three different methods, which were fed into DAS Tool to get integrated MAGs. MAGs were analyzed with GTDB-tk, DRAM and checkM. Hi-C reads that mapped viral contigs of sufficient length and viral gene content that linked with sufficiently characterized and uncontaminated bins were statistically analyzed for virus-microbe activity in the sample.

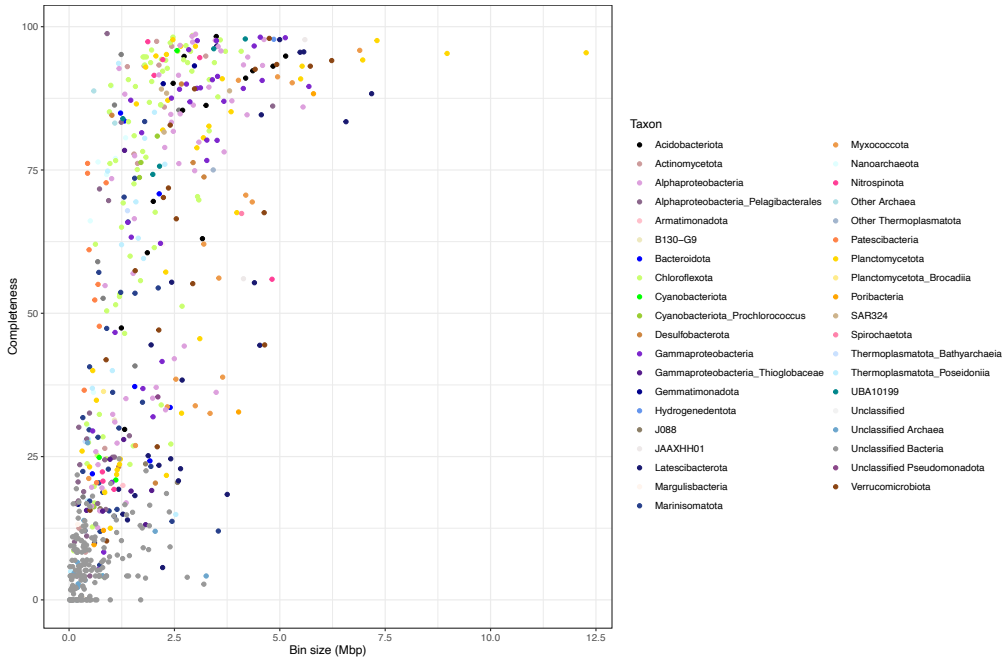


Figure 3. MAG completeness, and size of all non-contaminated bins, as measured by checkM. Each dot is a MAG plotted by checkM completeness and MAG size in millions of basepairs. Dots are colored based on their GTDB-tk taxonomical classification. Included in our medium quality bin analyses are 268 MAGs above 50% complete. All bins that are classified, regardless of completeness are included as potential microbial interaction partners in viral Hi-C analysis.

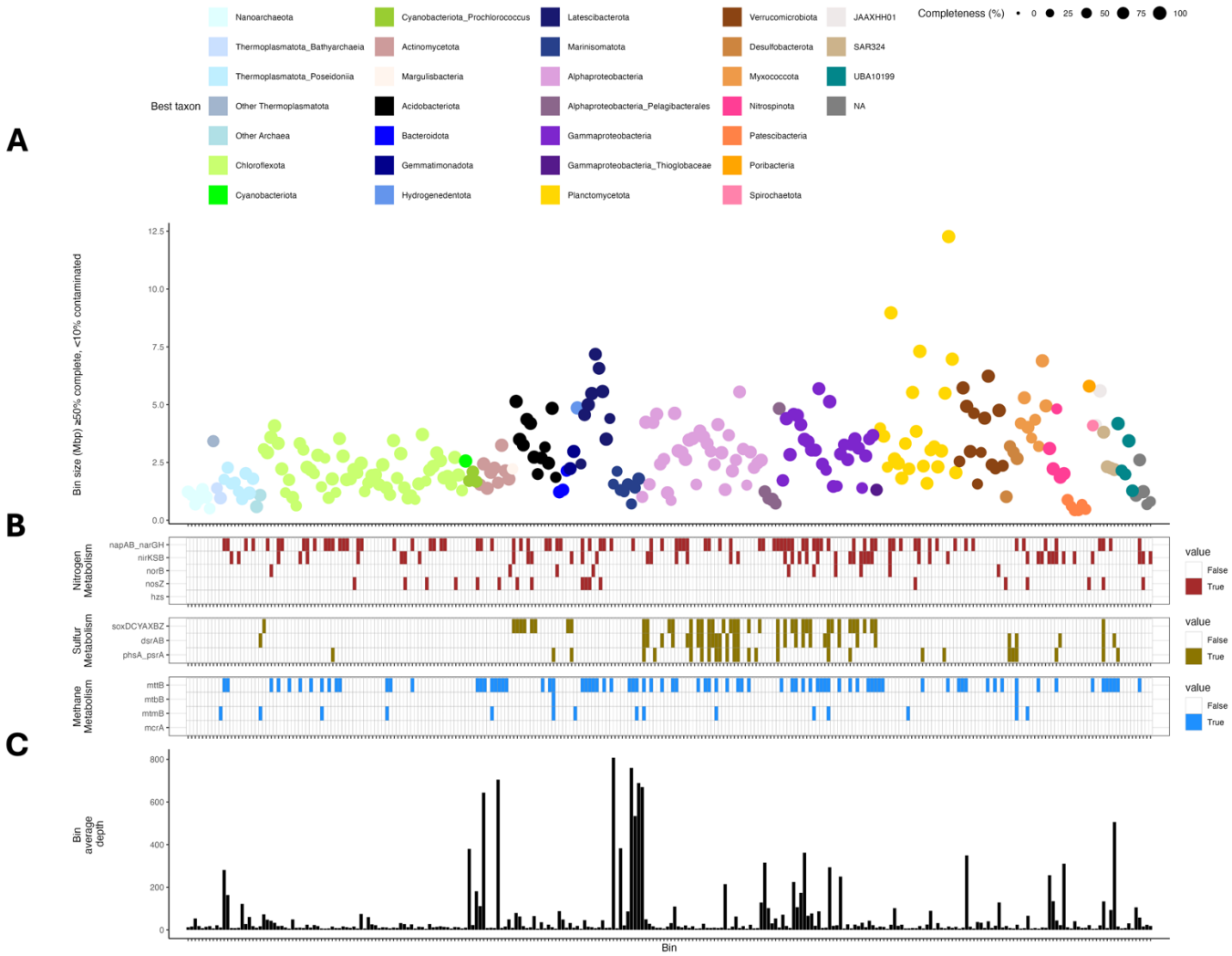


Figure 4. A. Each of the 268 medium-to-high quality bins, where circle size is determined by checkM completion, and genome size is plotted on the y-axis. Each dot is colored by broad taxonomic groupings, determined by GTDB-tk. **B.** The tile plots below the dot plot show the presence or absence of significant nitrogen, sulfur and methane metabolic genes in each of these MAGs, as determined by DRAM annotation. **C.** Average bin depth as determined by bowtie2 read mapping of short-read sequences.

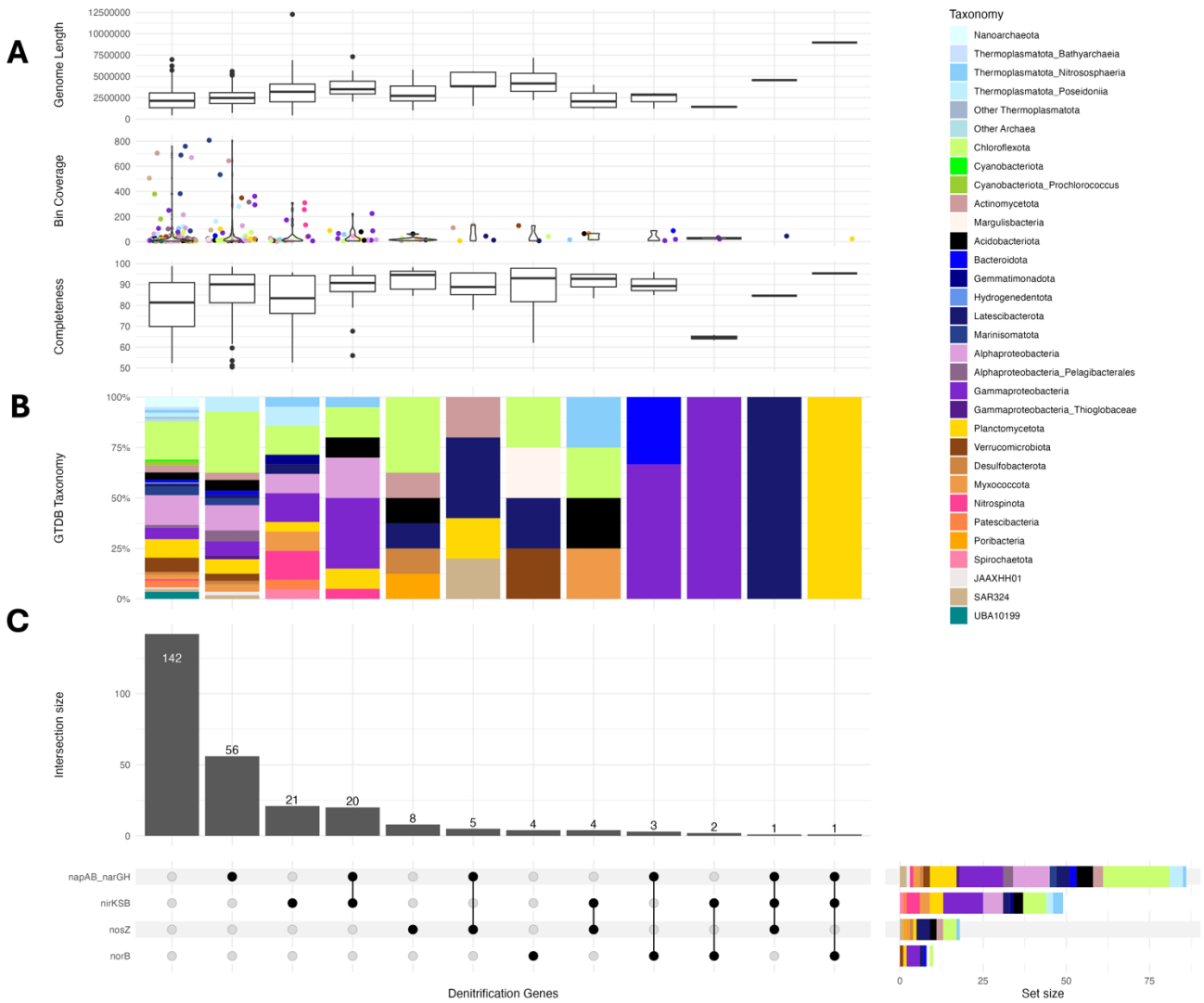


Figure 5 Intersection sets of groups of MAGs containing different combinations of denitrification and nitrogen cycling genes. **A.** Single copy core gene completeness, abundance in the short-read sequencing library, and total MAG length for the defined sets. **B.** MAG taxonomy per set. All taxonomic coloring is described in the legend on the right. **C.** Intersection sets and the quantity of medium-to-high quality MAGs contained in each group, with taxonomic abundance of each gene included by row shown to the right of the intersection sets.

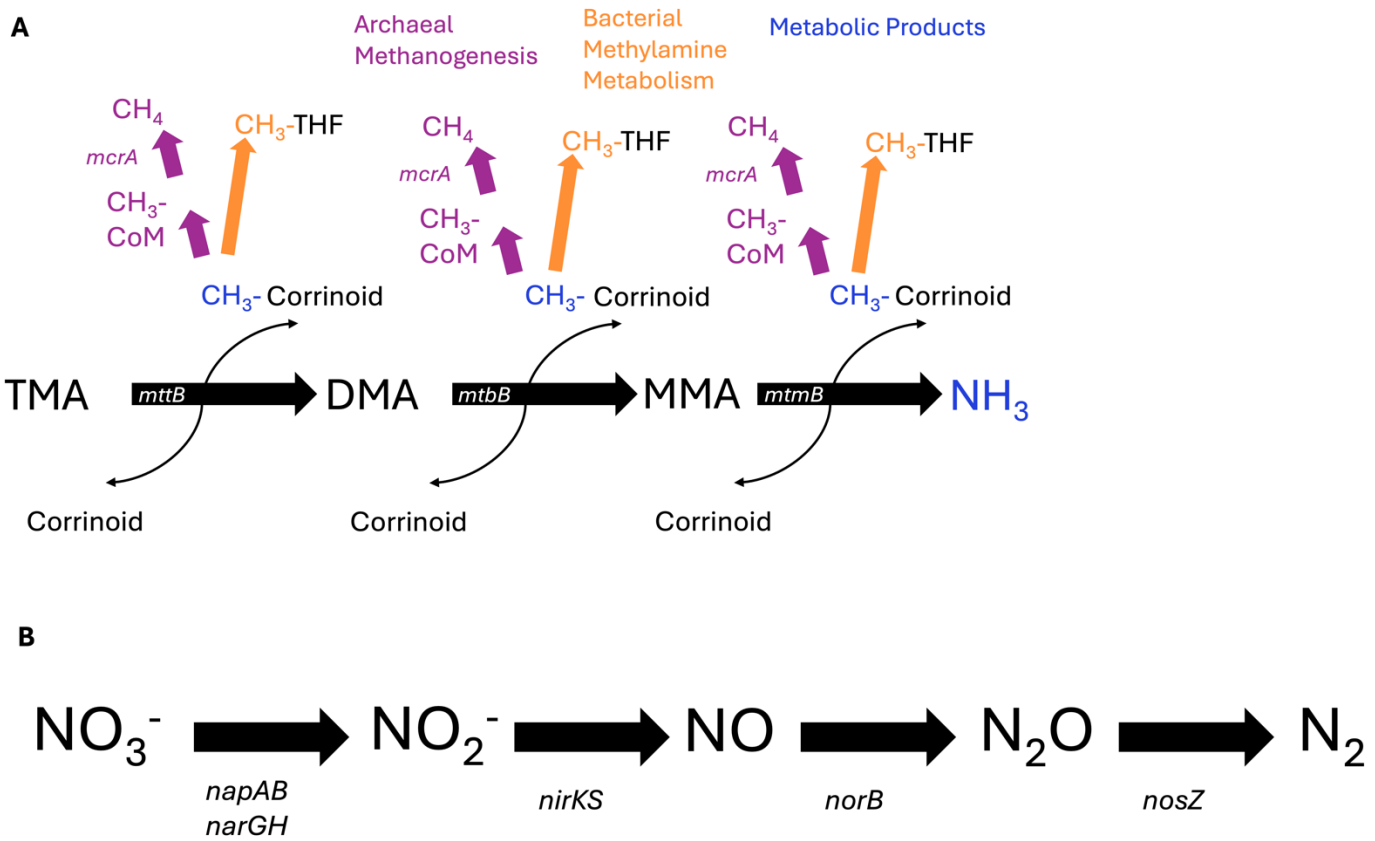


Figure 6A. Anaerobic metabolic pathway of trimethylamine (TMA) demethylation in archaeal methanogens (purple) and bacterial species (orange). Each methyl group is transferred in a two-step reaction to create a methyl-CoM (archaea) or some methylated cognate corrinoid (bacteria), which then makes the methyl group available for the methylation of important cofactors such as tetrahydrofolate (THF). Transferring the first methyl group produces dimethylamine (DMA) and the subsequent transfer creates monomethylamine (MMA). Removing the final methyl group, catalyzed by MtmB, generates ammonia. In methanogenic archaea the methyl-CoM is subsequently converted to methane, catalyzed by McrA. **B.** Modular denitrification, with each arrow representing the dissimilatory reduction of nitrogen species, with the gene name for the enzyme catalyst of each reaction below the arrow. Nitrate (NO_3^-) is respired to nitrite (NO_2-) with NapAB or NarGH, which can be respired to nitric oxide (NO) with NirKS, which can be respired to (N_2O) with NorB which can be respired finally to nitrogen gas (N_2) with NosZ.

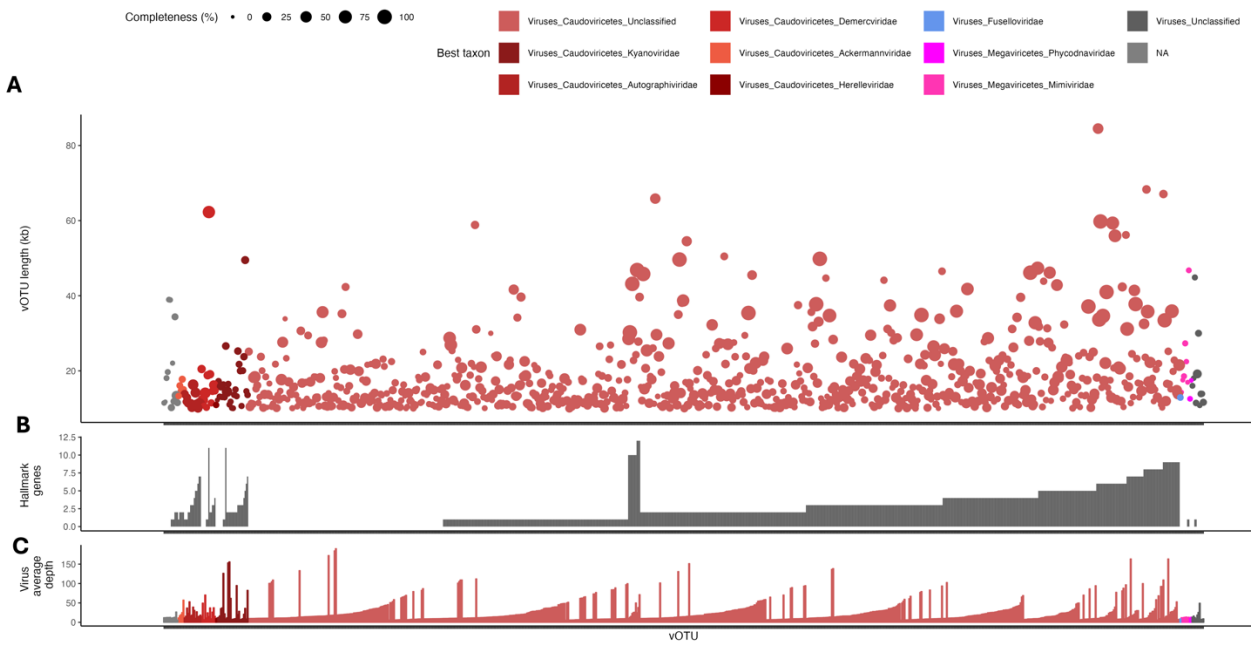


Figure 7. Final 861 viral sequences that were used to query virus-microbe interactions with Hi-C linkages. **A.** Virus contigs greater than 10 kb are shown with diameter dependent on checkV completeness and colored by geNomad taxonomic groupings. Total viral length is plotted on the y-axis. **B.** Hallmark gene presence for each virus contig as determined by geNomad. **C.** Virus coverage in the short-read sequencing library of the sample from the secondary chlorophyll maximum.

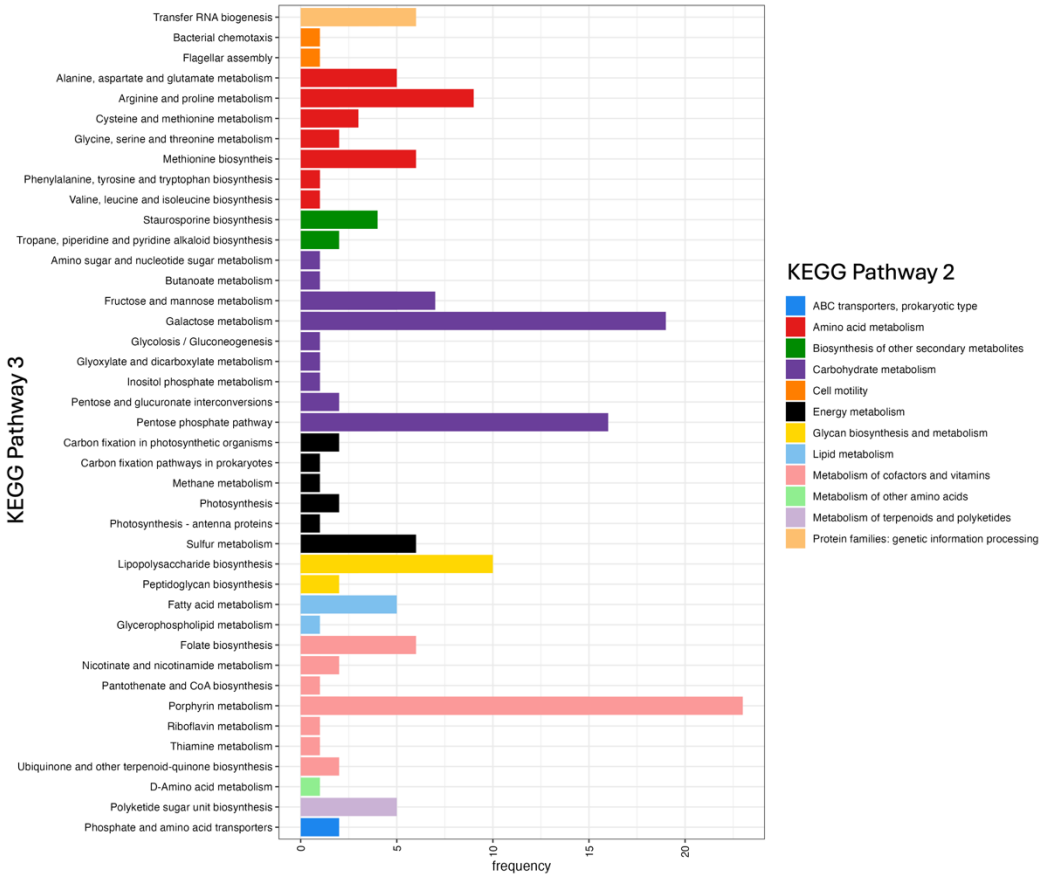


Figure 8. Auxiliary metabolic genes (AMGs) found in viral genomes. Viral genomes at least 10,000 bp in length were annotated with DRAM and VIBRANT and further curated manually to remove metabolic genes likely to have viral function, such as methyltransferases and peptidases. Remaining AMGs were named with KEGG Pathway terminology, with Pathway 1 being higher order divisions, and 2 and 3 more specific. Gene function color-coded by Pathway 2 divisions and grouped by Pathway 3 divisions.

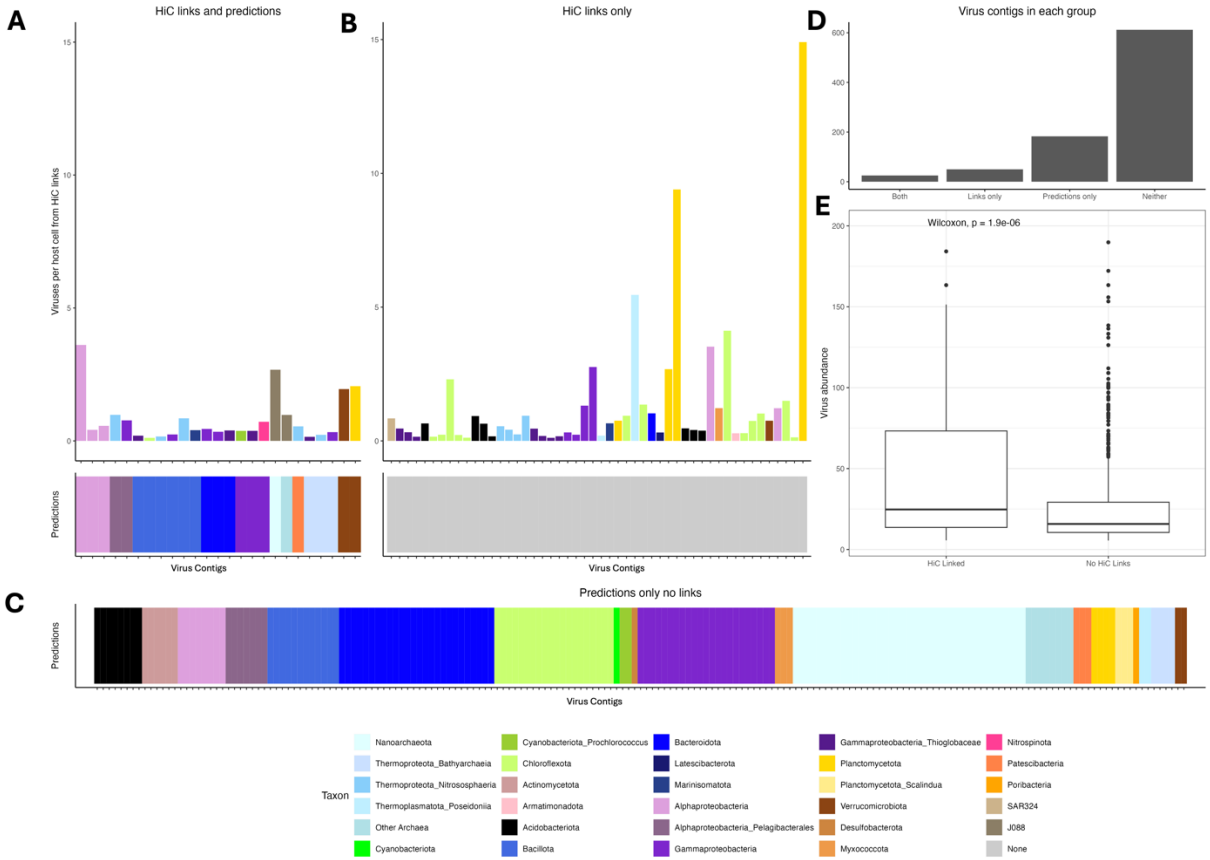


Figure 9. All virus-microbe associations in the sample from the secondary chlorophyll maximum as determined by Hi-C links and iPHoP host predictions. **A-B.** Upper bar plots show virus-microbe interactions determined with Hi-C links, and the lower panel shows best host predictions with iPHoP. The x-axis plots an individual virus contig on each tick and the y-axis shows the calculated viruses per host cell. **A.** The 25 virus contigs that had both iPHoP predictions and significant HiC links. **B.** The 50 viruses that had no iPHoP prediction but had significant HiC links. **C.** Host predictions for 183 viruses that had iPHoP predictions with no significant HiC links to microbial MAGs. **D.** The total number of viruses characterized by these methods and not characterized by either method. **E.** Average abundance plotted for virus contigs that had significant Hi-C links and those that had none, showing that Hi-C linked contigs were more abundant than those that had no links.

Figure 10. Data for each virus-host link determined by Hi-C to have an abundance of greater than one virus copy per host cell in the metagenome sample. **A.** Host predictions, Hi-C linked hosts, and virus per host (VPH) amounts are shown in the tiles to left. Colors for broad host taxonomy are included in the taxon legend below. Virus-per-host tiles are darker for a higher VPH, with the VPH value displayed in the tile. **B.** Gene calls and identities for each viral contig. Gene calls were determined through VIBRANT and DRAMv, which use KEGG, PFAM, VOG and viral RefSeq to identify gene content. Genes are colored by broad function, including nucleotide replication, structural, auxiliary metabolic genes (AMG) and other phage proteins.

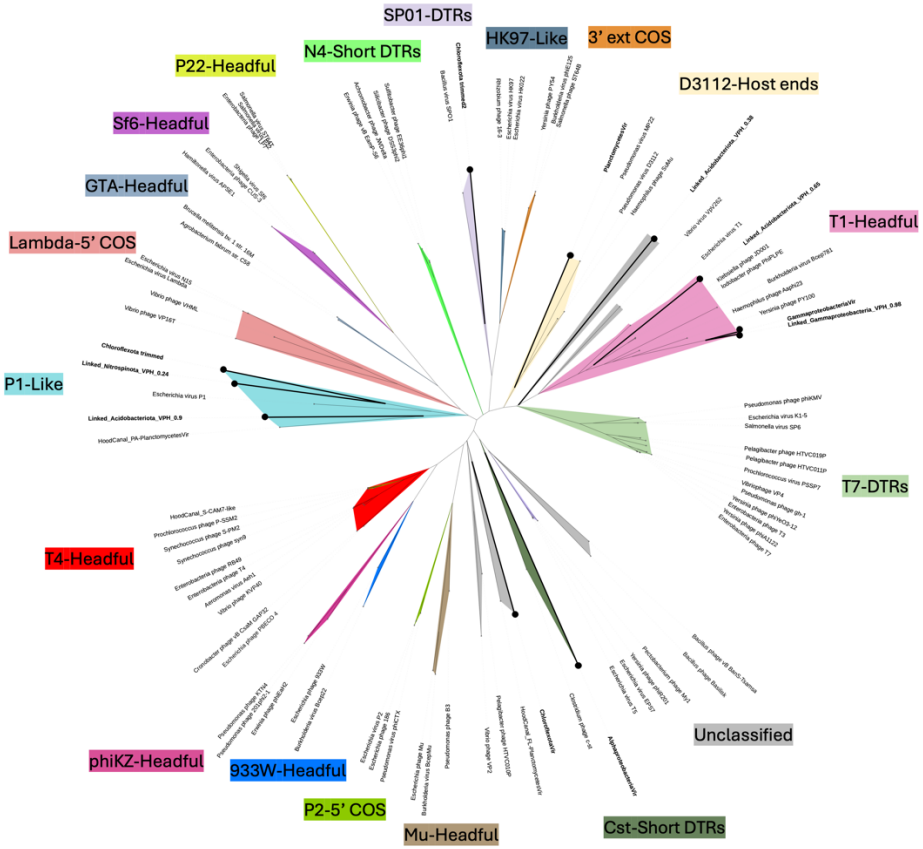


Figure 11. TerL phylogeny of linked viruses. A maximum likelihood phylogenetic tree (RaxML-ng) of the terL gene for each linked viral contig, where terL was annotated. Branches from this study are in bold, with a bold dot on the branch tip, and their annotations are also in bold. Clades are colored by packaging strategy, with colors described outside of the circle. Viruses from this study are named as *Taxon*Vir, where *Taxon* is the phylum or class (Proteobacteria) determined by GTDB-tk taxonomy of the linked microbial genome, for all viral sequences with a VPH of greater than 1. In linked viruses with a VPH less than 1, the name is stated as Linked_*Taxon*_VPH_*X,* where Taxon is the phylum or class (Proteobacteria) of the linked microbial genome, and the X is the VPH number determined for the virus-microbe pair.

3.8 Tables

Table 1. Sequencing and hybrid assembly statistics for all contigs greater than 1 kb

NovaSeq read pairs (all 2x150 bp reads)	309,018,748
PacBio HiFi SMRT number of reads	103,710
PacBio HiFi SMRT read N50	9,971
PacBio HiFi SMRT total length	1,006,013,643
Hybrid assembly length	2,020,378,287
Number of contigs	662,351
Longest contig	755,536
N50	4,375

Table 2. Binning results by method and final quality groupings

Method or Subset	Number of Bins	Number of Contigs binned
MetaBat2	72	207,734
MaxBin2	561	382,422
PhaseGenomics	180	23,590
DASTool	819	361,952
Final MQ Bins	268	73,171
Final HQ Bins	106	20,626

Table 3. Results of Various Virus Identification Methods

Method	Number of viral contigs
geNomad	14,900
VirSorter2 (>0.5)	52,553
VIBRANT	7,064
Contigs identified as viral by at least 2 of 3 of the methods above	12,284

Table 4. Genome statistics and taxonomy of all medium-to-high quality bins

Bin	Contigs	% Complete	% Contaminated	N50	GC	Genome Size	Longest Contig	GTDB Taxonomy
bin_119	44	98.8	0	34996	28.1	902806	72931	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__
bin_50	41	98.7	1.3	156216	58	3002251	657061	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA6615;f__UBA6615;g__;s__
2024_maxbin.086_sub	233	98.48	2.97	86660	53.6	2927985	298782	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1328;g__UBA1328;s__UBA1328 sp002501045
bin.409	115	98.33	0.56	40817	60.4	2918398	110757	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MPNO01;f__;g__;s__
2024_maxbin.074_sub	64	98.29	2.56	192596	67	3492524	510601	d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__UBA823;g__UBA2161;s__UBA2161 sp002730525
bin.266	27	98.18	0	119585	61.5	2465821	323420	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1151;g__GCA-2730555;s__GCA-2730555 sp002730555
bin_17	196	98.17	3.05	54116	55.5	4536705	225749	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA9619;s__UBA9619 sp002718155
bin_7	555	98.07	7.32	14178	53.6	5130675	77121	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA12578;s__UBA12578 sp002729195
bin.466	119	97.96	0.68	68194	60.4	4749378	182736	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__Roseibacillus_B;s__Roseibacillus_B sp002501065
2024_maxbin.194	243	97.86	1.71	28199	63.8	4181529	100381	d__Bacteria;p__UBA8248;c__UBA8248;o__JACPRG01;f__JACPRG01;g__JACPRG01;s__
bin_10_sub	383	97.8	4.4	18280	58.4	4857250	90789	d__Bacteria;p__Hydrogenedentota;c__Hydrogenedentia;o__Hydrogenedentiales;f__GCA-2746185;g__;s__
bin_15	154	97.76	1.74	53869	58.8	4623855	144600	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__GCA-2731375;f__GCA-2731375;g__GCA-2731375;s__GCA-2731375 sp002731375
bin.416	138	97.74	1.1	81423	51.1	5594468	256423	d__Bacteria;p__JAAXHH01;c__JAAXHH01;o__JAAXHH01;f__JAAXHH01;g__CAJXCW01;s__CAJXCW01 sp913051705
bin.107	113	97.74	1.1	75279	57.5	4997360	337149	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA2968;f__UBA2968;g__UBA2968;s__
bin_61_sub	88	97.73	1.7	69425	58.1	2464162	176743	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__UBA12567;s__
2024_maxbin.095_sub	299	97.69	3.17	34219	62.9	3588137	109146	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__CAJWBT01;s__
bin.114_sub	216	97.61	3.77	35988	53.9	3469233	125597	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__SMXS01;f__;g__;s__
bin.259_sub	265	97.58	2.65	43416	58.2	7305679	187356	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__;s__
bin_33_sub	367	97.56	3.32	15212	48	3497066	54217	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA5680;g__UBA5680;s__UBA5680 sp002420425
bin_48_sub	140	97.56	5.49	42640	55	3042827	159375	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA868;s__UBA868 sp002726415
bin.194	86	97.44	4.7	39478	45.1	2078062	122945	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__UBA8592;g__UBA8592;s__
bin.163	51	97.38	1.2	182176	41.1	1869080	362945	d__Bacteria;p__Nitrospinota;c__UBA7883;o__UBA7883;f__UBA7883;g__UBA7883;s__UBA7883 sp002500605
bin_51	165	97.32	2.61	27287	63.5	2930811	83965	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__SHVP01;f__SHVP01;g__;s__
bin.288	101	97.19	0	37127	53.1	2401197	291924	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1328;g__UBA1328;s__

bin_57	25	96.74	1.09	397128	52.4	2632693	601424	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__SMXP01;f__SMXP01;g__;s__
bin.213	47	96.7	1.98	151513	60.8	2773977	454450	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Tepidiformales;f__Tepidiformaceae;g__;s__
bin.88	171	96.64	0	31945	55.8	3507234	133684	d__Bacteria;p__Latescibacterota;c__MWCZ01;o__MWCZ01;f__;g__;s__
2024_maxbin.173	250	96.63	4.86	27295	58.6	4392592	130009	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Ga0077536;f__Ga0077536;g__;s__
bin.36	218	96.59	1.14	20762	52.2	2998618	70673	d__Bacteria;p__Planctomycetota;c__MHYJ01;o__MHYJ01;f__JACPRB01;g__;s__
bin_55	180	96.59	2.27	27122	60.2	2649212	101155	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__UBA12567;s__UBA12567 sp003520025
bin_63	144	96.58	5.56	33527	67.7	2435889	132889	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__S20-B6;s__S20-B6 sp002329075
bin.174_sub	225	96.52	4.48	25094	64.2	3536980	130331	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__CAMCUA01;g__;s__
bin.268	157	96.15	0	35451	59.9	3433532	121641	d__Bacteria;p__UBA8248;c__UBA8248;o__UBA8248;f__UBA8248;g__UBA8248;s__UBA8248 sp002685755
2024_maxbin.254	310	95.98	4.77	21316	51.5	2839409	121457	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Enterobacterales_A;f__Alteromonadaceae;g__Idiomarina;s__Idiomarina aquatica
bin.62_sub	99	95.93	8.91	64724	45.5	1957563	159709	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA6926;f__VFHL01;g__VFHL01;s__
bin.482	243	95.87	7.1	47907	56.2	6896384	147815	d__Bacteria;p__Myxococcota;c__XYA12-FULL-58-9;o__XYA12-FULL-58-9;f__JABJBS01;g__;s__
2024_maxbin.244	434	95.83	4.26	14965	56.9	2562106	74966	d__Bacteria;p__Cyanobacteriota;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__;s__
bin.251	220	95.77	1.49	25738	51.9	3604556	81077	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__UBA2165;g__GCA-2725625;s__GCA-2725625 sp002725625
bin.385	66	95.71	0.99	149979	65.1	3709676	373601	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Bin125;f__Bin125;g__;s__
bin.476	72	95.71	0.99	66773	64.3	2579820	164792	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__;g__;s__
2024_maxbin.108_sub	340	95.71	2.09	25422	60.7	4083672	143508	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA11138;g__UBA2160;s__
bin.312	50	95.71	2.97	103933	56.3	2881372	148052	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__UBA3495;g__UBA9611;s__UBA9611 sp002730485
bin.54	123	95.6	3.85	87959	60.2	5568559	307509	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA8231;f__UBA8231;g__UBA12574;s__UBA12574 sp002712305
bin.330_sub	161	95.54	1.1	64556	53.6	5481390	207807	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA2968;f__GCA-2709665;g__GCA-2709665;s__GCA-2709665 sp002709665
bin.440_sub	642	95.45	4.55	30200	56.5	12268032	97724	d__Bacteria;p__Planctomycetota;c__PUPC01;o__JACPCR01;f__JACPCR01;g__;s__
bin_1_sub	528	95.33	2.49	24946	56.6	8968811	97885	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__ARS98;s__ARS98 sp002685655
bin.318	38	95.16	1.08	118244	51.8	2309917	525309	d__Bacteria;p__Planctomycetota;c__UBA1135;o__UBA1135;f__UBA1135;g__UBA7888;s__UBA7888 sp002500505
2024_maxbin.208	112	95.15	2.73	63729	28	1236722	125778	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrososphaerales;f__UBA141;g__;s__
bin.83	39	94.89	1.7	80245	60	2061601	248879	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__;s__

2024_maxbin.060_sub	206	94.87	3.42	64438	65.4	5139058	194695	d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__UBA8438;g__NP936;s__NP936 sp002726085
bin.280	239	94.87	3.33	25204	55.3	3244020	151011	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__TK06;g__UBA2110;s__UBA2110 sp002388005
bin_53	108	94.82	1.71	175370	54.7	2730021	699157	d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__UBA823;g__UBA11600;s__UBA11600 sp002730735
bin.157	32	94.72	0	118870	51.7	1787965	279925	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__UBA2162;g__GCA-2685815;s__GCA-2685815 sp002685815
bin.477	97	94.72	0	31125	67.5	1833912	80038	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1127;f__UBA1127;g__;s__
bin.227	54	94.72	0	81158	54.8	1757408	158653	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA2985;f__UBA2985;g__JACQTZ01;s__
2024_maxbin.081	546	94.7	5.14	16305	66.4	4239349	65009	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA2966;f__UBA2966;g__GCA-2690215;s__
bin_46	228	94.6	2.69	40594	37.5	3101743	230760	d__Bacteria;p__Nitrospinota;c__CAJXCL01;o__CAJXCL01;f__CAJXCL01;g__CAJXCL01;s__CAJXCL01 sp913051775
bin.245_sub	136	94.31	1.52	52344	39.7	2191512	288049	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA2962;g__UBA2962;s__
bin_72_sub	135	94.25	4.33	31319	45.5	2224615	91579	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__Nitromaritima;s__Nitromaritima sp002731985
bin.55	83	94.22	0.99	46690	67.1	2718286	193897	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA2979;f__UBA2979;g__SHYR01;s__
bin.46_sub	330	94.19	2.3	36565	55.3	6967194	112686	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__GCA-2718815;s__GCA-2718815 sp002726515
bin_69_sub	65	94.14	0	105543	54.8	2293227	237487	d__Bacteria;p__Chloroflexota;c__UBA11872;o__UBA11872;f__UBA11872;g__UBA11872;s__
bin.353	139	94.09	4.65	83869	61	6231513	200020	d__Bacteria;p__Verrucomicrobiota;c__Lentisphaeria;o__JABSQK01;f__CAJWLW01;g__CAJWLW01;s__
bin.105_sub	146	93.73	5.94	41449	58.5	2812458	113619	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__CAJWER01;g__CAJWER01;s__
bin.307	172	93.73	1.98	28900	61.3	3447728	80899	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA11138;g__UBA2160;s__
bin.28_sub	85	93.73	0	68853	38.8	1935111	204911	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__GCA-2717385;g__GCA-2717385;s__GCA-2717385 sp002730795
bin.86	48	93.6	1.6	43051	63.6	1174650	115693	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__UBA8886;s__UBA8886 sp002731655
bin_9_sub	488	93.41	6.08	13219	57.5	4927725	107209	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__UBA1096 sp913030945
bin.346_sub	599	93.33	4.72	9896	58.8	4138210	55874	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA8366;f__;g__;s__
bin.23	141	93.32	0.99	17626	59.6	1795361	78370	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA2979;f__UBA2979;g__VYDC01;s__
bin.433_sub	131	93.3	2.14	26077	66.1	2133576	70746	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__UBA9410;s__UBA9410 sp003541675
2024_maxbin.251	573	93.24	4.84	14135	58.8	4590949	45783	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__;g__;s__
bin.468	58	93.16	2.14	47260	42.6	1781762	115150	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__MedAcidi-G1;s__MedAcidi-G1 sp002729765
bin.484	199	93.16	5.66	26313	61.7	2975171	121294	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Longimicrobiales;f__UBA6960;g__;s__

bin.302	104	93.11	3.42	142269	64.4	4840191	411992	d__Bacteria;p__Acidobacteriota;c__UBA890;o__UBA890;f__WLXD01;g__;s__
bin.203_sub	565	93.11	2.91	13269	58.5	5529565	86808	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__SM23-33;f__SM23-33;g__BS750m-G57;s__
bin_3_sub	363	93.11	4.76	23437	55.5	5723355	126360	d__Bacteria;p__Verrucomicrobiota;c__Kiritimatiellia;o__UBA8416;f__JADHWI01;g__;s__
bin_97	85	93.03	2.27	28399	37.8	1383025	129872	d__Bacteria;p__Actinomycetota;c__Humimicrobiia;o__Humimicrobiales;f__Humimicrobiaceae;g__M55B120;s__
bin_87	78	92.95	0.57	36457	44.6	1819339	90767	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__GCA-002718515;s__GCA-002718515 sp018656645
bin.443_sub	128	92.74	2.97	35926	51.2	2330390	198119	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1328;g__UBA1328;s__
bin_108	34	92.7	0	43192	29.6	1187057	105666	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__HIMB59;f__GCA-002718135;g__;s__
bin.345	356	92.57	4.12	20247	59.8	4414830	90760	d__Bacteria;p__Verrucomicrobiota;c__Kiritimatiellia;o__CAJWTS01;f__CAJWTS01;g__;s__
bin_20_sub	454	92.31	2.78	12448	67.2	4358886	59904	d__Bacteria;p__Acidobacteriota;c__Vicinamibacteria;o__Vicinamibacterales;f__;g__;s__
bin.474_sub	169	92.24	7.92	31199	45.8	2914223	115820	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA826;g__UBA11795;s__UBA11795 sp002708065
2024_maxbin.196_sub	41	92.24	1.98	126231	67	2220564	365107	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA2979;f__UBA2979;g__SHYR01;s__
bin.444	110	91.59	1.11	35341	58.5	2131803	128014	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__2-02-FULL-58-16;g__GCA-2686765;s__
bin.8	170	91.53	3.23	26513	61.7	2596799	100834	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA7887;f__UBA7887;g__UBA7887;s__UBA7887 sp002501105
bin.99_sub	87	91.51	3.42	34408	45.8	2017999	148413	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__Nitromaritima;s__Nitromaritima sp003451695
bin.360	97	91.33	0.19	58686	49.7	3522986	213445	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__UBA9145 sp002726595
bin.52_sub	245	91.25	1.96	32990	58.9	4946821	130379	d__Bacteria;p__Myxococcota;c__UBA796;o__UBA796;f__;g__;s__
bin_23	259	91.03	5.18	22549	57.9	4189509	89046	d__Bacteria;p__Acidobacteriota;c__Terriglobia;o__20CM-2-55-15;f__20CM-2-55-15;g__;s__
2024_maxbin.210	114	90.91	1.14	71790	60.9	3634835	213117	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__Phycisphaeraceae;g__ARS1224;s__ARS1224 sp002688155
bin.393_sub	791	90.89	4.44	8883	62.3	5494489	41406	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Planctomycetales;f__Planctomycetaceae;g__GCA-2726205;s__GCA-2726205 sp002726205
bin.292	47	90.76	0	45467	51.8	1407305	240010	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__RBG-16-60-22;g__GCA-2688435;s__GCA-2688435 sp002688435
2024_maxbin.178_sub	324	90.76	5.84	32254	61.1	3326416	106841	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA11138;g__JADFHV01;s__
bin_37_sub	77	90.74	1.81	74348	51.7	3400197	284252	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__UBA9145 sp002730035
bin.267	198	90.65	0	30762	66.5	4017986	125450	d__Bacteria;p__Myxococcota_A;c__UBA9160;o__UBA9160;f__SMWR01;g__GCA-2729515;s__GCA-2729515 sp002729515
bin_14_sub	248	90.62	8.01	26943	49	4581266	99881	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA4582;s__UBA4582 sp002725285
bin.283_sub	105	90.43	0.4	39912	36.1	1961414	122982	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__OSU-TB39;g__OSU-TB39;s__

bin.201	897	90.21	5.81	8091	70.4	5294675	45614	d__Bacteria;p__Myxococcota;c__Bradymonadia;o__;f__;g__;s__
bin.301_sub	121	90.12	3.85	43100	55.7	2469887	143682	d__Bacteria;p__Acidobacteriota;c__UBA890;o__UBA890;f__UBA890;g__UBA890;s__UBA890 sp002722645
bin.357_sub	58	90.05	2.2	61286	52.3	2234351	192052	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__Longimicrobiales;f__UBA6960;g__UBA1138;s__UBA1138 sp002501085
bin.81	166	90	0.65	28665	45.7	2666928	150303	d__Bacteria;p__Desulfobacterota;c__Desulfobacteria;o__Desulfobacterales;f__UBA2156;g__UBA2156;s__UBA2156 sp002328165
bin.424_sub	64	89.99	2.82	66633	53	2785140	187170	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__GCA-2729495;f__GCA-2729495;g__GCA-002729875;s__GCA-002729875 sp002729875
bin_112	6	89.77	1.32	755536	48.1	989615	755536	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__UBA2162;g__UBA2162;s__UBA2162 sp002335405
bin.336_sub	29	89.64	3.74	57504	34.6	1006059	119299	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__GW2011-AR4;g__GCA-2688925;s__GCA-2688925 sp002688925
bin.366	41	89.57	1.98	63512	43.4	1574131	192050	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA3495;f__GCA-2712585;g__GCA-2712585;s__
bin_5_sub	330	89.57	9.05	28533	54.4	5690528	113751	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__;s__
bin.29	380	89.45	1.85	13186	59.8	3318752	64763	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA828;f__UBA828;g__PTKB01;s__
bin.386_sub	113	89.33	1.83	51482	53.1	3117903	273815	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__UBA11791;s__UBA11791 sp012960345
bin_49	37	89.25	1.41	176520	58.3	3032393	418238	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA6522;f__UBA6522;g__JAHEKF01;s__
bin_24	379	89.22	1.85	14135	44.5	4133890	61811	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Enterobacterales_A;f__Alteromonadaceae;g__Alteromonas;s__Alteromonas macleodii
bin.100	216	89.15	2.7	23724	56.9	2976050	224138	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__MB11C04;g__GCA-2730975;s__
bin.70	266	89.12	0	11350	43.1	2191182	32224	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp002685535
bin.228	54	89.08	1.88	84609	55.7	2620122	232847	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__GCA-2729495;f__GCA-2729495;g__GCA-2729495;s__GCA-2729495 sp002729495
2024_maxbin.036	186	88.8	0.84	42506	46.9	3807543	118199	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA1014;s__UBA1014 sp002683785
bin.173	23	88.79	0	33778	40.7	587282	88604	d__Archaea;p__Undinarchaeota;c__Undinarchaeia;o__Undinarchaeales;f__Undinarchaeaceae;g__Undinarchaeum;s__Undinarchaeum sp002687935
bin.265	290	88.4	1.65	10082	57.8	2313624	27475	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__UBA8110;s__UBA8110 sp002726945
bin.457	787	88.31	1.1	11906	49.4	7178090	49438	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA2968;f__UBA2968;g__;s__
bin.91_sub	688	88.3	3.3	11363	47.2	5799041	46659	d__Bacteria;p__Poribacteria;c__WGA-4E;o__WGA-4E;f__UBA9662;g__GCA-2699335;s__
bin.410	114	88.23	0.65	19697	60.2	1323481	58706	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA11136;f__UBA11136;g__UBA11136;s__UBA11136 sp002686135
bin.197_sub	433	87.76	5.13	11705	58.3	3531397	60182	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1151;g__;s__
bin_62_sub	178	87.54	6.17	20234	51.7	2429705	63231	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Xanthomonadales;f__SZUA-36;g__SZUA-36;s__
bin.135	71	87.17	0	54962	35.8	1463994	132776	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA4486;f__UBA4486;g__UBA11869;s__UBA11869 sp913045195

bin.263_sub	472	87.16	1.24	5903	60.5	2343054	23851	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__SM23-33;f__SM23-33;g__;s__
bin.24	645	87.06	3.55	7977	63.1	3863492	42791	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__JAAXGF01;f__JAAXGF01;g__;s__
bin.455_sub	122	86.98	4.29	58763	50	3673359	210131	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__GCA-002726655;s__GCA-002726655 sp002726655
bin.164_sub	152	86.89	1.83	31142	52.6	2867475	90375	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__REDSEA-S09-B13;s__REDSEA-S09-B13 sp002731295
bin.49_sub	50	86.8	1.98	85549	61.3	1905069	186952	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA6926;f__UBA6926;g__;s__
bin.27_sub	116	86.52	7.61	31759	43.2	1597431	121232	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__GCA-002718515;s__
bin_74	97	86.36	1.09	38490	50.2	2179188	154491	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__Anaerolineales;f__UBA11657;g__;s__
bin.442	75	86.36	0.91	57378	41.7	2170178	159740	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__Anaerolineales;f__UBA11657;g__UBA11657;s__UBA11657 sp913040765
2024_maxbin.048	49	86.33	7.77	57822	32	1077665	121631	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;s__Nitrosopumilus sp002730325
bin.53_sub	201	86.32	0	23443	62	2947623	113164	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__CAMCUA01;g__;s__
2024_maxbin.179	178	86.27	3.42	78890	57.1	3249977	279412	d__Bacteria;p__Acidobacteriota;c__UBA6911;o__RPQK01;f__RPQK01;g__;s__
bin.299_sub	391	86.15	5.72	17296	66.1	4831412	65581	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__MPNO01;f__MPNO01;g__;s__
bin.485_sub	712	85.99	6.07	11040	65.2	5551893	73965	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__GCA-2731375;f__GCA-2731375;g__;s__
bin.200_sub	320	85.98	6.41	9812	60.2	2260991	43874	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__UBA5794;f__UBA11373;g__JACZWZ01;s__
2024_maxbin.078_sub	681	85.44	5.97	6981	36.1	2605401	53253	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrososphaerales;f__JACPRH01;g__;s__
bin.236_sub	84	85.42	4.27	81050	54.7	2694312	316856	d__Bacteria;p__Acidobacteriota;c__UBA890;o__UBA890;f__UBA890;g__UBA890;s__UBA890 sp002731215
bin.209	456	85.16	6.09	11417	55.4	3843950	48600	d__Bacteria;p__Planctomycetota;c__JACQVV01;o__JACQVV01;f__;g__;s__
bin.344	62	85.15	2.97	24664	45.5	957500	79167	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__ARS1013;g__ARS1013;s__ARS1013 sp913046255
bin.43	71	85.07	0.8	47847	47.9	2026373	108008	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Poseidoniaceae;g__UBA226;s__
bin_106	12	85.05	0	138034	35.4	1224714	206170	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__;g__;s__
bin_115	14	85.05	1.87	88067	32.5	961261	200858	d__Archaea;p__Thermoproteota;c__Bathyarchaeia;o__B26-1;f__UBA233;g__AD8-1;s__
bin.126	22	84.93	0	76520	29.7	1221393	180062	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Marisimplicoccus;s__Marisimplicoccus sp002729755
2024_maxbin.350	811	84.79	3.78	4130	54.1	2636864	23294	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__UBA5972;s__
bin.320_sub	324	84.7	3.65	10869	58.9	2414436	59821	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__Casp-alpha2;g__Casp-alpha2;s__
bin_16	78	84.62	2.2	71879	63.9	4561419	222449	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA8231;f__UBA8231;g__UBA8231;s__UBA8231 sp002726335

2024_maxbin.187_sub	625	84.62	4.39	13269	64.6	4223372	58913	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA2966;f__UBA2966;g__GCA-2690215;s__GCA-2690215 sp002690215
bin_111	26	84.55	1.02	70044	35.5	1016862	268303	d__Bacteria;p__Desulfobacterota_D;c__UBA1144;o__UBA1144;f__UBA1144;g__NP136;s__NP136 sp002730475
bin.417	211	83.96	0.94	7712	41.8	1276565	37412	d__Bacteria;p__UBA10199;c__UBA10199;o__2-02-FULL-44-16;f__2-02-FULL-44-16;g__;s__
bin.168_sub	147	83.55	0	14925	33.2	1313344	56584	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__BACL11;g__DUAL01;s__DUAL01 sp913047525
bin.453	78	83.45	0.1	38586	46.8	1794129	91397	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__S15-B10;g__UBA2125;s__UBA2125 sp002329125
bin.334_sub	43	83.42	0.11	60490	30.3	1410150	215377	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__GCA-002694895;g__GCA-002694895;s__
bin.464_sub	436	83.42	2.5	23721	60.6	6570033	156349	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA8231;f__UBA8231;g__;s__
bin_104	219	83.33	3.94	8258	30.7	1238871	35277	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__AG-414-E02;s__AG-414-E02 sp913050185
2024_maxbin.253	184	83.33	2.1	23484	66.3	2427241	118027	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__CAMCUA01;g__;s__
bin.472	12	83.18	0.93	317612	34.3	1085357	360365	d__Archaea;p__Hydrothermarchaeota;c__Hydrothermarchaeia;o__Hydrothermarchaeales;f__WAQR01;g__;s__
bin.368	371	82.83	1.35	8000	56.8	2400464	43155	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__MB11C04;g__UBA2993;s__
bin.124	691	82.65	2.27	5583	57.3	3318952	27729	d__Bacteria;p__Planctomycetota;c__MHYJ01;o__MHYJ01;f__JACPRB01;g__JACPRC01;s__
bin.20_sub	346	81.97	0	7835	58.9	2224936	31359	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__SM1A02;g__UBA12567;s__UBA12567 sp024230295
bin_38_sub	300	81.82	5.64	14538	49.3	3324826	61894	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__Anaerolineales;f__UBA11657;g__;s__
bin.73_sub	232	81.74	2.61	15084	67	2473548	56279	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__;f__;g__;s__
bin.419	292	81.61	1.68	10523	41.5	2254844	44951	d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp022448705
2024_maxbin.066_sub	152	81.5	4.41	72462	40.1	1722942	193950	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__UBA11654;s__UBA11654 sp002717445
bin.286_sub	156	80.97	4.95	15013	44.5	1545344	66294	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA2985;f__UBA2985;g__UBA2985;s__
2024_maxbin.141	383	80.91	1.86	12410	51.6	2093972	80056	d__Bacteria;p__Cyanobacteriota;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Prochlorococcus_C;s__Prochlorococcus_C sp000760295
bin.316	746	80.62	8.42	5102	64.6	3187142	17126	d__Bacteria;p__Planctomycetota;c__UBA1135;o__UBA1135;f__GCA-002686595;g__GCA-2720975;s__GCA-2720975 sp002720975
bin.226	161	80.61	0	11703	37.5	1341061	38837	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__JAGVYU01;g__;s__
bin.87	109	80.53	0	27302	43.4	1794813	80883	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-Q1;s__
bin.241_sub	568	80.21	8.31	7070	42.8	3270390	44456	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__UBA4486;f__UBA4486;g__JACVQI01;s__
bin.426_sub	130	80.17	4.2	42063	50.3	3512581	154795	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__GCA-002726655;s__GCA-002726655 sp002725995
bin.5_sub	602	79.86	2.26	7025	63.4	3080896	50784	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA828;f__UBA828;g__PTKB01;s__

bin.61	576	78.93	4.76	4673	64.4	2270387	18605	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Qipengyuania;s__Qipengyuania flava
bin.30_sub	296	78.87	0.63	15337	53.4	3033148	52262	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Pirellulales;f__UBA7805;g__UBA7805;s__
bin.379_sub	133	78.42	1.32	15699	37	1316182	63957	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__PS1;f__Thioglobaceae;g__SZUA-1055;s__
2024_maxbin.400	591	78.23	0.9	3933	61.2	1746747	18178	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__JACVQG01;f__;g__;s__
bin.436_sub	1026	78.15	5.74	4040	64.3	3679329	15306	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__;g__;s__
2024_maxbin.044_sub	112	77.78	1.28	57543	59.7	1541154	129285	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__UBA3125;s__UBA3125 sp002726315
2024_maxbin.122_sub	371	77.47	1.58	5600	54.4	1556103	25781	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__2-02-FULL-58-16_A;g__JADHSW01;s__
bin.31_sub	90	77.23	0.99	32052	63.2	1824128	143219	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA2962;g__;s__
bin.125	16	77.02	0	160411	46.8	1098814	243734	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__ARS49;g__GCA-2687275;s__
bin.21	732	76.66	1.79	5190	41.5	3268746	28070	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Enterobacterales_A;f__Alteromonadaceae;g__Pseudoalteromonas;s__Pseudoalteromonas shioyasakiensis
bin.295	148	76.64	1.98	8662	47.9	1105059	26058	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__JABGRK01;f__JABGRK01;g__JABGRK01;s__
bin.16_sub	76	76.4	0.93	21030	33.9	687996	61537	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__GW2011-AR4;g__JACPJP01;s__
2024_maxbin.015_sub	301	76.31	1.22	11923	50.8	1705434	84894	d__Bacteria;p__Cyanobacteriota;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Prochlorococcus_C;s__Prochlorococcus_C sp913047395
bin.64_sub	546	76.29	8.39	6271	54.3	2942353	25305	d__Bacteria;p__Desulfobacterota_B;c__Binatia;o__UBA9968;f__UBA9968;g__;s__
bin.232_sub	104	76.15	1.28	26812	64.9	1624379	120456	d__Bacteria;p__Actinomycetota;c__Acidimicrobiia;o__Acidimicrobiales;f__MedAcidi-G1;g__UBA3125;s__UBA3125 sp002365325
bin.171_sub	19	76.15	0	30874	37.7	439269	37742	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA9983_A;f__UBA2163;g__UBA2163;s__UBA2163 sp913041175
bin.15	164	76.13	5.28	6923	48.6	969668	23072	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__UBA2162;g__G020354275;s__
bin.447_sub	281	76.12	0	13341	49.7	2605441	59060	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__UBA2165;g__TMED23;s__TMED23 sp913041095
2024_maxbin.031_sub	424	76	6.92	12727	43.8	2272846	102691	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__CG-Epi1;s__CG-Epi1 sp002506485
bin_76_sub	327	75.66	0	7719	61.8	2146524	30366	d__Bacteria;p__UBA8248;c__UBA8248;o__UBA8248;f__UBA8248;g__;s__
bin.276_sub	130	75.08	0	19087	46.6	1618192	78836	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1328;g__UBA1328;s__UBA1328 sp002687335
2024_maxbin.331_sub	854	75.01	8.8	6082	46.3	3420788	37622	d__Archaea;p__Thermoplasmatota;c__JAEHCK01;o__;f__;g__;s__
bin.313_sub	466	74.87	1.9	8282	55.2	2980435	37549	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__Casp-alpha2;g__TMED8;s__
bin.407_sub	51	74.8	0	29923	57.6	915235	70670	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__MGIII;f__CG-Epi1;g__UBA8886;s__UBA8886 sp002731905
bin.380	28	74.43	0	29561	37.2	438729	70698	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__UBA9983_A;f__GCA-002724975;g__GCA-002724975;s__GCA-002724975 sp002724975

bin.133	196	74.23	0.9	16114	37.3	1987621	54716	d__Bacteria;p__UBA8248;c__UBA8248;o__UBA8248;f__UBA3496;g__UBA3496;s__
bin.212	64	74.22	0	26549	35.7	890671	51075	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Woesearchaeales;f__GCA-2687795;g__GCA-2687795;s__
bin_42_sub	423	73.78	3.98	8896	47.1	3201701	31348	d__Bacteria;p__Desulfobacterota;c__DSM-4660;o__Desulfatiglandales;f__HGW-15;g__;s__
bin_81_sub	120	73.69	2.13	20307	50.2	1670480	62895	d__Bacteria;p__Cyanobacteriota;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Prochlorococcus_C;s__
bin.234	72	73.56	0	32304	43.9	1555716	99584	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA1611;g__UBA2134;s__UBA2134 sp002471805
2024_maxbin.006_sub	262	73.49	8.43	6487	30.8	1010827	34088	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__HIMB59;f__GCA-002718135;g__JAGFWP01;s__JAGFWP01 sp017639935
2024_maxbin.217_sub	50	72.77	9.41	31117	39.5	879918	72476	d__Bacteria;p__Patescibacteria;c__ABY1;o__UBA2169;f__UBA2169;g__;s__
bin.496	338	72.59	2.48	5128	59	1541042	23330	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__JACVQG01;f__JAHKAY01;g__;s__
bin.488_sub	384	71.86	2.36	7639	59.2	2358244	39472	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__
bin_125	36	71.69	0	26867	28.1	722579	51061	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__
bin.317	102	70.84	0.54	33952	34.5	2139474	95046	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__F082;g__GCA-002708315;s__GCA-002708315 sp002686705
bin.162_sub	1148	70.61	8.23	4223	65.2	4193158	18450	d__Bacteria;p__Myxococcota;c__UBA796;o__UBA9615;f__UBA9615;g__;s__
bin_45_sub	437	70.36	1.32	8466	62.3	3047887	36211	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Tepidiformales;f__Tepidiformaceae;g__;s__
bin_100_sub	116	70.27	0	12794	44.3	1306648	43599	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA2128;g__UBA2128;s__UBA2128 sp002471865
bin.405_sub	285	70.2	4.71	9866	58.2	2238191	44297	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1096;g__UBA1096;s__
2024_maxbin.063_sub	454	69.77	3.08	45125	42.5	3079440	498651	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA826;g__UBA11996;s__UBA11996 sp002500485
bin_116	115	69.66	0.6	12312	28.7	937819	61655	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Pelagibacter;s__Pelagibacter sp913045275
bin.221	535	69.52	8.12	4126	42.2	1998451	20189	d__Bacteria;p__Acidobacteriota;c__UBA6911;o__RPQK01;f__RPQK01;g__;s__
bin.118	183	69.44	0	11290	50.4	1586412	45900	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-P;s__
bin.40_sub	377	69.42	1.94	16729	58	4346776	71977	d__Bacteria;p__Myxococcota;c__UBA9042;o__GCA-2863065;f__GCA-2863065;g__;s__
bin.69	254	69.25	2.97	6521	38.2	1284859	29502	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA6927;g__UBA6927;s__
2024_maxbin.155	51	67.89	0.97	78628	49.7	1386462	156345	d__Archaea;p__Thermoproteota;c__Bathyarchaeia;o__TCS64;f__TCS64;g__GCA-2726865;s__GCA-2726865 sp002726865
bin.254_sub	345	67.63	2.04	6930	55	2043520	25965	d__Bacteria;p__Chloroflexota;c__UBA11872;o__UBA11872;f__UBA11872;g__UBA11872;s__
2024_maxbin.206	227	67.56	3.41	26945	62.9	3980497	125007	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__Phycisphaeraceae;g__;s__
bin.137_sub	1019	67.55	3.63	5381	57.7	4629126	29253	d__Bacteria;p__Verrucomicrobiota;c__Kiritimatiellia;o__SLOV01;f__;g__;s__

2024_maxbin.518	2039	67.41	9.64	2168	53.2	4094460	9522	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__ARS1246;f__ARS1246;g__JAJRYZ01;s__
bin.2_sub	501	66.5	4.19	3346	66.2	1548939	13237	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Rhodospirillales_A;f__2-02-FULL-58-16;g__JADFTO01;s__
2024_maxbin.432_sub	858	66.48	3.77	3967	44.8	2544590	22766	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__UBA2987;g__PNEW01;s__
bin.399	28	66.12	0	44733	28.2	503449	145821	d__Archaea;p__Nanoarchaeota;c__Nanoarchaeia;o__Pacearchaeales;f__GW2011-AR1;g__NP1466;s__NP1466 sp002731115
bin.388	84	65.93	6.59	25900	41.6	1399364	103760	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__UBA2126;s__
bin.404	63	65.86	1.11	45566	36.4	1394041	167391	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__GCA-2730855;s__GCA-2730855 sp002730855
bin.272	251	65.02	1.25	5578	61.5	1244953	21529	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SM23-28-2;f__RBG-16-64-32;g__JACPQO01;s__
bin.155	137	63.28	8.23	21446	35.7	1476583	111260	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__GCA-2730855;s__GCA-2730855 sp913042085
bin.123	205	63.1	8.61	10681	53.6	1641921	34279	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin.273_sub	600	63.02	0	6286	66.4	3164377	24941	d__Bacteria;p__Acidobacteriota;c__UBA890;o__UBA890;f__UBA890;g__UBA890;s__
bin_68_sub	189	62.2	2.03	18188	49.3	2169101	68592	d__Bacteria;p__Pseudomonadota;c__Gammaproteobacteria;o__Arenicellales;f__UBA868;g__;s__
2024_maxbin.258_sub	403	62.07	5.17	19688	34.2	2219294	61813	d__Bacteria;p__Margulisbacteria;c__Marinamargulisbacteria;o__UBA817;f__UBA817;g__;s__
bin.408_sub	990	62.07	2.9	3529	62.3	3197140	16926	d__Bacteria;p__Myxococcota_A;c__UBA9160;o__UBA9160;f__UBA4427;g__;s__
bin_128_sub	66	62.05	1.19	11245	47.6	625266	29006	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__ARS1013;g__ARS1013;s__ARS1013 sp002690375
bin.242	142	61.96	5.6	12307	59.1	1245335	37331	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__
bin_82	272	61.46	0.93	10167	65.4	2021877	63127	d__Bacteria;p__Chloroflexota;c__UBA11872;o__UBA11872;f__;g__;s__
bin.467	19	61.08	0	40625	36.5	475495	69515	d__Bacteria;p__Patescibacteria;c__Paceibacteria;o__Paceibacterales;f__UBA6532;g__UBA6532;s__UBA6532 sp002434805
bin.494_sub	578	60.56	3.68	3459	42.2	1857311	17262	d__Bacteria;p__Acidobacteriota;c__Aminicenantia;o__Aminicenantales;f__Aminicenantaceae;g__;s__
bin.108	259	59.97	5.39	7171	46	1449059	28337	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__Dehalococcoidales;f__G020353235;g__G020353235;s__
2024_maxbin.021_sub	197	59.56	1.6	77940	55.9	1766911	306640	d__Archaea;p__Thermoplasmatota;c__Poseidoniia;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O3;s__MGIIb-O3 sp002722735
bin.319_sub	100	59	8.74	10270	37	680370	35399	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrososphaerales;f__JACPRH01;g__;s__
bin.219_sub	156	57.43	4.19	14840	51.9	1569137	46028	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Pedosphaerales;f__UBA1100;g__;s__
2024_maxbin.517_sub	1220	57.19	3.23	1943	61.6	2294445	9915	d__Bacteria;p__Planctomycetota;c__MHYJ01;o__MHYJ01;f__JACPRB01;g__JACPRC01;s__
bin.290	67	57.14	0	14985	39.2	706492	69572	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__UBA2126;s__UBA2126 sp002436185
bin.507_sub	586	56.95	1.99	2713	69.6	1520585	9280	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Pelagerythrobacter;s__Pelagerythrobacter marinus

bin.339_sub	937	56.14	1.75	4190	61.5	3554661	17798	d__Bacteria;p__Myxococcota;c__UBA796;o__UBA9615;f__UBA9615;g__UBA6601;s__
bin.14_sub	936	56.03	3.45	4880	55	4137974	19095	d__Bacteria;p__JAAXHH01;c__JAAXHH01;o__JAAXHH01;f__JABIQT01;g__JABIQT01;s__
bin_8_sub	803	55.94	5.9	6679	41	4816157	120679	d__Bacteria;p__Nitrospinota;c__Nitrospinia;o__Nitrospinales;f__Nitrospinaceae;g__Nitromaritima;s__
bin.395_sub	764	55.69	8.36	2173	70	1694563	6754	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA2979;f__UBA2979;g__SHYR01;s__
2024_maxbin.470_sub	1379	55.42	5.79	1776	55	2434516	10883	d__Bacteria;p__Latescibacterota;c__UBA2968;o__SICM01;f__SICM01;g__;s__
bin.98_sub	622	55.33	3.45	9454	56.8	4397804	34760	d__Bacteria;p__Latescibacterota;c__UBA2968;o__UBA8231;f__UBA8231;g__GCA-2724215;s__
bin.32_sub	230	55.17	3.45	23841	56.5	2934935	110711	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__MB11C04;g__UBA2993;s__
bin.439	124	55.04	1.71	7036	39.1	688206	20614	d__Bacteria;p__Patescibacteria;c__ABY1;o__Magasanikbacterales;f__UBA922;g__GCA-2716765;s__
bin_120_sub	88	54.83	0.43	11918	55.8	854699	34224	d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__UBA11136;f__UBA11136;g__;s__
2024_maxbin.286_sub	1063	54.42	6.9	2040	41.9	2115890	21515	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA1611;g__;s__
bin.202	193	53.65	0	10417	39.9	1225421	99208	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__UBA1611;g__GCA-002717915;s__
2024_maxbin.002_sub	236	53.5	8.52	11166	39.4	1564968	43497	d__Bacteria;p__Marinisomatota;c__Marinisomatia;o__Marinisomatales;f__TCS55;g__UBA2126;s__
bin.186_sub	110	53.37	1.98	17856	39.2	1203329	67021	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__;f__;g__;s__
2024_maxbin.242	136	52.91	8.64	17619	56.9	1196233	54831	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__Anaerolineales;f__UBA11657;g__CAJWGI01;s__
bin.84_sub	158	52.61	9.42	7593	33.4	808007	63422	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopelagicus;s__
2024_maxbin.326_sub	272	52.32	4.47	2588	36.4	610774	8485	d__Bacteria;p__Patescibacteria;c__ABY1;o__Magasanikbacterales;f__UBA922;g__GCA-2716765;s__
bin.349_sub	65	51.49	0	39815	47.2	1105025	98869	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__UBA1151;f__UBA1328;g__UBA1328;s__UBA1328 sp002725925
bin.12_sub	696	51.23	2.48	4167	57.3	2681708	18014	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__SAR202;f__UBA11138;g__UBA1123;s__
bin.364_sub	168	50.42	0	5682	53.8	893647	20757	d__Bacteria;p__Chloroflexota;c__Dehalococcoidia;o__JACPPY01;f__JACPPY01;g__;s__

CONCLUSION

The major theme of this thesis is the identification and classification of virus-host interaction in the marine environment. Viruses are astoundingly diverse and can have a great impact on the biogeochemical cycling of ocean microbial populations[18,19,30,36,265]. Viruses reshape abundance profiles of organisms, alter available pools of nutrients and lead to genetic diversity in microbial populations[93,265,283]. A massive community-wide effort to gather environmental viral sequences and delineate them taxonomically has been undertaken [8,30,35,284–286]. The clear next step is to determine what hosts they interact with, and ultimately how those viral-host interactions impact microbial communities and how interactions are affected by gradients in the environment.

When techniques are applied to look at virus-microbe interactions, beyond cultivation of lytic viruses on hosts, our view of highly specific virus-host interactions is shifted more. Cultured bacteriophage adsorb to a wide range of non-host membranes[260]. Single amplified genome sequencing of marine populations detected a cyanophage within a heterotrophic *Roseobacter* cell [66]. Hi-C sequencing applied within microbial mats of hydrothermal vents[225], groundwater[226], and soil[83] all reveal virus capacity for interactions across a wide range of microbial taxa. Yet, in culture, to isolate a virus that can infect broadly is rare[74]. Therefore, the question is not just which microbial populations are experiencing viral-induced lysis and under what conditions, but how else are viruses interacting with microbial populations? Furthermore, what impact does that interacting have on microbial diversity and biogeochemical cycling?

In Chapter One I used the then-new ProxiMeta technique to assess viral infection in a local hypoxic estuary. Specifically, I wanted to analyze virus-host interactions on particles at depth, because high rates of virus-infected prokaryotes have been documented on particles[287]. By

applying Hi-C technique we could infer that a *Synechococcus* population infected by SCAM-7, likely within shallower depths, had been delivered to hypoxic depths on particles, supporting the viral shuttle hypothesis[37]. However, in the free-living size fraction where *Synechococcus* cells and sample chlorophyll were both low, this S-CAM7 like phage was clustered with a Gammaproteobacteria. This evidence suggests that the viral shuttle[36] may deliver virus particles to depths where hosts are not abundant, and in these new locations viruses may interact with non-host or sub-optimal hosts. Particles are not the only mode of transportation for viruses in the ocean, as broad scale sequencing has shown that viral populations travel entire gyres via circulation and can be distributed across basins, occupying both coastal and oligotrophic regions, where the density of host populations certainly vary[34].

In Chapter Two, I applied Hi-C sequencing to a marine phytoplankton community to understand infection and virus interactions in a diverse marine community. Within the phytoplankton community of the deep chlorophyll maximum, we detected picoeukaryote populations experiencing infection including an *Ostreococcus*, unclassified Chlorophyta and a *Bathycoccus prasinos* population. The *Bathycoccus prasinos* population was linked with *Mimiviridae* as well as a relative to BpV1 virus. Previous metagenomics in the region had determined that *Bathycoccus* and their viruses were present in this region but had not associated the *Mimiviridae* sequences with this picoeukaryote population. Cyanophage containing photosynthesis AMGs linked with *Synechococcus* genomes within this sample showed the ability to link AMGs directly with hosts. Dominant prokaryotes were also linked with viral sequences, including Alphaproteobacteria, the Gammaproteobacteria order Arenicellales, and the Planctomycetes order Pirellulales. Most of the bacteriophage linking with each of these genomes were not classified. One classified bacteriophage, a CrAss-like phage was linking with a

Planctomycetes, a novel host for this phage, which is typically studied in the human gut. Furthermore, members of the PVC group show linkage patterns across major lineages of viruses, hinting at the ability of Hi-C linkages to detect DNA proximity by alternate routes than viral infection alone.

In Chapter 3 we brought Hi-C sequencing and long-read technology to an oxygen deficient zone (ODZ), a naturaly occurring anoxic region in the ocean where microbial metabolism accounts for a disproportionate amount of biogeochemical cycling by area, and almost all cells are uncultivated. Viruses in the ODZ have hosts that are mostly unidentified. Hi-C enhanced binning and produced 268 metagenome asssembled genomes (MAGs) of medium-to-high quality, indicating the role of microbial populations in methane, nitrogen and sulfur cycling within a secondary chlorophyll maximum in an ODZ. Hi-C can link large viral sequences within the metagenomic sample to a diversity of uncultivated microbial genomes. Almost 20 viruses from this sample had high virus-per-host ratios suggesting active lytic infection, the majority of which could not have been predicted with state-of-the-art computational prediction tools. Some viruses had host predictions and Hi-C links that were broadly dissimilar and varied across phyla, supporting the increasingly invoked hypothesis that virus-microbe interactions in environmental samples are more far-ranging than cultures would predict.

This thesis brings to light the necessity of utilizing new tools to study virus activity in the ocean. Hi-C brings unique insights to genomic sequencing projects, which both enhance our view of microbial communities and serve to distill which viral sequences may be of greatest importance at any given time *in situ*. It is my hope that the significance of the findings in this thesis motivates the broader scale collection of high biomass samples to run simultaneous Hi-C sequencing to metagenomes in the ocean environment. In order to confirm broad host ranges,

which defy conclusions made from decades of culturing work, a multifaceted approach should be employed. For example, studies with high-throughput SAG sequencing in conjunction with Hi-C sequencing could provide confirmation to the developing idea that viruses can interact with a wide range of bacterial and archaeal taxonomy in the environment.

The data recovered through Hi-C do not negate the necessity of culturing more viruses from the environment but rather suggest the need for each. The field will continue to benefit from gathering more cultures in the laboratory to classify the astounding number of unclassified viral genes and conduct controlled experiments on virus-host interactions. Additionally, a recently proposed theoretical framework suggests that the transcriptional regime of a viral genome helps define its ecological type more precisely than taxonomy[288]. Although increasing numbers of studies are utilizing both SAG and Hi-C sequencing to determine virus and host pairings in metagenomic samples, the interpretation of these data would benefit from laboratory-based analytical refinement. In SAG sequencing the speed of replication of a viral sequence versus the microbial genome has been suggested to relate to lysis of cells[66]. Application of Hi-C may experience varying signal due to the inherent complication a host genome and a viral genome in a bulk metagenome, which may be further complicated by the degradation of host genomes and the packaging of viral genomes during lytic infections. Both SAG and Hi-C sequencing could benefit from a time-point analysis of lytic and suboptimal host pairings in laboratory cultures. The assurances from a laboratory culture of the ability of Hi-C and SAG sequencing to discover non-host interactions as well as lytic interactions at various stages of infection could help bolster and parse the amassing environmental data from these studies. Hi-C techniques, however, offer a unique view of viral activity within a bulk community. The insights to be gained from applying this method in more environments and sampling efforts make these suggested calibration efforts

worthwhile. It is my hope that the field continues to build upon the massive metagenomic datasets collected by taking a closer look at viral activity using a variety of tools, including Hi-C[30,36].

APPENDIX A:LYSOGENIC INDUCTION EXPERIMENTS WITHIN OXIC AND ANOXIC DEPTHS OF THE EASTERN TROPICAL NORTH PACIFIC

Summary:

Both chapter 2 and chapter 3 look at viral interactions within chlorophyll maxima of the Eastern Tropical North Pacific (ETNP). Chapter 2 specifically assessed viral activity within the oxygenated chlorophyll maximum just above the anoxic waters of this oxygen deficient zone (ODZ). Chapter 3 analyzed viral activity of the chlorophyll maximum housed within the anoxic water just below. There is a discussion in chapter 3 about the potential adaptation of bacteriophage at colder less productive mesopelagic depths to be lysogenic. This hypothesis was explored with shipboard experiments to induce prophage from the surface water and the anoxic water within the ODZ, during a cruise on the R/V Sikuliaq in December 2016 to the region. Due to difficulty in reproducing a null effect with controls from anoxic depths these results are not conclusive, but data here may support the hypothesis that lysogeny is high within the anoxic water column.

Materials and Methods:

Sample Collection

Seawater samples were taken at an offshore site within the ETNP (P1: 20.12 °N, 106.5°W) from the research vessel *Sikuliaq*, during cruise SKQ201617S (cast 21, January 2, 2017, local time of 08:12). The water properties were determined using a CTD-Rosette equipped with conductivity, temperature, pressure, transmissivity, chlorophyll fluorescence and oxygen (SBE 043) sensors. Because chlorophyll maximums from oxic and anoxic water column varied in size, water outside the chlorophyll maxima were collected. Water for the oxic experiment was

collected the surface at 10 m depth located at above the chlorophyll maximum (21 m -26 m depth). Water for the anoxic experiment was collected from the 130 m which was within the anoxic water column (80 m -901 m depth), and below the secondary chlorophyll maximum (88 m - 100 m).

Maintaining anoxia for secondary chlorophyll maximum samples

Anoxic samples were collected directly from niskin with tubing, and by first overflowing glass bottles and collecting a bubble free sample and immediately sealed and flushed with nitrogen. Seawater from anoxic depths that had been prefiltered with 0.02 Anodisc filter, was also flushed with Nitrogen to prevent oxygen contamination for resuspension of bacteria. Bottles of water from the anoxic depth were sealed with septa lids and all handling of samples from anoxic depth was done in an inflated glove bag.

Induction of lysogens

Previously described protocols for the induction of environmental prophage were roughly followed[280]. Specifically, five liters of seawater from each depth were collected. Particles were prefiltered on 3.0 μm polycarbonate filters and free-living microbial community was concentrated on a 0.2 μm polycarbonate filter. This removes free viral particles and concentrates microbial cells. Microbial community was then resuspended in ≥0.02 μm seawater from the same depth. Resuspension occurred by carefully placing the filter, sample side facing in, into a 50 mL Falcon tube and gently shaking at 4 °C temperatures for 15 minutes. The whole community within the Falcon tube was sampled for a time zero sample. Immediately afterwards water was distributed equally into 100 mL glass bottles. Test bottles for each depth (two replicates) were injected with mitomycin C and control bottles (two replicates) were injected with ≥ 0.02 μm seawater of the same depth. All four bottles for each depth were then stored at *in*

situ temperatures (approximately 13°C for anoxic experiment and approximately 26°C for the oxic experiment) for 12 hours. At twelve hours another sample was collected for measurement. For each time point measurement, one milliliter from each bottle of either t=0 or t=12 hours sample was placed into a cryovial and microscopy grade paraformaldehyde was added to reach 1% final concentration. Vials were flash frozen in liquid nitrogen and placed in –80°C until processed.

Visualizing virus like particles and microbial cells

Epifluorescent microscopy was used to enumerate bacteria and viral cells before and after mitomycin or control treatments, and the change in bacterial or viral counts was calculated. The samples were thawed and filtered onto two 0.02 µm Anodisc filters. Dried filters were incubated in 2.5x concentration SYBR Gold for 25 minutes in the dark. Excess dye was removed, and filters were mounted on a slide using 50:50 Glycerol/1X TE. Samples were viewed at 1000x under excitation with a X-Cite 120 bulb. Images were captured of fields of view using the Nikon Eclipse 80i. Fluorescing regions were classified as microbial or virus like particles based on size and counted across 20 to 200 fields of view. Pixel to micrometer conversion was calculated using a 10 µm stage micrometer ruler at equal magnification. Sample counts were extrapolated to entire surface area of filter.

Results:

Virus like particles and bacterial cell abundance were compared from before injections and after 12 hours of incubation. In the water collected from the oxic chlorophyll maximum, and injected with filtered seawater of the same depth, microbial cells maintained abundance or slightly grew, with counts being on average 20% higher at 12 hours than at the start of the experiment. Virus like particles decreased with only 40% remaining at 12 hours in the oxic controls. With the

addition of mitomycin C viruses increased to 150% of their original abundance as measured in the starting point sample and microbial cells decreased by 20%.

In the water collected from the anoxic depths both control and mitomycin C treatments showed an increase in virus like particles and a decrease in microbial cells suggesting induction in both the control and the treatment samples. In the control, microbial cells dropped 20%, similar to the mitomycin C treatment of the oxic sample. Viruses in the anoxic control had the greatest increase, reaching 230% of their original counts by the end of 12 hours. In the mitomycin C treatment samples of anoxic water, microbial cells dropped to 40% of their original abundance, and viruses increased to 110% their original abundance.

Discussion:

The oxic sample had lysogenic bacteria that were effectively induced with addition of mitomycin C. Microbial cells increased in the control and decreased in the treatment with mitomycin C samples. Viruses increased by 50% with the induction of mitomycin C, and microbial cells decreased suggesting that almost a quarter of cells were lysed producing viral progeny.

In contrast, in the anoxic treatment, the increase in viral cells and reduction of microbial cells occurred in both control and treatment. The control results typically show the effect of the experimental protocol on the population. Within the control group for the anoxic microbial community there was an equivalent drop in microbial cells as in the mitomycin C treatment of the oxic microbial community, but over twice the increase in virus like particles. This suggests that the sampling of cells by passing them through the Niskin port into a bottle, filtering and concentrating bacterial cells and resuspending them could have sufficiently induced lysogenic bacteria to excise prophage sequences and begin reproducing them within the bacterial cells. An alternative explanation could be that lytic infections were already abundant within this

community and the experimental time of 12 hours was sufficient to see the release of their viral progeny. With either scenario, after 12 hours there was in increase in virus like particles and a decrease in the microbial cells.

With the addition of mitomycin C in the anoxic community we see an increase in the extracellular virus like particles and a reduction in microbial cells. However, it is the greatest reduction of microbial cells and the smallest production of viral cells in the experiment minus the oxic control. This means that as compared to the control this is a reduction in both viruses and microbial cells. It is possible that mitomycin C is too damaging an induction tool for anoxic communities and caused sufficient death of microbial cells that could not produce as much viable phage before their death.

With such large increases in virus like particles in our control, these conclusions are hard to confirm and suggest *in situ* inductions may be necessary for anoxic prophage, or an alternate way to sample or concentrate cells that does not induce prophage, in oxygen deficient zones. However, this high production of extra cellular virus like particles in the anoxic community suggests the abundance of phage capable of reproduction whether from a lysogenic origin or lytic one. It is possible that induction of phage was experimental be it from either increased concentrations of host, mechanical stress or exposure to oxygen, all of which were either inevitable or extremely difficult to avoid in processing.

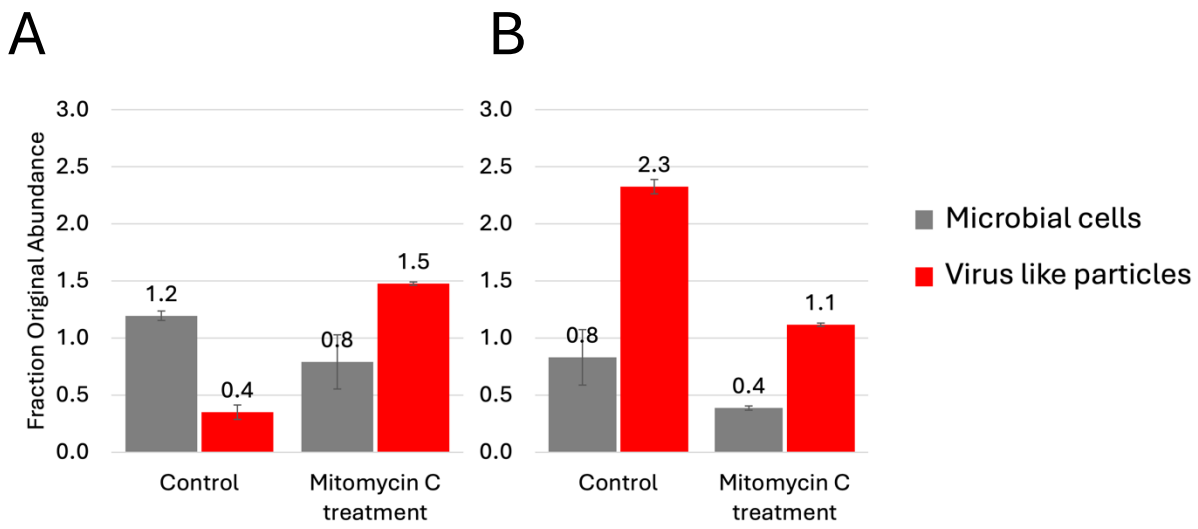


Figure 1. Change in virus like particle and microbial cell abundances as measured through epifluorescent microscopy after 12 hours of incubation of **(A)** samples taken from the oxic chlorophyll maximum and **(B)** samples taken from the anoxic chlorophyll maximum.

Table 1: Water properties at sampling depth for experimental microbial communities.

Depth	Temperature (°C)	Salinity (PSU)	Fluorescence (µg/L)	Oxygen (µmol/kg)
10m	25.8	34.7	0.33	199.1
130m	12.9	34.8	0.09	0.5

References

1. Suttle, C. The virosphere: the greatest biological diversity on Earth and driver of global processes. *Environ. Microbiol.* **7**, 481–482 (2005).

2. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284 (2005).

3. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* **117**, 5–16 (2006).

4. Villarreal, L. P. How viruses shape the tree of life. *Futur. Virol.* **1**, 587–595 (2006).

5. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. *Biol. Direct.* **1**, 29–29 (2006).

6. Gil, J. F. *et al.* Viruses in extreme environments, current overview, and biotechnological potential. *Viruses* **13**, 81 (2021).

7. Roux, S. & Emerson, J. B. Diversity in the soil virosphere: To infinity and beyond? *Trends Microbiol.* **30**, 1025–1035 (2022).

8. Zhong, Z.-P. *et al.* Viral ecogenomics of Arctic cryopeg brine and sea ice. *mSystems* **5**, e00246-20 (2020).

9. Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T. & Rohwer, F. Phage community dynamics in hot springs. *Appl. Environ. Microbiol.* **70**, 1633–1640 (2004).

10. Castelán-Sánchez, H. G. *et al.* Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. *Mar. Genom.*. **46**, 16–28 (2019).

11. SPENCER, R. A Marine Bacteriophage. *Nature* **175**, 690–691 (1955).

12. Baross, J. A., Liston, J. & Morita, R. Y. Incidence of *Vibrio parahaemolyticus* bacteriophages and other vibrio bacteriophages in marine samples. *Appl. Environ. Microbiol.* **36**, 492–499 (1978).

13. Spencer, R. Indigenous marine bacteriophages. *J. Bacteriol.* **79**, 614–614 (1960).

14. Wiebe, W. J. & Liston, J. Isolation and characterization of a marine bacteriophage. *Mar. Biol.* **1**, 244–249 (1968).

15. Torrella, F. & Morita, R. Y. 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 (1979).

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

17. Suttle, C. A. Enumeration and isolation of viruses. *Meth. Aquat. Microb. Ecol.* 121 (1993).

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

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

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

21. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).

22. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *Plos. Biol.* **3**, e144 (2005).

23. Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. & Hodson, R. E. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. *Appl Environ Microb* **68**, 1576–1584 (2002).

24. Schwalbach, M., Hewson, I. & Fuhrman, J. Viral effects on bacterial community composition in marine plankton microcosms. *Aquat. Microb. Ecol.* **34**, 117–127 (2004).

25. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea: Viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **49**, 781–788 (1999).

26. Culley, A., Lang, A. & Suttle, C. Metagenomic analysis of coastal RNA virus communities. *Science* **312**, 1795–1798 (2006).

27. Brussaard, C. P. *et al.* Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J* **2**, 575–8 (2008).

28. Brussaard, C. P. Viral control of phytoplankton populations—a review. *J. Euk. Microbiol* **51**, 125–138 (2004).

29. Hurwitz, B. L. & Sullivan, M. B. The Pacific Ocean Virome (POV): A marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **8**, e57355 (2013).

30. Coordinators, T. O. *et al.* Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).

31. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

32. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).

33. Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **4**, 470–483 (2009).

34. Brum, J. R. *et al.* Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).

35. Gregory, A. C. *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109-1123.e14 (2019).

36. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470 (2016).

37. Laber, C. P. *et al. Coccolithovirus* facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* **3**, 537–547 (2018).

38. Tréguer, P. *et al.* Influence of diatom diversity on the ocean biological carbon pump. *Nat. Geosci.* **11**, 27–37 (2018).

39. Agusti, S. *et al.* Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat. Commun.* **6**, 7608 (2015).

40. Broerse, A. T. C., Ziveri, P., Hinte, J. E. van & Honjo, S. Coccolithophore export production, species composition, and coccolith-CaCO3 fluxes in the NE Atlantic (34°N21°W and 48°N21°W). *Deep Sea Res. Part II: Top. Stud. Oceanogr.* **47**, 1877–1905 (2000).

41. Steinberg, D. K. & Landry, M. R. Zooplankton and the ocean carbon cycle. *Annu. Rev. Mar. Sci.* **9**, 413–444 (2014).

42. Wilson, W. H., Joint, I. R., Carr, N. G. & Mann, N. H. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl. Environ. Microb.* **59**, 3736–3743 (1993).

43. Dekel-Bird, N. P. *et al.* Evolutionary relationships of T7-like cyanopodoviruses. *Environ. Microbiol.* **15**, 1476–1491 (2013).

44. Dekel-Bird, N. P., Sabehi, G., Mosevitzky, B. & Lindell, D. Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. *Environ. Microbiol.* **17**, 1286–1299 (2014).

45. Maidanik, I. *et al.* Cyanophages from a less virulent clade dominate over their sister clade in global oceans. *ISME J* 1–12 (2022) doi:10.1038/s41396-022-01259-y.

46. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).

47. Crummett, L. T., Puxty, R. J., Weihe, C., Marston, M. F. & Martiny, J. B. H. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**, 219–229 (2016).

48. Fuchsman, C. A., Carlson, M. C. G., Prieto, D. G., Hays, M. D. & Rocap, G. Cyanophage host-derived genes reflect contrasting selective pressures with depth in the oxic and anoxic water column of the Eastern Tropical North Pacific. *Environ. Microbiol.* **23**, 2782–2800 (2021).

49. Luo, X.-Q. *et al.* Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* **10**, 190 (2022).

50. Zhong, Z.-P. *et al.* Viral potential to modulate microbial methane metabolism varies by habitat. *Nat. Commun.* **15**, 1857 (2024).

51. Gazitúa, M. C. *et al.* Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J* **15**, 981–998 (2021).

52. Howard-Varona, C. *et al.* Phage-specific metabolic reprogramming of virocells. *ISME J* **14**, 881–895 (2020).

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

54. Vik, D. *et al.* Genome-resolved viral ecology in a marine oxygen minimum zone. *Environ. Microbiol.* **23**, 2858–2874 (2021).

55. Jurgensen, S. K. *et al.* Viral community analysis in a marine oxygen minimum zone indicates increased potential for viral manipulation of microbial physiological state. *ISME J* **16**, 972–982 (2022).

56. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* **4**, e08490 (2015).

57. Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J. & Crosby, L. Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. *mSystems* **3**, e00055-18 (2018).

58. Tominaga, K., Morimoto, D., Nishimura, Y., Ogata, H. & Yoshida, T. In silico prediction of virus-host interactions for marine Bacteroidetes with the use of metagenome-assembled genomes. *Front. Microbiol.* **11**, 738 (2020).

59. Roux, S. *et al.* iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLOS Biol.* **21**, e3002083 (2023).

60. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).

61. George, N. A. & Hug, L. A. CRISPR-resolved virus-host interactions in a municipal landfill include non-specific viruses, hyper-targeted viral populations, and interviral conflicts. *Sci. Rep.* **13**, 5611 (2023).

62. Galiez, C., Siebert, M., Enault, F., Vincent, J. & Söding, J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113–3114 (2017).

63. Wang, W. *et al.* A network-based integrated framework for predicting virus–prokaryote interactions. *NAR Genom. Bioinform.* **2**, lqaa044 (2020).

64. Avrani, S., Schwartz, D. A. & Lindell, D. Virus-host swinging party in the oceans. *Mob. Genet. Elem.* **2**, 88–95 (2014).

65. Borodovich, T., Shkoporov, A. N., Ross, R. P. & Hill, C. Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterol. Rep.* **10**, goac012 (2022).

66. Labonté, J. M. *et al.* Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *Isme J* **9**, 2386–2399 (2015).

67. Roux, S. *et al.* Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* **3**, (2014).

68. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

69. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

70. Marbouty, M., Thierry, A., Millot, G. A. & Koszul, R. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. *Elife* **10**, e60608 (2021).

71. Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-level deconvolution of metagenome assemblies with Hi-C–based contact probability maps. *G3 Genes Genom. Genet.* **4**, 1339–1346 (2014).

72. Bickhart, D. M. *et al.* Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation. *Genome Biol.* **20**, 153 (2019).

73. Stalder, T., Press, M. O., Sullivan, S., Liachko, I. & Top, E. M. Linking the resistome and plasmidome to the microbiome. *Isme J* **13**, 2437–2446 (2019).

74. Børsheim, K. Y. Native marine bacteriophages. *FEMS Microbiol. Lett.* **102**, 141–159 (1993).

75. Hyman, P. & Abedon, S. T. Chapter 7: Bacteriophage host range and bacterial resistance. *Adv. Appl. Microbiol.* **70**, 217–248 (2010).

76. Hanson, C. A., Marston, M. F. & Martiny, J. B. H. Biogeographic variation in host range phenotypes and taxonomic composition of marine cyanophage isolates. *Front. Microbiol.* **7**, 983 (2016).

77. Stocker, R. Marine microbes see a sea of gradients. *Science* **338**, 628–633 (2012).

78. Ross, A., Ward, S. & Hyman, P. More is better: Selecting for broad host range bacteriophages. *Front. Microbiol.* **7**, 1352 (2016).

79. Crill, W. D., Wichman, H. A. & Bull, J. J. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* **154**, 27–37 (2000).

80. Elena, S. F., Agudelo-Romero, P. & Lalić, J. The evolution of viruses in multi-host fitness landscapes. *Open Virol. J* **3**, 1–6 (2009).

81. Vos, M., Birkett, P. J., Birch, E., Griffiths, R. I. & Buckling, A. Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* **325**, 833–833 (2009).

82. Rathwell, C., McKay, C. & Rocap, G. Hi-C assembled genomes of estuarine populations reveal virus-microbe associations and a broad interaction range of a cyanophage. *bioRxiv* 2023.12.06.570405 (2023) doi:10.1101/2023.12.06.570405.

83. Wu, R. *et al.* Hi-C metagenome sequencing reveals soil phage–host interactions. *Nat. Commun.* **14**, 7666 (2023).

84. Ivanova, V. *et al.* Hi-C metagenomics in the ICU: Exploring clinically relevant features of gut microbiome in chronically critically ill patients. *Front. Microbiol.* **12**, 770323 (2022).

85. DeMaere, M. Z. *et al.* Metagenomic Hi-C of a healthy human fecal microbiome transplant donor. *Microbiol. Resour. Announc.* **9**, e01523-19 (2020).

86. Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–206 (1998).

87. Siegel, D. A. *et al.* Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochem. Cy.* **28**, 181–196 (2014).

88. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).

89. Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W. & Weitz, J. S. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat. Rev. Microbiol.* **12**, 519–528 (2014).

90. Yamada, Y., Tomaru, Y., Fukuda, H. & Nagata, T. Aggregate formation during the viral lysis of a marine diatom. *Frontiers Mar. Sci.* **5**, 167 (2018).

91. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).

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

93. Brito, I. L. Examining horizontal gene transfer in microbial communities. *Nat. Rev. Microbiol.* **19**, 442–453 (2021).

94. Sheyn, U. *et al.* Expression profiling of host and virus during a coccolithophore bloom provides insights into the role of viral infection in promoting carbon export. *ISME J.* **12**, 704–713 (2018).

95. Rosenwasser, S. *et al.* Rewiring host lipid metabolism by large viruses determines the fate of *Emiliania huxleyi*, a bloom-forming alga in the ocean. *The Plant Cell* **26**, 2689–2707 (2014).

96. Huang, S., Sun, Y., Zhang, S. & Long, L. Temporal transcriptomes of a marine cyanopodovirus and its Synechococcus host during infection. *MicrobiologyOpen* **10**, e1150 (2021).

97. Carlson, Michael. C. G. *et al.* Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nat. Microbiol.* **7**, 570–580 (2022).

98. Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* **12**, 3035–3056 (2010).

99. Matteson, A. R. *et al.* High abundances of cyanomyoviruses in marine ecosystems demonstrate ecological relevance. *FEMS Microbiol. Ecol.* **84**, 223–234 (2013).

100. Tully, B. J., Sachdeva, R., Graham, E. D. & Heidelberg, J. F. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* **5**, e3558 (2017).

101. Tully, B. J., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data.* **5**, 170203 (2018).

102. Royo-Llonch, M. *et al.* Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat. Microbiol.* **6**, 1561–1574 (2021).

103. Alneberg, J. *et al.* Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. *Microbiome* **6**, 173 (2018).

104. Bertagnolli, A. D., Padilla, C. C., Glass, J. B., Thamdrup, B. & Stewart, F. J. Metabolic potential and *in situ* activity of marine Marinimicrobia bacteria in an anoxic water column. *Environ Microbiol* **19**, 4392–4416 (2017).

105. Reji, L. & Francis, C. A. Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J* **14**, 2105–2115 (2020).

106. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.*. **3**, 804–813 (2018).

107. Vossenberg, J. van de *et al.* The metagenome of the marine anammox bacterium '*Candidatus Scalindua profunda*' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ. Microbiol.* **15**, 1275–1289 (2013).

108. Plominsky, A. M. *et al.* Metabolic potential and in situ transcriptomic profiles of previously uncharacterized key microbial groups involved in coupled carbon, nitrogen and sulfur cycling in anoxic marine zones. *Environ. Microbiol.* **20**, 2727–2742 (2018).

109. Ganesh, S. *et al.* Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria. *ISME J* **12**, 2706–2722 (2018).

110. Tsementzi, D. *et al.* SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* **536**, 179–183 (2016).

111. Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol*. **38**, 1079–1086 (2020).

112. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

113. Beitel, C. W. *et al.* Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* **2**, e415 (2014).

114. Marbouty, M. *et al.* Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* **3**, e03318 (2014).

115. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

116. Sanyal, A., Lajoie, B., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).

117. Stewart, R. D. *et al.* Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **9**, 870 (2018).

118. Bickhart, D. M. *et al.* Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat. Biotechnol.* 1–9 (2022) doi:10.1038/s41587-021-01130-z.

119. Waite, D. W. *et al.* Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* **70**, 5972–6016 (2020).

120. Wiegand, S. *et al.* Cultivation and functional characterization of 79 planctomycetes uncovers their unique biology. *Nat. Microbiol.* **5**, 126–140 (2020).

121. Wiegand, S., Jogler, M. & Jogler, C. On the maverick Planctomycetes. *FEMS Microbiol. Rev.* **42**, 739–760 (2018).

122. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* **1**, e323 (2021).

123. Thompson, L. R., Zeng, Q. & Chisholm, S. W. Gene expression patterns during light and dark infection of *Prochlorococcus* by cyanophage. *PLoS One* **11**, e0165375 (2016).

124. Nguyen, A. A. *et al.* CpeT is the phycoerythrobilin lyase for Cys-165 on β-phycoerythrin from *Fremyella diplosiphon* and the chaperone-like protein CpeZ greatly improves its activity. *BBA - Bioenergetics* **1861**, 148284 (2020).

125. Chain, P. S. G. *et al.* Genome project standards in a new era of sequencing. *Science* **326**, 236–237 (2009).

126. Boedeker, C. *et al.* Determining the bacterial cell biology of Planctomycetes. *Nat. Commun.* **8**, 14853 (2017).

127. Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z. & Ettema, T. J. G. Innovations to culturing the uncultured microbial majority. *Nat. Rev. Microbiol.* **19**, 225–240 (2021).

128. Wigington, C. H. *et al.* Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.* **1**, 15024 (2016).

129. Hershey, A. D., Dixon, J. & Chase, M. Nuceleic acid economy in bacteria infected with bacteriophage T2. *J. Gen. Physiology* **36**, 777–789 (1953).

130. Mackinder, L. C. M. *et al.* A unicellular algal virus, *Emiliania huxleyi virus 86*, exploits an animal-like infection strategy. *J. Gen. Virol.* **90**, 2306–2316 (2009).

131. Zborowsky, S. & Lindell, D. Resistance in marine cyanobacteria differs against specialist and generalist cyanophages. *Proc. National Acad. Sci.* **116**, 16899–16908 (2019).

132. Schneiker, S. *et al.* Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat. Biotechnol.* **24**, 997–1004 (2006).

133. Correa, A. M. S. *et al.* Revisiting the rules of life for viruses of microorganisms. *Nat. Rev. Microbiol.* **19**, 501–513 (2021).

134. Brown, J. M. *et al.* Single cell genomics reveals viruses consumed by marine protists. *Front Microbiol* **11**, 524828 (2020).

135. Devol, A., Ruef, W., Emerson, S. & Newton, J. *In situ* and remote monitoring of water quality in Puget Sound: the ORCA time-series. *2006 IEEE US EU Baltic Int. Sympos.* 1–6 (2006) doi:10.1109/baltic.2006.7266156.

136. Noble, R. & Fuhrman, J. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* **14**, 113–118 (1998).

137. Chen, F., Lu, J., Binder, B. J., Liu, Y. & Hodson, R. E. Application of digital image analysis and flow cytometry to enumerate marine viruses stained with SYBR Gold. *Appl. Environ. Microb.* **67**, 539–545 (2001).

138. Press, M. O. *et al.* Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *bioRxiv* 198713 (2017) doi:10.1101/198713.

139. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

140. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).

141. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

142. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119–119 (2010).

143. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

144. Du, Y. & Sun, F. HiCBin: binning metagenomic contigs and recovering metagenome-assembled genomes using Hi-C contact maps. *Genome. Biol.* **23**, 63 (2022).

145. Du, Y., Laperriere, S. M., Fuhrman, J. & Sun, F. Normalizing metagenomic Hi-C data and detecting spurious contacts using zero-inflated negative binomial regression. *J. Comput. Biol.* **29**, 106–120 (2022).

146. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* **7**, 11257 (2016).

147. Ashtiani, M., Mirzaie, M. & Jafari, M. CINNA: an R/CRAN package to decipher Central Informative Nodes in Network Analysis. *Bioinformatics* **35**, 1436–1437 (2018).

148. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

149. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).

150. Klemetsen, T. *et al.* The MAR databases: development and implementation of databases specific for marine metagenomics. *Nucleic Acids Res.* **46**, D692–D699 (2018).

151. Keeling, P. J. *et al.* The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

152. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

153. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).

154. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).

155. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, gkaa913- (2020).

156. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421–421 (2009).

157. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).

158. Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* **11**, 2500 (2020).

159. Darriba, D. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol* **37**, 291–294 (2020).

160. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

161. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, gkab301- (2021).

162. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

163. Riser, S. C. & Johnson, K. S. Net production of oxygen in the subtropical ocean. *Nature* **451**, 323–325 (2008).

164. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).

165. Dunigan, D., Fitzgerald, L. & Etten, V. J. Phycodnaviruses: a peek at genetic diversity. (2006).

166. Scola, B. L. *et al.* A giant virus in amoebae. *Science* **299**, 2033–2033 (2003).

167. Claverie, J.-M. & Abergel, C. Mimiviridae: An expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic eukaryotes. *Viruses* **10**, 506 (2018).

168. Kaneko, H. *et al.* Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience* **24**, 102002 (2021).

169. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).

170. Camargo, A. P. *et al.* Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* 1–10 (2023) doi:10.1038/s41587-023-01953-y.

171. Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).

172. Meijenfeldt, F. A. B. von, Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* **20**, 217 (2019).

173. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).

174. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).

175. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

176. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).

177. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

178. Uritskiy, G. *et al.* Accurate viral genome reconstruction and host assignment with proximity-ligation sequencing. *bioRxiv* 2021.06.14.448389 (2021) doi:10.1101/2021.06.14.448389.

179. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–403 (2004).

180. Moreau, H. *et al.* Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J. Virol.* **84**, 12555–12563 (2010).

181. Hiraoka, S. *et al.* Diverse DNA modification in marine prokaryotic and viral communities. *Nucleic Acids Res.* **50**, gkab1292- (2022).

182. Bachy, C. *et al.* Viruses infecting a warm water picoeukaryote shed light on spatial co-occurrence dynamics of marine viruses and their hosts. *ISME J.* **15**, 3129–3147 (2021).

183. Yutin, N. *et al.* Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* **3**, 38–46 (2018).

184. Shkoporov, A. N. *et al.* ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 4781 (2018).

185. Ramos-Barbero, M. D. *et al.* Characterization of crAss-like phage isolates highlights Crassvirales genetic heterogeneity and worldwide distribution. *Nat. Commun.* **14**, 4295 (2023).

186. Yutin, N. *et al.* Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features. *Nat. Commun.* **12**, 1044 (2021).

187. Choi, A., Kang, I., Yang, S.-J. & Cho, J.-C. Complete genome sequence of bacteriophage P8625, the first lytic phage that infects Verrucomicrobia. *Stand. Genom. Sci.* **10**, 96 (2015).

188. DeLong, J. P., Etten, J. L. V., Al-Ameeli, Z., Agarkova, I. V. & Dunigan, D. D. The consumption of viruses returns energy to food chains. *Proc. Natl. Acad. Sci.* **120**, e2215000120 (2023).

189. Cho, A. *et al.* Genomic analyses of Symbiomonas scintillans show no evidence for endosymbiotic bacteria but does reveal the presence of giant viruses. *PLoS Genet.* **20**, e1011218 (2024).

190. Husnik, F. *et al.* Bacterial and archaeal symbioses with protists. *Curr. Biol.* **31**, R862–R877 (2021).

191. Petroni, G., Spring, S., Schleifer, K.-H., Verni, F. & Rosati, G. Defensive extrusive ectosymbionts of Euplotidium (Ciliophora) that contain microtubule-like structures are bacteria related to Verrucomicrobia. *Proc. Natl. Acad. Sci.* **97**, 1813–1817 (2000).

192. Wernegreen, J. J. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann. N. York Acad. Sci.* **1360**, 16–35 (2015).

193. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).

194. Arthofer, P., Delafont, V., Willemsen, A., Panhölzl, F. & Horn, M. Defensive symbiosis against giant viruses in amoebae. *Proc. Natl. Acad. Sci.* **119**, e2205856119 (2022).

195. Bertagnolli, A. D. & Stewart, F. J. Microbial niches in marine oxygen minimum zones. *Nat. Rev. Microbiol.* **16**, 723–729 (2018).

196. DeVries, T., Deutsch, C., Rafter, P. A. & Primeau, F. Marine denitrification rates determined from a global 3-D inverse model. *Biogeosciences* **10**, 2481–2496 (2013).

197. Lam, P. & Kuypers, M. M. M. Microbial nitrogen cycling processes in Oxygen Minimum Zones. *Annu. Rev. Mar. Sci.* **3**, 317–345 (2011).

198. Deutsch, C., Brix, H., Ito, T., Frenzel, H. & Thompson, L. Climate-forced variability of ocean hypoxia. *Science* **333**, 336–339 (2011).

199. Horak, R. E. A., Ruef, W., Ward, B. B. & Devol, A. H. Expansion of denitrification and anoxia in the Eastern Tropical North Pacific from 1972 to 2012. *Geophys. Res. Lett.* **43**, 5252–5260 (2016).

200. Graf, D. R. H., Jones, C. M. & Hallin, S. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. *PLoS One* **9**, e114118 (2014).

201. Kahle, M., Appelgren, S., Elofsson, A., Carroni, M. & Ädelroth, P. Insights into the structure-function relationship of the NorQ/NorD chaperones from *Paracoccus denitrificans* reveal shared principles of interacting MoxR AAA+/VWA domain proteins. *BMC Biol.* **21**, 47 (2023).

202. Bianchi, D., Weber, T. S., Kiko, R. & Deutsch, C. Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nat. Geosci.* **11**, 263–268 (2018).

203. Helen, D., Kim, H., Tytgat, B. & Anne, W. Highly diverse *nirK* genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics* **17**, 155 (2016).

204. Fuchsman, C. A., Devol, A. H., Saunders, J. K., McKay, C. & Rocap, G. Niche partitioning of the N cycling microbial community of an offshore Oxygen Deficient Zone. *Front. Microbiol.* **8**, 2384 (2017).

205. Sun, X. & Ward, B. B. Novel metagenome-assembled genomes involved in the nitrogen cycle from a Pacific oxygen minimum zone. *ISME Commun.* **1**, 26 (2021).

206. Zhang, I. H. *et al.* Partitioning of the denitrification pathway and other nitrite metabolisms within global oxygen deficient zones. *ISME Commun.* **3**, 76 (2023).

207. Ravishankara, A. R., Daniel, J. S. & Portmann, R. W. Nitrous oxide (N_2O): The dominant ozone-depleting substance emitted in the 21st century. *Science* **326**, 123–125 (2009).

208. Kuypers, M. M. M. *et al.* Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**, 608–611 (2003).

209. Lam, P. *et al.* Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc. Nat. Acad. Sci.* **106**, 4752–4757 (2009).

210. Dalsgaard, T., Thamdrup, B. & Canfield, D. E. Anaerobic ammonium oxidation (anammox) in the marine environment. *Res. Microbiol.* **156**, 457–464 (2005).

211. Dalsgaard, T., Thamdrup, B., Farías, L. & Revsbech, N. P. Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnol. Oceanogr.* **57**, 1331–1346 (2012).

212. Thamdrup, B. *et al.* Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnol. Oceanogr.* **51**, 2145–2156 (2006).

213. Vliet, D. M. van *et al.* The bacterial sulfur cycle in expanding dysoxic and euxinic marine waters. *Environ. Microbiol.* **23**, 2834–2857 (2021).

214. Chronopoulou, P.-M., Shelley, F., Pritchard, W. J., Maanoja, S. T. & Trimmer, M. Origin and fate of methane in the Eastern Tropical North Pacific oxygen minimum zone. *ISME J.* **11**, 1386–1399 (2017).

215. Ferry, J. G. & Lessner, D. J. Methanogenesis in marine sediments. *Ann. N. York Acad. Sci.* **1125**, 147–157 (2008).

216. Sansone, F. J., Popp, B. N., Gasc, A., Graham, A. W. & Rust, T. M. Highly elevated methane in the Eastern Tropical North Pacific and associated isotopically enriched fluxes to the atmosphere. *Geophys. Res. Lett.* **28**, 4567–4570 (2001).

217. Thamdrup, B. *et al.* Anaerobic methane oxidation is an important sink for methane in the ocean's largest oxygen minimum zone. *Limnol. Oceanogr.* **64**, 2569–2585 (2019).

218. Padilla, C. C. *et al.* NC10 bacteria in marine oxygen minimum zones. *ISME J.* **10**, 2067–2071 (2016).

219. Jørgensen, B. B., Findlay, A. J. & Pellerin, A. The biogeochemical sulfur cycle of marine sediments. *Front. Microbiol.* **10**, 849 (2019).

220. Canfield, D. E., Glazer, A. N. & Falkowski, P. G. The evolution and future of Earth's nitrogen cycle. *Science* **330**, 192–196 (2010).

221. Raven, M. R., Keil, R. G. & Webb, S. M. Microbial sulfate reduction and organic sulfur formation in sinking marine particles. *Science* **371**, 178–181 (2021).

222. Sun, X. *et al.* Uncultured Nitrospina-like species are major nitrite oxidizing bacteria in oxygen minimum zones. *ISME J* **13**, 2391–2402 (2019).

223. Babbin, A. R., Keil, R. G., Devol, A. H. & Ward, B. B. Organic matter stoichiometry, flux, and oxygen control nitrogen loss in the ocean. *Science* **344**, 406–408 (2014).

224. Cassman, N. *et al.* Oxygen minimum zones harbour novel viral communities with low diversity. *Environ. Microbiol.* **14**, 3043–3065 (2012).

225. Hwang, Y., Roux, S., Coclet, C., Krause, S. J. E. & Girguis, P. R. Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. *Nat. Microbiol.* **8**, 946–957 (2023).

226. Cheng, Z. *et al.* Virus impacted community adaptation in oligotrophic groundwater environment revealed by Hi-C coupled metagenomic and viromic study. *J. Hazard. Mater.* **458**, 131944 (2023).

227. Moore, L. R. *et al.* Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol. Oceanogr.: Methods* **5**, 353–362 (2007).

228. Antipov, D., Korobeynikov, A., McLean, J. S. & Pevzner, P. A. Hybrid SPAdes : an algorithm for hybrid assembly of short and long reads. *Bioinformatics* **32**, 1009–1015 (2016).

229. Coutinho, F. H. *et al.* RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. *Patterns* **2**, 100274 (2021).

230. Zhang, R. *et al.* SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics* **37**, 3364–3366 (2021).

231. Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A. & Sun, F. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* **45**, 39–53 (2017).

232. Lu, C. *et al.* Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biol.* **19**, 5 (2021).

233. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).

234. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).

235. Nishimura, Y. & Yoshizawa, S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Sci. Data* **9**, 305 (2022).

236. Liu, Y. *et al.* Metagenome-assembled genomes reveal greatly expanded taxonomic and functional diversification of the abundant marine Roseobacter RCA cluster. *Microbiome* **11**, 265 (2023).

237. Hugerth, L. W. *et al.* Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**, 279 (2015).

238. Lidbury, I., Murrell, J. C. & Chen, Y. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *Proc. Natl. Acad. Sci.* **111**, 2710–2715 (2014).

239. Seibel, B. A. & Walsh, P. J. Trimethylamine oxide accumulation in marine animals: relationship to acylglycerol storage. *J. Exp. Biol.* **205**, 297–306 (2002).

240. Lidbury, I. D., Murrell, J. C. & Chen, Y. Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: implications for marine carbon and nitrogen cycling. *ISME J.* **9**, 760–769 (2015).

241. Mesquita, C. P. B. de, Wu, D. & Tringe, S. G. Methyl-based methanogenesis: an ecological and genomic review. *Microbiol. Mol. Biol. Rev.* **87**, e00024-22 (2023).

242. Bojanova, D. P. *et al.* Well-hidden methanogenesis in deep, organic-rich sediments of Guaymas Basin. *ISME J.* **17**, 1828–1838 (2023).

243. Patterson, J. A. & Hespell, R. B. Trimethylamine and methylamine as growth substrates for rumen bacteria and *Methanosarcina barkeri*. *Curr. Microbiol.* **3**, 79–83 (1979).

244. King, G. M. Metabolism of trimethylamine, choline, and glycine betaine by sulfate-reducing and methanogenic bacteria in marine sediments. *Appl. Environ. Microbiol.* **48**, 719–725 (1984).

245. Müller, V., Kozianowski, G., Blaut, M. & Gottschalk, G. Methanogenesis from trimethylamine + H_2 by *Methanosarcina barkeri* is coupled to ATP formation by a chemiosmotic mechanism. *BBA - Bioenerg.* **892**, 207–212 (1987).

246. Widner, B., Fuchsman, C. A., Chang, B. X., Rocap, G. & Mulholland, M. R. Utilization of urea and cyanate in waters overlying and within the Eastern Tropical North Pacific oxygen deficient zone. *FEMS Microbiol. Ecol.* **94**, (2018).

247. Kivenson, V., Paul, B. G. & Valentine, D. L. An ecological basis for dual genetic code expansion in marine Deltaproteobacteria. *Front. Microbiol.* **12**, 680620 (2021).

248. Paul, L., Ferguson, D. J. & Krzycki, J. A. The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *J. Bacteriol.* **182**, 2520–2529 (2000).

249. Borrel, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood–Ljungdahl pathway: An ancient, versatile, and fragile association. *Genome Biol. Evol.* **8**, 1706–1711 (2016).

250. Ragsdale, S. W. Enzymology of the Wood-Ljungdahl pathway of acetogenesis. *Ann. N. York Acad. Sci.* **1125**, 129–36 (2008).

251. Wood, H. G. Life with CO or CO_2 and H_2 as a source of carbon and energy. *FASEB J.* **5**, 156–163 (1991).

252. Ljungdhal, L. G. The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annu. Rev. Microbiol.* **40**, 415–450 (1986).

253. Thauer, R. K. Dissimilatory sulphate reduction with acetate as electron donor. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **298**, 467–471 (1982).

254. Howard-Varona, C. *et al.* Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *ISME J.* **12**, 1605–1618 (2018).

255. Laguna-Castro, M. & Lázaro, E. Propagation of an RNA bacteriophage at low host density leads to a more efficient virus entry. *Front. Virol.* **2**, 858227 (2022).

256. Mruwat, N. *et al.* A single-cell polony method reveals low levels of infected Prochlorococcus in oligotrophic waters despite high cyanophage abundances. *ISME J.* **15**, 41–54 (2021).

257. Kuhn, J. & Campbell, A. The bacteriophage λ attachment site in wild strains of Escherichia coli. *J. Mol. Evol.* **53**, 607–614 (2001).

258. Fogg, P. C. M., Allison, H. E., Saunders, J. R. & McCarthy, A. J. Bacteriophage Lambda: a paradigm revisited. *J. Virol.* **84**, 6876–6879 (2010).

259. Rutkai, E., Dorgai, L., Sirot, R., Yagil, E. & Weisberg, R. A. Analysis of insertion into secondary attachment sites by phage λ and by *int* mutants with altered recombination specificity. *J. Mol. Biol.* **329**, 983–996 (2003).

260. Jonge, P. A. de *et al.* Adsorption sequencing as a rapid method to link environmental bacteriophages to hosts. *iScience* **23**, 101439 (2020).

261. Sakaguchi, Y. *et al.* The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proc. Natl. Acad. Sci.* **102**, 17472–17477 (2005).

262. Łobocka, M. B. *et al.* Genome of bacteriophage P1. *J. Bacteriol.* **186**, 7032–7068 (2004).

263. Heo, Y.-J., Chung, I.-Y., Choi, K. B., Lau, G. W. & Cho, Y.-H. Genome sequence comparison and superinfection between two related *Pseudomonas aeruginosa* phages, D3112 and MP22. *Microbiology* **153**, 2885–2895 (2007).

264. Laganenka, L. *et al.* Quorum sensing and metabolic state of the host control lysogeny-lysis switch of bacteriophage T1. *mBio* **10**, e01884-19 (2019).

265. Zimmerman, A. E. *et al.* Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* **18**, 21–34 (2020).

266. Labonté, J. M. *et al.* Single cell genomics-based analysis of gene content and expression of prophages in a diffuse-flow deep-sea hydrothermal system. *Front. Microbiol.* **10**, 1262 (2019).

267. Williamson, S. J. *et al.* Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME J.* **2**, 1112–1121 (2008).

268. McDaniel, L. & Paul, J. H. Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine *Synechococcus* species. *Appl. Environ. Microbiol.* **71**, 842–850 (2005).

269. Otsuji, N., Sekiguchi, M., Iijima, T. & Takagi, Y. Induction of phage formation in the lysogenic *Escherichia coli K-12* by mitomycin C. *Nature* **184**, 1079–1080 (1959).

270. Tuttle, M. J. & Buchan, A. Lysogeny in the oceans: Lessons from cultivated model systems and a reanalysis of its prevalence. *Environ. Microbiol.* **22**, 4919–4933 (2020).

271. Knowles, B. *et al.* Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).

272. Jiang, S. C. & Paul, J. H. Significance of lysogeny in the marine environment: Studies with isolates and a model of lysogenic phage production. *Microb. Ecol.* **35**, 235–243 (1998).

273. Barnhart, B. J., Cox, S. H. & Jett, J. H. Prophage induction and inactivation by UV light. *J. Virol.* **18**, 950–955 (1976).

274. Weinbauer, M. G., Brettar, I. & Höfle, M. G. Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters. *Limnol. Oceanogr.* **48**, 1457–1465 (2003).

275. Casjens, S. *et al.* A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.* **35**, 490–516 (2000).

276. Casjens, S. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* **49**, 277–300 (2003).

277. Márquez-Artavia, A. *et al.* A suboxic chlorophyll-a maximum persists within the Pacific oxygen minimum zone off Mexico. *Deep Sea Res. Part II: Top Stud. Oceanogr.* **169**, 104686 (2019).

278. Wright, J. J., Konwar, K. M. & Hallam, S. J. Microbial ecology of expanding oxygen minimum zones. *Nat. Rev. Microbiol.* **10**, 381–394 (2012).

279. Little, J. W. Phages. 37–54 (2020) doi:10.1128/9781555816506.ch3.

280. Jiang, S. & Paul, J. Occurrence of lysogenic bacteria in marine microbial communities as determined by prophage induction. *Mar. Ecol. Prog. Ser.* **142**, 27–38 (1996).

281. Margolskee, A., Frenzel, H., Emerson, S. & Deutsch, C. Ventilation pathways for the North Pacific oxygen deficient zone. *Glob. Biogeochem. Cycles* **33**, 875–890 (2019).

282. Ji, Q., Babbin, A. R., Jayakumar, A., Oleynik, S. & Ward, B. B. Nitrous oxide production by nitrification and denitrification in the Eastern Tropical South Pacific oxygen minimum zone. *Geophys. Res. Lett.* **42**, 10,755-10,764 (2015).

283. Weinbauer, M. G. & Rassoulzadegan, F. Are viruses driving microbial diversification and diversity? *Environ Microbiol* **6**, 1–11 (2004).

284. Schulz, F. *et al.* Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436 (2020).

285. Ren, J. *et al.* Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77 (2020).

286. Zhao, J. *et al.* Novel viral communities potentially assisting in carbon, nitrogen, and sulfur metabolism in the upper slope sediments of Mariana Trench. *mSystems* **7**, e01358-21 (2022).

287. Proctor, L. & Fuhrman, J. Roles of viral infection in organic particle flux. *Mar. Ecol. Prog. Ser.* **69**, 133–142 (1991).

288. Clokie, M. R. J., Blasdel, B. G., Demars, B. O. L. & Sicheritz-Pontén, T. Rethinking phage ecology by rooting it within an established plant framework. *PHAGE* **1**, 121–136 (2020).