Diversity, activity, and adaptations of phage communities in anoxic hydrocarbon-rich marine sediments

Thesis by

Aditi Kalpagam Narayanan

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2025

(Defended April 15, 2025)

Aditi Kalpagam Narayanan
ORCID: 0000-0003-0627-1859

ACKNOWLEDGEMENTS

I must thank, first and foremost, my advisor Victoria Orphan. You let someone who was three years removed from any microbiology and had no background in marine sciences or viruses into your lab. Thank you for making the childhood dream of being a marine biologist come true. I must also thank my committee, Dianne Newman, Joe Parker, and Rob Phillips for their input, enthusiasm, and thought-provoking questions.

In our quest to learn about and ostensibly protect the ocean floor, we repeatedly tear it up with pushcores and robots. I am grateful to the world's largest habitat for its patience and for what it does in keeping the planet livable for the rest of us. May it remind us that unseen and unheard things can make the biggest difference.

To my scientific collaborators for their input and help: Mason Mackey from UC San Diego for teaching me how to use an electron microscope, Igor Antoshechkin from the Caltech Genomics Facility, Andres Collazo from the Biological Imaging Facility, Songye Chen and Mark Ladinsky from the Cryo-EM facility, Yunbin Guan from the Caltech MicroAnalysis Center, Fenfang Wu from Caltech GPS, and Gunter Wegener from the University of Bremen.

To the Orphan Lab, you are an amazing group of people to work with, scientifically or otherwise. Being in an environment where people are collaborative rather than competitive and where they want to see everyone succeed made such a profound difference to my graduate school experience. In no particular order, I wanted to especially thank the following people for their friendship and guidance over the years:

My honorary committee members and good friends Ranjani Murali and Alon Philosof. Listening to your conversations, and eventually developing the confidence to join them, has taught me how to think critically about my research. Plus both of you are always up for a laugh and a rant (not always in that order).

Usha Lingappa, Rodney Tollerson, Kriti Sharma, Eryn Eitel, Magdi Mayr, Dan Utter, and Kelly Luis: you have all been incredible friends and mentors, and talking to you has helped me think harder about what it means to be a socially-engaged scientist. Usha, I am so grateful to that we've had each other to help pull through every existential crisis.

Stephanie Connon, I think you know this by now, but nothing in this thesis or anyone else's happens without you keeping the lab running. I will treasure every conversation on gardening, candlemaking, and cooking.

To the past and present denizens of North Mudd 100 and 102: Rebecca Wipfler, Grayson Chadwick, Hannah Dion-Kirschner, Sujung Lim, Sergio Parra, Calvin Rusley, Makayla Betts, Hannah Way, Phillip Woods, Shaelyn Silverman, Yongzhao Guo, Forrest McCann, James Mullahoo, Renee Wang, Sean Mullin, and Kyle Metcalfe. Thanks for the moral support, for always being a scientific sounding board, and for your legendary group procrastination skills. Special thanks to Rebecca for always telling me where to find the free snacks in the department

and to Hannah DK, Serg, Shae, Makayla, and Yongzhao for the support as we all wrote our theses together.

To Jeremy Schrier, Shep Schroer, Justus Fink, Francisco Martinez-Hernandez, and Madison Dunitz, thank you for your company, your scientific input, and the always-interesting conversation.

To the early labbies: Ranjani, Usha, Gray, Sujung, Daan Speth, Yamini Jangir, Sean Mullin, Kyle Metcalfe, Haley Sapers, Hank Yu, my rotation mentor Connor Skennerton, and John Magyar, you all welcomed a very green first year into your midst. Thank you for making me part of the team from the jump.

Liz Ayala, thank you for making me feel like I belonged here from the moment I set foot on campus, and for me helping navigate the journey both personally and professionally.

To Mauricio, who was always up for a chat and told me on more than one occasion that he knew I could do it and he was rooting for me.

To UAW Local 2478, formerly known as Caltech Grads and Postdocs United. You have shown me that it is possible to make the world better, and that every single one of us can and must play an active role. Thank you to Sam, Nadia, Jasmine, Jessica, Piril, Natasha, Jakob, Jack, Alex VW, Alex C, Sean, Sophie, David, Sim, Rohan, Ranjani, Kriti, Elizabetta, Tom, Mike, and so many others. All of you have given me the community that I yearned for. Special shoutout to comms crew.

To the people who've been along for the whole ride: Lev, Yana, Julian, Sharan, David, and John. Your friendship has been a lifeline. I am so grateful for the camaraderie, and I wish the pandemic hadn't ended our annual camping trip tradition.

A very special thank you to Jamilla Akhund-Zade and Raghu Sathyamurthy for having been there and done that.

To Pam, Steve, Addy, Sam, Emma, Hope, and Felix, thank you for welcoming me into your family. I met Tom in grad school, and you have all been a part of this process with both of us. Thank you for your support and your steady belief that we would get this done.

To my parents, my sister, my brother-in-law, and my new niece, it's the small things that add up over the course of a lifetime that tell me I am valued, loved, and worthy. My principles have come from your guidance. Thank you for reminding me that I can do this, and also that it's okay if ever a time comes where I can't.

Tom, everyone should be lucky enough to find a life partner like you. When faced with the task of describing what you mean to me, I have finally run out of words.

ABSTRACT

The viruses of the global ocean, especially those infecting prokaryotic taxa, are known to play an important role in maintaining the genetic and taxonomic diversity of their host communities and in the cycling of atmospheric carbon and key nutrients like nitrogen and iron. However, the vast majority of these conclusions are drawn from the surface ocean and upper water column, while the sediments, which constitute one of the largest biomes on earth, are understudied in comparison. Of special interest are areas on the ocean floor where methane and other hydrocarbons are produced and released by geological activity and oxidized by a consortium of archaea and bacteria. Using direct genomic sequencing of the viruses from a variety of simplified sediment-free enrichments of hydrocarbon oxidizers, I compare viral communities sampled from different locations and incubated under a range of temperatures to understand the role these parameters might play in shaping distribution and community structure. I then present the most comprehensive picture thus far of viral diversity and distribution from a methane cold seep and discuss whether the viral assemblages are influenced by the steep geochemical gradients that characterize seep sediments. From these datasets, I propose that viral communities in methane-oxidizing sediments are tailored specifically to the physical constraints of the sediment matrix rather than to the dominant members of the cellular community or to other physicochemical parameters such as temperature, sampling location, or depth below the seafloor. I then outline the development of two methods, stable-isotope probing coupled to nanoSIMS and biorthogonal non-canonical amino acid tagging, to work in heterogenous sediment virus samples rather than the liquid pure cultures on which they had previously relied. Implementation of these methods, which allow us to temporally constrain viral production and virus-influenced nutrient flow, resulted in the hypothesis that viral production likely responds to shifts in the major metabolic processes within an ecosystem and may influence cellular community composition.

PUBLISHED CONTENT AND CONTRIBUTIONS

Narayanan, A.K., et al. (*In prep*). Viral communities from long-term anaerobic alkane-oxidizing enrichments may promote cell surface adhesion.

A.K.N. participated in the conception of the project, performed all DNA extraction and sequencing library preparation, analyzed the data, and wrote the manuscript.

Philosof, A., Narayanan, A.K., et al. (*In prep*). Phage infection networks spanning distant taxonomic groups are a defining feature of deep marine methane seeps.

A.K.N. participated in the conception of the project; performed all sample preparation, DNA extractions, and sequencing library preparation; and analyzed the cellular metagenome data.

Narayanan, A.K.*, Philosof, A.* (2020). Opti-Prep density separation of viruses for metagenomics V.1. *Protocols.io*, doi: dx.doi.org/10.17504/protocols.io.261geobqwl47/v1

A.K.N. and A.P. developed the method jointly, and A.K.N. tested it extensively and wrote the protocol.

Narayanan, A.K.*, Philosof, A.* (2024). SYBR Gold staining of viral-like particles on Anodiscs. *Protocols.io*, doi: dx.doi.org/10.17504/protocols.io.8epv5rd85g1b/v1

A.K.N. and A.P. developed the method jointly. A.K.N. wrote the protocol.

Narayanan, A.K.*, Philosof, A.* (2025). Viral nanoSIMS sample preparation. *Protocols.io*, doi: dx.doi.org/10.17504/protocols.io.8epv52dp5v1b/v1

A.K.N. and A.P. developed the method jointly. A.K.N. wrote the protocol.

Narayanan, A.K.*, Philosof, A.* (2025). Viral extraction from Sediment V.1. *Protocols.io*, doi: dx.doi.org/10.17504/protocols.io.6qpvr97zbvmk/v1

A.K.N. and A.P. developed and tested the method jointly. A.K.N. wrote the protocol.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

NOMENCLATURE

16S ribosomal RNA (16S rRNA): A component of the prokaryotic ribosome. The gene encoding the 16S rRNA is an excellent taxonomic marker; as a result, sequencing this section alone from a sample is a fast and effective way to get a general sense of which organisms are present in a sample.

ANME: Anaerobic Methanotroph. Specifically refers to known archaeal taxa that oxidize (See *Oxidation* below) methane for energy. Often perform this process in partnership with sulfate or nitrate reducing bacteria (see *SRB*), though specific taxa can grow without a bacterial partner.

Chemosynthesis: A form of autotrophy involving the generation of food for growth and storage using chemical compounds as energy sources. Similar to the more well-known process of photosynthesis, which relies on light instead of chemicals.

Chronic: A viral lifestyle in between *lytic* and *lysogenic* (see below) in which the virus slowly reproduces and progeny bud off of the host cell without killing it. Phage genomes may or may not integrate into the host genome.

Enrichment culture: A technique favored in environmental microbiology in which laboratory conditions are made as hospitable as possible to encourage the growth of a target organism, while impeding and/or willfully neglecting the growth of others. This technique is useful since many organisms are resistant to growing in pure culture (see *Pure culture* below).

Hydrothermal vent: Fissures in the seafloor where seawater comes into contact with magma rising from undersea geological activity. Results in superheated water and the release of an impressive array of chemicals, including trace metals.

Isotope: A variation on the canonical form of an element's atoms. Isotopes of an element will have the same number of protons and electrons as the standard atomic form, but will have fewer or more neutrons, resulting in a higher mass while maintaining the same charge.

Lytic: A viral lifestyle in which replication of phage progeny occurs rapidly after infection. Progeny phage then lyse, or break open, their host cell and move on to find a new host.

Lysogenic: A viral lifestyle in which the viral genome integrates itself into the host genome after infection. Lysogenic viral genomes in this state are called *prophages*, and they are replicated as part of the host genome and passed down to daughter cells during cell division. An integrated

lysogenic virus can remove itself from the host genome and switch to a lytic lifestyle if triggered by certain conditions.

Metagenome: The result of extracting and sequencing microbial DNA in bulk from a complex sample like marine sediment or the human gut. The DNA sequences obtained from a metagenome allow us to generate hypotheses about potential functional and metabolic niches of the microbes in the sample based on the genes that are present.

Metagenome-Assembled Genome (MAG): A species-level microbial genome constructed from metagenomic data. See *Metagenome*. Viral MAGs are often called vMAGs.

Methane cold seep: Areas of high methane emissions from the seafloor as a result of high geological activity. Often found in areas of high tectonic activity such as subduction zones and continental margins. Note that methane production in seeps is not exclusively geological and can also be a direct result of biological processes.

Microbe: A biological entity that is too small to be seen with the naked eye. Microbes can be eukaryotic or prokaryotic. Some consider viruses to be microbes, which is why this thesis specifically refers to cellular microbes when discussing non-viral entities.

Oxidation: A chemical reaction that involves the removal of electrons from an atom, resulting in a more positive charge for that atom. Paired with a reduction reaction, which acts as a sink for the electrons removed from the aforementioned atom. Organisms use these Reduction-Oxidation, or *Redox* reactions to generate energy. See *Reduction* below.

Phage: A virus specifically infecting prokaryotes, i.e., bacteria or archaea.

Pseudolysogenic: A form of the lytic or lysogenic phage lifestyle in which the phage infects the host, but neither reproduces immediately nor integrates into the genome. Upon infection, phages may simply wait in stasis if conditions are not right for reproduction or reproduce in sync with the host cell. The phage is likely to switch to a true lytic or lysogenic lifestyle once conditions improve.

Pure culture: A laboratory population of a single species of microorganism, i.e., one growing without any other species present.

Pushcore: A tool used to sample sediment or soil while preserving the vertical stratification of the environment. Consists of a plastic tube that is pushed into the sediment and retrieved, with stoppers on both the bottom and the top preventing sample loss. The term may also apply to the sediment sample that is collected, and not only to the tool.

Reduction: A chemical reaction that involves the addition of electrons from an atom, resulting in a more negative charge for that atom. Paired with an oxidation reaction, which provides the electrons donated to the atom in question. (see *Oxidation* above).

SRB: Sulfate-Reducing Bacteria. There are many bacterial taxa that reduce sulfate (see *Reduction*), using it as a terminal electron acceptor in their metabolic processes. In the context of this thesis, SRB generally refers to the bacteria that live with methanotrophic archaea (see *ANME*) and pair the reduction of sulfate to methane oxidation.

Syntrophy: A form of mutualism in which two cellular microbes are metabolically entangled, usually involving the direct transfer of metabolites from one organism to another. Also called cross-feeding.

Virus: The simplest known biological entity. At their most basic, viruses are pieces of genetic material (DNA or RNA) encapsulated by protein. Viruses do not grow, metabolize, or generally respond to their environment, and rely on the metabolic machinery of their hosts to reproduce. They are closely related and sometimes indistinguishable from selfish genetic elements, existing only to reproduce and transmit themselves even at the expense of their hosts.

VLP: A Virus-Like Particle. Used especially in the context of viral fluorescence microscopy, where the tiny particles visible through the eyepiece are suspected to be viral but could also be vesicles or autofluorescent particles.

Chapter 1

INTRODUCTION

The deep ocean is a remarkable place. It sustains life at pressures that would flatten a human instantly, and the creatures that live there thrive on noxious poisons at temperatures we could not endure. The rapid pace of growth and consumption in the Global North is sowing the seeds for the destruction of an environment we do not value because we cannot understand it. We lay underwater cables through quiet habitats, mine the seafloor for ever more mineral wealth, and send our trash to the deepest areas of the ocean—out of sight, out of mind.

This thesis seeks to move us a step closer to understanding and respecting the abyss for what it is, not what we think we can extract from it. I hope that by adding a little more to our knowledge of the deep ocean, we can come to value how profoundly we are connected to it.

The Ocean Floor

The vast energy-limited sediments of the deep ocean are punctuated by nutrient-rich oases for the creatures that live there. Whalefalls, marine snow, and hydrocarbon-fueled chemosynthesis provide energy for diverse microbial and megafaunal communities. This thesis will focus on the benthic chemosynthetic ecosystems of the methane cold seep and the hydrothermal vent.

Early scientific conceptions of the seafloor were not unlike my own before my foray into marine biology: a desert-like abyss, too far from sunlight, too cold, with too little oxygen, and under too much pressure for life to exist. Dredging expeditions showed that this was not the case as early as the mid-1800s, but these conceptions were truly changed in the 1970s, when scientists from the Woods Hole Oceanographic Institute discovered the first known hydrothermal vent (Ballard, 1977). They found fissures in the seafloor where seawater percolated deep into the Earth's crust, coming into contact with magma rising from undersea volcanic activity. We know now that this process results in an anoxic "chemical soup" superheated as high as 400°C (Lubofsky, 2018) that provides energy to billions upon billions of microbes, which in turn feed multicellular organisms like clams and giant tubeworms. Where surface ocean life is dependent on photosynthesis by marine algae, i.e., harvesting energy from sunlight to produce organic

carbon, the deep ocean depends on chemosynthesis, which relies on chemical energy for the same purpose. The chemicals fueling this ecosystem include metals, hydrocarbons, sulfur compounds, and various types of minerals.

Not long after, in 1984, scientists found the first cold seep (Paull et al., 1984). Unlike the superheated waters in hydrothermal vents, the anoxic methane-rich waters of the cold seep are generally at ambient temperatures (from barely above freezing to about 10˚C, though warmer sites exist). The cold temperature and high pressure sometimes result in the formation of gas hydrates, which are solid complexes of water ice and methane gas. Like hydrothermal vents, seep formation is also the result of geological activity, and they are most often found at continental margins and subduction zones (Barry et al., 1996; Levin et al., 2016; Paull et al., 1984). The methane produced at these cold seeps is usually a result of both the aforementioned geological processes, like the thermal decomposition of organic matter (Schoell, 1988), and biological methanogenesis.

Seabed methane release is estimated at 20Tg of methane per year (Judd, 2004), a noticeable contribution to the atmospheric concentration of this potent greenhouse gas. It would no doubt be higher if not for the microbial processes keeping it in check. Hydrocarbon seepage in marine sediment ecosystems is an important source of energy for microbial and multicellular life in the deep sea. In the case of methane, as methane diffuses or is advected upwards, it is anaerobically oxidized to CO2 by anaerobic methanotrophic archaea (ANME), which couple this reaction to the reduction of naturally occurring sulfate in seawater by a sulfate-reducing bacterial partner (SRB) (**Figure 1.1**). This process occurs in a variety of temperature and depth regimes, including low-temperature cold seeps (Boetius et al., 2000; Orphan et al., 2001), sandy mesophilic sediments (Ruff et al., 2016), and hydrothermally influenced sediments (Holler et al., 2011; Merkel et al., 2013; Schouten et al., 2003; Teske et al., 2002; Wankel et al., 2012). Microbially-mediated anaerobic oxidation of methane (AOM) is thought to consume anywhere from less than 20% to 80% of the methane released from the seafloor (Boetius & Wenzhöfer, 2013). Similar mechanisms underlie the anaerobic oxidation of alkanes like propane and butane in hydrothermal vents (Hahn et al., 2020; Laso-Pérez et al., 2016; Wegener et al., 2022). Such processes are crucial to providing the organic carbon that sustains a diverse ecosystem of

megafauna, including crabs, worms, and vesicomyid clams (Mullin, 2020). While the cellular mechanisms contributing to this globally important process are an area of active research, less is known about the diversity and possible roles of viruses found in these sediments.

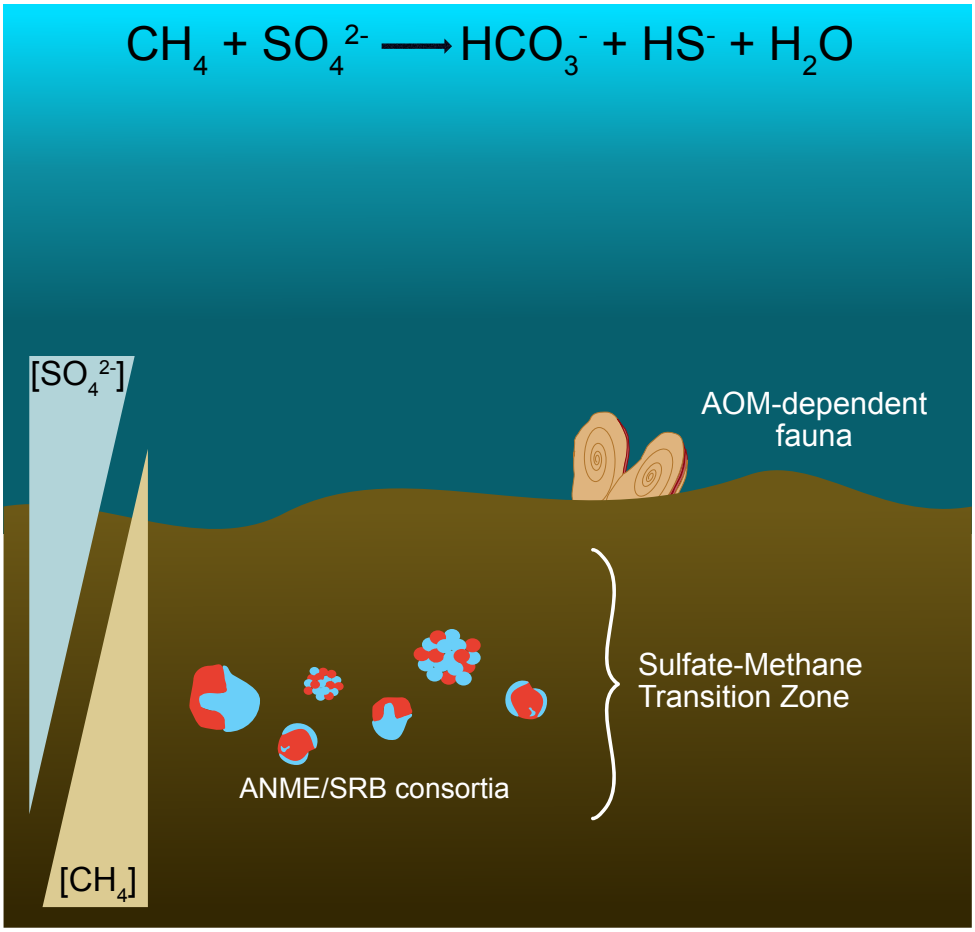

Figure 1.1. Schematic of a methane cold seep. As methane diffuses or is advected upwards from deep in the sediment, it meets an opposing gradient of sulfate diffusing downwards from the water column. In the sulfate-methane transition zone, the oxidation of methane is coupled to the reduction of sulfate by consortia of anaerobic methanotrophic archaea (ANME, shown in red) and sulfate-reducing bacteria (SRB, shown in blue). Similar mechanisms result in the oxidation of longer-chain alkanes in hydrothermal vent systems. These consortia can be found in a variety of morphological arrangements and underpin a thriving megafaunal ecosystem.

What are viruses? Why are they relevant in marine ecology?

A virus is one of the simplest biological entities known. Usually, it consists of genetic material (DNA or RNA) tightly packed into a protein capsid. In some cases, the virus is also enveloped by a lipid layer—it was this property that made the COVID-19 virus so susceptible to simple

soaps and detergents. Viruses do not eat, breathe, or react to their environment[1], leading to questions of whether they can truly be considered alive. They also rely entirely on their hosts to reproduce; upon infection, viruses hijack cellular metabolic processes and divert them towards making even more viruses. Some viruses do this rapidly, multiplying and bursting out of their host cell and killing the host in the process (lytic viruses). Others take their time, some by embedding their genes in the host's chromosome and waiting for the opportune moment to hijack its metabolism (lysogenic viruses), and others by releasing themselves from the host in a nondestructive "budding" fashion, thus prolonging the host's utility as a virus factory (chronic or carrier viruses) (Hobbs & Abedon, 2016; Mäntynen et al., 2021). In the environment, we usually think the greatest threat to an organism is predation by something at a higher trophic level; a shark will eat a sardine and use the energy from the food to grow and hopefully produce offspring. Viruses are predators of a different sort, taking a much more direct path to acquiring the building blocks for reproduction.

Viruses are the most abundant genetic entities in the ocean, with estimated numbers as high as 10^{30} total viral particles, and up to 15 times the number of viral particles as bacteria and archaea (Suttle, 2007). They are known to play a substantial role in community structure, ecology, microbial evolution, and geochemical cycling. In a process known as the viral shunt, viral lysis of autotrophs in the upper ocean results in the cells' conversion to particulate and dissolved organic matter (POM and DOM, respectively), disrupting carbon fixation and short-circuiting carbon sequestration via predation and sinking (Wilhelm & Suttle, 1999) (**Figure 1.2**). However, the resulting POM and DOM may also be taken up by other cells in the vicinity, promoting growth and re-sequestering carbon in cellular biomass. The geochemical importance of viruses reaches beyond the fixation and sequestration of carbon. Viral lysis of cellular hosts releases fixed nitrogen (Shelford et al., 2012) and bioavailable forms of iron (Poorvin et al., 2004), the latter being a limiting nutrient in marine environments. Viruses are also known to play a role in the collapse of cyanobacterial and algal blooms (Beltrami & Carroll, 1994; Bratbak et al., 1993),

[1] The *Acidianus* two-tailed virus finishes morphogenesis outside of its host cell; the two tails in question are formed after exiting the host as long as ambient temperatures are between 75°C and 90°C (Häring, Vestergaard, et al., 2005). When phage particles were stored at 4°C, no tail appeared until the phages were transferred to 75°C. I think you could reasonably argue that this constitutes a response to the environment. Perhaps a more accurate way to phrase this idea is that viruses do not respond transcriptionally, translationally, or metabolically to environmental changes.

which, in addition to recycling nutrients, fundamentally shapes microbial community structure and diversity by re-opening ecological niches. This phenomenon is described by the "kill-the-winner" model (Thingstad & Lignell, 1997), in which the fastest-growing and most abundant organisms are most vulnerable to viral predation. However, in another case, phages actually encouraged the growth of dinoflagellate blooms by killing a bacterial species that was harmful to the dinoflagellate (W. Cai et al., 2011).

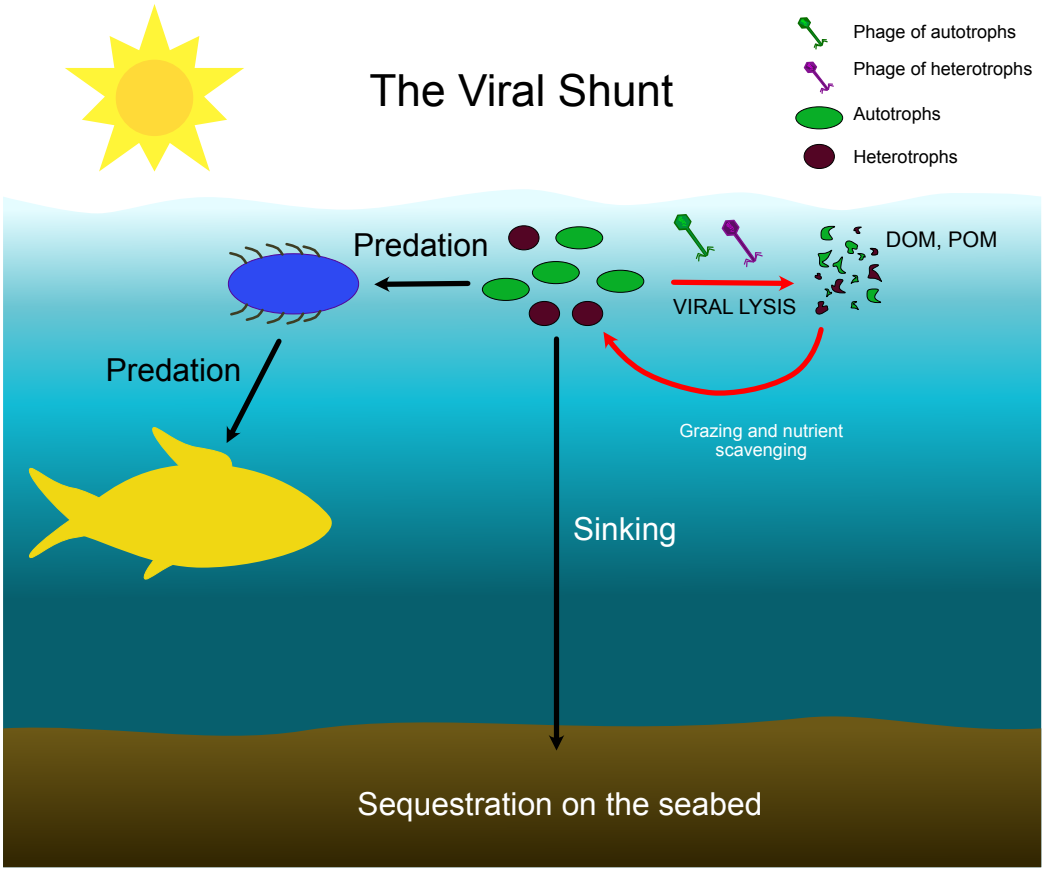

Figure 1.2. Schematic of the viral shunt. In a world without viruses, we might expect photosynthetic cyanobacteria (green ovals) to continue to convert carbon dioxide to sugars using the power of sunlight, thus removing it from the atmosphere and feeding heterotrophic microorganisms (dark red ovals). As they are preyed upon by larger organisms or sink to the ocean floor, the carbon is sequestered either in the sediments or in the bodies of predators. Viral infection and lysis disrupts that process, sending the host biomass into the dissolved or particulate organic matter pool (DOM/POM), preventing sequestration via sinking but allowing other organisms to scavenge nutrients and continue to grow.

Viruses in deep ocean sediments

Much of the existing work in marine virology concerns the upper ocean, while the deep ocean and sediments are generally understudied. It has been estimated that there are 10-1000 times

more viruses in surface and subsurface sediments than in the pelagic zone on a per volume basis (Danovaro et al., 2008), and though this number may decay rapidly (on the order of meters) with depth, the decrease corresponds to a drop in cellular host numbers (Engelhardt et al., 2014; Middelboe et al., 2011). Given the greater abundance of viruses and the relative lack of top-down predation in sediment systems, it is not unreasonable to hypothesize that they might occupy an ecosystem role on par with or larger than that of their pelagic cousins.

There are several axes on which viruses might impact deep sea ecosystems. We might expect that lower energy availability in the deep ocean compared to the surface would result in higher probability of lysogeny; because a virus might not have the ability to reproduce as quickly or abundantly when the host is probably scrounging for nutrients, it might choose to wait until a more opportune time. Finding an appropriate host might also be more difficult in a dense sediment matrix, and lysogenic lifestyles would allow a phage integrated into the host genome to be carried along as the cell divides. Indeed, prior work in the subsurface has shown that viruses are more likely to adopt a lysogenic rather than lytic lifestyle in such environments (Liang et al., 2020; Williamson et al., 2008; Yi et al., 2023). On the other hand, large studies in soils have found a lack of lysogeny-related genes, seemingly contradicting our previous assumptions (Santos-Medellín et al., 2023; Trubl et al., 2018). This phenomenon may be explained by pseudolysogeny (Hazard et al., 2025; Łoś & Węgrzyn, 2012; Ripp & Miller, 1995; Romig & Brodetsky, 1961), in which lytic viral infection is simply arrested until the opportune moment arrives: infection occurs, but phage are not produced, the cell is not lysed, and the viral genome is not integrated into the host genome.

Other roles for viruses pertain to their ability to generate genetic diversity. A suspected ANME virus found in methane seep metagenomic data was observed to contain a diversity generating retroelement (Paul et al., 2015). Viral infection is also thought to have played a role in the diversification of the ANME-1 clade via the horizontal transfer of a gene involved in DNA synthesis, thymidylate synthase *ThyX*, which is non-homologous to the *ThyA* found in other ANME lineages (Laso-Pérez et al., 2023). When cellular doubling times are on the order of months and opportunities for random mutation are thus lower, viruses likely play a key role in increasing genetic diversity in a population.

I also speculate that aggregation in the subsurface could be a part of a host defense strategy. Cells in AOM consortia are surrounded by a thick extracellular matrix that helps them stick to the sediment and to each other; I have wondered if the presence or absence of viruses might result in changes to aggregate morphology. Perhaps aggregates would be easier to break apart in the absence of viruses, resulting in smaller aggregate sizes. As we will see in Chapter 2 of this thesis, viruses are indeed likely to influence cell-cell adhesion, though whether ANME/SRB aggregate morphology is affected is still untested. Viruses might even have a direct influence on alkane degradation rates and sulfur metabolism via the presence of the auxiliary metabolic genes discussed above. Broader examinations of auxiliary metabolic genes in methane-rich environments have shown that viruses can encode parts of the assimilatory sulfate reduction pathway (Kieft et al., 2021; L. Wang et al., 2022) and several other cofactor-associated genes (Z. Li et al., 2021; Peng et al., 2023). Recently, viruses encoding genes exclusive to methane metabolism were found in a variety of methane-rich habitats, including marine and freshwater ecosystems, cow rumen, and permafrost (Zhong et al., 2024).

In sites where the anerobic oxidation of alkanes is an active process, existing examination of the viral community has mostly focused on methane-oxidizing archaea to the exclusion of the sulfate reducing bacteria and other community members, and results are derived from viral contigs mined from bulk metagenomes. This thesis examines the following open questions in marine viral ecology:

1. How diverse are the viruses in alkane-oxidizing communities? Who do they infect, and what can we learn about their adaptations and patterns of distribution in relation to various physico-chemical parameters?
2. What are the possible functions of viruses in alkane-oxidizing communities? What kinds of auxiliary metabolic genes (AMGs) do we find?
3. How do nutrients flow through seep microbes into their viral predators, and on what timescales?

Answering these questions has involved leveraging and refining existing techniques and developing new ones altogether. We used to assume that if we just found the right combination of nutrients, we could isolate and grow any microbe in pure culture. The sequencing revolution showed us how untrue that was; by unveiling the full genetic diversity of the microbial world, we realized our culture-based methods were looking at only a fraction of the microorganisms on

the planet. In short, most organisms are uncultured and possibly indefinitely unculturable because we just do not know what makes them tick. The field of microbial ecology has instead learned how to make discoveries in complex, heterogenous environmental samples without culture-based techniques.

Chapters 2 and 3 answer the first two questions using the power of modern DNA sequencing technology to examine viral gene-level diversity in alkane-degrading laboratory incubations. While pure cultures of alkane-oxidizing organisms continue to evade us, we have successfully created enrichment cultures, i.e., increased in the abundance of the organism of interest in a soup of other taxa. These enrichments are sediment-free after several rounds of dilution in an artificial seawater medium, and removing the sediment also removes some of the methodological and ecological complexities of working with a bottle full of mud. However, we also use laboratory incubations of environmental sediment samples[2], more closely preserving the *in situ* abundances of the members of the microbial community. Our studies greatly expand known viral diversity, including dramatically increasing our catalog of viruses capable of infecting both archaea and bacteria, and show that physicochemical parameters of their environment have a greater impact at the community level than at the individual virus level. Comparisons to water column viral samples illustrate the importance of sediment-focused studies, as we find distinct communities and a possible viral role in mediating surface interactions of microbial cells in sediment environments.

With a firmer grasp on the question of who is present, Chapter 4 then discusses the development of methods to begin quantitatively constraining the impact of viruses on the nutrient flow in the methane seep. Chapter 5 examines the data produced using the techniques from Chapter 4 to constrain the flow of fixed nitrogen into the fraction of viral biomass over time. Our preliminary data suggest that differences in substrate preferences in the host community may result in differences in estimates of the timescale of viral production. The data also suggest that there may be distinct populations whose dominance waxes and wanes successively. Though much is left

[2] For an explanation of shipboard sample collection and processing, see Appendix II.

to be studied, we will examine how these results suggest a distinct role for viral infection in the physiology of microbial hosts and nutrient movement in the deep sea.

A new perspective

Most people with an interest in ecology and environmental science quickly learn the extent to which all things influence all others. But like so many of us, I live in a system that works hard to forget that concept, and my understanding of it was and likely still is superficial at best. Science, when practiced intentionally, is one potential avenue by which to regain our sense of connection. As a surface-dweller, researching and writing this thesis has forced me to walk a mile in the shoes of the average subterranean denizen. I hope I can take you on an abridged version of the journey here.

Chapter 2

GLOBALLY-DISTRIBUTED VIRAL COMMUNITIES MAY PROMOTE CELL SURFACE ADHESION

Aditi K. Narayanan (1), Alon Philosof (1), Ranjani Murali (2), Stephanie A. Connon (3), Gunter Wegener(4), Victoria J. Orphan (1,3)

1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA
2. School of Life Sciences, University of Nevada Las Vegas, USA
3. Division of Geology and Planetary Sciences, California Institute of Technology, Pasadena, California, USA
4. Center for Marine Environmental Sciences, University of Bremen, Germany

ABSTRACT

The microbially-mediated degradation of alkanes on the ocean floor is a well-studied metabolic lifestyle, often involving syntrophic interactions between alkane-oxidizing archaea and sulfate-reducing bacteria. However, the viral component of alkane-rich ecosystems and their associated microbial communities has historically been overlooked. We analyzed the viral community in long-term sediment-free enrichments of alkane-degrading organisms and found that abiotic factors such as incubation temperature had a greater correlation with community composition than with the phylogenetic patterns of individual viral species. Analysis of auxiliary metabolic genes also revealed little to no relative abundance patterns based on temperature or sampling location, but the presence of AMGs involved in heme synthesis pathways common in methane oxidizers hints at a possible viral role in alkane degradation. We also found evidence supporting the presence of a viral genus infecting a number of phyla across both the bacterial and archaeal domains, including one of the sulfate-reducing bacterial partners in the alkane-oxidizing syntrophy. Lastly, we report the presence of *nosD*-like proteins in viruses from sediment-derived systems that are not present in water column datasets; their distribution, genomic context, and lack of canonical *nosD* characteristics suggest an alternate adhesion-related role in sediment communities. The number of viruses obtained from these enrichment cultures and their potential roles in mediating host physiology illustrate the importance of studying the viral component in laboratory and environmental systems.

INTRODUCTION

Hydrocarbon seepage in marine sediment ecosystems is an important source of energy for microbial and multicellular life in the deep sea (Kellermann et al., 2012; Levin et al., 2016). In the case of methane seepage, microbially-mediated anaerobic oxidation of methane (AOM) is a check on atmospheric concentrations of this potent greenhouse gas. As geologically- and biologically-produced methane diffuses or is advected upwards through marine sediments, methane oxidation to CO_2 is carried out by anaerobic methanotrophic archaea (ANME), which couple this reaction to sulfate reduction via a sulfate-reducing bacterial partner (SRB) (Boetius et al., 2000; Orphan et al., 2001). AOM is found in a wide range of environments, including cold methane seeps (Knittel et al., 2005), hydrothermally influenced sediments (Holler et al., 2011; Teske et al., 2002; Vulcano et al., 2022), and freshwater ecosystems (Takeuchi et al., 2011). Related archaea performing propane and butane oxidation coupled to sulfate reduction have been discovered in hydrothermally-influenced sediments from the Gulf of California (Laso-Pérez et al., 2016; Wegener et al., 2022).

An often-overlooked component of marine seep ecosystems is the viral community. Viruses are the most abundant biological entities in the ocean, with estimated numbers as high as 10^{30} total viral particles and up to 15 times the number of viral particles as bacteria and archaea (Suttle, 2007). They play a substantial role in community structure, ecology, microbial evolution, and geochemical cycling in the surface ocean and water column (Beltrami & Carroll, 1994; Bratbak et al., 1993; Shelford et al., 2012; M. B. Sullivan et al., 2006; Wilhelm & Suttle, 1999). For example, viral lysis of microbial hosts is thought to strongly influence the concentrations of bioavailable iron, a limiting nutrient for microbial growth, in the ocean (Poorvin et al., 2004). It has been estimated that surface and subsurface sediments contain 10-1000 times more viruses than in the pelagic zone on a per volume basis (Danovaro et al., 2008). In low-energy sediments lacking higher-order predators, viruses are considered the primary agents of microbial mortality (L. Cai et al., 2019). Ignoring the role of viruses in hydrocarbon seeps results in an incomplete picture of the ecology and geochemical cycling in these critical deep ocean ecosystems controlling greenhouse gas emissions.

Recent work regarding viruses of diverse methane-oxidizing archaea hints at the breadth of potential influence on their ecology and evolution. For example, viral infection is thought to have played a role in the diversification of the ANME-1 clade via the horizontal transfer of the distinct thymidylate synthase *ThyX*, which is non-homologous to the *ThyA* found in other ANME lineages (Laso-Pérez et al., 2023). Other work in methane seep systems has found viruses with essential metabolic genes, including phosphoadenosine phosphosulfate reductase (*cysH*), which is critical in assimilatory sulfate reduction (L. Wang et al., 2022); O-antigen synthesis, which may be involved in cell-cell adhesion (L. Wang et al., 2022); and several other cofactor-associated genes (Z. Li et al., 2021; Peng et al., 2023). Recently, viruses encoding genes exclusive to methane metabolism were found in a variety of methane-rich habitats, including marine and freshwater ecosystems, cow rumen, and permafrost (Zhong et al., 2024). However, all current viral genomes isolated from hydrocarbon-rich sediments have been mined from bulk metagenomes, and few explore the diversity of habitats and metabolic niches in which these organisms are found. Existing work is also limited by the physical complexity of sediment environments and the slow-growing nature of the organisms in their low-energy environments.

Here, we leverage sediment-free enrichments of organisms from multiple sampling locations at different in-situ temperatures to directly sequence the viral community (**Figure 2.1**) that is maintained in the presence of these keystone microbial species. These enrichments select for various anaerobic hydrocarbon degraders and their sulfate-reducing partners via the addition of specific substrates. By reducing the complexity of the sediment matrix, these enrichments allow targeted investigations of viruses infecting members of this keystone syntrophy.

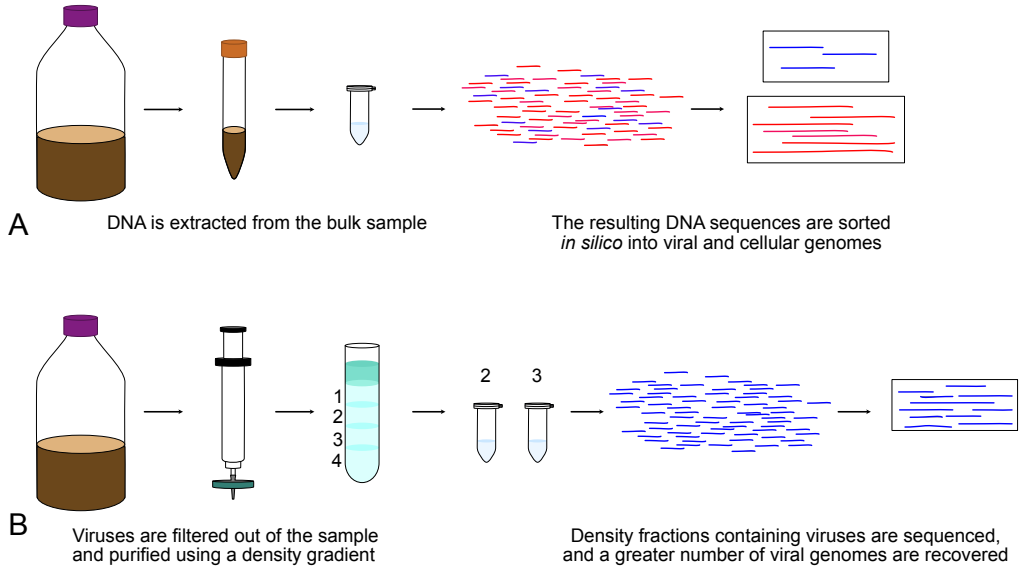

Figure 2.1 Comparison between sequencing methods for environmental viruses. A) Standard method for studying environmental viruses, in which viral sequences are mined computationally from bulk metagenomes. As discussed in the first chapter, the majority of known marine viral diversity and ecology comes from the surface ocean and the water column, which are easier to access and sample. Attempting to understand the major viral players in the deep ocean and the sediments is a rather recent undertaking, and most papers rely on mining existing published datasets; most of what we know about viral diversity in the deep ocean and its sediments comes from the bulk sequencing of all the biological entities in a sample, a technique known as shotgun metagenomics. Viral sequences are then pulled out of these metagenomes using a variety of published algorithms and characterized using their gene content. However, these bulk metagenomes are not well-suited to capturing viral diversity. **B)** Examining viruses directly by extracting and purifying them from the sample before sequencing yields a far more detailed picture of the viral community in an ecosystem (Kosmopoulos et al., 2024; Santos-Medellin et al., 2021).

Our findings indicate that the impacts of incubation temperature and original sampling location influence community composition rather than the phylogenetic and whole-genome patterns of individual viral species. We also found a putative cross-domain virus whose potential hosts include only one member of the methane oxidizing syntrophy, unlike previous findings from hydrothermal microbial mats (Hwang et al., 2023). Despite the extreme nature of the environments from which these viruses originate, we detected little difference in the identities and abundances of auxiliary metabolic genes between different conditions. Putative AMGs relating to heme synthesis in methanotrophs suggest the potential for viral influence on fundamental metabolism and syntrophic associations for these alkane-oxidizing microorganisms. In addition, several viral contigs encoded a *nosD*-like protein lacking the

conserved motifs required for the canonical copper-shuttling function. The proximity of several of these *nosD*-like proteins to cell-surface related proteins and their conspicuous absence in water column datasets suggests a role in attachment in biofilm-dominated communities.

MATERIALS AND METHODS

Development and maintenance of incubations

Anaerobic alkane-oxidizing enrichments from Guaymas Basin, Elba, Menes Caldera, and the Black Sea were developed via dilution of sediment slurries with anoxic artificial seawater medium: 150mL slurries (150 ml) were transferred in 256 ml serum bottles sealed with butyl rubber stoppers as described in Laso-Pérez et al., 2018. Enrichments from Santa Monica Basin were developed and maintained as described in Yu et al., 2022. All incubations are maintained in an artificial seawater medium containing sulfate as the sole electron acceptor. Sulfide concentrations were measured to track metabolic activity, and active cultures were consecutively diluted. Spent media from 100mL incubations was replaced by allowing the particles to settle and removing 80mL of medium and replacing the same volume with new anoxic artificial seawater. The *D. auxilii* and *Archaeoglobus* incubations were maintained under 0.2 MPa of 80:20 hydrogen:carbon dioxide, while the AOM enrichments were provided 0.2MPa of 100% methane. Propane and butane enrichments were produced as described above from Guaymas Basin, and slurries and their dilutions were kept at 0.15 MPa propane/butane. A full description is provided in **Supplementary Table S2.1**, including references for media composition.

Media recipe for Guaymas Basin Methane and Hydrogen incubations

1. Begin with the following in 800mL of water, for a final volume of 1L:

Reagent	**g or mL/L**	**g/mol**	**final mM concentration**
$MgCl_2 \cdot 6H_2O$	5.67g	203.3	27.9
$CaCl_2 \cdot 2H_2O$	0.22g	147.0	1.5
NaCl	26.37g	58.4	451.2
KCl	0.60g	74.6	8.0
Na_2SO_4	1.50g	142.0	10.6
K_2HPO_4	0.17g	174.2	1.0

NH_4Cl	0.11g	53.5	2.0
Se/W solution (see below)	0.10mL		0.01/0.0067 uM
Trace elements solution (see below)	1.00mL		
250mM HEPES buffer (see Appendix II)	100.00mL		25.0

2. The pH should be between 7.2-7.45. Filter through a 0.22µm pore size filter to sterilize and flush with nitrogen gas for 20-30 minutes to make the media anoxic.
3. Add the following after the media has been flushed:

Reagent	**g or mL/L**	**final mM concentration**
200mM HS^- solution, flushed with nitrogen gas	2.5mL	0.5
1M $NaHCO_3^-$ solution, filter sterilized and flushed with nitrogen gas	5mL	5

Collection and purification of viruses

Spent media from the enrichment incubations were removed and filtered through a 0.22µm pore size polyethersulfone membrane (Cat. SLGP033N, Millipore Sigma; St. Louis, MO, USA) to remove cells and cell debris. Viruses in the filtrate were concentrated in an Amicon Ultra-15 Centrifugal Filter Unit with a 100kDa cutoff (Cat. UFC910024, Millipore Sigma; St. Louis, MO, USA) spun at 2000×g for 3 minutes (Allegra X-I5R Centrifuge, Beckman Coulter; Brea, CA, USA). To increase yield of viruses from the filtration membrane, the cartridge was wrapped with parafilm and vortexed at the maximum setting for 10 seconds (Vortex Genie 2, Scientific Industries; Bohemia, NY, USA). Concentrated viruses were then purified using ultracentrifugation (Ultima MAX-E Ultracentrifuge; Beckman Coulter, Brea, CA, USA) with Optiprep Density Gradient Medium (Cat. D1556, Sigma Aldrich; St. Louis, MO, USA), according to the protocol adapted from Kauffman et al., 2018 and detailed on protocols.io (Narayanan & Philosof, 2020). Density fractions were then stained with SYBR™ Gold Nucleic Acid Gel Stain (Cat. S11494, ThermoFisher; Waltham, MA, USA) and visualized with fluorescence microscopy to determine which fractions contained viral particles (Narayanan &

Philosof, 2024). The fractions marked by 30%, 35%, and 40% were found to contain the largest number of viral particles. Note that this method does not account for the presence of non-viral particles containing nucleic acid, such as vesicles, but vesicles are thought to affect total viral counts by less than an order of magnitude in seawater samples (Biller et al., 2017).

Transmission Electron Microscopy

Spent media from incubations was collected, filtered, and concentrated as described under *Collection and purification of viruses*. Five microliters of virus concentrate were spotted on a glow-discharged carbon-coated 300 mesh copper TEM grid (Cat. 01813, Ted Pella Inc., Redding, CA, USA). Samples were incubated on the grid at room temperature for five minutes, and excess liquid was wicked off with filter paper. Once dry, grids were stained with 2% uranyl acetate solution in 1X PBS for 30 seconds; the staining procedure was repeated twice for a total of 60 seconds of treatment. Excess stain was removed and grids were air dried completely before imaging with an FEI Tecnai T12 at 120kV. Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech.

Viral DNA extraction and sequencing

Density fractions were treated with RNase-free DNaseI (Cat. M0303L, New England Biolabs, Ipswich, MA, USA) to remove free DNA. Viral DNA was extracted using Promega Wizard PCR Preps DNA Purification Resin (Cat. A7181, Promega, Madison, WI, USA) and Wizard Minicolumns (Cat. A7211, Promega, Madison, WI, USA) in accordance with Poulos, 2012. Sequencing libraries were prepared using the Illumina DNA Prep library kit (Cat. 20018705, Illumina, San Diego, CA, USA) with IDT® for Illumina® DNA/RNA UD Indexes Set C (Cat. 20042666). The DNA from the two Menes Caldera incubations was pooled to ensure sufficient input for library preparation. Paired-end sequencing was performed at the University of Southern California Keck School of Medicine using an S1 flow cell for 300 cycles on the Illumina NovaSeq6000 sequencing platform (v1.5 reagents).

Cellular DNA extraction and sequencing

Approximately 15mL of each enrichment was collected and centrifuged at 5000×g and 10°C for 30 minutes. The supernatant was removed and the resulting cell pellet was extracted using Qiagen's DNeasy Blood and Tissue Kit according to their Gram positive protocol (Cat. 69504,

Qiagen, Hilden, Germany). 16S rRNA archaeal and bacterial genes were amplified in duplicate with Q5 Hot Start Master Mix (Cat. M0492S, New England Biolabs, Ipswich, MA, USA) according to the manufacturer's directions, using 515F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-GTGYCAGCMGCCGCGGTAA-3') and 926R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-CCGYCAATTYMTTTRAGTTT-3') primers (Parada et al., 2016) that include Illumina adapters for 32 PCR cycles. PCR duplicates were then pooled and barcoded with the Illumina Nextera XT index2 primers that include unique 8-bp barcodes. Amplification was performed with Q5 Hot Start PCR mixture and 3 μL of the pooled duplicate product was added to a 30 μL reaction volume, annealed at 66°C, and cycled 11 times. Products were run on a 1.5% agarose gel and quantified by band intensity. These barcoded PCR products were then combined in equimolar amounts into a single tube and 300uL of this pooled sample was run on a 1.5% low melt agarose gel (Cat BP165-25, Fisher Scientific, Waltham, MA, USA) and purified using Promega's Wizard SV Gel and PCR Clean-up System (Cat. A9281, Promega). The sample was sequenced by Laragen (Culver City, CA, USA) using the MiSeq Reagent Kit v3 (600-cycle, Cat. MS-102-3003, Illumina) on Illumina's MiSeq platform with the addition of 15-20% PhiX. Figure 1 was generated after dropping singletons and organisms classified as "uncultured" at the family level. The 16S rRNA gene sequencing data was not included for Guaymas Basin Methane 1G and Guaymas Basin Methane G due to a significant temporal gap between virus sampling and 16S rRNA gene sampling, in which the bottles were diluted into larger containers. 16S rRNA amplicon data was processed with the DADA2 v.1.22 workflow (Callahan et al., 2016). A reproducible workflow is available (see Data Availability statement). SILVA SSU database r138 (Quast et al., 2012) with in-house sequences added was used to assign taxonomy to amplicon sequence variants. Whole genome library preparation and paired-ended sequencing was performed at the Millard and Muriel Jacobs Genetics and Genomics Laboratory, California Institute of Technology, using a P1 flowcell for 300 cycles with the Illumina NextSeq2000 sequencing platform.

Cellular metagenome assembly and binning

Reads were trimmed with bbduk v38.81 (Bushnell, n.d.) and filtered such that reads with <150 nucleotides were eliminated. Reads that passed this step were assembled using metaSPAdes version 3.15.3 (Nurk et al., 2017). Reads were aligned to the assembled contigs using the Burrows-Wheeler Alignment tool version 0.7.17-r1188 (H. Li & Durbin, 2009) before contigs were binned using MetaBAT version 2.12.1 (Kang et al., 2019) with a minimum contig length of 1500 bp. The CheckM v1.1.3 lineage workflow (Parks et al., 2015) was used to determine the quality of each bin. Taxonomy was assigned with GTDB-Tk v2.1.0 (Chaumeil et al., 2022) using the GTDB R207 database (Parks et al., 2018, 2020, 2022; Rinke et al., 2021). All reads were also mapped back to the full set of contigs using bbmap v38.81 (Bushnell, n.d.); unmapped reads were reassembled with metaSPAdes.

vMAG assembly, annotation, and phylogeny

Viral reads were trimmed using bbduk v38.81 with kmer lengths of 23, minimum kmer length of 11, maximum hamming distance of 1, and parameters set to trim evenly at the overlap of paired reads. Contigs were assembled using metaViralSPAdes v3.15.3 (Antipov et al., 2020) with default parameters. To improve the quality of assembly and number of complete vMAGs, all Guaymas Basin methane incubation reads were pooled before assembly, as were Guaymas Bain hydrogen reads and Santa Monica reads. Viral contigs were identified and quality checked using CheckV (Nayfach et al., 2021), viralVerify (Antipov et al., 2020), and seeker (Auslander et al., 2020). viralVerify was also set to distinguish plasmids. All virome-assembled contigs, regardless of whether they were identified as viral by the preceding algorithms, were subject to annotation by Vibrant (Kieft et al., 2020); Vibrant filters based on probable viral identity before annotation, adding a fourth set of checks to the three algorithms cited above. All cellular contigs (see *Cellular Metagenome Assembly and Binning*) were checked to identify viruses in the cellular metagenomes; these viral contigs were also annotated with Vibrant and combined with the vMAGs from the viral sequencing for downstream work.

Terminase large subunit (*TerL*) and major capsid protein (*MCP*) phylogenies were constructed using the annotations obtained from Vibrant. Relevant amino acid sequences were aligned using MAFFT v7.505 (Katoh & Standley, 2013) and trimmed with trimAl v1.4.rev15 (Capella-

Gutiérrez et al., 2009). Alignments were manually curated to remove short sequences or sequences with large gaps. Phylogenetic trees were generated using IQ-TREE multicore version 1.6.12 (Nguyen et al., 2015).

Auxiliary metabolic genes were included in the analysis if classified as such by Vibrant. Because annotations and open reading frame calls are not always reliable without a great deal of manual curation, AMG abundances per sample were quantified as the abundance of the viral contig on which the AMG was found (see *Viral Abundance Calculations and Diversity*) rather than the abundance of the AMG itself. Within a given sample, a contig was included in the AMG calculations if the centered log ratio of the contig's abundance was >1.

Viral classification and clustering

Viral contigs that were 98% complete or higher according to CheckV were classified at the family level using GRAViTy: Genome Relationships Applied to Virus Taxonomy (Aiewsakun et al., 2018), with ICTV Virus Metadata Resource VMR_16-180521_MSL36 (Lefkowitz et al., 2018). The whole-genome protein clustering network was generated using vConTACT2 (Bolduc et al., 2017) using Refseq release 207, pcs mode MCP, vcs mod ClusterONE. The network was visualized using Cytoscape (Shannon et al., 2003). ANME-1 phages from Laso-Pérez et al., 2023 were included in the network calculation due to the similarities of the sampling site and host taxonomy to those used in this study. T7 control sequences were included in vConTACT2 analysis as an internal check.

Viral abundance calculations and ordination

Viral reads were quantified per viral cluster and per contig using salmon v1.10.0 (Patro et al., 2017) and summarized using tximport v1.18.0 (Soneson et al., 2015). DESeq2 v1.30.1 (Love et al., 2014) was used to quantify the abundance of each cluster/contig per library. Abundances were centered log ratio (CLR) transformed with ALDEx2 v1.25.1 (Fernandes et al., 2013). Diversity metrics were calculated using the transcripts per million value for each viral cluster, as output by tximport. Principal components, corresponding PERMANOVA (Adonis, analysis of dissimilarities) values, and rank abundances were calculated using the CLR-transformed data.

Host community nonmetric multidimensional scaling (NMDS) ordination plots were generated using the vegan v2.6-4 (Oksanen et al., n.d.) metaMDS function (Bray-Curtis dissimilarity metric on relative abundance data) at the amplicon sequence variant (ASV) level (McArdle & Anderson, 2001). Viral community principal component analysis (PCA) ordination plots were generated using the prcomp function (Euclidean distance on centered log ratio transformed abundance data) at the viral genus level; principal component analysis (PCA) was used for the viral community instead of NMDS because the centered log ratio transformation moves compositional data into Euclidean space (Quinn et al., 2018). For both host and virus communities, corresponding PERMANOVA values were calculated with the vegan adonis2 function.

CRISPR and tRNA identification and host matching

Potential cellular host genomes were obtained from NCBI from Projects PRJNA276404 (Krukenberg et al., 2016; Wegener et al., 2015); PRJNA318983 & PRJNA319143 (Laso-Pérez et al., 2016); PRJNA418316 (Krukenberg et al., 2018); PRJEB36446 & PRJEB36096 (Hahn et al., 2020); PRJNA713414 (Speth et al., 2022); PRJNA758896 (Yu et al., 2022); PRJNA875076 & PRJNA721962 (Laso-Pérez et al., 2023); members of *Ca. Desulfofervidaceae* from PRJNA762493 (Murali et al., 2023); and personal communications with Dr. Daan Speth at the University of Vienna and Dr. Rafael Laso Pérez at the Museo Nacional de Ciencias Naturales, Madrid. These genomes were generated from bottles used in this study or were sampled from closely-related sites. Sixty-three additional *Archaeoglobus* genomes were obtained from GTDB under the search term "archaeoglobus". Additional cellular genomes were generated from three of the incubation bottles (Guaymas Basin Methane 1G, Guaymas Basin Methane G, and Guaymas Basin Hydrogen 3A) as detailed above.

CRISPRs were called using CRISPRDetect 3.0 (Biswas et al., 2016) and CRISPRCasTyper v1.1.4 (Russel et al., 2020). Host genome CRISPR spacers were combined and dereplicated with MMseqs2 (Steinegger & Söding, 2017) version 13.45111 easy-cluster workflow at 99% minimum sequence identity and 80% coverage. BLASTn v2.12.0+ and SpacePharer v5.c2e680a (R. Zhang et al., 2021) were used to compare the viral contigs and the host spacers. BLAST hits were considered high-quality with at least 98% sequence similarity and no more than 2 mismatches.

tRNAs were identified in putative host genomes and viral genomes using tRNAscan-SE v2.0.5 (Chan et al., 2021) and Aragorn v1.2.41 (Laslett, 2004). Viral and host tRNAs were combined and dereplicated, respectively, as done for the CRISPRs. Host tRNAs were aligned against viral contigs and viral tRNAs using BLAST. Matches were considered high-quality only if they were exact matches: 100% sequence identity, 100% coverage, and no mismatches.

Host/phage network diagrams were made using adjacency matrices of viral clusters that shared hosts, and hosts infected by the same viral clusters, respectively. Where viruses were not clustered at the genus level by vConTACT2, the contig was used on its own. Networks were calculated using the networkX from_pandas_adjacency function (Hagberg et al., 2008), and self-loops were removed. The network was visualized using Cytoscape (Shannon et al., 2003).

Identification of nosD-like proteins using a hidden Markov model

Input sequences for the *nosD* Hidden Markov Model were derived from Müller et al., 2022, Extended Data Figure 7A. Sequences for *Geobacillus thernodenitrificans* (WP_008880109.1), *Shewanella denitrificans* (WP_011496651.1), *Stutzerimonas stutzeri* (WP_036990777.1), *Marinobacter nauticus* (WP_342631055.1), *Wolinella succinogens* (WP_011138819.1), *Hydrogenobacter thermophilus* (WP_012962810.1), and *Deslfitobacterium* hafniense (WP_018214548.1) were aligned against a curated set of genomes (See Data Availability) using BLAST v2.9.0+. Hits with e-value<0.0005 and bitscore>200 were retained and aligned with the selected sequences from Müller et al., 2022 using MUSCLE v3.8.31 (Edgar, 2004). The alignment was manually curated and dereplicated at 70% identity in Jalview (Waterhouse et al., 2009). The HMM was built from this alignment; all viral proteins output by Vibrant were searched for *nosD*-like proteins using HMMER v3.3.2 (Eddy, 2007). *nosD*-like proteins were included in this study if they had an hmmsearch e-value<0.001, came from contigs that were confirmed to be viruses using the criteria under *vMAG Assembly and Annotation*, came from contigs at least 5000 basepairs in length, and had a greater proportion of viral genes than host genes according to checkV. Vibrant *nosD* annotations that were not found in the hmmsearch results were included in the study if they were >200 amino acids in length and appeared to span the conserved His-Met-Met residues of the canonical *nosD* proteins in a MUSCLE-generated alignment. These open reading frames were truncated to remove any additional protein domains before inclusion in the phylogeny; the presence of

additional domains and the location within the ORF of the nosD-like protein were determined using NCBI's Conserved Domain Search. Viral *nosD*-like proteins from the Tara Oceans viral study (Roux et al., 2016), Costa Rica methane cold seeps (Li et al., 2021; Philosof et al., *in prep*), lake sediment and water column (Jaffe et al., 2023), and soils (Santos-Medellín et al., 2023) were similarly searched using the *nosD* HMM and thresholded at e-value<0.001. For published datasets that were only available as nucleic acid fasta files, genes were predicted with Prodigal V2.6.3: February, 2016 (Hyatt et al., 2010) before hmmsearch was run. Alignments were generated with MUSCLE v3.8.31 (Edgar, 2004) and trees were inferred using IQ-TREE2 (Nguyen et al., 2015) using the model finder mode. Protein structures were predicted using Alphafold2 (Jumper et al., 2021) in monomer mode with the reduced set of databases (MGnify, PDB70 (with PDBx/mmCIF and PDB SEQRES), BFD, Uniclust30, Uniprot, and Uniref90).

RESULTS AND DISCUSSION

Alkane-oxidizing enrichment cultures are rich in viruses

Sediment-free enrichments of alkane-oxidizing/sulfate-reducing consortia and an additional enrichment of one of the sulfate-reducing partners (*Desulfofervidus*, formerly HotSeep-1) were established between 2003 and 2017 from seep sediment samples collected from Guaymas Basin, Gulf of California in 2009 (Hahn et al., 2020; Holler et al., 2011; Krukenberg et al., 2016; Laso-Pérez et al., 2016); the Black Sea in 2004; waters off of the west coast of Elba, Italy in 2010 (Ruff et al., 2016); Menes Caldera, Egypt in 2003; and Santa Monica Basin, California in 2013 (Yu et al., 2022) and maintained continuously. All enrichments, except for Santa Monica, were generated via dilution of bulk seep sediment. Inoculum for the Santa Monica enrichments was made using Percoll density separation to capture methane-oxidizing consortia followed by serial dilution and maintenance under a methane headspace. Incubations were placed under temperatures similar to those in the environment and given energy substrates intended to enrich for alkane degraders and their sulfate-reducing syntrophic partners (**Figure 2.2a**). The substrates used were hydrogen, methane, propane, and butane; and incubation temperatures spanned from 10˚C to 60˚C (**Supplementary Table S2.1**).

16S rRNA amplicon sequencing confirmed the dominant archaeal and bacterial taxa within each incubation bottle (**Figure 2.2b**). Major archaeal groups included known hydrocarbon oxidizers

such as members of the ANME-1 subgroup, ANME-2 subgroup, and *Syntropharchaeaceae*, along with members of the *Bathyarchaeia* and *Methanosarcinales*. *Archaeoglobaceae* was most abundant in those incubations supplied with hydrogen and sulfate. Major bacterial groups included sulfur cycling lineages such as *Desulfofervidaceae* and *Dissulfuribacteraceae*, as well as other taxa common in seeps and other anaerobic and/or sedimentary marine environments such as *Fermentibacteraceae*, the JS1 lineage, and *Aminicenantales* (Dombrowski et al., 2017; Savvichev et al., 2023; Yu et al., 2022). In general, family-level groups were primarily composed of one or two amplicon sequence variants (ASVs) in each enrichment (**Supplementary Table S2.2**), matching previously described patterns in sediments (Peña-Salinas et al., 2024) in which ASVs are broadly limited to one or two variants at a single depth horizon. Other work has found that a few ANME operational taxonomic units tend to dominate in seep and vent sediments, while accounting for less than 0.6% of global ANME diversity (Ruff et al., 2015).

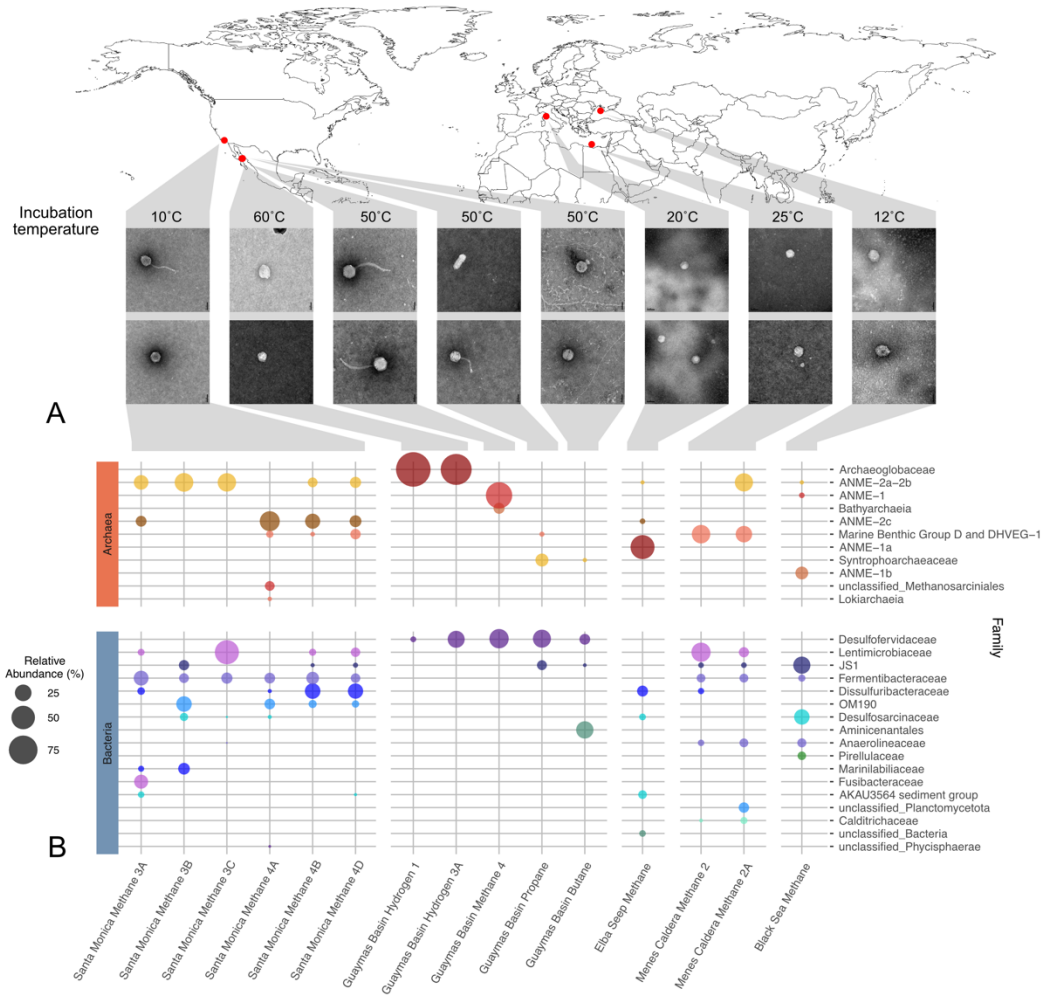

Figure 2.2. Prokaryotic community composition of enrichment cultures. **A)** Example transmission electron microscopy of viral-like particles from the enrichment incubations used in this study, with corresponding laboratory incubation temperature; original sampling location marked with red points on the map above. Scale bars = 50nm. **B)** Family-level taxa present in each incubation bottle at relative abundance >3%, determined by 16S rRNA amplicon sequencing. Samples are grouped by sampling location and were incubated with the indicated energy source.

Viral ecological patterns differ between the community and OTU levels

Transmission electron microscopy confirmed that all enrichment bottles contained viruses even after several years under controlled laboratory conditions (**Figure 2.2a**). Past work in enrichment cultures has shown that bacteriophages are present in bioreactor systems (Gambelli et al., 2016) and can influence the abundance of the target organism(s) (Muniesa et al., 2005). To construct viral genome libraries, we concentrated and purified viral-like particles (VLPs) from the spent media of the enrichment cultures. In total, 989 dsDNA viral contigs greater than 5000 nucleotides (nt) in length were obtained. Contigs were considered viral if they were identified as

such by either ViralVerify (Antipov et al., 2020), seeker (Auslander et al., 2020), or Vibrant (Kieft et al., 2020), and if more than 50% of the genes predicted were categorized as viral rather than cellular by checkV (Nayfach et al., 2021). Of the 989 contigs, 341 were at least 98% complete according to checkV. We taxonomically classified these 341 viral contigs using GRAViTy (Aiewsakun et al., 2018) and obtained 60 novel viral families, though 121 contigs could not be classified at the family level. Genus-level clustering with vConTACT2 (Bolduc et al., 2017) of all 989 contigs yielded 231 genus-level clusters that contained only 642 contigs; the remaining 347 could not be clustered. Despite the simplification of the environment and microbial community via sediment removal, a remarkable amount of viral richness is still present (**Supplementary Table S2.3**). However, the process of enrichment has still resulted in lower taxonomic evenness and richness than previously reported from *in situ* seep viral communities (Li et al., 2021; Philosof et al., *in prep*).

The range of origins and incubation conditions of the enrichment cultures provides a unique framework to examine how factors such as temperature, substrate, and original sampling location shape the viral community structure. Dimensionality reduction (nonmetric multidimensional scaling, NMDS) of host abundance (16S rRNA-based) found that the sampling location, incubation temperature, and methane versus non-methane energy sources were all statistically significant factors influencing host community composition (PERMANOVA, **Figure 2.3a, Supplementary Table S2.4**), though the effects of these factors cannot be disentangled from each other. Similar analysis of viral abundance data also found that both temperature and sampling location were significant factors (**Figure 2.3b, Supplementary Table S2.4**). Interestingly, the source of electron donor (methane vs. non-methane substrates) did not appear to strongly influence the relationship between the different viral communities. Analysis at different taxonomic resolution, i.e., species level OTU vs. amplicon sequence variant (ASV) for the microbial host community and the genus level or individual contigs for the associated viruses, yielded similar results (**Supplementary Figure S2.1**). An important caveat to these observations is the lack of replication available for the Menes Caldera, Elba Seep, and Black Sea sampling sites.

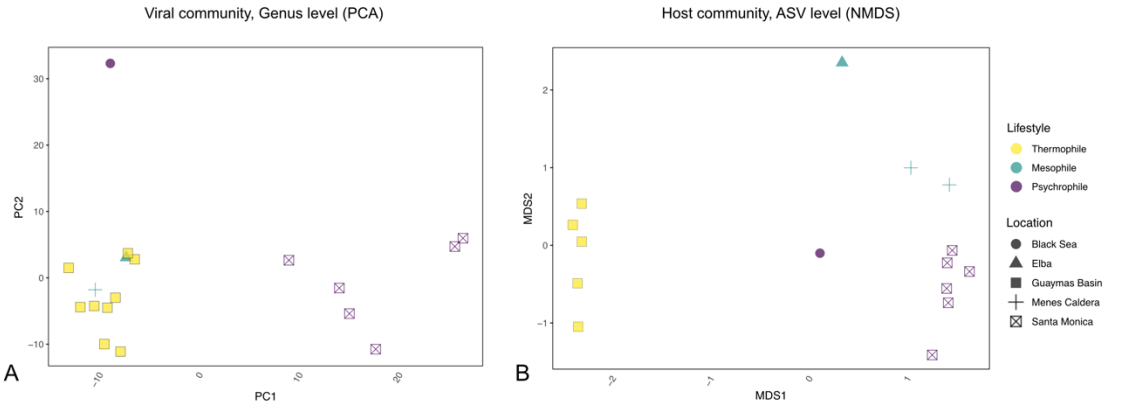

Figure 2.3. Community-level comparisons between enrichment cultures. **A)** Ordination plot of principal component analysis (PCA) of viral genera. PC1 importance: 0.3349. PC2 importance: 0.1638. **B)** Ordination plot of nonmetric multidimensional scaling (NMDS) of 16S rRNA genes at ASV level. Stress: 0.065. The communities that show the clearest differences between host and virus ordinance plots are the mesophilic Menes Caldera and and Elba Seep, and the psychrophilic Black Sea; additional data from mesophilic communities is necessary to better constrain the degree to which mesophilic viral and host communities differ in their clustering patterns. Note that for the viral community, DNA from the two Menes Caldera incubations was combined for the sequencing library (see methods), resulting in one data point for the viral community though there are two for the host.

To constrain the influence of various physico-chemical parameters on the response of individual viruses to their environments, we examined phylogenetic and protein clustering relationships between phage genomes from the different cultures. Unlike the community level analysis in Figure 2, the terminase large subunit (*TerL*) and major capsid protein (*MCP*) phylogenies showed no discernable pattern related to incubation temperature, original sampling location, or energy source (**Figure 2.4a, Supplementary Figure S2.2**). Because phages possess no universal marker analogous to the 16S rRNA gene, *TerL*, which is present in tailed viruses, the DNA Polymerase gene, which was not well-annotated in this dataset, and *MCP* are commonly-used proxies. Visualization of vConTACT2 protein clustering networks, generated at the whole-contig level, similarly showed little to no clear pattern based on physicochemical parameters for the enrichment cultures (**Figure 2.4b**). However, there were differences in how vMAGs clustered between the whole genome network and the *TerL* phylogeny. For example, the groups circled in the whole-genome network diagram (Figure 3B), composed of viruses that share genome-level similarities, are often scattered across the *TerL* tree (**Figure 2.4a**) with no consistent pattern (corresponding arrow colors). Some of these groups identified in the vConTACT clustering network showed greater homogeneity with respect to location and

temperature, while other groups included a mixture of phage genomes from different sampling sites and incubation temperatures (**Figure 2.4b**).

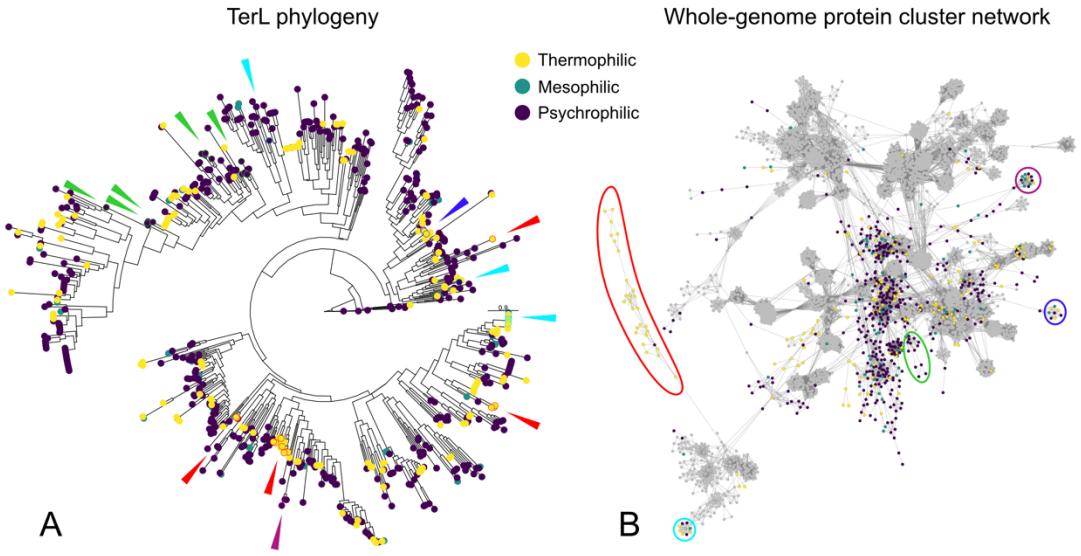

Figure 2.4. Single-gene and whole-genome comparisons of phages from enrichment cultures. **A)** Terminase large subunit phylogeny. Both figures are colored by incubation temperature: thermophilic (≥50˚C), mesophilic (<50˚C and ≥ 20˚C), or psychrophilic (<20˚C). Groups circled in the vConTACT2 diagram in part B are labeled with arrows in the corresponding color on the *TerL* phylogeny; not all contigs contained an annotated *TerL* gene, and thus group sizes in the protein cluster network do not correspond perfectly with the number of arrows on the phylogeny. **B)** vConTACT2 whole-contig protein cluster network. Grey nodes are reference sequences. The groups circled in red and magenta are largely homogenous in terms of incubation temperature, while the groups in blue, light blue, and green are highly mixed. The complete network is available as Supplementary Figure S3.

Further supporting the general lack of clustering by shared physico-chemical parameters, we found that such mixed groups were not unusual in this dataset; of the viral genus-level clusters found by vConTACT2, approximately 29% contain contigs that were assembled from enrichment cultures of different incubation temperatures (ranging from 10-60 ºC), while 36% contain viral contigs from different sampling locations. Despite the varied provenances and natural histories of the microorganisms in these enrichment cultures, highly related viruses were recovered across incubations. For example, viral genus-level cluster VC25 (dark blue group, **Figure 2.4b**) contains viral contigs assembled from the Elba Seep (mesophilic), Menes Caldera (mesophilic), the Black Sea (psychrophilic), Santa Monica (psychrophilic), and Guaymas Basin (thermophilic). Contigs within this cluster share multiple non-core genes, such as a *chaB* cation transport regulator (K06197) and *yfbK* Ca-activated chloride channel homolog (K07114).

Temperature-related gene and gene content signatures in viruses are not well-constrained, and this study represents one of the first attempts to find patterns in viral gene content spanning multiple temperature regimes. In bacteria and archaea, biases in GC content are thought to be indicative of temperature adaptation, though this is currently a contentious hypothesis in the field (Hu et al., 2022; Hurst & Merchant, 2001; Sharp et al., 2010). Evaluation of GC content within each viral genome showed a clear lack of response to temperature (not shown due to inclusion of embargoed data in the analysis). We then evaluated amino acid composition within the TerL and MCP proteins to confirm a lack of temperature-based bias in amino acid composition within these core genes. Bacterial and archaeal thermophiles can show distinct amino acid composition signatures compared to mesophiles and psychrophiles (Di Giulio, 2005; Kreil, 2001; Reed et al., 2013); within our phages, amino acid composition data of the core viral proteins TerL and MCP showed no corresponding temperature response (**Supplementary Figure S2.4 B and D**). However, specific amino acids did show significant differences in use between phages from psychrophilic and thermophilic environments, including valine, threonine, histidine, aspartic acid, serine, and glutamine (**Supplementary Table S2.5**). In prokaryotes, valine and aspartic acid have been highlighted as hallmarks of thermostability, while glutamine, histidine, threonine, and serine are less frequent in thermophiles (Fukuchi & Nishikawa, 2001; Singer & Hickey, 2003). Surprisingly, arginine, an amino acid that is well-established in thermophile literature as a stabilizing amino acid (Argos et al., 1979; Haney et al., 1999; Mrabet et al., 1992), was not significantly different.

These phages are polyextremophilic, covering extremes in temperature, but also in pressure and salinity; these overlapping factors confound our analysis of temperature-specific responses. The above amino acid patterns may also be a result of bottlenecks during distribution, since temperature and geographic location are overlapping parameters in these enrichment cultures. It is additionally important to remember that viruses do not metabolize the way cells do; there are no respiration activities, no building and repair of biomass, and no breakdown of unnecessary or foreign molecules. Viruses may thus have distinct patterns of response to temperature when compared to living cells.

In sum, these results suggest that evolutionary pressures on individual viral species/genera may have been greater before their distribution across this range of habitats, thus explaining the of temperature and location-related patterns at the species/genus level. Furthermore, current selection pressures may act on "accessory" genes rather than highly conserved, core genes such as *TerL* and *MCP*, resulting in some visible clustering in gene sharing networks that is lacking in the single-gene phylogenies (though further investigation of the amino acid abundance patterns is warranted); these findings agree with previous research in dsDNA phages (Comeau et al., 2007; Koonin et al., 2020). However, as seen with VC25, this is not a universal rule as contigs in this cluster share many non-core genes but were assembled from enrichments of diverse provenance. When taken together with the community-level results, these data may also indicate that the rate of selection on the viral community composition is faster than the rate of convergence of individual viral OTUs to a specific environment.

A potential cross-domain, broad host range virus

To link viruses with their microbial hosts in the enrichment cultures, we matched CRISPR spacers and tRNAs between host and viral genomes. In total, 19 of the 231 viral genus-level clusters and an additional 16 unclustered viruses were matched to microbial hosts, though many of the hosts found were often not the most abundant in these incubations. Of the prokaryotic phyla that have viral matches, 13 are bacterial and 4 are archaeal (**Supplementary Figure S2.5**). All of these phyla are commonly found in deep-sea alkane-rich sediments (Dombrowski et al., 2017; Savvichev et al., 2023; Wankel et al., 2012; Yu et al., 2022). A single match was found to a member of the methanotrophic ANME-2c (*Ca. Methanogaster*) enriched from a seep in Elba, and two matches to *Desulfofervidus auxilii*, a sulfate-reducing syntrophic partner of both the methane-oxidizing ANME-1 and propane/butane oxidizing *Syntrophoarchaeum* from Guaymas Basin (Krukenberg et al., 2016; Laso-Pérez et al., 2016; Wegener et al., 2022). Viral cluster 25, which was highlighted above for its presence across multiple incubation and sampling conditions, likely infects members of candidate phylum *Omnitrophota* (formerly OP3). Ca. *Omnitrophota* includes nanobacteria, with traits suggestive of a host-associated lifestyle (Seymour et al., 2023) and predicted roles in sulfur, carbon, and nitrogen cycling in aquatic ecosystems (Perez-Molphe-Montoya et al., 2022). Phage infection in members of the *Omnitrophota* has not yet been documented.

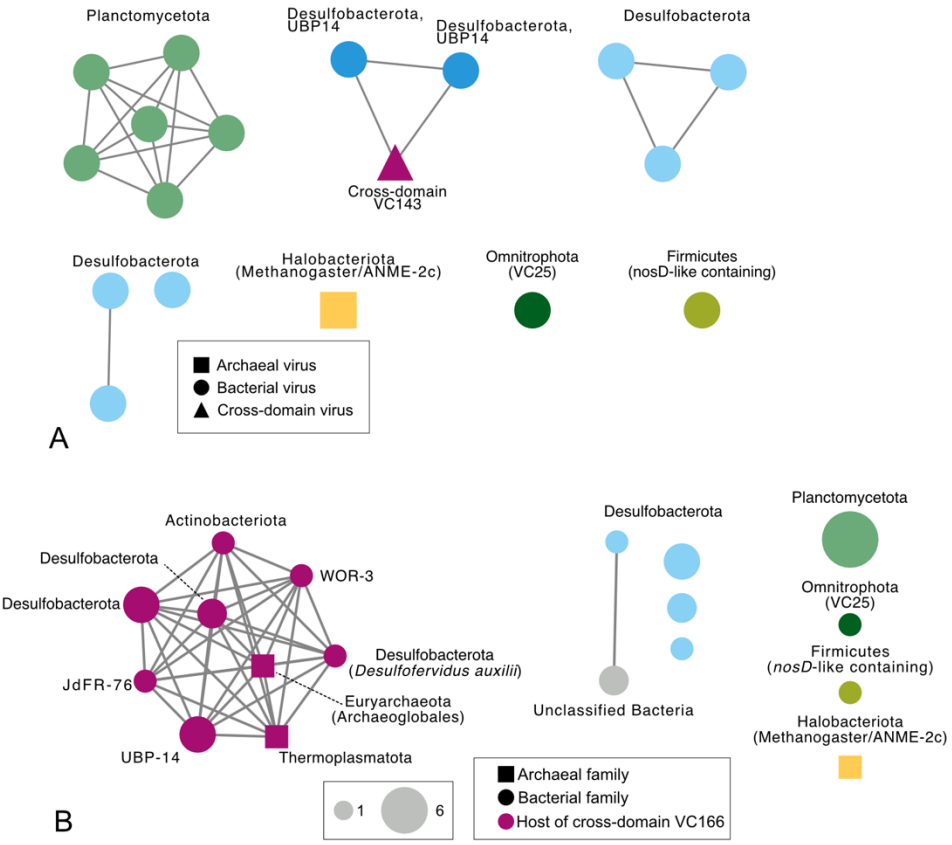

Figure 2.5. Host-virus matching. **A)** Network with phage genus-level clusters as nodes; edges represent whether they share a family-level microbial host. Nodes are labeled by host phylum; with the exception of the cross-domain virus, any viral cluster infecting *Desulfobacterota* is colored in shades of blue. **B)** Network with hosts at family taxonomic level as nodes; edges represent whether the hosts are infected by the same viral genus-level cluster, and number of edges represent the number of shared viral clusters. Nodes are labeled by host phylum. Magenta nodes are families that are infected by the broad host range viral cluster VC143, though VC143 may not be the only virus infecting the given family. The node size corresponds to the number of viral clusters matched to each family via CRISPR spacers or tRNAs. These networks have been truncated to remove most singletons and taxonomic families not addressed in the main text. For the full networks, see **Supplementary Figure S2.4**.

While CRISPR-Cas systems are present in most prokaryotes (Makarova et al., 2020), certain lineages have been found in which they are nearly absent (Burstein et al., 2016). The limited matches with the most dominant microbial lineages in the enrichment cultures were unexpected, but since CRISPR spacers in hosts are subject to high turnover rates (Andersson & Banfield, 2008; Minot et al., 2013; Pride et al., 2011; Tyson & Banfield, 2008), and our host matching was primarily based on genomes assembled from public databases and earlier time points in the multi-year enrichment cultures, it is possible that host spacer arrays have turned over. In addition, CRISPR spacer matching as a tool for host identification is highly specific, but prone

to false negatives (Edwards et al., 2016). We also note the possibility of biases in the viral community due to founder effects. It is therefore unsurprising that large numbers of spacer matches were not found. However, of the matches found, one viral cluster (VC143, Guaymas Basin methane enrichment cultures) appears particularly promiscuous in its range of hosts and is of interest due to its potential for cross-domain infections between bacteria and archaea (**Figure 2.5**).

CRISPR and tRNA-identified microbial hosts of VC143 include *Desulfofervidus auxilii* (a sulfate-reducing bacterial partner of Guaymas Basin methane, propane, and butane oxidizers), other lineages of *Desulfobacterota*, members of the archaeal phylum *Thermoplasmatota*, *Archaeoglobus sp.*, and members of several other bacterial phyla; all predicted hosts are common in deep-sea thermophilic environments, including Guaymas Basin (Teske et al., 2002).

The two contigs in VC143, at approximately 75kb in length, are longer than the average marine dsDNA phage genome size of 50kb (Perez Sepulveda et al., 2016; G. F. Steward et al., 2000). VC143 is also predicted by Vibrant (Kieft et al., 2020) to be lytic and have a circular genome. No *TerL* or *MCP* annotations are available, making phylogenetic placement difficult. Of the 117 predicted open reading frames in the VC143 genome, only 14 could be annotated (**Figure 2.6**). These include magnesium chelatase *comM* (K07391), DNA (cytosine-5)-methyltransferase 1 DNMT1 (K00550), and sulfur carrier protein adenlylyltransferase *ThiF* (K03148). The taxonomy of the nearest NCBI BLASTp hit to *comM* was *Deltaproteobacteria* and to DNMT1 was *Nitrospirota,* while the nearest hit for *thiF* was to the methanogenic archaeon *Methanogenium*. An additional five annotated genes are all related to sugar metabolism and cell wall functions (identified by either Vibrant or NCBI's Conserved Domain Search). These included *waaF*/*rfaF* (K02843, heptosyltransferase), a closely-related vonWillibrand factor type A, *plyA* (K12547, polysaccharidase), *lpxH* (K03269, UDP 2,3 diacylglucosamine hydrolase), and *amiABC* (K01448, N-acetylmuramoyl-L-alanine amidase). A protein BLAST search of each of these ORFs against GTDB r214, NCBI (non-redundant protein sequences), and the cellular genomes used in this study showed similarity both to archaeal genomes and bacterial genomes across a wide number of phyla, including Desulfobacterota, Euryarchaeota, Omitrophota, and Thermoplasmatota (**Supplementary Table S2.6**). The diversity of sugar metabolism genes, commonly co-opted

for viral adhesion to the cell surface (Ciccarelli et al., 2002), may facilitate this viral genus' potentially broad host range.

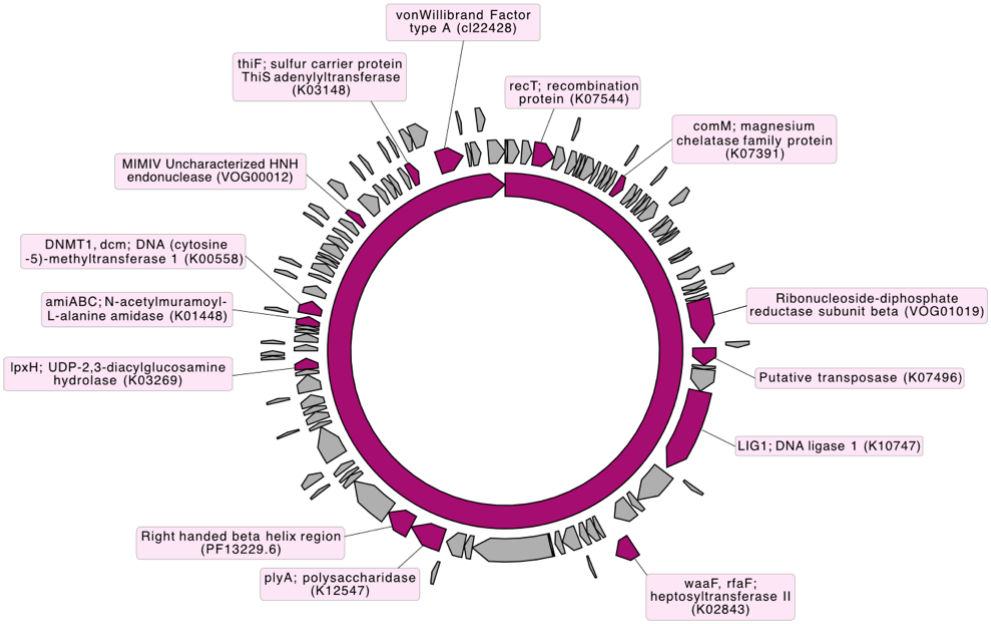

Figure 2.6. Genome map of a contig from VC143. Annotations from either Vibrant or Conserved Domain Search are highlighted in pink. The grey segments are hypothetical proteins. The map was visualized with DNA Features Viewer Python package (Zulko, 2016).

Cross-domain viruses have previously been hypothesized to exist in a variety of methane-rich environments: an analysis of methanotrophic and methanogenic archaea found viruses predicted to infect methanogens, members of *Chloroflexi*, and members of *Deltaproteobacteria*. The cross-domain infectivity of these viruses is predicted to be the result of close mutualistic relationships between the host groups (L. Wang et al., 2022). Additional work in hydrothermal microbial mats from Guaymas Basin found that CRISPR spacers for a single virus have been found in both syntrophic partners, and were proposed to occur via one of four models: 1) viral entry into a non-primary host or viral DNA uptake by a non-primary host, resulting spacer gain, 2) Direct DNA transfer between two organisms of different domains, including CRISPR spacer sequences, 3) Viral host switching, or 4) A fully-fledged broad host range virus, capable of infection across domains (Hwang et al., 2023). Interestingly, the alkane-oxidizing archaeal partners of *D. auxilii* do not appear to be hosts for the VC143 viruses, with the caveat that the

host genomes used in this study may simply have lacked a spacer that matched the viral genomes due to CRISPR spacer turnover.

Given the breadth of prokaryotic taxa associated with VC143 and the lack of spacer or tRNA matching to the archaeal partners of *D. auxilii*, it is unlikely that these CRISPRs were transferred via direct contact between cells. We also predict that a complete switch in host would result in the loss of unnecessary genes, since viral genomes are highly constrained in size (Hatfull & Hendrix, 2011). We caveat this proposal with the understanding that extracellular DNA is thought to be an important source of phosphorus in the deep ocean (Dell'Anno & Danovaro, 2005), and we cannot eliminate the possibility that CRISPR spacers are gained via scavenging of extracellular viral DNA. However, we think it unlikely that DNA scavenging alone is responsible for the entry and insertion of these spacers into the correct section of the host genome.

Auxiliary metabolic genes are not adapted to cellular metabolic lifestyles

Organisms in the deep ocean thrive despite extreme temperatures, high pressure, and low energy inputs compared to the surface ocean. Our dataset uniquely captures a broad range of temperatures (10–60˚C) and sampling locations, providing an opportunity to examine whether the auxiliary metabolic genes (AMGs) present in virus genomes extracted from the studied cultures were shaped by their unique environmental and incubation conditions. Given the importance of methane oxidation to climate outcomes, we were particularly interested in any evidence of unique AMGS involved in alkane degradation or sulfur cycling, following the trend of reports from other sulfur-rich (Kieft et al., 2021) and methane-rich (Zhong et al., 2024) environments.

Of the AMGs annotated in this study, heme synthase (*ahbD*; K22227) and anaerobic Mg-protoporphyrin IX monomethyl ester cyclase (*bchE*; K04034) are of particular interest. *ahbD* encodes a radical SAM heme synthase, and its homologs are primarily known for their roles in the biosynthesis of heme via siroheme-dependent pathways, particularly in methanogenic and methanotrophic archaea and sulfate-reducing bacteria (Buchenau et al., 2006; Fix et al., 2023; Ishida et al., 1998; Kühner et al., 2014). More specifically, *ahbD* catalyzes the conversion of iron-coproporphyrin III to heme through a radical SAM mechanism (Bali et al., 2011). The enzyme contains auxiliary [4Fe-4S] clusters that likely participate in intermediate binding and electron

transfer during the reaction. NCBI Conserved Domain Search broadly confirmed that these proteins are radical SAM superfamily members. *bchE* is generally involved in chlorophyll synthesis in anaerobic phototrophs, though homologs have been found in deep subsurface sediments in non-photosynthetic organisms, including Asgard archaea (Liu et al., 2024), suggesting a broader function. Protoporphyrin IX is also an intermediate in the protoporphyrin-dependent heme biosynthesis pathway, utilizing iron where chlorophyll synthesis would use magnesium as the metal in the tetrapyrrole (Layer, 2021). NCBI Conserved Domain Search showed similarities between the sequences annotated as *bchE* and cobalamin-dependent radical SAM proteins.

Viral replication may be enhanced through manipulation of heme-dependent respiratory processes found in both the ANME and SRB. As previously proposed, genes like *ahbD* (Heyerhoff et al., 2022) and other radical SAM proteins may influence host energy availability via manipulation of cellular redox states. In alkane-degrading systems, *ahbD* is of particular interest due to its role in the synthesis of heme *b* (Lobo et al., 2014), a component of the cytochrome *c* protein essential to direct interspecies electron transport (DIET) (Lovley, 2017; Zhang et al., 2023). DIET is hypothesized to underlie the syntrophic alkane oxidation/sulfate reduction reaction (Laso-Pérez et al., 2016; McGlynn et al., 2015; Meyerdierks et al., 2005; Skennerton et al., 2017; Wegener et al., 2015), in which electrons from the alkane-oxidizing archaea are transferred directly to the bacterial partner for sulfate reduction. While hosts of the viruses carrying these genes have not been identified, the high demand for both iron and sulfur by alkane-oxidizing archaea and their partner sulfate-reducing bacteria (Johnson et al., 2021) and their reliance on heme for energy generation might offer clues as to which organisms are most susceptible to viral manipulation with *ahbD* and *bchE*. Though phylogenetic and biochemical characterization is necessary, it is possible that viruses containing such as *ahbD* and *bchE*-like proteins may interact with alkane-degrading processes in the sediment via manipulation of ANME and SRB metabolisms.

Despite the extreme nature of the host metabolisms and the environments in which they are found, the viral AMGs in this study do not appear to have a great deal of specificity to their incubation temperatures or original sampling locations (**Figure 2.7**). The most abundant AMG

in this dataset, annotated as DNA (cytosine-5)-methyltransferase (DNMT, K00558), was nearly ubiquitous across the assembled viromes (**Figure 2.7**) and is hypothesized to be involved in viral evasion of host defenses in a variety of environments (Baranyi et al., 2000; Murphy et al., 2013). Earlier work has shown that AMGs can be specific to an environment (Heyerhoff et al., 2022), though the comparison occurred between sediment and water column samples rather than across geographical locations or temperature. The AMGs found here play general roles in the biosynthesis of amino acids and cofactors, are broadly important for the maintenance of cellular metabolism, and have been previously found in a variety of systems, including methane cold seep sediments (Z. Li et al., 2021; Peng et al., 2023); cow rumen (Yan et al., 2023); compost systems (Chao et al., 2023); Baltic Sea sediments and water column (Heyerhoff et al., 2022); and a variety of marine, freshwater, terrestrial, and engineered systems (Kieft et al., 2021). Though further study is needed, perhaps the selection pressure exerted by the need to find a host in a heterogeneous sediment matrix is greater than the competitive advantage available from specializing to a particular host metabolic lifestyle.

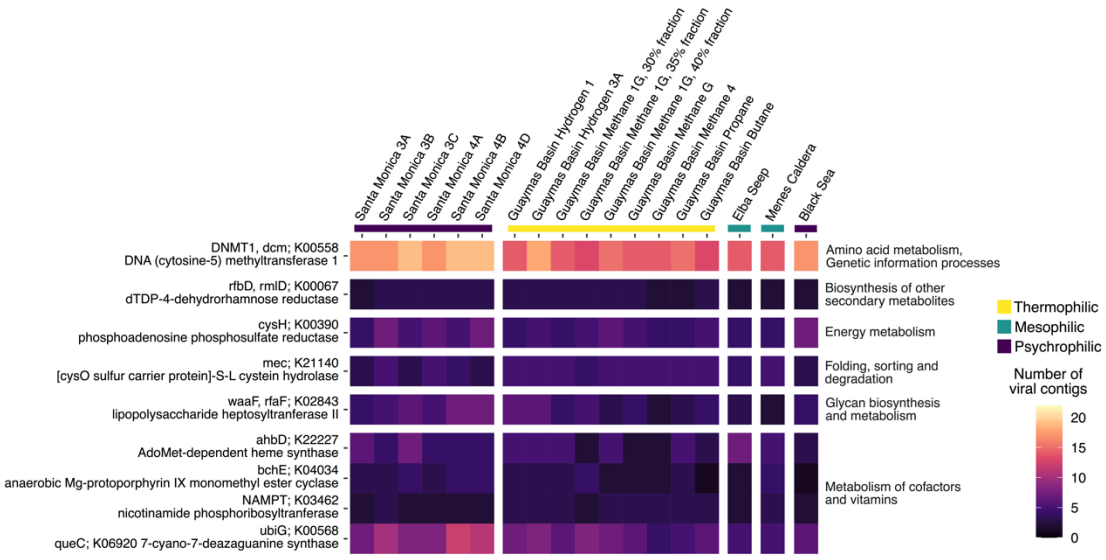

Figure 2.7. Heatmap of the number of contigs in each sample containing the AMG. Contigs were included in each sample's analysis only if the centered log ratio of their abundance was greater than one. AMGs were included in this figure if at least one sample had more than two contigs containing that AMG; the full heatmap is available as **Supplementary Figure S2.6**. Colored lines under each incubation name indicate the broad category of incubation temperature. For Guaymas Basin Methane 1G, we sequenced each density fraction independently after density separation to see if different fractions would yield different AMG abundances; this proved not to be the case.

nosD-like proteins likely involved in adhesion

While examining AMGs, we noticed that NCBI's Conserved Domain Search found similarities between certain ORFs annotated by Vibrant as a gene encoding sugar metabolism and the nitrous oxide reductase D subunit (*nosD*). Canonical *nosD* protrudes into the periplasm and anchored to the *nosY* inner membrane subunit, and shuttles a copper ion to the active site of the *nosZ* subunit of nitrous oxide reductase (Müller et al., 2022). As nitrous oxide reductase plays an important role in bacterial denitrification and nitrogen cycling more broadly, we sought to understand the distribution and potential function of the *nosD* homolog among the viruses in the alkane enrichments. An HMM search, in conjunction with Vibrant annotations, revealed *nosD*-like proteins in four genus-level vMAG clusters and two additional unclustered contigs. No pattern was observed regarding incubation temperature or sampling location; the vMAGs containing *nosD*-like proteins were assembled from Santa Monica Basin (10˚C), Guaymas Basin (methane, propane, and butane incubations, all 50˚C), and the Black Sea (12˚C). We also searched the sediment incubation-based Costa Rica methane seep virome from Philosof et al., *in prep* and found several *nosD*-like proteins within the vMAGs (**Figure 2.8a**).

Notably, none of the *nosD*-like proteins encoded by viral contigs contain the conserved histidine-methionine-methionine motif, a transient binding site for copper ions that may also play a role in sulfur shuttling (Müller et al., 2022; Zumft & Kroneck, 2006). Several of the viral proteins were quite short, between 90-200 amino acids, and understanding their precise function is beyond the scope of this study. However, several longer sequences (>200 amino acids) were found, many of which span the conserved active site of the canonical *nosD* domain but appeared to have lost the His-Met-Met motif, and all of which lack the residues necessary to bind to *nosY* (**Figure 2.8a, Supplementary Figure S2.7**).

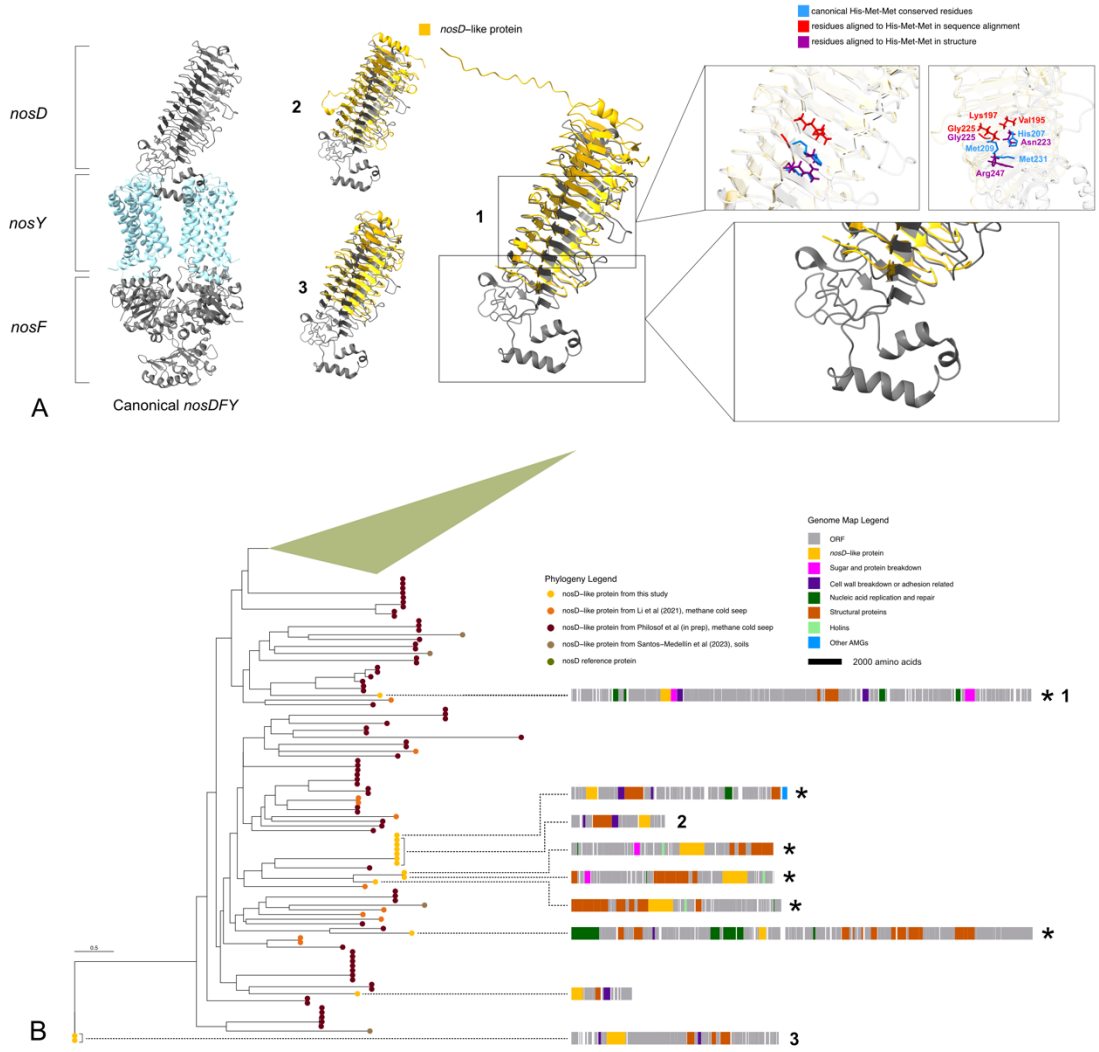

Figure 2.8. Structural and phylogenetic analysis of nosD-like protein. **A)** Three example structures of viral *nosD*-like proteins from this study, shown in yellow and overlaid on the canonical *nosD* (dark grey). On the left is the canonical *nosDFY* structure. Insets on the right highlight the two major structural differences between these nosD-like proteins and the canonical protein. The upper inset panels show the loss of the His-Met-Met motif, with the red residues corresponding to the motif in the sequence alignment, and the purple residues corresponding to the motif in the structural alignment. Gly225 corresponds to an active site residue in both the sequence and structure alignments. The lower panel shows the loss of the structures necessary to bind to the *nosY* subunit, suggesting an alternative role for the *nosD*-like proteins. Structures of all remaining *nosD*-like proteins from this study are available in **Supplementary Figure S2.8**, and those referenced in Figure 2.8b are available from the Caltech Research Data Repository (see Data Availability). Structures were visualized with ChimeraX (Goddard et al., 2018; Meng et al., 2023; Pettersen et al., 2021) **B)** Phylogeny of *nosD*-like proteins from this study, some of which were found using hmmsearch and others of which were annotated by Vibrant, and from other methane cold seep datasets (Li et al., 2021; Philosof et al., *in prep*) and terrestrial soils (Santos-Medellín et al., 2023). The green clade consists of known canonical *nosD* proteins. *nosD*-like proteins from this study are connected by dotted lines to a gene map of all open reading frames found in the assembled contig (in some cases, this does not represent the full length of the assembled

contig as ORFs were not necessarily found across the entire length); categorization of proteins besides the *nosD*-like sequences are based on Vibrant annotations. Approximate length scale of genomes is shown under the Genome Map Legend. Asterisks indicate contigs where the Vibrant-called ORF containing the *nosD*-like protein also contains additional adhesion or sugar breakdown related domains. The numbered genomes (1, 2, and 3) correspond to the structures labeled in figure A.

Prior work has hypothesized that AMGs related to O-antigen synthesis in viruses of methanotrophic and methanogenic archaea are important for enhancing cell-cell adhesion, potentially resulting in both improved host survival rates via aggregation and higher probability that the phage will encounter another host (L. Wang et al., 2022). Such proteins, including *galE* (K01784) and *rfbD* (K00067), are also present in this dataset (**Figure 2.7 and Supplementary Figure S2.6**). We hypothesize that these viral *nosD*-like proteins have a similarly important alternative role related to cell-surface binding and sugar metabolism. The *nosD* domain has known similarities to the CASH domain, which is a widespread carbohydrate and sugar binding domain with homologs to archaeal S-layer proteins (Ciccarelli et al., 2002). The three-sided right-handed beta-helix fold in *nosD*, involved in glycan interactions like depolymerization, is also thought to potentially function in a glycan binding role (Burnim et al., 2024). In several of the viral contigs in which *nosD*-like proteins are found, these proteins are in an ORF that also contains attachment and sugar depolymerization domains that are typically extracellular (**Figure 2.8b**) such as pectate lyase and concanavalin A type lectin/glucanase. Other ORFs on these viral contigs encode additional cell wall and sugar related proteins such as glycosyltransferases involved in cell wall biosynthesis and mannosyltransferases (annotated by Vibrant and confirmed using NCBI Conserved Domain Search).

One notable contig (labeled "1" in **Figure 2.8**) contained a *nosD*-like protein that shares its ORF with a dockerin domain. Other portions of the contig encode an archaeal S-layer protein and additional dockerin domains. Dockerin domains have been previously described in ANME-1 and ANME-2 genomes (Chadwick et al., 2022; Timmers et al., 2017) and are predicted to be involved in cell-cell adhesion and attachment of proteins such as transporters to the outside of the cell. These genes were most closely related to genes from ANME-2 genomes (BLASTp, NCBI non-redundant protein sequences database), suggesting this may be a virus capable of infecting these methanotrophic archaea. Because ANME-2 archaea generally live in tightly packed consortia with sulfate-reducing bacteria, the presence of cell-surface proteins in this virus

near the *nosD*-like protein lends credence to *nosD*-like protein's hypothesized role in adhesion. Another notable *nosD*-like-containing viral cluster putatively infects members of the bacterial class *Desulfotomaculia* (phylum *Firmicutes*, **Figure 2.5**). These observations suggest that viruses encoding *nosD*-like domains are likely associated with diverse microbial hosts.

The *nosD*-like proteins found in these enrichments are variably located in the viral genomes, with some located within areas of high AMG concentration and many located near viral structural and assembly genes such as terminases, prohead proteases, and tail tube proteins (**Figure 2.8a**); this suggests that *nosD*-like proteins take on diverse functions that likely vary among phages. We propose three possible roles for virus-encoded *nosD*-like proteins: 1) *nosD*-like proteins are involved in viral contact and adhesion to the host at the start of infection, 2) *nosD*-like proteins are involved in lysis of the host cell at the end of infection via adhesion and breakdown of the peptidoglycan layer and/or extracellular matrix, and 3) *nosD*-like proteins are active during infection and influence adhesion between the host and other cells, potentially providing viral progeny with immediate access to a new host upon lysis (**Figure 2.9**). The first and second hypotheses are supported by the presence of homologous right-handed beta-helix domains in both *nosD* and phage tailspike proteins, the latter of which are generally polysaccharide lyases and glycosyl hydrolases (Burnim et al., 2024).

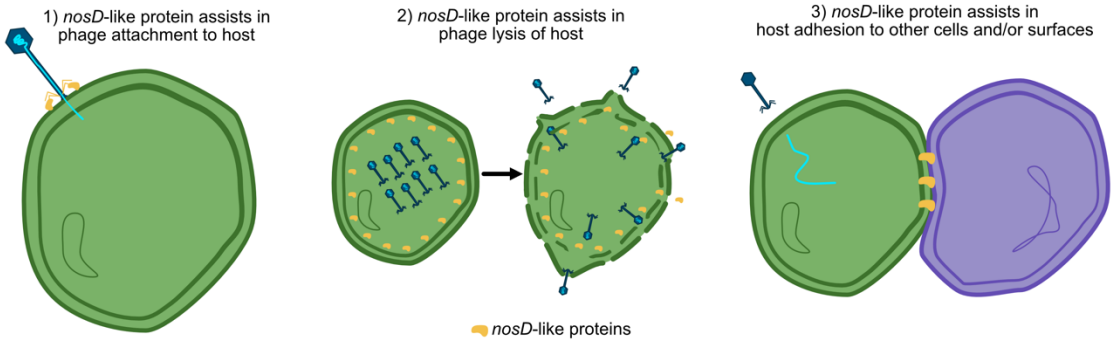

Figure 2.9. Schematic of the three hypotheses for the role of these *nosD*-like proteins. In hypothesis 1, *nosD*-like proteins are involved in the attachment of the phage to the host at the very beginning of infection. In 2, *nosD*-like proteins are expressed near the end of infection to assist with lysis of the host via breakdown of the cell wall. In 3, *nosD*-like proteins play a greater role in cell physiology, influencing host tendency and ability to attach to other cells and surfaces.

Further support for all three hypotheses comes from the lack of evidence for the presence of these longer *nosD*-like proteins in the water column. We searched viral genomes from the Tara Oceans Project (Roux et al., 2016) using our *nosD* HMM. Though several shorter hits were found, none were longer than approximately 200 amino acids, and none reliably spanned the conserved copper-shuttling motif in a multiple sequence alignment against canonical *nosD* proteins. We also examined both sediment and water column phages from the meromictic lake Lac Pavin (Jaffe et al., 2023); though all hits were shorter than 200 amino acids, they were exclusively found among sediment phages. Searches of additional datasets from a previous methane seep study (Z. Li et al., 2021) and terrestrial soils (Santos-Medellín et al., 2023) also yielded possible nosD-like proteins (**Figure 2.8b**). As sediment and soil ecosystems have more opportunities for surface attachment than the water column, viruses in these systems may require additional sugar breakdown and adhesion proteins to infect their hosts due to the extracellular matrix (Sutherland et al., 2004), or they may carry *nosD*-like proteins to influence their hosts' adhesion to other cells and substrates.

CONCLUSIONS AND FUTURE WORK

This study represents the first use of sediment-free enrichment incubations to study the viruses in a variety of slow-growing, hydrocarbon-oxidizing systems. Removing the sediment matrix and selecting for organisms of interest via temperature and substrate greatly reduces the complexity of deep-sea sediment environments and allows for targeted investigations. We find that despite a variety of temperature regimes, sampling locations, and energy sources, there appear to be few adaptations at the level of individual viruses. Though the impacts of incubation temperature and original sampling location cannot be decoupled in this study, these parameters may have more influence at a community composition level.

The discovery of a putative cross-domain virus in this system adds to existing knowledge of sediment-based cross-domain phages. This raises the question of whether cross-domain phages are more common in low-energy sediments than they are in the water column, where the odds of collision with a host are likely higher and higher-energy host metabolisms enable faster reproduction. Given the results above showing that individual viruses are not necessarily adapted to specific environments, it is perhaps not surprising to see that some also lack specificity in their

hosts. Isolation of such phages coupled to long-term evolutionary experiments and targeted analysis of existing phage datasets are necessary for further testing of these hypotheses.

Since our current body of knowledge regarding marine phages is strongly biased towards the upper water column, working in sediment and sediment-derived communities is crucial to understanding the role phages play in these important carbon sinks. The presence of unique *nosD*-like proteins in these sediment-derived enrichments, sediment incubations from methane cold seeps, and terrestrial soils, but their conspicuous absence in water column datasets from the Tara Oceans project suggests novel adhesion roles in communities dominated by aggregation and attachment-based lifestyles. Taken together, the presence of viruses infecting both archaea and bacteria, the lack of pattern in AMG abundances, and the presence of *nosD*-like proteins suggest that viruses are adapted more to the sediment environment itself than to other physicochemical parameters or specific host metabolic processes.

Further investigation of *nosD*-like proteins would be greatly aided by their expression in genetically tractable model systems followed by biochemical characterization. Such investigations may prove interesting to the medical field, where an estimated 80% of human bacterial infections are thought to be biofilm-dominated (Da Silva et al., 2021; Römling & Balsalobre, 2012); searching for nosD-like proteins in viral communities of such infections could shed light on dynamics of biofilm construction and chronic infection. If these *nosD*-like proteins are present in viruses of chronic biofilm-based infections, and depending on which, if any, of the hypothesized functions of *nosD*-like proteins are relevant, we might expect a couple of possible phage-based contributions to biofilm stability. If the *nosD*-like protein is indeed involved in cell-to-cell or cell-to-surface adhesion, phage infection might enhance the physical integrity of the biofilm and prevent its dissolution. However, if the *nosD*-like protein is instead important for adhesion to or lysis of the biofilm-embedded cell, perhaps it can be used in future therapeutics to assist in breaking down the structure of the biofilm.

The presence of viruses in these sediment-free incubations is a reminder that our usual techniques to understand laboratory or environmental communities may result in a less-than-complete picture. As the field of microbial ecology develops more powerful techniques to enrich

and isolate organisms of interest, we should be careful not to overlook the possible contributions of viruses to enrichment composition and stability.

CONTRIBUTIONS

The authors would like to acknowledge the following individuals and groups who helped make this work possible. From the University of Southern California, Dr. Zarko Manojlovic and Ashley Jauregui for performing the metaviromic sequencing; Dr. Daan R. Speth and Dr. Rafael Laso-Pérez for allowing us to use their unpublished genomes in our hunt for host-virus pairs; Dr. Igor Antoshechkin at the Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech for library preparation and sequencing of the cellular metagenomes; Dr. Hang Yu for initial incubation development from Santa Monica and long-term maintenance of other laboratory enrichments used in this study; Anja Seebeck and Susanne Menger at the University of Bremen for help with cultivation and incubation maintenance; the crews of the RV Poseidon exp. 317/3, RV Atlantis AT15-45, RV Atlantis AT37-06, RV L'Atalante NAUTINIL Expedition, and the MBARI RV Western Flyer 2013 Southern California Expedition; Hydra Station in Fetovaia, Italy; Dr. Daniel Utter for bioinformatics advice, especially in the use of the DADA2 workflow; and Dr. Justus W. Fink and Dr. Jeremy E. Schreier for manuscript feedback.

FUNDING

1) The NOMIS Foundation, Switzerland. 2) This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0020373 and DE-SC0022991. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DATA AVAILABILITY

Raw fastqs from the metagenome, metavirome, and 16S rRNA sequencing are under NCBI PRJNA1226641. All MAGs generated in this study for which a virus CRISPR or tRNA match was found, all 989 concatenated viral contigs, viral protein annotations, *nosD*-like sequences used in 2.8, pdb files of all *nosD*-like sequences from Figure 2.8, and the *nosD* HMM file are available on the Caltech Library Data Repository at doi: 10.22002/7sp3k-ewh62.

SUPPLEMENTAL DATA

Sample	Sampling Lat, Long	Incubation Temp. (˚C)	Energy substrate	Collection year	Start of incubation	References for media recipe
Santa Monica 3A	33.78888889, -118.6680556	10	0.2MPa Methane	2013	2013	Yu et al (2022), *AEM*
Santa Monica 3B	33.78888889, -118.6680556	10	0.2MPa Methane	2013	2013	Yu et al (2022), *AEM*
Santa Monica 3C	33.78888889, -118.6680556	10	0.2MPa Methane	2013	2013	Yu et al (2022), *AEM*
Santa Monica 4A	33.78888889, -118.6680556	10	0.2MPa Methane	2013	2013	Yu et al (2022), *AEM*
Santa Monica 4B	33.78888889, -118.6680556	10	0.2MPa Methane	2013	2013	Yu et al (2022), *AEM*
Santa Monica 4D	33.78888889, -118.6680556	10	0.2MPa Methane	2013	2013	Yu et al (2022), *AEM*
Guaymas Basin Hydrogen 1A	27.01213889, -111.4152222	60	0.2MPa 80:20 Hydrogen:CO_2	2009	2012	Krukenberg et al (2016), *Environ. Microbiol.*, Methods from this study
Guaymas Basin Hydrogen 3	27.01213889, -111.4152222	60	0.2MPa 80:20 Hydrogen:CO_2	2009	2012	Krukenberg et al (2016), *Environ. Microbiol.*, Methods from this study
Guaymas Basin Methane 1G	27.01666667, -111.6833333	50	0.2MPa Methane	2009	2010	Holler et al (2011), *ISME,* Methods from this study
Guaymas Basin Methane G	27.01666667, -111.6833333	50	0.2MPa Methane	2009	2010	Holler et al (2011), *ISME*, Methods from this study
Guaymas Basin Methane 4	27.01666667, -111.6833333	50	0.2MPa Methane	2009	2010	Holler et al (2011), *ISME*, Methods from this study

Guaymas Basin Propane	27.01213889, -111.4152222	50	0.15MPa Propane	2009	2010	Laso-Pérez et al (2016), *Nature*
Guaymas Basin Butane	27.01213889, -111.4152222	50	0.15MPa Butane	2009	2010	Laso-Pérez et al (2016), *Nature*
Elba Seep	42.73507778, 10.2	20	0.2MPa Methane	2010	2010	Ruff et al (2016), *Front. microbiol.*
Menes Caldera 2	32.30091667, 28.43519444	Room temperature	0.2MPa Methane	2003	2003	Wegener et al (2016), *Front. microbiol.*
Menes Caldera 2A	32.30091667, 28.43519444	Room temperature	0.2MPa Methane	2003	2003	Wegener et al (2016), *Front. microbiol.*
Black Sea	44.77777778, 31.97222222	12	0.2MPa Methane	2000	2004	Wegener et al (2016), *Front. microbiol.*

Table S2.1. Incubation bottle metadata. All bottles include sulfate as an electron acceptor. A full description of how the enrichment cultures (excepting Santa Monica) were set up is contained in Laso-Pérez et al., 2018. Santa Monica enrichments are described in Yu et al., 2022.

Sample	**Organism (Family level)**	**Relative Abundance (Family level)**	**Number of ASVs**	**ASVs comprising >1% of reads in the sample (% total reads in sample)**
Santa Monica 3A	Fermentibacteraceae	17.4%	13	ASV_16 (13.6%), ASV_15 (1.4%)
	ANME-2a-2b	16.8%	17	ASV_10 (14.7%)
	Fusibacteraceae	16.5%	7	ASV_40 (14.1%)
	ANME-2c	9.9%	2	ASV_12 (8.1%), ASV_81 (1.8%)
	Dissulfuribacteraceae	6.0%	8	ASV_25 (4.1%)
	Lentimicrobiaceae	5.3%	13	ASV_7 (4.8%)
	AKAU3564 sediment group	5.0%	20	ASV_193 (1.6%), ASV_224 (1.2%)
	Marinilabiliaceae	4.7%	6	ASV_50 (3.4%)
Santa Monica 3B	ANME-2a-2b	27.1%	9	ASV_10 (19.8%), ASV_6 (6.1%)
	OM190	18.9%	5	ASV_32 (11.8%), ASV_22 (2.5%), ASV_124 (2.1%), ASV_128 (2.0%)
	Marinilabiliaceae	11.3%	10	ASV_43 (7.2%), ASV_50 (3.4%)
	JS1	9.2%	7	ASV_14 (5.4%), ASV_58 (2.8%)
	Fermentibacteraceae	8.5%	7	ASV_15 (5.8%), ASV_31 (2.3%)
	Desulfosarcinaceae	6.6%	18	ASV_45 (3.3%), ASV_147 (1.2%)
Santa Monica 3C	Lentimicrobiaceae	45.0%	30	ASV_7 (43.9%)

	ANME-2a-2b	26.3%	6	ASV_6 (25.7%)
	Fermentibacteraceae	10.6%	11	ASV_15 (9.7%)
	Anaerolineaceae	3.1%	12	ASV_86 (2.9%)
	Desulfosarcinaceae	3.1%	18	ASV_45 (1.9%)
Santa Monica 4A	ANME-2c	31.0%	7	ASV_12 (16.7%), ASV_74 (6.7%), ASV_77 (6.3%)
	Fermentibacteraceae	9.8%	7	ASV_31 (9.6%)
	OM190	9.7%	5	ASV_22 (8.9%)
	Unclassified Methanosarciniales	8.4%	8	ASV_57 (8.2%)
	Marine Benthic Group D and DHVEG-1	5.7%	11	ASV_49 (4.0%)
	Lokiarchaeia	3.7%	11	ASV_60 (3.5%)
	Desulfosarcinaceae	3.6%	10	ASV_114 (2.2%)
	Dissulfuribacteraceae	3.6%	4	ASV_25 (2.3%)
	Unclassified Phycisphaerae	3.2%	6	ASV_110 (2.8%)
Santa Monica 4B	Dissulfuribacteraceae	19.0%	6	ASV_25 (8.7%), ASV_27 (3.6%), ASV_79 (3.4%), ASV_82 (3.2%)
	ANME-2c	18.0%	5	ASV_12 (15.5%), ASV_81 (2.1%)
	Fermentibacteraceae	13.0%	17	ASV_16 (11.8%)
	ANME-2a-2b	8.5%	21	ASV_10 (3.7%), ASV_6 (3.4%)
	OM190	6.5%	4	ASV_22 (5.4%)
	Lentimicrobiaceae	5.5%	19	ASV_7 (1.2%)
	Marine Benthic Group D and DHVEG-1	3.8%	22	ASV_49 (1.3%)
	JS1	3.6%	5	ASV_14 (2.9%)
Santa Monica 4D	Dissulfuribacteraceae	18.4%	8	ASV_25 (5.8%), ASV_27 (5.7%), ASV_79 (3.5%), ASV_82 (3.3%)
	ANME-2c	11.6%	5	ASV_12 (9.0%), ASV_81 (1.7%)
	ANME-2a-2b	10.4%	17	ASV_10 (5.0%), ASV_6 (3.3%)
	Marine Benthic Group D and DHVEG-1	9.3%	21	ASV_49 (2.8%), ASV_28 (1.2%)
	Fermentibacteraceae	8.2%	11	ASV_16 (7.5%)
	Lentimicrobiaceae	8.0%	21	ASV_7 (2.8%), ASV_164 (1.0%)
	OM190	5.7%	4	ASV_22 (4.8%)

	JS1	4.1%	5	ASV_14 (3.3%)
	AKAU3564 sediment group	3.3%	17	None (all ASVs <1%)
Menes Caldera 2	Lentimicrobiaceae	28.7%	7	ASV_17 (27.6%)
	Marine Benthic Group D and DHVEG-1	25.6%	27	ASV_28 (9.2%), ASV_69 (6.4%), ASV_80 (4.0%), ASV_67 (2.0%), ASV_89 (1.6%)
	Fermentibacteraceae	7.2%	4	ASV_39 (7.1%)
	Anaerolineaceae	5.0%	22	ASV_64 (2.9%)
	Dissulfuribacteraceae	4.8%	7	ASV_87 (4.2%)
	JS1	4.4%	5	ASV_14 (3.7%)
	Calditrichaceae	3.5%	6	ASV_41 (1.5%), ASV_150 (1.2%)
Menes Caldera 2A	ANME-2a-2b	25.7%	9	ASV_36 (15.6%), ASV_10 (4.7%), ASV_100 (3.2%), ASV_184 (1.3%)
	Marine Benthic Group D and DHVEG-1	21.2%	26	ASV_28 (7.2%), ASV_67 (3.1%), ASV_89 (2.3%), ASV_80 (2.2%), ASV_131 (1.3%), ASV_49 (1.1%)
	Unclassified Plactomycetota	9.3%	3	ASV_68 (8.6%)
	Lentimicrobiaceae	9.0%	3	ASV_85 (5.2%), ASV_17 (3.8%)
	Fermentibacteraceae	7.4%	4	ASV_39 (7.4%)
	Anaerolineaceae	7.3%	18	ASV_64 (6.4%)
	Calditrichaceae	5.5%	5	ASV_41 (5.1%)
	JS1	4.4%	4	ASV_14 (4.2%)
Black Sea	JS1	23.0%	14	ASV_24 (17.1%), ASV_14 (3.7%)
	Desulfosarcinaceae	18.7%	21	ASV_23 (18.4%)
	ANME-1b	13.4%	2	ASV_35 (13.2%)
	Anaerolineaceae	7.6%	29	ASV_54 (6.2%)
	Pirellulaceae	7.0%	24	ASV_91 (4.6%)
	Fermentibacteraceae	5.6%	23	ASV_16 (1.8%), ASV_202 (1.1%)
	ANME-1	4.4%	12	ASV_3 (4.2%)
	ANME-2a-2b	3.6%	15	ASV_133 (2.3%)
Elba Seep	ANME-1a	44.7%	1	ASV_5 (44.7%)
	Dissulfuribacteraceae	10.1%	11	ASV_33 (6.3%), ASV_27 (3.3%)
	AKAU3564 sediment group	7.1%	11	ASV_38 (5.9%)
	Desulfosarcinaceae	5.2%	11	ASV_53 (4.2%)

	Unclassified Bacteria	5.1%	7	ASV_44 (5.0%)
	ANME-2c	4.4%	1	ASV_51 (4.4%)
	ANME-2a-2b	3.6%	2	ASV_66 (3.6%)
Guaymas Basin Methane 1G	Bathyarchaeia	34.2%	30	ASV_8 (17.1%), ASV_9 (14.8%)
	ANME-1	18.8%	12	ASV_3 (15.8%), ASV_47 (2.8%)
	Aerophobales	16.5%	7	ASV_13 (16.3%)
	Desulfofervidaceae	7.9%	11	ASV_30 (4.1%), ASV_2 (3.1%)
	Thermodesulfobacteriaceae	7.5%	2	ASV_19 (7.5%)
	GN01	3.0%	1	ASV_56 (3.0%)
Guaymas Basin Methane G	Bathyarchaeia	35.9%	24	ASV_8 (17.8%), ASV_9 (10.8%), ASV_55 (5.0%)
	Desulfofervidaceae	13.2%	12	ASV_2 (7.2%), ASV_30 (5.0%)
	Aerophobales	12.2%	8	ASV_26 (10.1%), ASV_13 (1.7%)
	ANME-1	7.4%	4	ASV_3 (4.3%), ASV_47 (3.1%)
	Thermodesulfobacteriaceae	6.1%	2	ASV_19 (5.8%)
	Aminicenantales	5.0%	4	ASV_72 (2.8%), ASV_11 (2.1%)
	Caldatribacteriaceae	3.6%	1	ASV_76 (3.6%)
Guaymas Basin Methane 4	ANME-1	53.8%	16	ASV_3 (52.1%)
	Desulfofervidaceae	29.0%	13	ASV_2 (28.2%)
	Bathyarchaeia	10.3%	24	ASV_9 (4.7%), ASV_8 (4.5%)
Guaymas Basin Hydrogen 1	Archaeoglobaceae	93.3%	3	ASV_1 (93.0%)
	Desulfofervidaceae	4.5%	1	ASV_2 (4.5%)
Guaymas Basin Hydrogen 3A	Archaeoglobaceae	74.9%	2	ASV_1 (74.8%)
	Desulfofervidaceae	22.7%	5	ASV_2 (22.5%)
Guaymas Basin Propane	Desulfofervidaceae	25.3%	22	ASV_2 (23.3%)
	Syntrophoarchaeaceae	13.6%	6	ASV_20 (13.3%)
	JS1	8.7%	4	ASV_21 (8.6%)
	Marine Benthic Group D and DHVEG-1	4.1%	5	ASV_102 (2.2%), ASV_144 (1.2%)
Guaymas Basin Butane	Aminicenantales	23.4%	4	ASV_11 (23.4%)
	Desulfofervidaceae	9.9%	9	ASV_2 (9.6%)
	Syntrophoarchaeaceae	3.7%	2	ASV_107 (2.2%), ASV_20 (1.5%)
	JS1	3.4%	3	ASV_21 (3.4%)

Table S2.2. ASVs comprising at least 1% of total reads in a given sample, with their family-level classification. Only families >3% relative abundance are shown. The majority of families are comprised of no more than two ASVs.

	Richness		Simpson's Evenness		EQ		Evar	
Sample	**Viral**	**Cellular**	**Viral**	**Cellular**	**Viral**	**Cellular**	**Viral**	**Cellular**
Santa Monica 3A	313	259	0.0032	0.0516	0.0603	0.1009	0.0733	0.1894
Santa Monica 3B	395	230	0.0071	0.0580	0.0503	0.0913	0.0495	0.1595
Santa Monica 3C	384	209	0.0160	0.0177	0.0488	0.0948	0.0457	0.1815
Santa Monica 4A	412	260	0.0155	0.0565	0.0547	0.0957	0.0591	0.1810
Santa Monica 4B	422	363	0.0174	0.0490	0.0504	0.1094	0.0493	0.2320
Santa Monica 4D	417	342	0.0239	0.0873	0.0510	0.1128	0.0500	0.2345
Guaymas Basin Hydrogen 1	333	40	0.0376	0.02880	0.0614	0.0759	0.0750	0.1156
Guaymas Basin Hydrogen 3A	319	66	0.03382	0.02478	0.0621	0.0889	0.0772	0.2014
Guaymas Basin Methane 1G, 30% optiprep	331	N/A	0.0041	N/A	0.0590	N/A	0.0694	N/A
Guaymas Basin Methane 1G, 35% optiprep	325	N/A	0.0042	N/A	0.0522	N/A	0.0567	N/A
Guaymas Basin Methane 1G, 40% optiprep	320	N/A	0.0094	N/A	0.0522	N/A	0.0551	N/A
Guaymas Basin Methane G	330	N/A	0.00642	N/A	0.0515	N/A	0.0533	N/A
Guaymas Basin Methane 4	240	116	0.0275	0.0224	0.0505	0.0952	0.0505	0.1764
Guaymas Basin Propane	287	124	0.0040	0.0413	0.0585	0.0894	0.0714	0.1467
Guaymas Basin Butane	260	101	0.0092	0.0272	0.0577	0.0887	0.0711	0.1699
Elba Seep	376	202	0.0257	0.0223	0.0527	0.0905	0.0554	0.1682
Menes Caldera 2	321	204	0.0110	0.0461	0.0492	0.0931	0.0492	0.1656
Menes Caldera 2A		197		0.0812		0.0904		0.1678
Black Sea	377	460	0.0402	0.0234	0.0450	0.1117	0.0398	0.2455

Table S2.3. Diversity metrics for host and viral communities.

	Host 16S NMDS, ASV level	Host 16S NMDS, Species level	Viral PCA, Viral genus level	Viral PCA, Viral contig level
Unique energy source	0.00039996	0.00089991	0.9259074	0.9429057
Methane vs non-methane energy source	0.00179982	0.00189981	0.39986	0.4427557
Temperature lifestyle	9.999e-05	0.00019998	9.999e-05	9.999e-05
Original sampling location	9.999e-05	9.999e-05	9.999e-05	9.999e-05

Table S2.4. ADONIS values for ordination plots in **Figure 2.3**.

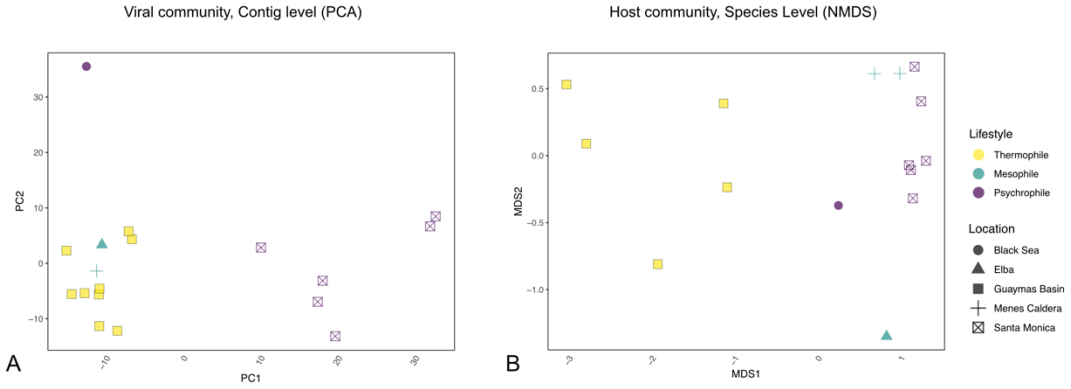

Figure S2.1. Ordination of viral and prokaryotic communities. **A)** Viral CLR-transformed abundance-based PCA, at individual viral contig level. PC1 importance: 0.3511, PC2 importance: 0.1547 **B)** 16S relative abundance-based NMDS, at species taxonomy level. Stress=0.076.

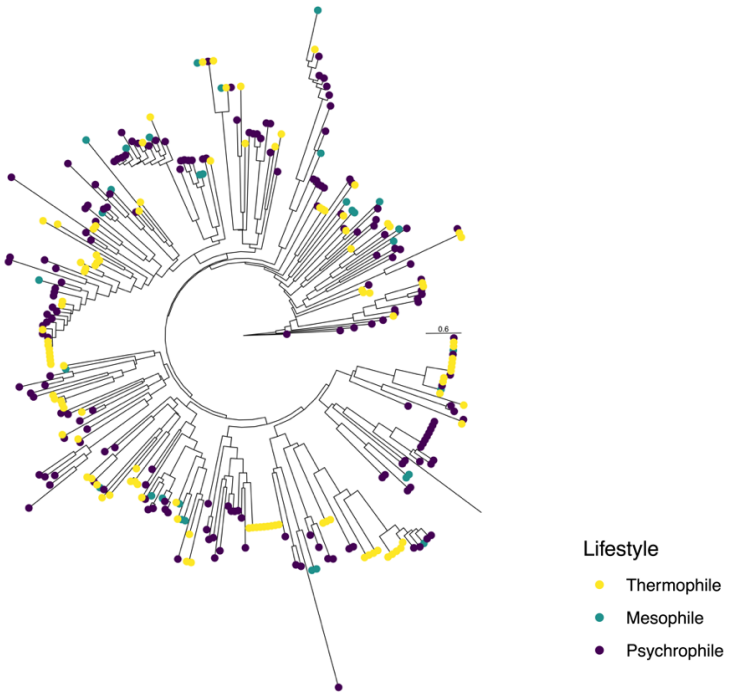

Figure S2.2. Major Capsid Protein Tree, colored by incubation temperature.

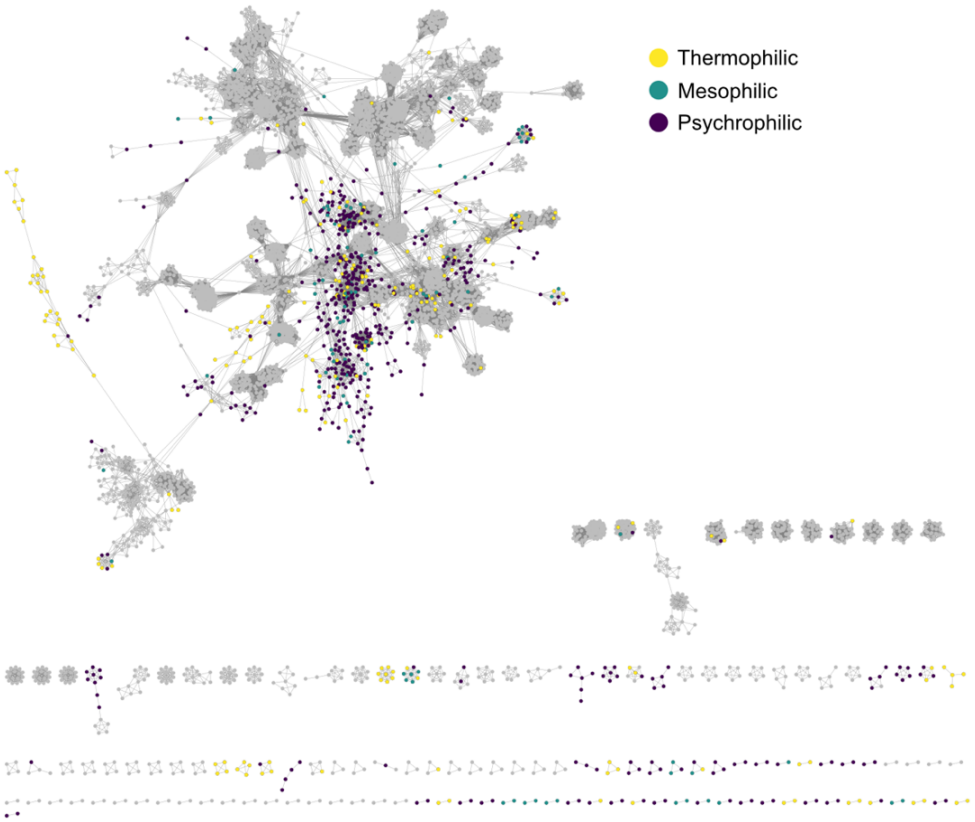

Figure S2.3. Complete vConTACT2 network

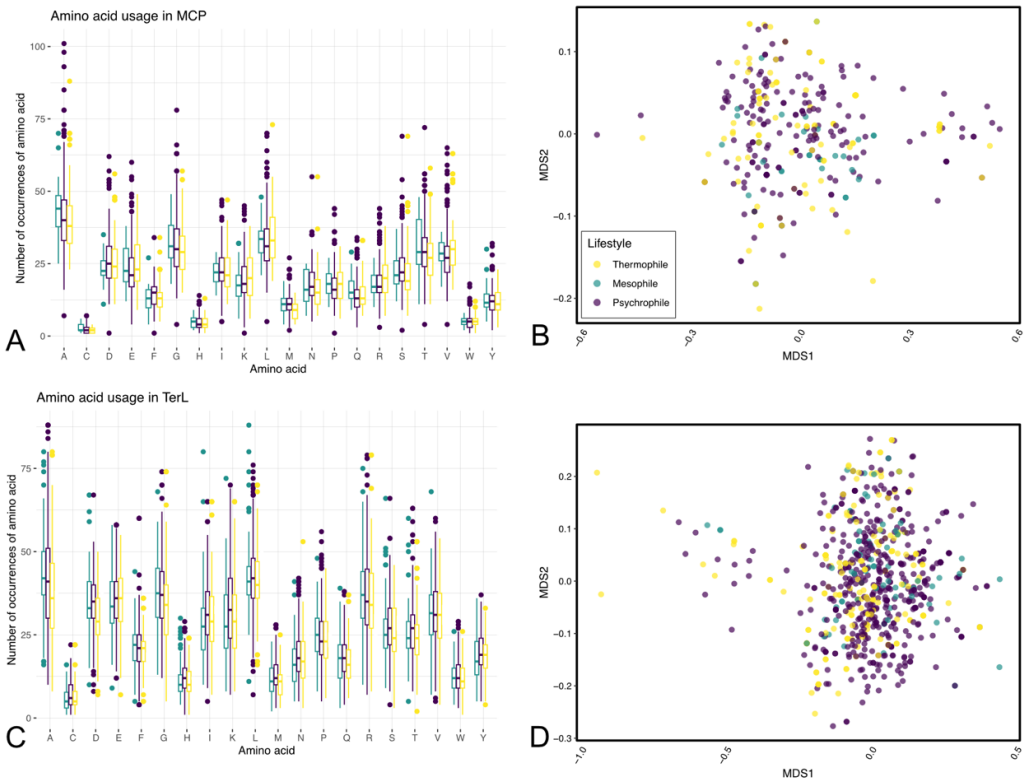

Figure S2.4. Amino acid composition of major capsid proteins (MCP) and terminase large subunits (TerL). **A)** Boxplot showing the number of occurrences of a given amino acid in viral MCP proteins. MCP proteins are grouped by their incubation temperature categories. **B)** Amino acid abundance-based NMDS (Bray-Curtis dissimilarity) of viral MCP proteins, colored by their incubation temperature. Stress=0.138. **C)** Boxplot showing the number of occurrences of a given amino acid in viral TerL proteins. **D)** Amino acid abundance-based NMDS (Bray-Curtis dissimilarity) of viral TerL proteins, colored by their incubation temperature. Stress=0.116. Figures B and D show no apparent bias in amino acid composition based on incubation temperature.

Amino Acid	Mann-Whitney Statistic	p-value	Protein
Valine	11812	0.00180526	MCP
Asparagine	7268.5	0.00039737	MCP
Methionine	6784	2.04e-05	MCP
Glycine	37509.5	6.98e-05	TerL
Aspartic acid	28836.5	7.71e-15	TerL
Threonine	35769	2.13e-06	TerL
Serine	39629.5	0.00228157	TerL
Histidine	37199.5	3.75e-05	TerL
Glutamine	38518	0.00040036	TerL
Methionine	33382	6.92e-09	TerL

Table S2.5. Significantly different amino acids between phages from thermophilic (50–60˚C) and psychrophilic (10–12˚C) incubations. Mann-Whitney U test with Bonferroni-corrected p-value cutoff, $p<0.0025$ (0.05/20 amino acids, with the alternative hypothesis that psychrophilic and thermophilic phages are indeed different in amino acid use). The protein column indicates whether the difference was seen in the TerL protein or the MCP protein. **Note:** Methionine is the start codon. There may be assembly biases related to the quality of sequencing libraries or DNA extraction from different incubations that falsely bias quantification of methionine.

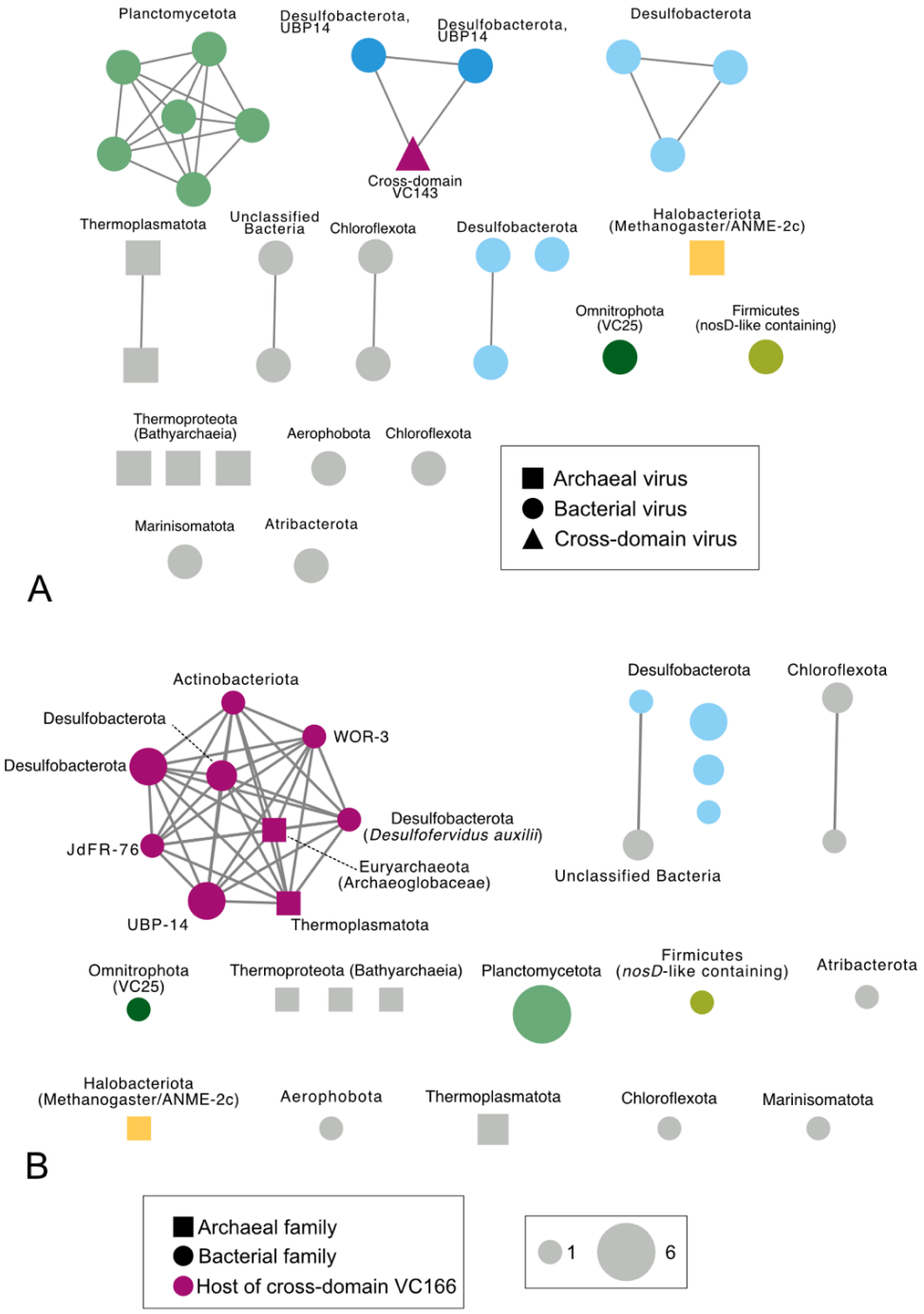

Figure S2.5. Full network of host-phage pairings. **A)** Network with phage genus-level clusters as nodes and shared host at the family level as edges. Nodes are labeled by host phylum; any viral cluster infecting *Desulfobacterota* is colored in shades of blue. **B)** Network with hosts at family taxonomic level as nodes, and shared viral clusters as edges. Nodes are labeled by host phylum. Magenta nodes are families that are infected by broad host range viral cluster VC143, though VC143 may not be the only virus infecting that family. Node size corresponds to the number of viral clusters matched to each family via CRISPR spacers or tRNAs.

ORF	Vibrant Annotation	CD Search	GTDB r214 BLASTp top phyla	NCBI BLASTp top phyla
45	K02843 (*waaF*/*rfaF*)	*rfaF* superfamily (COG0859)	Desulfobacterota	Ca. Methanomethylicota, Thermodesulfobacteriota
57	K12547 (polysaccharidase *plyA*)	*nosD* (cl34609) with a choice-of-anchor Q domain (cl49486)	KSB1, Fermentibacterota, OLB16	KSB1, Calditrichota Patescibacteria group
74	K03269 (*lpxH*)	Metallophosphatase superfamily (cl13995 and cl42652)	Thermoplasmatota, UBP14	Bathyarchaeota, Thorarchaeota
82	K01448 (*amiABC*)	*amiC* family (COG0860)	Methanobacteriota, Bacillota	Thermodesulfobacteriota, Euryarchaeota
112	Hypothetical protein	Uncharacterized conserved protein w/ vonWillibrand factor type A (cl27002)	Spirochaetota	Synergistota, Euryarchaeota, Ignavibacteriota
109	K03148 (*thiF*)	E1 activating enzymes of ubiquitin-like proteins (cl22428)	Myxococcota	Euryarchaeota
16	K07391 (*comM*)	*YifB* superfamily (cl33973)	Omnitrophota	Deltaproteobacteria
84	K00558 (DNMT1)	*Dcm* DNA-cytosine methylase (COG0270)	Halobacteriota	Nitrospirota

Table S2.6. BLAST and conserved domain search results at the phylum level for open reading frames from contig f9bf2df06235b506f7396f56f61038ed in VC143. Though the *plyA* annotation was found to have some conserved domain similarities to *nosD*, it was not found to be a high quality *nosD*-like sequence in a search with the *nosD* HMM.

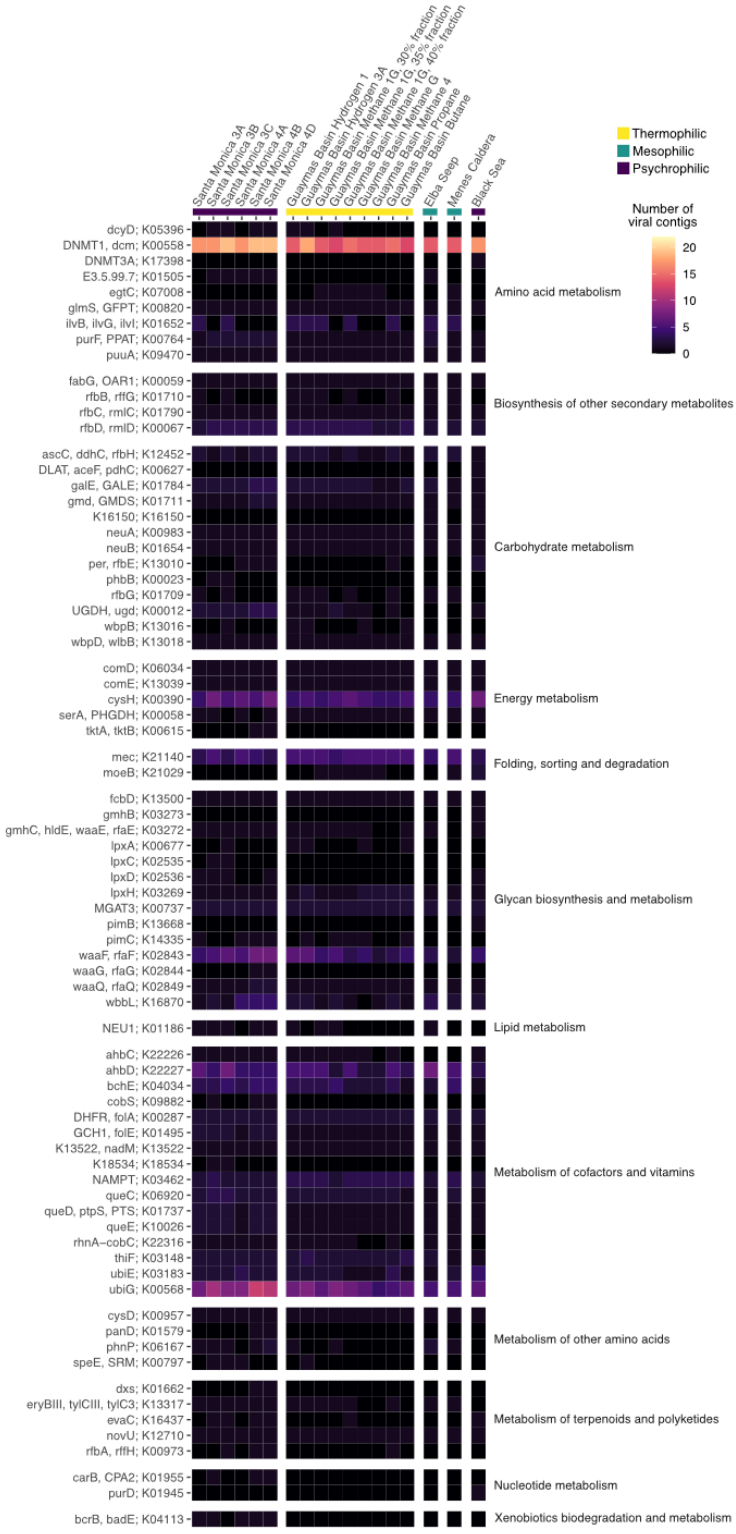

Figure S2.6. Complete list of AMGs and their abundance in each sample. See main text **Figure 2.7** for information on abundance calculations.

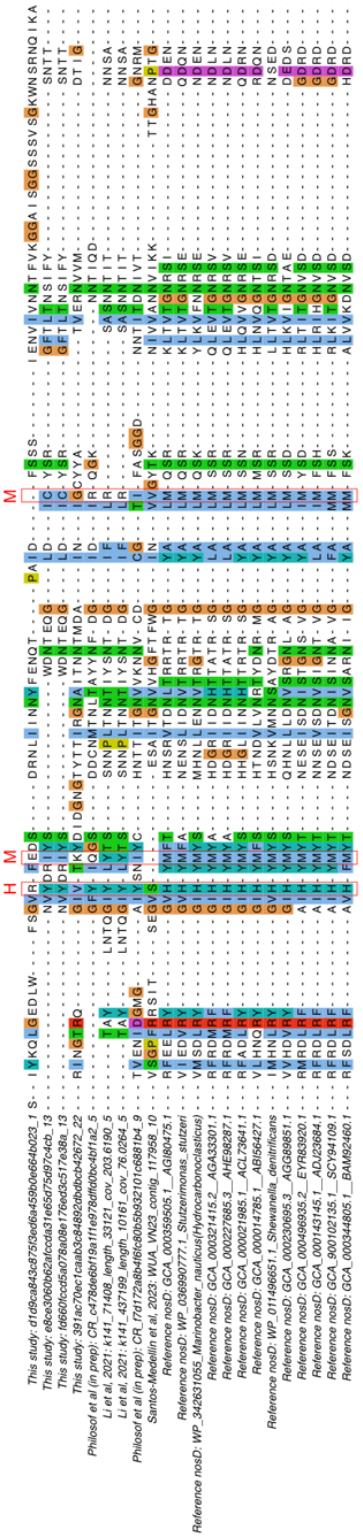

Figure S2.7. Examples of *nosD*-like proteins spanning the conserved His-Met-Met motif (outlined with red boxes). Sequence names begin with the provenance of the sequence, with "Reference *nosD*" referring to canonical *nosD* proteins and all others referencing the study of origin.

Figure S2.8. Structures of *nosD*-like proteins from this study, labeled by the fasta header of the viral contig from which they came. The *nosD*-like protein is in yellow and is superimposed on the structure of the canonical *nosD* in grey.

Chapter 3

DISTRIBUTION PATTERNS AND INFECTION NETWORKS OF THE VIRAL COMMUNITY IN A COSTA RICA COLD SEEP

Alon Philosof (1), Aditi K. Narayanan (1), Ranjani Murali (2), Victoria J. Orphan (1,3)

1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA
2. School of Life Sciences, University of Nevada Las Vegas, USA
3. Division of Geology and Planetary Sciences, California Institute of Technology, Pasadena, California, USA

ABSTRACT

Methane seep sediments are highly spatially complex and often characterized by steep geochemical gradients based on depth in the sediment. Until now, correlating the changes in the viral community with these geochemical gradients and across sampling sites has not been attempted due to a lack of data. Here, we present a detailed metaviromics-based investigation of viruses within deep sea sediments supporting active anaerobic methane oxidation (AOM); this dataset is the most comprehensive picture of the viral community in a methane seep to date. Through direct sequencing of viruses from laboratory sediment microcosms, we recovered thousands of complete or nearly complete viral contigs, with the majority from viral families that have not been previously described. The viral communities, collected from four sediment pushcores sampled within methane seeps off of Costa Rica between 991-1811m below sea surface, tended to vary more with sampling location than with sediment depth. Each pushcore also contained a "core" viral community that was consistently present despite variation in sediment depth and cellular microbial community, possibly underscoring the role of fluid advection in seep virus distribution. As in the sediment-free enrichment dataset described in Chapter 2, auxiliary metabolic genes do not appear to be highly specific to methane-oxidizing or sulfate-reducing metabolism and are highly enriched in lipopolysaccharide synthesis and other sugar metabolism pathways. With the results of Chapter 2, we propose that viral communities of the methane seep are less adapted to the specific metabolic niches of their hosts or the local geochemical conditions; instead, adaptation is driven primarily by the complexity of the sediment itself and by the prevalence of low energy metabolisms.

INTRODUCTION

In the previous chapter, we looked across multiple alkane-rich deep ocean habitats and found that viral communities retained in long-term alkane-oxidizing enrichment cultures show a great deal of variation in their composition across geographic location and temperature. In contrast, comparisons among individual viral taxa rather than at the level of community composition showed no signals in response to these parameters. Surprisingly, we also found that the distribution and types of auxiliary metabolic genes in the viral genomes hardly varied between the different sampling sites, even though we might have expected that incubation temperature or energy substrate might have affected the metabolic pathways that would be most advantageous for viral genomes to carry, or that geographic bottlenecks might have resulted in greater divergence. Functionally speaking, we found a novel means by which viruses might influence cell-cell adhesion, a particularly interesting role in light of the cellular aggregation necessary for the syntrophic interaction between anaerobic methanotrophic aarchaea nd sulfate reducing bacteria.

We will now both narrow and broaden our scope. This chapter examines the viral community of a single location of cold methane seepage off the west coast of Costa Rica; however, we examine samples from multiple sampling sites and attempt to understand viral distribution across these sites and with increasing sediment depth. We utilize pushcores collected from a 2017 research cruise in the waters west of Costa Rica (see Appendix II for general information on pushcore collection); by sampling from methane-fed microcosm incubations that preserve the complexity of the sediment community, we create a more ecologically relevant picture of viral communities in methane cold seeps.

We present the most comprehensive picture of the methane cold seep viral community to date, from eighty viral sequencing libraries from twenty sediment microcosm incubations and two additional water column samples. These libraries produced thousands of complete viral genomes from hundreds of novel viral families. We find that viral community composition is more similar among seep sampling location rather than sediment depth the sampling site matters more to viral community composition than depth in the sediment, e.g., the topmost sediment horizons are more similar to the deeper horizons of their respective cores than they are to each other. We

also find high abundances of AMGs related to lipopolysaccharide synthesis and sugar metabolism, consistent with the previous chapter and with prior research in methane seep viruses (Z. Li et al., 2021; Zhong et al., 2024). The presence of cross domain viruses, as in Chapter 2, hints at a possible viral adaptation to the slow growth and physical complexity of the sediment ecosystem. Comparisons to the overlying water column indicate that sediment communities are distinct and must be better incorporated into marine viral ecology.

MATERIALS AND METHODS

Many of the molecular and analytical methods are as described in Chapter 2. Exceptions and differences are described below.

Sample collection

Sediment pushcores were collected in 2017 and processed as described in Mullin, 2020. Sediment horizons were stored in an artificial seawater medium (see Appendix II) in glass Pyrex bottles (Cat. #1395-500, Corning Life Sciences, Corning, NY, USA), sealed with butyl rubber stoppers, and kept at 4˚C with a 0.14MPa methane headspace. A brief explanation of how pushcores are collected from the seafloor is also available in Appendix II.

Cellular DNA extraction and sequencing

Whole genome library preparation and sequencing was performed at the University of Southern California Keck School of Medicine using an S1 flow cell for 300 cycles on the Illumina NovaSeq6000 sequencing platform (v1.5 reagents).

Clustering of cellular community data

NMDS analysis used the python scikit-learn MDS function with the Bray-Curtis dissimilarity metric and the following parameters: n_components=3, dissimilarity='precomputed', n_init=35, random_state=42, max_iter=500, eps=1e-6, metric=False.

Clustering of viral community data

Hierarchical clustering of viral communities was performed using the R package DynamicTreeCut (Langfelder et al., 2008) with the cutreeDynamic method on centered log ratio (CLR)-transformed abundance data.

Host-virus pair network calculation

The host-virus infection network was constructed using the networkX (Hagberg et al., 2008) community.greedy_modularity_communities function on binary host-phage interaction tables. This function implements the Clauset-Newman-Moore greedy modularity maximization algorithm for detecting communities in networks (Clauset et al., 2004). The function iteratively merges communities to maximize the modularity gain until no more modularity improvement is possible or all nodes are in one community. The sizes of nodes were scaled based on the number of interactions.

Linear regression analysis

We plotted the pairwise comparison of CLR abundance between horizon groups (top, middle, bottom) within the same pushcore and calculated the linear regression. The modified Z-score was calculated as follows:

Modified z-score = $0.6745(x_i - \tilde{x})/MAD$

where

- x_i: A given data value
- $\tilde{x}$: The median of the dataset. In this case, we first calculate the rank abundance of every vOTU in each dataset (top, middle, or bottom). We then do a pairwise comparison of each vOTU's rank between two datasets (top vs middle, middle vs bottom, top vs bottom) and calculate the difference in rank for each vOTU, using the dataset on the x-axis (see **Figure 3.4**) as the reference values. $\tilde{x}$ is then the median of the change in rank of each vOTU.
- MAD = median($|x - \tilde{x}|$); in other words, the median of the absolute value of every data point's deviation from the data median.
- 0.6745 is the scaling factor used when the underlying distribution is half-normal, which is the closest approximation of the rank abundance curves.

A modified z-score of >=3.5 was considered significant, as it is more robust and less sensitive to outliers (Iglewicz & Hoaglin, 1993). Linear regression was calculated with the function stat_smooth in R, with the method lm (linear model).

RESULTS AND DISCUSSION

Sample and Site description

Sediment pushcores from Costa Rica were collected in 2017 on the *R/V Atlantis* cruise AT37-13 (**Figure 3.1a**, Costa Rica Seeps Project, Woods Hole Oceanographic Institute), and incubated in Pyrex bottles in the laboratory from 2017-2021. The cores were sectioned and incubated in 3cm horizons, as described in Chapter 1; incubations are kept at 4˚C, as close to *in situ* temperature as possible, under 0.2MPa of 100% methane headspace. The sulfate-containing media (Appendix II) and headspace were replaced every three months to maintain active anaerobic methane oxidation coupled to sulfate reduction. Four cores were selected for analysis of the viral community. Three cores were taken from different seep areas around Mound 12, a mud volcano approximately 1000m below sea surface and 800m in diameter that regularly releases dissolved compounds including methane (Linke et al., 2005). Cores were taken from sites colloquially named Yetisburg (covered by grey bacterial mat and populated by the yeti crab *Kiwa puravida*), Mussel Beach (covered by grey bacterial mat and populated by bathymodiolin mussels, approximately 80m from Yetisburg), and Little Knoll (covered by black and grey microbial mat, approximately 1100m from Mussel Beach and Yetisburg) (Mullin, 2020). An additional core came from the center of a white bacterial mat in Jaco Scar, a hydrothermal seep site approximately 60km from Mound 12 that was formed by the subduction of a seamount (Levin et al., 2012). Viruses were also collected from four Niskin water samples that were collected using the *ROV SuBastian* above Mound 12 at 1004m below sea surface and Quepos Slide at 396-398m below sea surface in 2018 (*R/V Falkor* FK190106). Quepos Slide is a bacterial mat-covered submarine landslide, thought to have been formed by intense geological fluid emanation from below (Karaca et al., 2010). The presence of sulfur-oxidizing filamentous bacterial mats, yeti crabs, mussels, and vesicomyid clams are all hallmarks of an active seep site.

Viral-like particles (VLPs) were collected from the 20 incubation bottles in July 2019 and October 2019 (biological replicates) by concentrating viruses from seawater fraction of the incubated sediment slurry. VLPs were purified using a density gradient and ultracentrifugation (see methods in Chapter 2); most VLPs were in the 30% and 35% density fractions, and DNA extraction was performed only on those two fractions. Individual sequencing libraries were generated from viral DNA extracted from each density fraction for each of the two biological

replicates; 72 libraries were sequenced in total for the sediment horizons. Eight additional libraries were generated from the Niskin samplers: two deployed above Mound 12 and two above Quepos Slide, with the 30% and 35% density fractions separated for each.

Analysis of 16S rRNA gene sequencing of the sediment host community in each of these incubations shows that they are largely dominated by members of the ANME-2 clade, especially ANME2a_2b, and their sulfate reducing bacteria partners SEEP-SRB1, family *Desulfosarcinaceae* (**Figure 3.1c**). Cores from Mussel Beach and Jaco Scar appeared to show less variation in host taxonomy with depth, while the top centimeter of Little Knoll and Yetisburg are vastly different from the rest of the core. However, clustering analysis shows that host communities are generally most similar within a core rather than across depth horizons (**Figure 3.1b**). The geochemical gradients present in seep sediments (Jørgensen & Boetius, 2007; Joye, 2020; Joye et al., 2004) may influence relative abundances as depth changes, but the taxonomic identities of the organisms in question appear to vary laterally rather than vertically.

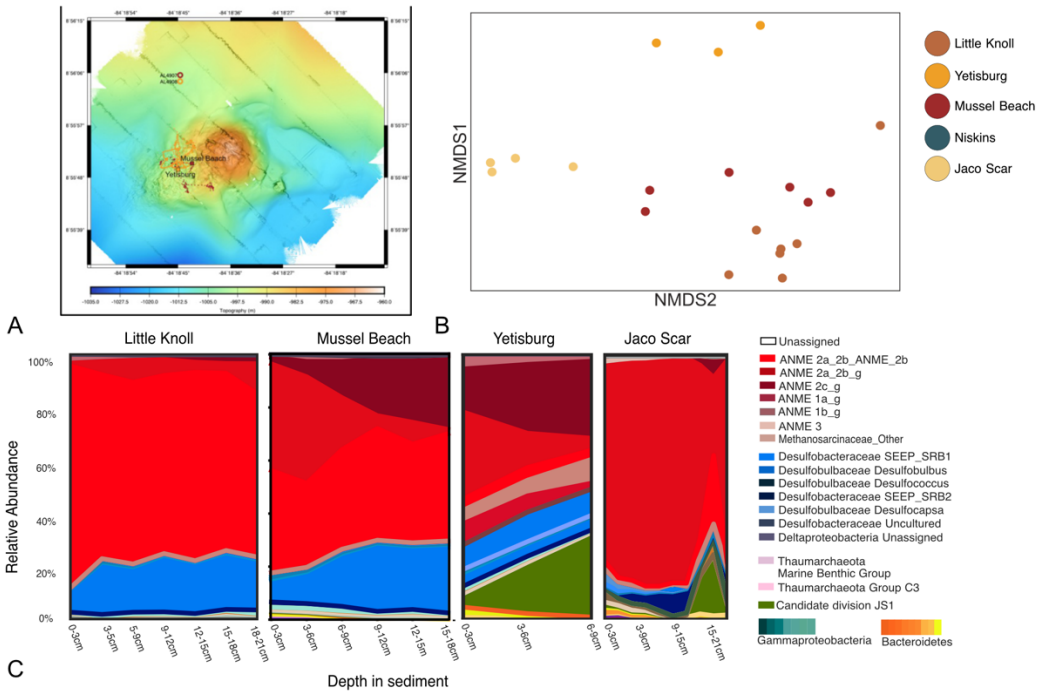

Figure 3.1 Prokayotic community composition and sampling location of the incubations from which the viruses were sequenced. **A)** Bathymetric Map of Mound 12 showing sampling sites. **B)** NMDS plot showing clustering of cellular communities from the horizons of each pushcore. Communities tend to cluster by the pushcore from which they came. Based on relative abundance data from 16S rRNA amplicon sequencing. Stress=0.0183 **C)** Relative abundance of taxa at the phylum level from each sediment horizon. *Figure credit: Dr. Alon Philosof.*

Viral diversity

The 80 viral libraries yielded over 350,000 viral contigs that were at least 5kb in length. Of these, over 28,000 are at least 50% complete and 3,400 are at least 98% complete. The vast majority of the viruses in these bottles are novel; GRAViTy (Aiewsakun et al., 2018), which relies on gene homology and genome organization to predict viral taxonomy at the family level, found 167 novel viral families within the complete/nearly complete viruses, and placed less than half of the viruses into known viral groups. The majority of the viruses classified into known families are tailed, with Siphoviruses being particularly abundant. Supporting the conclusions from the family-level classification, gene sharing networks generated using vConTACT2 (Bolduc et al., 2017) show that most contigs do not cluster with known viruses (**Figure 3.2a**). Tailed phages are often overrepresented at the expense of abundant non-tailed groups in viral datasets (Benler & Koonin, 2021; Brum et al., 2013; Kauffman et al., 2018). Given that the terminase large subunit (*TerL*) gene, which is found in tailed viruses, was only present in 13% of viral metagenome assembled-genomes (vMAGs), the viral community appears to be well-sampled in these datasets. In the absence of a single marker gene analogous to 16S rRNA in prokaryotes, *TerL* and DNA polymerase subunit B (*PolB*) are commonly used hallmark genes in dsDNA phages. Phylogenetic analysis of the *TerL* and *PolB* genes, coupled to data regarding genome completeness, lends further support to the belief that the viral community is well-sampled; complete genomes have been assembled from clades distributed across the phylogenies, rather than favoring specific subgroups (**Figure 3.2b and c**). We believe purifying and directly sequencing the viruses, in tandem with high-resolution library preparation which preserved the separation between depth horizons, biological replicates, and density fractions, allowed us to sample the viral community with unprecedented depth. Our dataset provides an unprecedented increase to the known phage diversity.

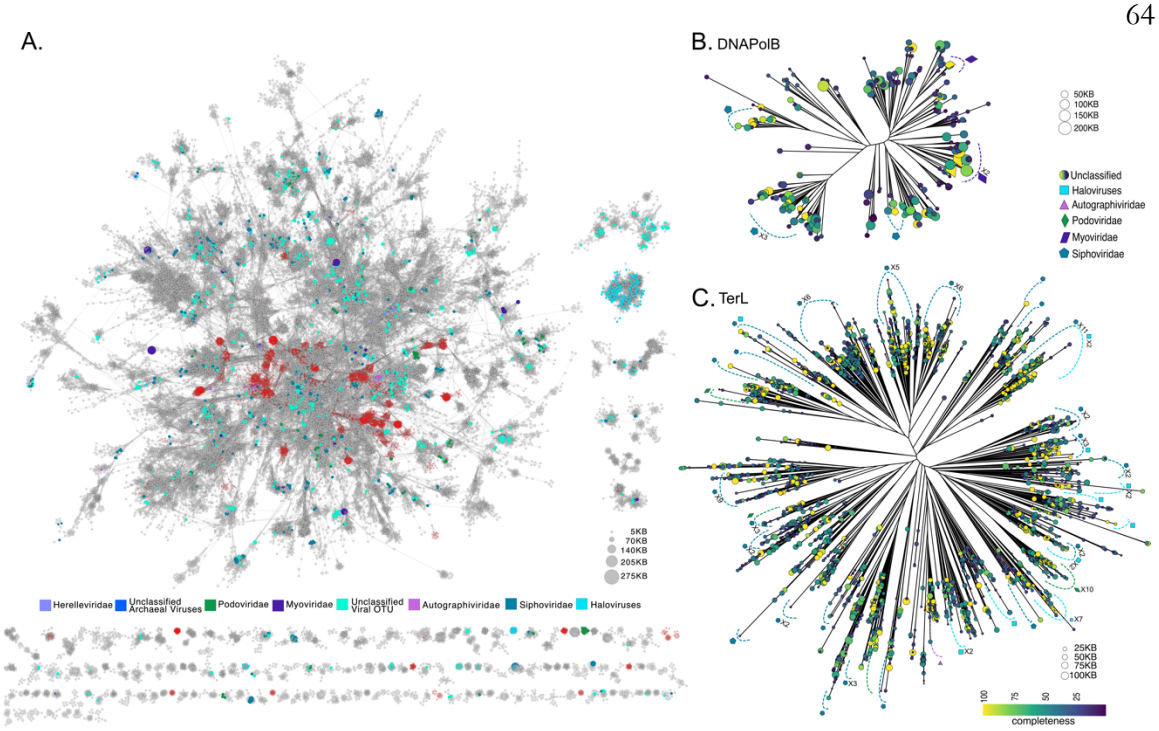

Figure 3.2 Whole-genome and single-gene patterns among viral genomes. **A)** Gene-sharing networks of all viral contigs, colored by the taxonomic family. Red nodes are reference viruses, grey nodes are vMAGs without family-level classification, and all other colors refer to family-level classification. Note the prevalence of large clusters with no apparent connection to the reference sequences, once again indicating the degree of viral diversity captured. **B)** Terminase large subunit phylogeny and **C)** DNA polymerase B phylogeny, both with nodes colored by the completeness of the viral genome of origin. We are successfully capturing viruses from across various phylogenetic clades, as genome completeness does not appear to show a phylogenetic bias. *Figure credit: Dr. Alon Philosof.*

Despite the large number of viral genomes we have recovered, the fact that less than 10% of them were at least 98% complete indicates that the enrichment of viruses before sequencing is a double-edged sword. While we have greatly expanded known viral diversity, we are also forced to confront the number of genomes that we could not fully assemble; many of those may never have been found at all by mining from bulk metagenomes, but one could argue that ignorance is bliss when confronted with the sheer enormity of phage diversity. Deeper sequencing, more replication, and an unlimited budget would likely go a long way towards allowing us to make additional inferences on questions of intra-family diversity, variation in intra-family diversity across sites and down a pushcore, and more.

We also find that the vast majority of the 2708 viral MAGs that could be assigned a lifestyle by Vibrant (Kieft et al., 2020) are classified as lytic (2421 out of 2708), rather than the lysogenic

(287 out of 2708) lifestyle we might expect in a low-energy, physically constrained sediment. Our datasets are inherently biased towards lytic viruses, as they are more likely to be found in the free virus pool and finding prophages in cellular metagenome-assembled-genomes is not trivial. However, as we discussed in Chapter 1, this may also be due to pseudolysogenic lifestyles, in which the lytic viral infection cycle is arrested after the viral genome has entered the cell but before phage replication can begin in earnest. Though the mechanisms by which pseudolysogeny occurs are unknown and distinguishing pseudolysogeny from canonical lysis is difficult, the phenomenon is hypothesized to explain the high prevalence of lytic viruses in soils and subsurface environments (Hazard et al., 2025).

Viral abundance and distribution

Given the observations above, in which host communities vary with geographic location rather than sediment depth, we sought to understand the corresponding influences on the viral community. The water column Niskin samples showed that despite being collected from two distinct locations, at different depths and above different fauna (i.e., the Mound 12 Niskin was collected above a clam bed at 1004m below sea surface, while the Quepos Slide Niskin was collected over bacterial mats nearly 400m below sea surface), the viral communities clustered together. Though additional water column samples are necessary, it appears that sediment-hosted viral communities present enough of a difference from water column communities such that distinct water column samples are most similar to each other rather than the sediments above which they were collected. This further underscores that importance of exploring viral communities in sediment ecosystems and the danger in overreliance on water column samples in attempting to generate comprehensive pictures of marine viral ecology.

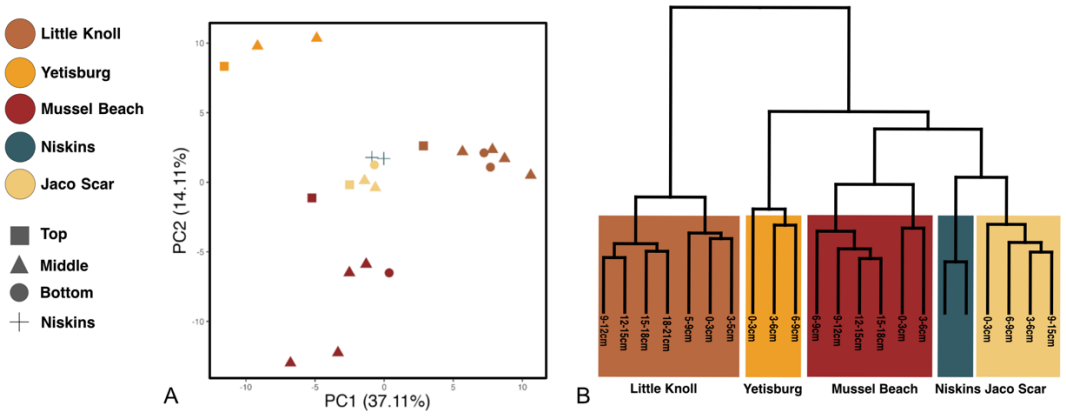

Figure 3.3 Similarity analysis of viral community. **A)** Viral community clustering using principal component analysis, showing that viral communities tend to cluster based on the core from which they came rather than by equivalent depth in the sediment. Based on centered log ratio (CLR)-transformed abundance data. **B)** Hierarchical clustering of CLR-transformed abundance data recapitulates the pattern in figure A. *Figure credit: Dr. Alon Philosof.*

In examining the spatial patterns within the sediment pushcores, we found that clustering of CLR-transformed abundance data shows the same pattern as the host community: viral communities are most similar within a core rather than across depths (**Figure 3.3A and B**). The persistence of such patterns within viral assemblages is remarkable given that the original pushcores had been sampled and sectioned over two years prior to the time of viral sampling.

Linear regression also shows that viral communities within the top (0-3 cm below sea floor) and middle (3-12 cmbsf) portions of a core tend to be less similar than the middle and bottom (12-21 cmbsf) portions of a core (**Figure 3.4**). We believe this difference is driven primarily by the mixing and direct contact with the water column in the top horizon of the core.

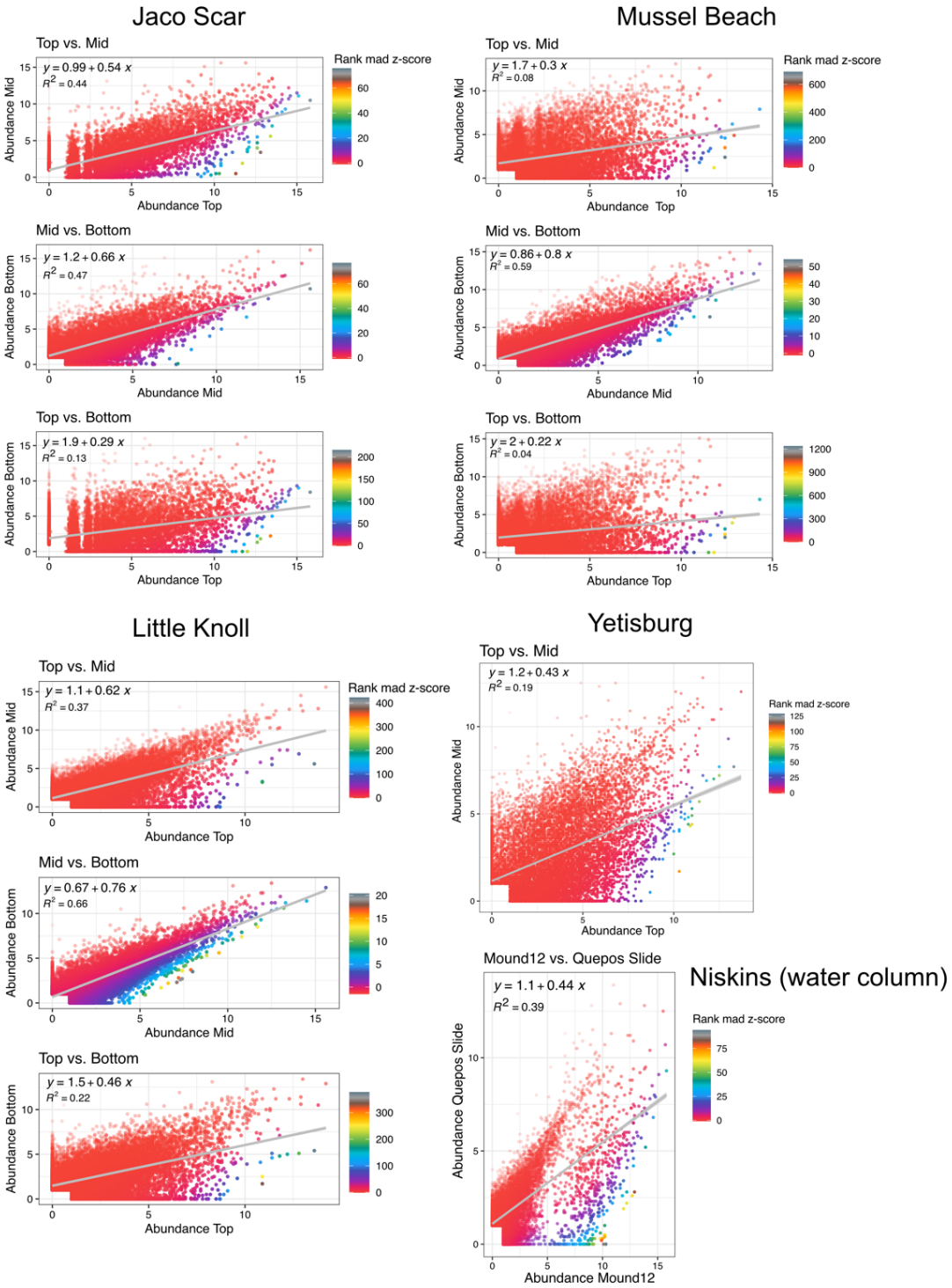

Figure 3.4 Pairwise comparisons between horizons within pushcores show different abundance (CLR-transformed, see methods) patterns within of assemblages. Overall, middle and bottom horizons show greater correlation in viral abundances than middle and top or middle and bottom. Presumably this result is because the topmost horizons are exposed to the water column. Little Knoll shows comparatively high correlation between the middle and bottom horizons, followed by Mussel Beach and Jaco Scar. Pairwise comparisons of matching horizon groups (e.g., middle

vs. middle across two cores) showed no correlation in viral abundances (data not shown, based on personal communication with Dr. Alon Philosof), further supporting the conclusion that viral assemblages are more closely related with a core than within sediment depth groups. Surprisingly, the Niskin samples were not well-correlated despite their proximity in both the PCA and hierarchical clustering analysis. Colors represent deviations from the median Z-score of the change in rank of each vOTU between samples (>=3.5 considered significant). *Figure credit: Dr. Alon Philosof.*

Rank abundance analysis indicates that viral assemblages are low in evenness, but with very heavy tails, in agreement with previous work in viral communities (Hevroni et al., 2020; Z. Li et al., 2021). This indicates that low-abundance viruses are dominant in the community, though outranked by relatively few high-abundance members. Closer examination of the change in rank of different vOTUs (viral operational taxonomic units) between different sections of the core resulted in a clearer picture of a "core" virome, i.e., viruses that persist along across sediment depth and sampling location. In the Little Knoll pushcore, 46-49% of the vOTUs found were part of the "core" virome, while less than a third of the vOTUs are members of the core in Jaco Scar and Mussel Beach. Given that the Little Knoll pushcore is the longest of the pushcores in this study, we were somewhat surprised to see that there was more consistency within the viral community in that core than in some of the shorter ones. Fluid seepage from below is thought to facilitate the movement of cellular microbes from the deeper subsurface to shallower sediments (Chakraborty et al., 2020) and may play a similar role in viral communities. It may be that fluid advection is more forceful in Little Knoll than at the other sites. This advection could result in a higher concentration of methane throughout the core, which may explain why the cellular community is similarly consistent throughout. The similarity may also be due to cellular community distribution, in which we see that the dominant taxa are consistent throughout the core.

Our results raise the question of how the role of fluid advection compares to the role of steep geochemical gradients of the methane seep ecosystem (Levin et al., 2016; Mullin, 2020) in determining the structure of viral communities. The regression analysis above shows greater similarity of the viral community between the top and middle sections of pushcores, generally low similarity between top and bottom, and higher similarity between middle and bottom. This indicates a level of viral responsiveness to host community, which is hugely dependent on

geochemical redox gradients for metabolism. However, the evidence for a "core" virome suggests that fluid flow plays a non-negligible role in shaping the viral assemblages of cold seeps.

It is important to note that this particular dataset has some limitations. For one, we have only one pushcore per sampling location, limited as we were by sample availability and the sheer time and expense of generating and sequencing viral DNA libraries. More interestingly, this dataset also does not span sediment depths below the sulfate-methane transition zone (SMTZ). If such data were available, I would hypothesize that the patterns in viral community composition we have observed above would be quite different due to the decrease in sulfate availability below the SMTZ. Rather than seeing similarities within a pushcore, where the top horizon is more similar to the bottom horizon of the same core, we might see more similarities *across* pushcores depending on depth in the sediment. For example, viral communities from sediment horizons below the SMTZ, where methane oxidation coupled to sulfate reduction is no longer the dominant metabolic process, might be more similar to each other than they are to top horizons within the same core, which would have access to both sulfate, methane, and the water column. Currently, we are limited in our ability to fully explore the spatial distribution of viral communities in these sediments with the samples we have available. However, recent cruises in the Orphan Lab have obtained longer pushcores roughly one meter in length, and future work should include a focus on viral communities below the SMTZ.

Interaction networks spanning diverse hosts

An important part of understanding viral impacts on an ecosystem is to know which organisms are being infected. For example, as discussed in Chapter 1, phytoplankton blooms often end as a result of widespread viral infection (Beltrami & Carroll, 1994; Bratbak et al., 1993). We searched potential host genomes for CRISPR spacers and tRNA sequences that matched the vMAGs in this dataset. The host genomes used were derived from metagenomes from samples 9077-9086 and 10133-10143 and a curated in-house database of metagenome-assembled genomes (MAGs) from methane cold seeps and single ANME/SRB aggregates (Hatzenpichler et al., 2016; Yu et al., 2018).

In total, we found 2224 possible unique host-virus pairs, though most of the host could not be assigned taxonomic classification. Of those that were classified, two-thirds were predicted to be

bacterial and the rest were predicted to be archaeal. Several of the groups are common in methane seeps, including members of the family *Desulfosarcinaceaea, Methanocomedenaceae*, *Methanoperedenaceae,* and ANME-1*;* and the orders *Pseudomonadota*, *Bacillota*, *Paatescibacteria*, and *Fermentibacterota* (Dombrowski et al., 2017; Savvichev et al., 2023; Yu et al., 2022) (**Figure 3.5a**). We then searched for patterns (Clauset et al., 2004) within this host-virus infection network and detected over 360 modules, 21 of which have at least 10 pairings between a host MAG and a phage genus (**Figure 3.5b**). Of the 1205 host-virus pairs in which taxonomy could be assigned to the host cell, 34 viral taxa infect both bacteria and archaea. An additional 97 target different phyla within a domain, and 94 target different classes within a phylum. As seen in Chapter 2, Hwang et al., 2023, and Wang et al., 2022, viruses capable of cross-domain infection have been found in marine sediments before, and may speak to a tendency towards broader host ranges in environments in which collisions with a host are less likely.

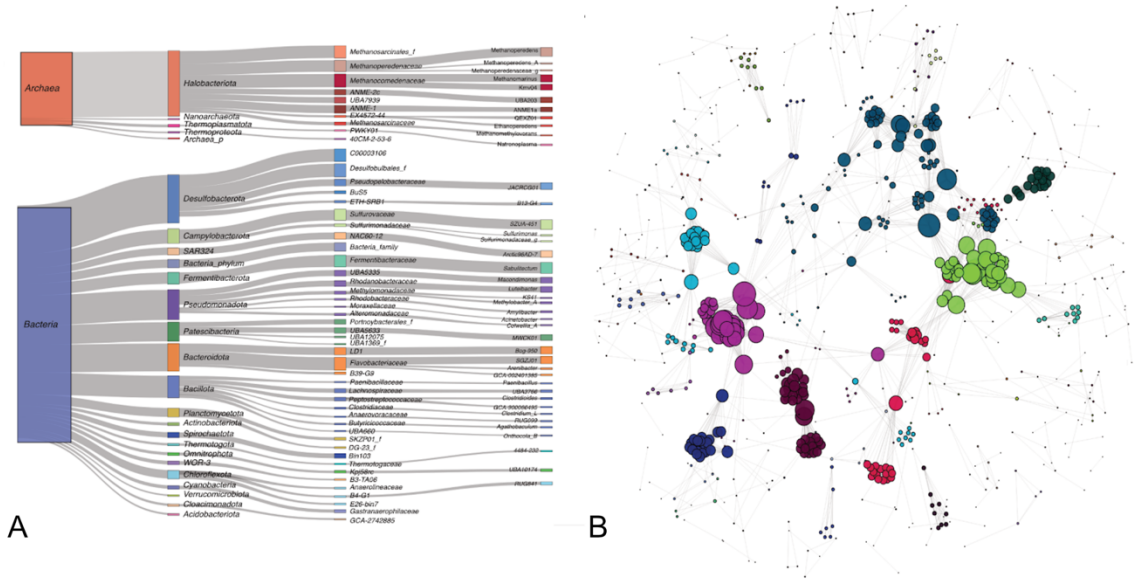

Figure 3.5 Host taxonomy and clustering. **A)** Sankey diagram showing the taxonomic identity of the known hosts for which taxonomy could be assigned. The majority of this subset of hosts are bacterial. **B)** Clustering of host virus pairs in which nodes are hosts and edges are viruses. The size of the node corresponds to the number of viral genera infecting it. In general, the larger the module, the more diverse the hosts within that module tend to be; smaller modules tend to have hosts of more closely related taxonomy. Colors refer to subnetworks within the larger pattern. *Figure credit: Dr. Alon Philosof.*

AMGs are rich in lipopolysaccharide synthesis pathway components

Our Costa Rica metaviromic data is the largest collection of methane seep viruses in existence. The auxiliary metabolic genes found here provide the most comprehensive picture to date of

the potential metabolic impact of viruses in seeps. As in Chapter 2, the most common gene in our dataset is DNMT1 (K00558), mirroring results from other methane seep research (Z. Li et al., 2021; Peng et al., 2023; L. Wang et al., 2022). Also prevalent are parts of the ubiquinone biosynthesis pathway (*ubiG* K00568), the folate biosynthesis pathway (*folE* K01495, *queE* K10026, *queF* K09457), and phosphoadenosine phosphosulfate reductase (*cysH* K00390), all of which have been found in sediment and seep environments before (Heyerhoff et al., 2022; Z. Li et al., 2021; Peng et al., 2023); *cysH* is hypothesized to regulate organosulfur metabolism, including cysteine and methionine degradation (Kieft et al., 2021). We also once again find the putative *ahbD* heme synthase gene (K22227), noted in Chapter 2, as the third most dominant AMG in this dataset (**Figure 3.6**).

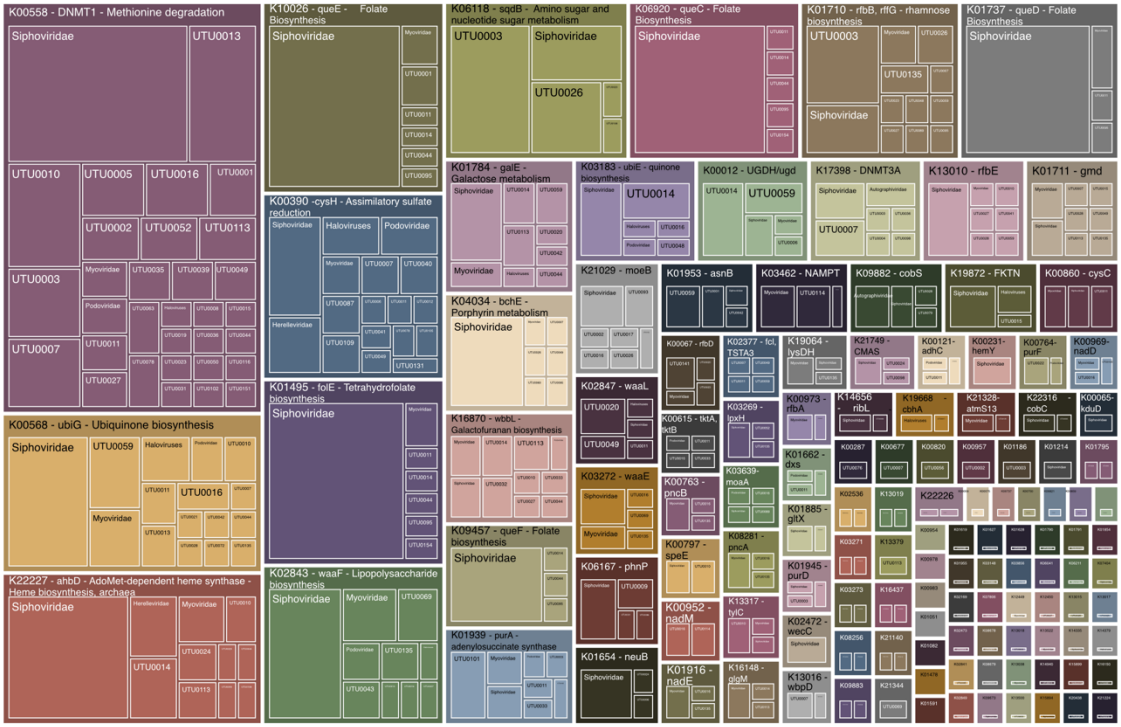

Figure 3.6 AMG abundances across different viral families. The size of the outer box for each AMG corresponds to the number of viral families in which the AMG is found. The inner boxes each represent a viral family, and the size is scaled by the number of contigs in which the AMG is found within that family. *Figure credit: Dr. Alon Philosof.*

Of particular note is the dominance of AMGs from lipopolysaccharide and glycoprotein synthesis pathways. These include the O-antigen and lipid A, which are both cell membrane-related molecules with possible roles in adhesion (L. Wang et al., 2022), and a number of glycan and other sugar related proteins such as *galE* (K01784), *wbbL* (K16870), *waaL* (K02847), *waaEF*

(K12984, K12983), *rfbB* (K01710), *rfbE (*K13010), and *rfbD* (K00067), several of which were also found in the dataset from Chapter 2. A comparison across sediment, soil, and water viral communities found that sediments tend to be particularly enriched in carbohydrate metabolism genes (Luo et al., 2022). The presence of adhesion and sugar metabolism genes may be indicative of adaptation to biofilm and surface-attachment host lifestyles; the extracellular matrix of biofilms has been found to contain a variety of polysaccharides and adhesin proteins (Blanco-Romero et al., 2024; Flemming & Wingender, 2010). For example, *galE* is known to play an required for biofilm formation in *Thermus spp.* (Niou et al., 2009) and *Vibria cholerae* (Nesper et al., 2001), while *galE* knockouts in *Porphyromonas gingivalis* resulted greater levels of biofilm formation (Nakao et al., 2006).

SYNTHESIS: The sediment matrix drives viral specialization

The two studies presented in Chapter 2 and this chapter paint a fascinating picture of viral ecology when examined at different scales. In the first study using sediment-free enrichments, we find that while communities are clearly distinct across distant sampling locations and their associated temperatures, phylogenies of core genes like the terminase large subunit (*TerL*) shows relatively little convergence along these parameters at the level of the individual virus. Perhaps the evolution of core viral genes is so ancient that there is relatively little flexibility left for modern viruses to modify such a gene; this is supported by the presence of viruses that, while in wildly different places in a *TerL* phylogeny, appear to share certain non-core genes that result in clustering in a gene-sharing network. The second study focuses on the sediment environment at a single site but examines intra-site variability and tackles differences along the vertically stratified sediment. We found that the geographic location at which the viral community is found played a greater role in determining similarity than the depth in the sediment. Though the effects of temperature and location cannot be decoupled in the sediment-free study, the results of the Costa Rica study are remarkably similar in highlighting the importance of geography. The vast differences between the sediment communities and the Niskin bottles sampled directly above them also show us that the existing marine viral data, which is mostly water column focused, is not capturing the fullness of the ocean's viral community.

Both studies greatly expand our knowledge of viral diversity, discovering hundreds of new viral families while showing us, in all our incomplete genomes and low-quality contigs, just how much is left to be found. When we examine the 16S rRNA data from these samples immediately after they were collected (**Figure S3.1**), we see how much more diverse the cellular community was before multiple years of incubation in the laboratory. Presumably the viral community has been similarly whittled down, and our bottles contain only a fraction of cold seep viral taxa. These studies also show us how very common cross-domain infections may be in marine sediments, where the amount of energy a cell can harvest from its surrounding substrates is far less than can be harvested from the sun in the surface ocean. It is possible that such an environment, in which the energy available for phage reproduction is low and the sediment matrix reduces the chances of collision with a host, forces phages to adopt a "kill anybody" strategy analogous to the "kill the winner" lifestyle hypothesized to exist in aquatic environments (Thingstad & Lignell, 1997). If cross-domain phages could be isolated or enriched, testing their infectivity in different substrates and matrices might shed light on their host preferences.

This "kill anybody" strategy may also explain why the results of our auxiliary metabolic gene analysis show so few specific adaptations, by temperature, location, depth in sediment, or host community composition. For that matter, the most abundant AMG in both datasets is hardly specific to methane seeps, being most likely involved in evading host defenses (DNMT1; K00558) (Baranyi et al., 2000; Murphy et al., 2013). Choosing to be a metabolic generalist gives a phage a higher probability of being able to successfully infect and influence whichever cell it happens to collide with. If the above hypothesis is correct, phages in lower energy and/or sediment environments might be expected to contain a greater number and diversity of sugar metabolism and adhesion genes in order to successfully adhere to a host, as suggested for the cross-domain phage genus in Chapter 2. This does appear to be the case, as work comparing community-wide AMG abundance across a series of sediment, particle-attached, and free-living viruses finds that carbohydrate and LPS-associated genes are enriched in the lytic sediment community, and genes for secretion systems related to biofilm formation are enriched in the temperate sediment community (Luo et al., 2022).

However, we also find phages carrying AMGs that may play a role in methane-based energy generation, such as a putative heme synthase (*ahbD*; K22227) and a putative anaerobic Mg-protoporphyrin IX monomethyl ester cyclase (*bchE*; K04034). This is quite reasonable in ecosystems that rely on methane oxidation as a source of electrons, and might appear to directly contradict the preceding paragraph, in which I suggest that phages are not tailored with great specificity to their environments. I do not necessarily believe these are contradictory ideas for a few reasons: 1) The abundance of the various AMGs, including *ahbD* and *bchE*, is similar across different sediment cores in Costa Rica and across the much broader range of habitats in our enrichment cultures; *bchE* has also been found in millipede gut viromes (Nweze et al., 2024), 2) The viral sequences of the two genes I highlighted require much more characterization to be sure of their actual roles, and 3) While we hypothesize that the two aforementioned genes may be related to alkane oxidation processes, they are also not the core genes involved in the central pathways of alkane oxidation or sulfate reduction, such as methyl co-enzyme reductase A (*mcrA*) or dissimilatory sulfite reductase (*dsrAB*). In the surface ocean, certain phages of cyanobacteria are rather famously in possession of core components of the photosynthetic machinery (Lindell et al., 2004; Puxty et al., 2016), indicating that viruses are fully capable of carrying key metabolic genes. Methane metabolism AMGs including the particulate methane monooxygenase subunit C (*pmoC*) have been found in prior work examining viruses across various methane-rich habitats, but all were found in cow rumen (Zhong et al., 2024) or freshwater lakes (Chen et al., 2020) rather than sediments.

While we do not see metabolism-based genes that are specific to certain energy generating pathways, we *do* see other physiological adjustments to life in the deep subsurface. For example, the *nosD*-like genes related to cell adhesion in Chapter 2 are not found in water column samples. We also find that some of the more common AMGs in both datasets are related to sugar metabolism and lipopolysaccharide synthesis, which I also hypothesized to play a part in cross-domain infectivity in Chapter 2. Viruses may not find a strong advantage in carrying highly specific pieces of metabolic pathways, but they may find it useful to carry additional genes that help them infect and lyse hosts living a surface-attached lifestyle. Phages appear to be in a delicate balancing act between being generalist enough to have their pick of hosts should the opportunity arise, and specialist enough to successfully reproduce within the constraints of their habitat.

Perhaps AMGs, like the viruses that carry them, must also be generalists—genes with a more general role in energy generation, like *ahbD* and *bchE*, are able to maintain that balance better than a highly specific protein like *mcrA*.

Should it be true that these viruses are more tailored to the physical sediment environment than things like host and ambient temperature, it might be interesting to know which of the criteria discussed above drive this pattern. Is it that the sediment ecosystem happens to have less access to energy than surface water ecosystems, so viruses are compelled to infect any host that is currently managing to grow? Or is it the physical complexity and dense nature of the sediment itself? Perhaps both? Decadal observation of viruses in our sediment-free enrichment cultures or experiments in oligotrophic surface waters would allow us to test at least the sediment complexity theory. At least for now, I believe that the sediment matrix plays a larger role. ANME/SRB consortia make up such a large proportion of cellular activity in the seep ecosystem that some adaptation to their metabolisms would be expected; methane metabolism AMGs have been described before, precluding the idea that genes from low energy metabolisms are inherently not advantageous to phages.

We have made an effort to understand the *who*, the *what*, and a little bit of the *how* of viruses in marine sediments, though an indescribable amount remains to be discovered. The next chapters will move us out of the bioinformatics world and into the laboratory to ask the *when* and a bit more of the *how*— specifically, *how much*.

CONTRIBUTIONS

The analysis of the dataset from Costa Rica was led by Dr. Alon Philosof. For that reason, I have described the major findings here but will not reproduce the in-preparation manuscript as I did for Chapter 2. I performed all sample preparation, DNA extraction, and sequencing library preparation for this study, as well as computationally processing the cellular metagenomes detailed above to the point of binning and taxonomic classification. Dr. Philosof performed all computational analysis of the viral data and produced the figures. The discussion and interpretation are both Dr. Philosof's and my own. The final synthesis section is my own work.

SUPPLEMENTAL DATA

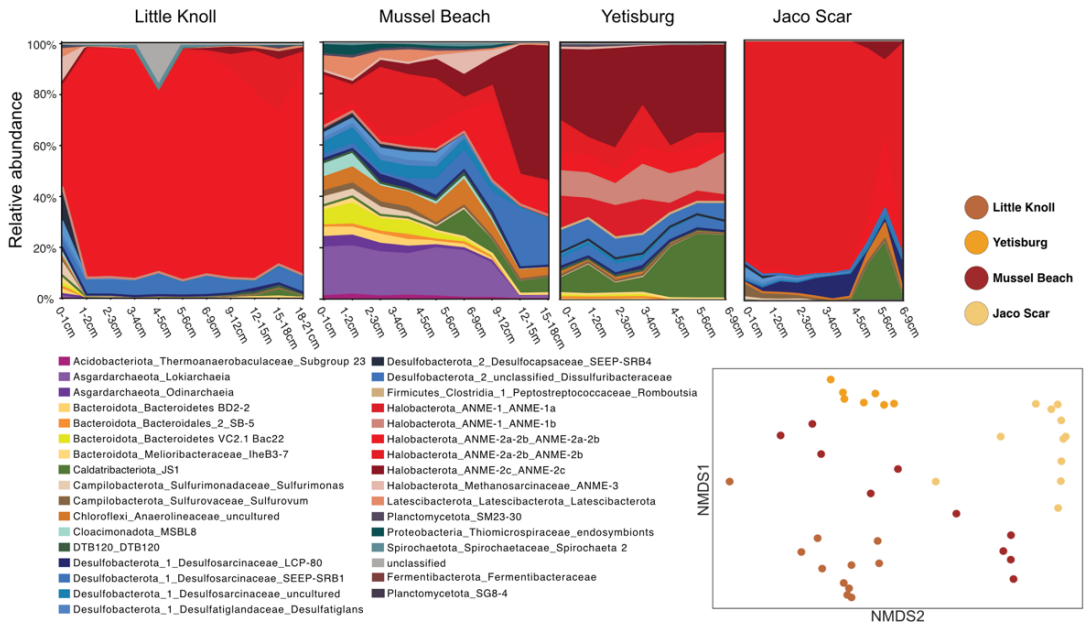

Figure S3.1 Relative abundance data for the cellular community for each of the cores, collected immediately after the cores were sampled in Costa Rica. Based on 16S rRNA amplicon sequencing. The diversity of taxa in this dataset was much broader than it was when sampled after two years of laboratory incubation (**Figure 3.1**). NMDS stress=0.0766. *Figure credit: Dr. Alon Philosof.*

Chapter 4

METHODS FOR THE STUDY OF ENVIRONMENTAL VIRUSES

Aditi K. Narayanan* (1), Alon Philosof* (1), Ranjani Murali (2), Francisco Martinez-Hernandez (3), Yunbin Guan (4), Victoria J. Orphan (1,4)

1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA
2. School of Life Sciences, University of Nevada Las Vegas, USA
3. University of Alicante, Spain
4. Division of Geology and Planetary Sciences, California Institute of Technology, Pasadena, California, USA

INTRODUCTION

Common microbiological techniques to understand physiology and genetic regulation in an organism include treating it with various substrates or knocking out various genes and measuring the subsequent effect on cell growth and metabolism. These powerful methods have several important requirements, such as a pure culture of the organism, genetic tractability, and doubling times on the order of minutes to hours. Most microbes, including ANME and SRB consortia, fit one or none of these criteria. Much of what we know about microbial life comes from a tiny number of culturable taxa who may or may not be representative, and who certainly do not exist in nature isolated from other microbial species. With roadblocks like these, how do you begin to unlock the secrets of organisms that double every few months and cannot be stably maintained in culture?

Microbial ecologists have developed a variety of techniques to address these questions, including microscopy, stable isotope probing, and genetic sequencing. Ecology is a spatial science, and many of these techniques are deployed across sampling sites, depth, and habitat (see Chapters 2 & 3). Thus far, many non-sequencing methods that are well-utilized in cellular and organismal systems have only recently been optimized for model host/phage systems: viral biorthogonal noncanonical amino acid tagging (BONCAT) allows us to fluorescently label newly-produced viruses and quantify them using fluorescence microscopy, while stable isotope probing followed by nanoscale secondary ion mass spectrometry (nanoSIMS) can quantitatively trace nutrient flow into viral particles (Hellwig et al., 2024; Pasulka et al., 2018). Both methods rely on cellular uptake

of specific molecular labels resulting in the eventual incorporation of the labels into newly-produced viral particles. Nevertheless, methods that work well in model systems often struggle when tasked with answering questions in complex environmental samples. We redeveloped both BONCAT and nanoSIMS for environmental viruses from marine sediments, which pose unique challenges of heterogeneity and sample purification.

EXTRACTION OF VIRUSES FROM SEDIMENT SAMPLES

In collaboration with Dr. Alon Philosof

As reported in *Viral extraction from Sediment v1* (Narayanan & Philosof, 2025b)

Viruses like to stick to things. It is well-documented that phages form complexes with clay and iron minerals, and that these complexes are hypothesized to increase the lifespan of a viral particle (Babich & Stotzky, 1980). Marine sediment is nothing if not a thick stew of minerals and organic matter, so the existing filtration and flocculation protocols (B. T. Poulos et al., 2018) used to extract viruses from the water column are not sufficient. The following protocol is adapted from Pan et al., 2019. Sediment in the laboratory is stored in an artificial seawater medium (See Appendix II) at a 1:1 ratio. This protocol has been tested on this sediment slurry, not on undiluted sediment.

Prepare 4X stock solutions of 10% sodium chloride (CAS #7647-14-5) and 20mM tetrasodium pyrophosphate (CAS #7722-88-5, Cat. #P8010, Millipore-Sigma, Burlington, Massachusetts, USA). Filter through a Millex 33mm 0.22µm pore size PES syringe filter (Cat. #SLGPM33RS, Millipore Sigma) to remove larger debris, followed by filtration with a Whatman Anotop 0.02µm syringe filter (Cat. #6809-2102, Cytiva, Marlborough, Massachusetts, USA) to obtain virus-free reagents. A syringe pump is recommended since reasonable force is needed to push liquids through the 0.02µm pore size filter. In a 15mL Falcon tube, mix 1mL of sediment slurry with 1.5mL of the NaCl solution, 1.5mL of the $Na_4P_2O_7$ solution, and 1mL of virus-free Nanopure water for a total volume of 6mL. The final concentrations of the reagents should be 2.5% NaCl and 5mL pyrophosphate. Sonicate the mixture on ice at 25 amps (QSonica CL-188, Newton, Connecticut, USA) for ten seconds, then rest for ten seconds. Repeat two more times for a total of 30 seconds of sonication spread over the course of one minute.

The viruses are then separated from the sediment using density purification, as summarized in **Figure 4.1**. Prepare 50% and 30% weight/volume solutions of iohexol (Nycodenz, also available as Histodenz, CAS #66108-95-0, Cat. #D1258-100G, Millipore Sigma) and sterile water. Filter as above to obtain virus-free density medium. Use an underlay approach to create a density gradient. Draw up 1mL of the 30% medium into a 1mL luer lock syringe using a 1.5 inch 23 gauge needle, and remove all air bubbles from the syringe and needle. Dispense into the bottom of a 15mL Falcon tube. Draw up an additional 1mL of the 50% medium, repeating the steps to remove air bubbles. Air bubbles are particularly disruptive as subsequent layers of the medium are added since they can travel upwards and distort the layers. Place the tip of the needle at the bottom of the falcon tube and dispense the 50% medium slowly, taking care to keep the 50% and 30% layers distinct. Use a thin marker to mark the boundary between the layers.

Mix the sonicated sediment mixture to re-suspend particulate matter. Draw up the mixture into a Pasteur pipette. Gently touching the tip of the pipette to the top of the 30% layer, slowly dispense the sediment sample such that it floats on top of the density gradient. Heavier mineral particles may fall to the bottom. Centrifuge in a swinging bucket rotor at 2900xg for 30 minutes (e.g., Beckman Coulter Allegra X-15R Centrifuge with SX4570 swinging bucket rotor). Use a needle and syringe to remove the liquid fraction: the overlying aqueous medium, the 30% fraction, and the liquid portion of the 50% fraction. Pass the sample through a Millex 33mm 0.22μm pore size PES syringe filter to remove debris and cells. We tested removing only the aqueous layer and the 30% fraction per the protocol in Pan et al., but we found that additional viral-like particles were present in the small portion of liquid in the 50% fraction. All fractions were stained with 25X concentration SYBR Gold Nucleic Acid Gel Stain (Cat. #S11494, ThermoFisher Scientific, Waltham, MA, USA) and examined under a fluorescence microscope to determine the presence of viral particles. As acknowledged in Pan et al., this protocol does not select for viruses specifically, and vesicles and gene transfer agents may be included in viral counts. However, previous work in water samples shows that vesicles affect SYBR Gold viral counts by less than an order of magnitude (Biller et al., 2017). We also found that repeating the sonication and centrifugation protocol a second time on the remaining sediment pellet yielded an additional ~50% more viral-like particles.

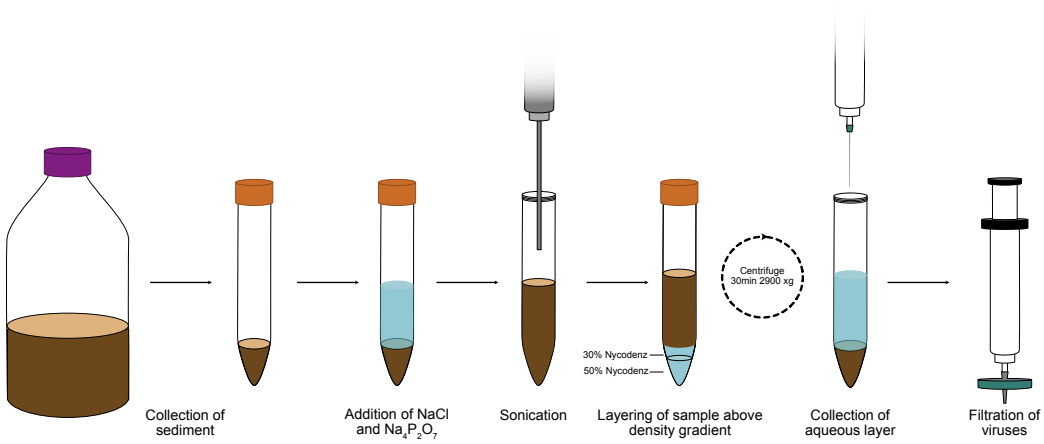

Figure 4.1. Schematic of the viral collection and purification process. A sodium chloride and tetrasodium pyrophosphate solution is added to sediment slurry. The mixture is sonicated, then layered over a Nycodenz density step gradient. After centrifugation, the liquid fraction is collected and filtered.

Once the viruses have been extracted and filtered, they will store well at 4˚C for about a month. We recommend using glass vials since anecdotal evidence suggests that viruses do not store well in plastic at these temperatures. For long term storage, we fixed the viral samples in 2mL microcentrifuge tubes (e.g., Cat. #05-408-138, Fisher Scientific, Waltham, Massachusetts, USA) 4˚C for 30 minutes with a final concentration of 0.1% glutaraldehyde, before flash freezing in liquid nitrogen and storing at -80˚C.

Samples must be cleaned of all density separation medium before any downstream application. Samples stored at -80˚C should thaw overnight at 4˚C before being placed in a centrifugal ultrafiltration device with a 100kDa molecular weight cutoff (e.g., Cat. #UFC5100, Amicon Ultra Centrifugal Filter, Millipore Sigma, or Cat. #VS0142, Vivaspin 500, Sartorius, Göttingen, Germany). The samples should be spun slowly to prevent damage to the viral particles. A speed of 900xg for 20 minutes is recommended in a microcentrifuge, adding additional sample and repeating the centrifugation to remove as much liquid volume as possible. Usually, a total of 1mL of fixed viral suspension is filtered in a single device, though filters occasionally become clogged and may require the sample to be transferred to a new filter device. When working with volumes larger than 1mL, we used larger models of the same ultrafiltration devices and a Beckman Coulter Allegra X-15R centrifuge with a SX4750 swinging bucket rotor. The filtrate

collected in the bottom tube is discarded, with the sample remaining in the upper filtration cartridge. Samples are then washed with salt mix (SM) buffer (M. Sullivan, 2015); SM buffer is added to the cartridge and gently mixed into the sample by pipetting, and the sample is spun again for another 20 minutes at 900xg. The wash step is repeated three times to remove as much density medium as possible. The final sample is reconstituted to a 200µL volume. Before removing the sample from the filtration cartridge, the cartridge is wrapped in parafilm and vortexed for 15 seconds to loosen viral particles trapped in the membrane. The suspension is then transferred to a 0.6mL microcentrifuge tube (Cat. #MCT-060-C, Corning Life Sciences-Axygen, Union City, California, USA).

A note on cleaning viral samples from sediment after density separation

Viruses extracted from sediment need to be cleaned; in the section above, the use of centrifugal ultrafiltration devices with a 100kDa molecular weight cutoff (MWCO) is detailed as a method to remove density separation medium, remaining cellular debris, and other non-viral particulate matter from viral suspensions. This is a slow, but parallelizable, method requiring constant monitoring of the ultrafiltration devices. To reduce hands-on time, we also tested dialysis devices like the Thermo Fisher Slide-A-Lyzer MINI Dialysis devices (20kDa MWCO, Cat #88405), which allow the overnight exchange of buffers and removal of particles smaller than the molecular weight cutoff (MWCO) of the device. Unfortunately, the maximum available MWCO on the market is 20-30kDa, which is insufficient to remove non-viral particulates from the samples. Unpublished fluorescence-activated sorting data generated by Dr. Francisco Martinez-Hernandez from samples prepared by Dr. Alon Philosof and me showed that using centrifugal devices with 100kDa MWCO resulted in more easily discernable viral populations than the dialysis devices at 20-30kDa MWCO.

VIRAL NANOSIMS FOR ENVIRONMENTAL SAMPLES

In collaboration with Dr. Alon Philosof and Dr. Yunbin Guan

Nanoscale secondary ion mass spectrometry is a technique that allows us to combine stable isotope probing with spatial resolution to quantitatively trace the movement of nutrients at the single-cell level. It has been employed with great success in the Orphan Lab to correlate taxonomic identity to activity (Dekas et al., 2016), examine the consequences of decoupling the

ANME/SRB syntrophy (Scheller et al., 2016), and discover patterns of silica biomineralization in ANME/SRB consortia (Osorio-Rodriguez et al., 2023). The method works as follows: a primary ion beam of positively charged cesium ions moves across a predetermined area. The ablation of the sample by the primary beam releases secondary ions, which are channeled to a series of detectors by magnets that separate them based on mass and charge. The study detailed in the following chapter primarily employed nitrogen isotope probing; for nitrogen, the detectors are tuned to pick up CN^- ions, which are separated by mass to $^{12}C^{15}N^-$ or $^{12}C^{14}N^-$. The instrument correlates ion detection with primary beam location to show the spatial distribution of the isotopic signature in the sample.

As discussed in this chapter's introduction, a former member of the lab, Dr. Alexis Pasulka, and her co-authors extended nanoSIMS to viral particles in their 2018 paper *Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods*. ^{15}N-labeled ammonium was fed to liquid cultures of the coccolithophore *Emiliania huxleyi* and its phage EhV207, the cyanobacterium *Synechococcus* and its phage Syn1, and *Escherichia coli* and T7. As virus-infected cells take up labeled ammonium, they incorporate the labeled nitrogen into the newly-produced viruses, allowing us to track the movement of fixed nitrogen into the viral pool within the system of interest. Pasulka et al. placed liquid suspensions of the cultured phage on a silicon wafer, which has the conductive surface required for nanoSIMS. They mapped the viral locations on the wafer using SYBR Gold staining and fluorescence microscopy, then gently washed off the microscopy mounting fluid. The washing process naturally results in some sample loss, which is more acceptable in cultured systems than in diverse environmental samples due to sample homogeneity in the former. Our sediment incubations are not so forgiving. Dr. Alon Philosof and I devised a method to embed environmental viral suspensions in resin to fix their position in space.

To state the obvious, viruses are tiny. The Tara Oceans project estimated the average diameter of a non-tailed marine phage to be approximately 54nm±12 (Brum et al., 2013). The primary cesium ion beam of the Cameca NanoSIMS 50L is 50nm, which places viruses at the limit of the instrument's resolution. The instrument needed precise tuning to detect individual viruses.

Dr. Yunbin Guan, Director of the Caltech Micoranalysis Center, provided essential instrument tuning and sample preparation advice to help maximize viral signal.

Sample embedding and sectioning

As reported in *Viral nanoSIMS sample preparation v1* (Narayanan & Philosof, 2025a)

Standard NanoSIMS preparation requires embedding samples in resin and cutting thin sections of around 2μm thickness. This fixes viral-like particles in position so that fluorescence microscopy can be correlated to NanoSIMS images. The existing protocols in the lab are based on high-density samples with large cellular aggregates; typically, cell suspensions are mixed in a 1:1 ratio with 3% Difco Agar Noble (BD Biosciences Ref. 21430) in a 0.2mL PCR tube (for example, Cat. # PCR-02-C, Corning Life Sciences-Axygen), producing an agar plug of 1.5% agar before resin infiltration. Viruses are small and more dispersed, and a 1:1 mix of sample and agar greatly dilutes particles. In addition, the conical base of PCR tubes results in strong vertical dispersal of viruses, reducing the number of viruses per field of view in the thin sections produced downstream. A stock concentration of 6% Agar Noble instead of 3% solves both problems. The 6% agar forms a cushion that fills the conical section of the tube; the sample itself fills the cylindrical section of the tube, reducing vertical dispersion of the viruses, and is mixed in a 3:1 ratio with the agar to increase final concentration of viral particles. The steps are as follows: 6% melted agar is used to fill the conical portion of the tube, then allowed to harden. To distinguish the cushion from the sample, 5μL of methylene blue are mixed with 150μL of liquid sample. Seventy-five microliters of the methylene blue/sample mixture are dispensed above the hardened 6% agarose cushion, to which 25μL of melted 6% agar are added and rapidly mixed for a final 1.5% agar concentration (**Figure 4.2a**). This requires speed, as 6% agar hardens extremely quickly. We recommend warming the viral sample in your hands before adding the agar.

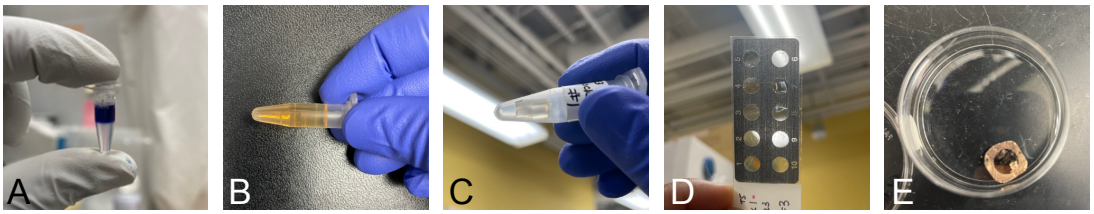

Figure 4.2. NanoSIMS sample preparation. **A)** A methylene blue-dyed 1.5% agar plug containing the viral suspension, sitting on a clear 6% agar cushion that fills the conical space of the tube. **B)** The agar plug embedded in Technovit resin. The yellowish color is due to oxygen exposure and does not affect the sample quality, but the embedding procedure can be done anaerobically to prevent discoloration**. C)** The technovit block cut out and transferred to a larger tube for storage. **D)** Thin sections dried onto 10-well poly-lysine slides. **E)** A cut and gold-coated well containing a thin section. The scratches on the bottom right of the section form the "G"-like symbol used for orientation.

After embedding in agar, standard procedures for the Technovit 8100 embedding system (Cat. #14654, Electron Microscopy Sciences, Hatfield, Pennsylvania, USA) apply. The tube containing the agar is carefully cut at the tip and the plug is pushed into a 1.5mL microcentrifuge tube. Separation between the 6% cushion and the 1.5% sample blocks is not uncommon and does not hinder downstream work. The 6% cushion can be disposed of in this event. The agar block is then serially dehydrated by incubating for two minutes in each of the following solutions: ethanol 50:50 1X phosphate buffered saline (PBS), followed by ethanol 75:25 1X PBS, ethanol 90:10 1X PBS, two incubations with pure ethanol. Once the liquid from the final ethanol incubation has been removed, allow the remaining ethanol to evaporate. The methylene blue is ethanol soluble, but the high concentration should allow the sample to be distinguished from the 6% cushion even after dehydration.

Add 1mL of Technovit 8100 infiltration solution to the tube containing the agar plug and incubate at 4° C for one hour. Replace the infiltration solution and incubate at 4° C overnight. Remove the infiltration solution the next day and push the agar plug into a petri dish. Follow the instructions enclosed in the Technovit 8100 kit to make the polymerization solution. Place 100µL of the polymerization solution in a 0.6mL tube and push the plug into the solution. If the 6% cushion is still attached, orient the agar plug such that the 1.5% sample is facing the bottom of the 0.6mL tube. The sample should fill the majority of the cross section of the tube, while ensuring that it is surrounded completely by polymerization solution. Use an additional 100-300µL of polymerization solution to cover the plug entirely. Work quickly and avoid

bubbles. Mark the tube at the interface between the infiltrated agar and the pure resin so that thin sections are cut from the correct part of the resin block. Allow the resin to harden overnight (**Figure 4.2b**) before removing cutting the block out of the tube (**Figure 4.2c**).

Cut 2μm thin sections from the resin using a microtome; the Orphan Lab uses a Reichert Jung Ultracut E Ultramicrotime. As each section is cut, it should be transferred to a water droplet contained in the well of a ten-well poly-lysine glass slide (PTFE/Teflon Printed Slides with Poly-L-Lysine adhesive surface chemistry, Tekdon Incorporated, Myakka City, Florida, USA). Technovit is a hydrophilic resin, and the water droplet helps unfurl/flatten the thin section, while the poly-lysine coating helps the section adhere to the slide (**Figure 4.2d**)**.** We cut an average of forty thin sections per sample to ensure at least two sections with high viral density. Once the sections dry, stain each well with a 25X solution of SYBR Gold for 30 minutes. Remove the SYBR Gold stain and wash each slide gently by pipetting water into each well and wicking with a kimwipe. Use VectaShield Antifade Mounting Medium (Cat. #H-1000-10, Vector Laboratories, Newark, CA, USA) to mount a glass coverslip for imaging. To image the sections, search for areas of high viral density under a 100X oil objective with the FITC filter cube. The difference in texture between pure resin and agar infiltrated by resin is visible under the microscope (**Figure 4.3a & b**). Tilescan high-density areas at 100X under the FITC filter and under brightfield. The output of the CCD camera in the Cameca NanoSIMS 50L is most visually similar to the brightfield function on a microscope, and a brightfield map of the sample is necessary to understand sample orientation in the SIMS sample holder. Move up to the 10X objective and take images of the larger geographic area from which your 100X images came. By taking care to image the edges of the thin section, we were able to quickly locate the regions of high viral density under the NanoSIMS CCD camera. Because the CCD camera also inverts the image, we manually cut a shape similar to the letter "G" into each glass slide—choosing a symbol that lacks symmetry allows us to determine how the CCD camera transformed the image and transform our tilescan map accordingly.

We used the Olympus BX51 microscope (Olympus Corporation, Tokyo, Japan) to manually tilescan high-density regions instead of using the automatic tilescan functions on more high-powered microscopes in the core facilities due to the sheer volume of samples and resulting

cost limitations. The images were then stitched together to form a map (**Figure 4.3c**). Another challenge in working with viruses is that the thin sections themselves are 2μm in thickness. Given that the average marine viral capsid is around 54nm, viruses may be embedded deep in the resin that cannot be seen in fluorescence images. Where cost-effective, we recommend confocal microscropy for its speed and resolution advantages; if available, the Zeiss Airyscan (Carl Zeiss AG, Oberkochen, Germany) detector can provide additional improvements in z-resolution.

After imaging, each slide is placed into a petri dish with deionized water gently flowing over and between the slide and coverslip. Once the coverslip is detached, water should continuously flow over the slide to remove remaining mounting medium, though the slide should be kept outside the direct stream of water to prevent sample damage. The slide can then air dry. The final step before NanoSIMS is to cut the well with the mapped section out of the glass slide by scoring and breaking the glass. The edges of the well are smoothed with a Dremel to remove sharp edges and fit into the NanoSIMS sample holder. The well containing the section is then coated with 15nm of gold to make the surface conductive (**Figure 4.2e**). Because of the large number of samples we screened, we used the more cost-effective glass slides rather than the naturally conductive silicon wafers.

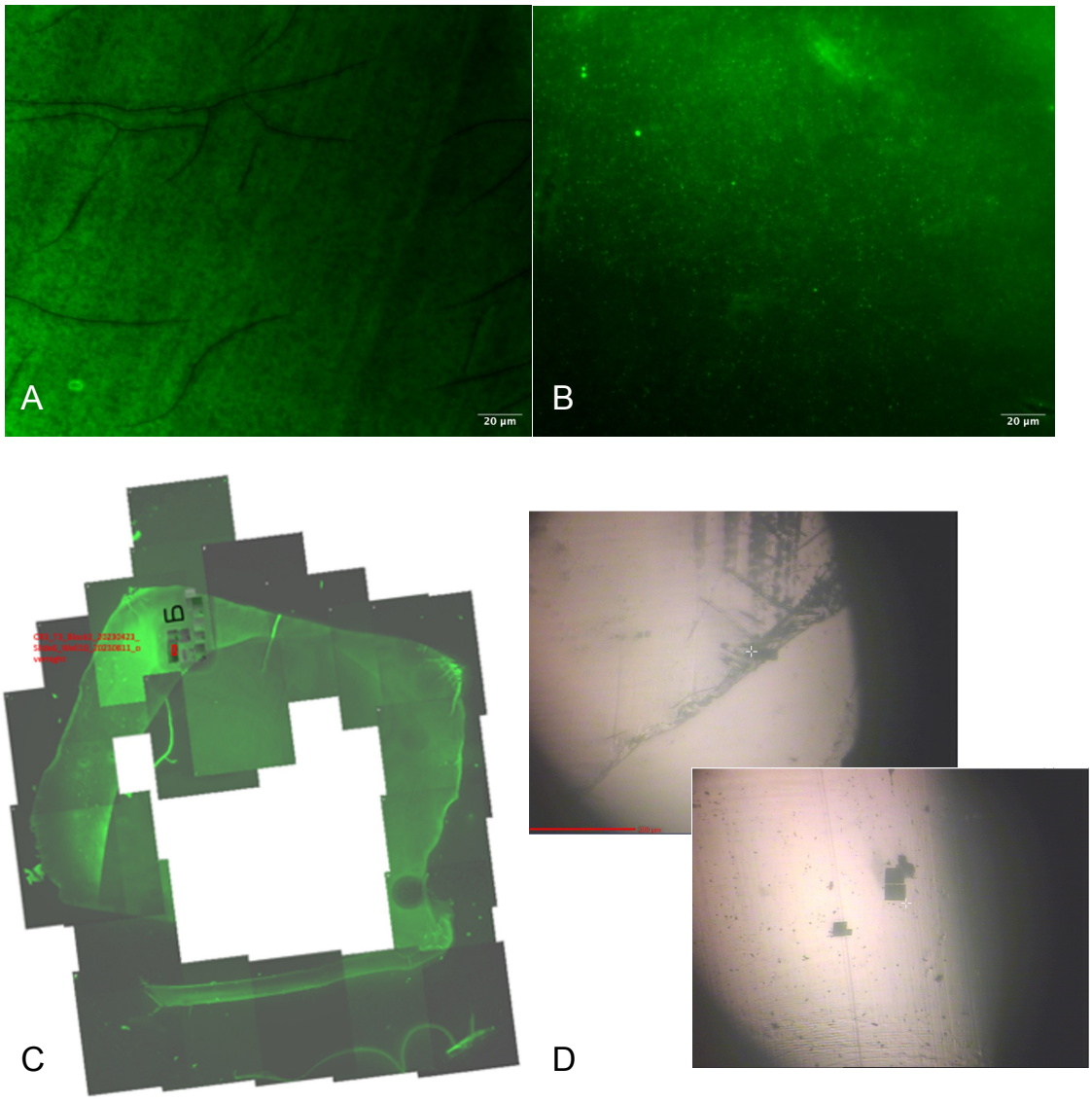

Figure 4.3 Imaging of thin sections for viral nanoSIMS. **A)** An example of a resin thin section where no viruses are present. **B)** An example thin section where viruses are abundant. **C)** An example of a hand-stitched map of a thin section at 10X magnification, with the 100X images superimposed in the top left corner. The red text was used to mark the area where the NanoSIMS data was collected. **D)** Top: An example of how the thin section appears under the Cameca NanoSIMS 50L CCD camera. Bottom: The thin section after the data is collected. NanoSIMS is a destructive technique, and the dark rectangular shapes indicate where the sample has been rastered away.

NanoSIMS data collection

In collaboration with Dr. Yunbin Guan of the Caltech Microanalysis Center and Dr. Alon Philosof. A very special thank you to Dr. Yunbin Guan for his invaluable suggestions and for tuning and running the instrument.

- **Instrument:** Cameca NanoSIMS 50L
- **Presputter:** Variable. This is the setting to ensure that leftover mounting medium, the gold coating, etc. are removed from the surface before data collection commences.
- **Dwell time:** 7.5 microseconds. This setting determines how long the cesium ion beam stays on a single spot.
- **Current:** 13 picoAmps, adjusted as needed for each sample. Effectively the force of the cesium ion beam. Because nanoSIMS is a destructive technique, current that is too high will obliterate your sample too quickly, while current that is too low may result in slow and insufficient data collection.
- **Raster size:** The area over which data is collected. In our data collection, we used a square of 10μm x 10μm or 15μm x 15μm. This was not ideal (see Chapter 5 for details)—rasters should be on the order of 5μm x 5μm to properly resolve viruses smaller than about 100nm. Smaller raster sizes allow for more targeted data collection and are faster, while larger raster sizes may take longer and run into problems with the beam drifting but allow you to collect more data at a time. In hindsight, we can see that larger rasters are ideal for samples like ANME/SRB aggregates, where you might want to collect data from an entire aggregate at once, and where the size of the target means that a small amount of beam drift makes little difference.

There are also several assumptions that must be considered when examining nanoSIMS data. 1) We assume that the viruses produced over the given time period are highly enriched in the isotope label. This is not unreasonable given previous work showing that a substantial fraction of the elements incorporated into newly produced phage are derived from outside the host cell (Cohen, 1948; Waldbauer et al., 2019). 2) We assume that all cellular hosts have the same preference for isotope-labeled nutrients, and thus all viruses that are newly produced in the given time period will be isotopically labeled. In a mixed environmental viral community, neither assumption 1 or 2 may be universally true, but are what we can work with at the moment. 3) We assume that smaller viruses can actually be resolved using nanoSIMS. As previously discussed, the cesium beam on the nanoSIMS is 50nm wide, and smaller viruses certainly exist. Depending on the power of the primary Cs^+ ion beam, these viruses may also simply be vaporized *en masse*, which complicates detection and requires individual optimization of each sample.

VIRAL BONCAT FOR ENVIRONMENTAL SAMPLES

In collaboration with Dr. Alon Philosof, Dr. Ranjani Murali, and Dr. Francisco Martinez-Hernandez

Existing methods of quantifying viral production rates involve taking samples at regular time intervals, performing a series of dilutions, and counting them under a fluorescence microscope with SYBR Gold nucleic acid stain (Dell'Anno et al., 2009). The advantage of this method is that many samples can be processed in parallel, making high-throughput estimates possible. While this method offers fast bulk quantification, it cannot tell us which viruses are being produced in the given experimental period. We sought to expand the use of existing viral biorthogonal non-canonical amino acid tagging (viral-BONCAT) protocols developed in the Orphan Lab (Pasulka et al., 2018), based on the BONCAT method first developed in mammalian cells (Dieterich et al., 2006). The method involves incubating the cells of interest with a methionine analog containing an azide or alkyne group. The cells take up the methionine analog and incorporate it into newly-translated proteins; in the case of viral-BONCAT, infected cells will incorporate it into newly-synthesized viral capsids. A fluorophore can be attached through a copper-catalyzed click reaction, allowing us to visualize individual viruses produced during the HPG treatment period (**Figure 4.4**).

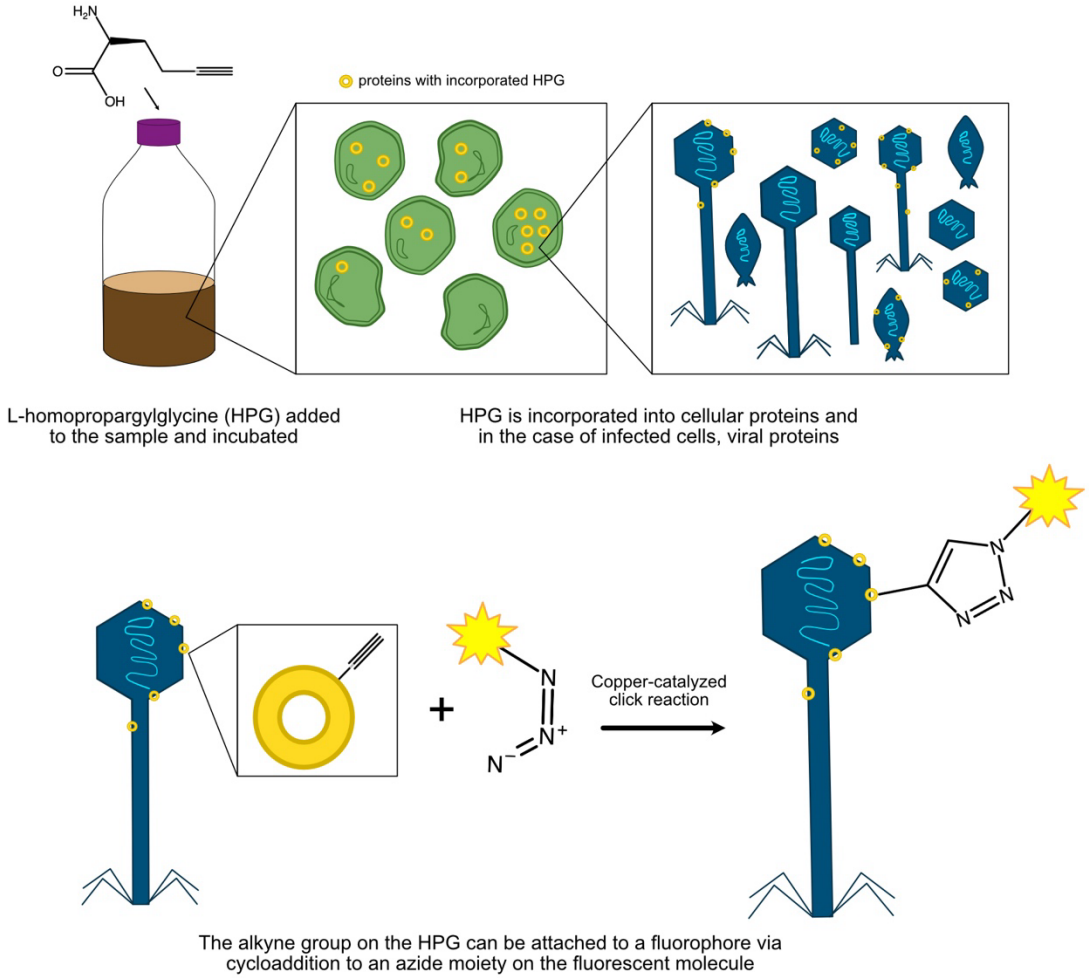

Figure 4.4. Schematic of BONCAT. A non-canonical amino acid (in this case, the methionine analog HPG) is added to the sample of interest and incubated. HPG is characterized by the presence of an alkyne moiety, which is important for downstream applications. Metabolically active cells take up the HPG and incorporate it into their proteins, shown as yellow rings. Cells that are undergoing active viral infection also incorporate the HPG into the newly-produced viral capsids. Note that pre-existing viruses do not contain HPG. The alkyne moiety on HPG can be attached to a fluorescently labeled azide group via a copper-catalyzed cycloaddition reaction. The fluorophore-labeled viruses can now be visualized with fluorescence microscopy or undergo fluorescence-activate viral sorting.

Prior work in the lab was exclusively in cultured virus/host pairs. We have expanded this method to complex environmental samples, including marine sediment. Because these newly-produced "active" viruses are fluorescently tagged, they can theoretically be sorted via fluorescence-activate sorting. BONCAT-based fluorescence sorting has been successfully implemented in complex heterogenous cellular communities (Reichart et al., 2020; Valentini et al., 2020). Experimental evidence shows that this is also possible with viral communities (Martínez

Martínez et al., 2020; Martinez-Hernandez et al., 2017), allowing for high-throughput counting as well as downstream single-virus genomics to identify active viruses.

Initial methods development

Our methods development required optimization of the following: 1) stabilizing the sample, 2) reduction of background noise, and 3) increasing signal intensity from viruses. Sample stability was generally straightforward; filtered viral particles are stable in a liquid suspension for at least one month. For longer term storage, fixation with glutaraldehyde and storage at -80˚C is preferable, as outlined in the above section entitled "Extraction of Viruses from Sediment Samples". Paraformaldehyde as a fixative for long-term storage resulted in loss of phage (data not shown), as did storage at -20˚C without fixative.

Reduction of background noise and increasing signal intensity from labeled viruses are closely related methodological difficulties. Initial experiments were carried out on water taken from the Caltech Turtle Pond before sunrise and incubated with 50μM of the methionine analog L-homopropargyl glycine (HPG, Cat. #CCT-1067, Vector Laboratories, Newark, California, USA), which contains an alkyne group not typically found in methionine (**Figure 4.4**). Typically, samples were taken every two hours until twelve hours had elapsed from HPG addition. The first tests of viral-BONCAT, performed by Dr. Alon Philosof and Dr. Ranjani Murali, involved filtering viral suspensions onto an Whatman Anodisc aluminum oxide wafer with 0.02μm pore size (Millipore Sigma, Cat. #WHA68096002), then incubating the Anodisc for 30 minutes at room temperature with the click reagents, including the azide-containing fluorophore (AZDye 647 picolyl azide, Cat. #CCT-1300, Vector Laboratories). The disc was then stained with 25X SYBR Gold to fluorescently label the entire viral pool (Narayanan & Philosof, 2024). This method was effective in distinguishing HPG-labeled viruses from non-HPG labeled viruses, but often resulted in high levels of background noise despite repeated washes after the click reaction and after SYBR Gold counterstaining.

The following recipe is for the Click reaction cocktail with 647 picolyl azide, though other azide reagents, such as those described in the following subsection, can be substituted. This recipe makes 250μL, enough for approximately two Anodisc samples.

Mix the following and incubate for 3 minutes in the dark:
1.25 µL 20 mM $CuSO_4$
2.5 µL 50 mM tris-hydroxypropyltriazolylmethylamine (THPTA)(SKU #CCT-1010, Vector Laboratories)
0.5 µL of 10 mM 647 picolyl azide

Separately mix the following:
12.5 µL of 100 mM Ascorbate
12.5 µL 100 mM Aminoguanidine
221 µL 1 X PBS

Add the azide/THPTA mixture to the ascorbate solution and keep it protected from the light.

Leveraging the binding affinity between streptavidin and biotin

To eliminate the unlikely possibilities that the background noise was the result of the fluorophore not being stably bound to the HPG-labeled viral particle and that there was nonspecific binding of the fluorophore-azide to the Anodisc, we leveraged the well-established binding affinity between biotin and streptavidin (Chivers et al., 2011). We replaced the fluorophore-azide with a diazo-biotin-azide complex (SKU #CCT-1041, Vector Laboratories), later replaced by the more efficient biotin picolyl azide (SKU #CCT-1167, Vector Laboratories) during the click reaction step. The biotin picolyl azide was then incubated with streptavidin conjugated to a fluorophore.

After incubating each anodisc with 100µL of the click cocktail in the dark for 30 minutes at room temperature, we washed the anodiscs by filling a petri dish with 1X PBS and incubating the discs for 3 minutes to remove excess biotin. The discs were removed and excess PBS was wicked away with a kimwipe. At this point, the click reaction has attached the biotin molecule to the HPG, but no fluorescent molecule is bound. The discs were then incubated for an additional 30 minutes on a drop of 40ng/µL streptavidin conjugated to AlexaFluor (Cat. #S11227 or #S21374, ThermoFisher Scientific). The streptavidin-fluorophore would bind to the biotin, causing HPG-labeled viruses to fluoresce. We then washed with 1X PBS as described and repeated the washing step with water before allowing the Anodiscs to dry and staining with SYBR Gold. We did not find significant improvement in the signal to noise ratio, so we returned to the less arduous method that relied only on a fluorophore azide; however, we added additional wash steps after the incubation to remove excess reagents, with three incubations in 1X PBS for two minutes each, and an additional two incubations with autoclaved water. This resulted in a

reasonably clean image in which newly produced viral progeny could be clearly identified (**Figure 4.5**).

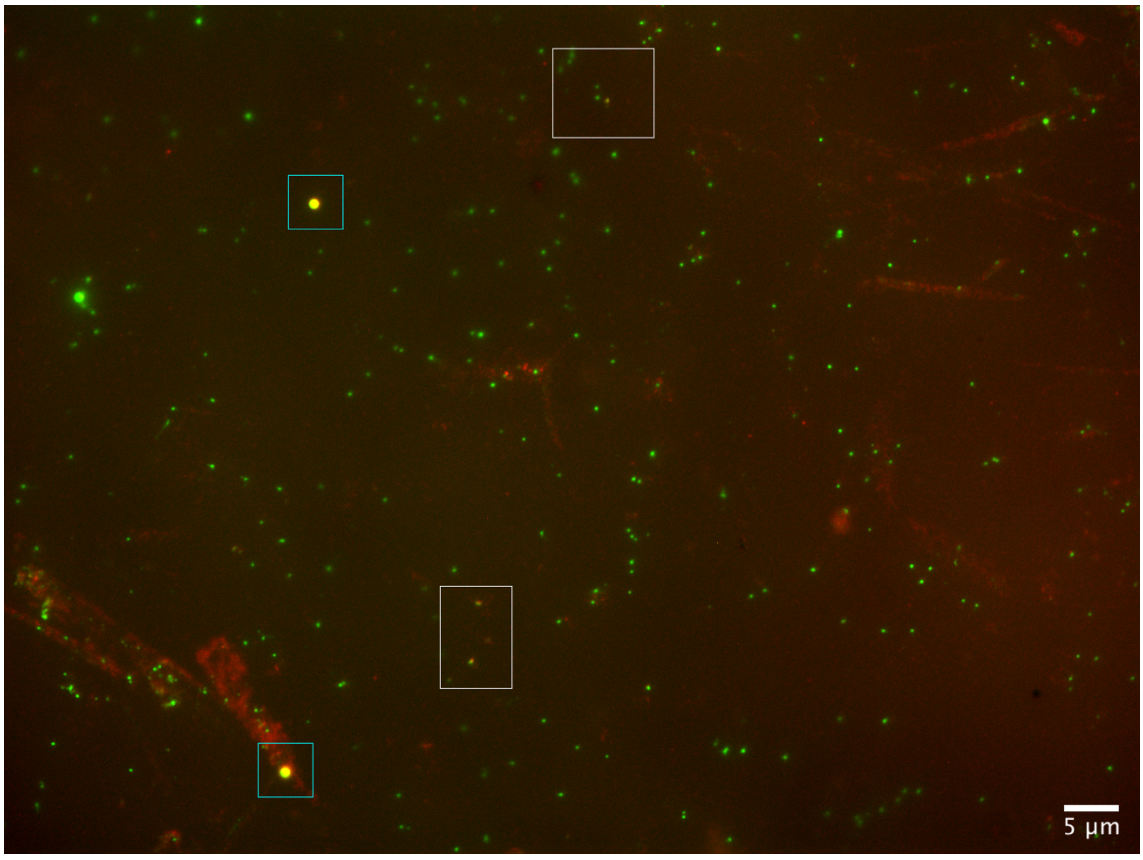

Figure 4.5. Viral BONCAT using Picolyl azide 647, performed on a sample taken from a Costa Rica methane seep sediment incubation. False color image with SYBR-Gold-stained viruses shown in green and HPG-containing (i.e., BONCAT positive) viruses shown in red. Particles that show signal in both channels are yellow. Blue boxes outline 1µm magnetic beads that contain an alkyne group, which served as a positive control (Cat. #CCT-1035, Vector Laboratories, Newark, CA, USA). White boxes outline viral-like particles that were newly produced over the timescale of this incubation.

During this methods development phase, a new type of fluorophore that does not fluoresce unless bound via a click reaction became commercially available (Shieh et al., 2015) (CalFluor 647 Azide, SKU #CCT-1372, Vector Laboratories). By eliminating the fluorescence from unbound dye, Dr. Alon Philosof found that the viral-BONCAT signal to noise ratio was substantially improved. This also removed the need for additional washes of the sample after the click reaction.

Viral-BONCAT in suspension

Viruses filtered onto an Anodisc cannot be fluorescently sorted; while Anodiscs allow for precise counts of newly-produced viruses, they remove one of the major advantages of the viral-BONCAT method over existing dilution methods. Once we had a reasonably optimized protocol on Anodisc filters, we worked to optimize this protocol in suspension. This work was largely spearheaded by Dr. Alon Philosof and refined by Dr. Francisco Martinez-Hernandez. In brief, they found that incubating the click reaction under argon gas to prevent oxidation of reagents and allowing the reaction to run overnight resulted in visibly improved signal. Remaining click cocktail and SYBR Gold could be washed from the suspensions in preparation for sorting with 100kDa MWCO Amicon ultrafiltration devices. In combination with the techniques described earlier in this chapter to extract viruses from sediments, viral-BONCAT can now be applied in the complex marine subsurface.

A note on the limitations of BONCAT

A question not addressed in this thesis but of concern to any user of the BONCAT method is the degree to which HPG's replacement of methionine affects protein function in a cell. Changes to cell function can easily throw results into question, especially in time- or rate-based experiments. A previous study has shown that despite some visible changes in cell physiology upon addition of noncanonical amino acids, none of the changes observed were statistically significant (K. F. Steward et al., 2020). However, unpublished work by Dr. Alon Philosof and undergraduate student Matticus Brown in the Orphan lab has found differences in phage plaque morphology between *Escherichia coli* grown on HPG and infected with P1 phage and *E. coli* grown without HPG and infected with P1. Further investigation is warranted to understand the extent to which noncanonical amino acids may influence phage infection and production.

A further confounding factor of special importance in environmental samples is nutrient uptake rates. We cannot confirm that all species in a given environmental system will take up HPG at the same rates, if they take it up at all. It is important to acknowledge that while this tool is extremely powerful and allows us to obtain classes of data that we could not before (i.e., a temporally-constrained understanding of viral replication rates with the ability to identify the newly-produced viruses), there are certain built-in assumptions regarding the distribution of

HPG. As we will see in the following chapter, these constraints highlight the importance of using multiple techniques in estimating growth rates, community sizes, and doubling times in environmental samples.

CONCLUSIONS

These methods allow us to collect temporally-constrained ecological data that are not available to us via sequencing alone, such as the number of viruses produced, how long it takes for certain nutrients to flow through the system, and the identities of viruses produced in a given time period. They also work effectively in environmental samples, precluding the need for pure cultures and providing us with a more realistic picture of complex ecosystems. In the next chapter, we will put into practice some of the techniques discussed above using sediment samples from Costa Rica cold seeps.

Chapter 5

TEMPORALLY CONSTRAINING VIRAL PRODUCTION AND THE TRANSFER OF NITROGEN TO THE VIRAL POOL IN COLD SEEPS

Aditi K. Narayanan (1), Alon Philosof (1), Yunbin Guan (2), Victoria J. Orphan (1,2)

1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA
2. Division of Geology and Planetary Sciences, California Institute of Technology, Pasadena, California, USA

ABSTRACT

Despite extensive genomic data, the ecological impacts of viral lysis and production are entirely unconstrained in methane cold seeps. This study implements ^{15}N stable isotope probing coupled to nanoSIMS to examine the timescale on which host-assimilated ammonium is transferred into the pool of free viruses in seep sediments. We correlate the isotopic analysis with calculations of dissolved inorganic carbon (DIC) production from $^{13}CH_4$ as a proxy for anaerobic methane oxidation activity, observations regarding cellular community composition over time, and viral-BONCAT assays using the methionine analog HPG to place tighter constraints on viral production timescales. Our viral nanoSIMS and BONCAT data show distinct trends, with nanoSIMS data showing a decrease in ^{15}N enrichment below natural abundance initially before climbing by Day 35 of incubation, while BONCAT data shows immediate incorporation of HPG continuing until approximately day 28, matching the trend in host DIC production and shifts in the prokaryotic community. While data from the cellular community is needed to corroborate these patterns, they may suggest either a clear hierarchy of substrate preferences within the entire community or distinct populations that consume substrates differently. The initial depletion of ^{15}N may also be a result of cell lysis that returned large amounts of ^{14}N to the fixed nitrogen pool, possibly caused by stress from the experimental setup procedures. These results demonstrate the effectiveness of the viral-nanoSIMS and viral-BONCAT techniques in environmental samples and show us how best to refine them. They also provide the impetus for future work in which we examine deeper questions of how shifts in host activity might skew viral activity on a collective scale.

INTRODUCTION

As discussed in Chapter 1, viruses play a well-established role in geochemical cycling in the upper water column. The biological pump, the sequestration of carbon at the bottom of the ocean via sinking or in higher trophic levels via predation, can be disrupted by viral lysis of the photosynthetic microbes that underpin energy availability at the surface (Wilhelm & Suttle, 1999 and **Figure 1.2**). Experiments in model systems have also shown that viral lysis results in the availability of new carbon sources for otherwise nutrient-limited members of the microbial community (Middelboe et al., 2003). This viral lysis is also known to facilitate the release of critical nutrients and reopen ecological niches, though less is known about the degree to which viruses themselves serve as a sink for carbon, nitrogen, and phosphorus. Prior stable isotope proteomics work has shown that ^{15}N fed to infected cells of the model marine cyanobacterium *Synechococcus* is almost exclusively incorporated into the proteins of the progeny phage, with little to no incorporation into the proteins of the cellular host (Waldbauer et al., 2019). They also found that up to 41% of the nitrogen in the phage pool was derived extracellularly during infection, rather than being repurposed from existing nitrogen stores in the cell. Up to 70% of the phosphorus in the DNA of T2 and T4 phages of *E. coli* was also found to be derived from the surrounding media rather than from host phosphorus (Cohen, 1948).

In keeping with one of the themes in this thesis, there is comparatively little work in deep ocean sediments regarding the impacts of viral lysis. Marine sediments are thought to contain higher rates of lysogeny at 48.1%, in contrast to surface seawater where lysogeny rates are estimated at 38.4% (Yi et al., 2023), though other work in terrestrial soils (Santos-Medellín et al., 2023; Trubl et al., 2018) and our own data from Costa Rica has found that lytic or pseudolysogenic phage may dominate (Chapter 3). From the standpoint of the virus, choosing lysogeny or pseudolysogeny makes perfect sense. In sediment environments with lower flow rates, the odds of collision with a new host are low; a virus may increase its chances of survival and propagation if it adopts a more temperate lifestyle. Geochemically speaking, lower rates of viral lysis could result in lower rates of turnover for nutrients and a slower rate of change in the composition of the prokaryotic community. At the global level, the decomposition of benthic viruses is thought

to release up to 50 megatons of fixed carbon per year (Dell'Anno et al., 2015)[3]. However, little data exists that quantifies fixed nitrogen in benthic viral mass, and there are no current estimates of viral production rates in methane seeps.

The intention behind this study was to begin unravelling viral production rates and the role of viral communities in the movement of nitrogen through cold seep ecosystems. We were further interested how these factors might change when the keystone microbial taxa, i.e., the methane oxidizing consortia, were perturbed. The ANME/SRB syntrophy is thought to be based on direct interspecies electron transfer, in which electrons generated from methane oxidation are directly transferred to the sulfate reducing bacteria via multiheme cytochromes (Chadwick et al., 2022; Skennerton et al., 2017; Wegener et al., 2015). Previous work in the lab found that the humic acid analog anthraquinone-2,6-disulfonate (AQDS) could decouple the ANME/SRB syntrophy by replacing sulfate as the electron acceptor in the methane oxidation reaction, rendering the SRB metabolically inactive (Scheller et al., 2016) (**Figure 5.1**).

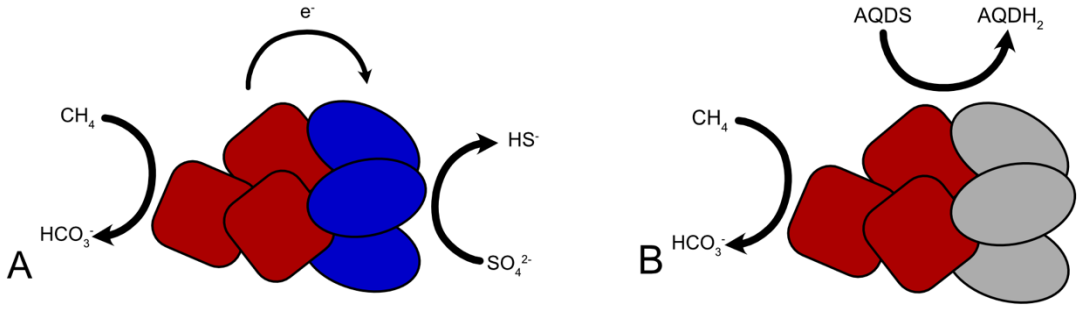

Figure 5.1. Schematic of methane oxidation processes by electron acceptor. **A)** Hypothesized mechanism by which the ANME/SRB syntrophy works, in which the electrons generated by the ANME (red) from anaerobic methane oxidation are transferred directly to the SRB (blue) for sulfate reduction **B)** The mechanism by which AQDS is thought to work, in which the electrons that would normally be transferred to the bacteria are shunted to AQDS instead.

Viral capsid morphologies from preliminary transmission electron microscopy (TEM) data that I collected indicated that there was a shift in the viral community between a laboratory sediment

[3] It is worth noting that the global carbon storage in marine sediments is around 2300 Pg, or 2300000 megatons (Atwood et al., 2020). In comparison, viral decay seems totally negligible. As we will see when we examine our BONCAT data, however, viral decay may still visibly affect nitrogen sourcing for organisms. I would also suggest that since variations on the order of parts per million and parts per billion in atmospheric carbon and NO_x have had profound impacts on the climate, discounting seemingly small sources of any nutrient may not serve us well.

incubation treated with AQDS and one without (**Figure 5.2a and 5.2b**, incubations developed and maintained by Dr. Hang Yu and Yongzhao Guo). Most spectacularly, the fraction of spindle shaped viruses in the decoupled bottles is much greater than their fraction in the bottles with normal methane oxidation/sulfate reducing activity. Spindle shaped viruses have thus far only been described in archaea (Ackermann & Prangishvili, 2012), and their greater presence in bottles where the keystone bacterial taxa are rendered inactive while the archaea continue to function is striking. I also measured capsid areas in FIJI (Schindelin et al., 2012, 2015; Schneider et al., 2012) as a proxy for the size of the viral genome to see if there were any differences in viral size between the AQDS and non-AQDS conditions (**Figure 5.2c and 5.2d**). Viral capsids are packed so tightly as to render the DNA nearly crystalline (Nurmemmedov et al., 2007), so the size of the capsid is a decent first approximation for the size of the genome. The Kolmagarov-Smirnov test showed a lack of significance ($p=0.084$) in capsid sizes when tails are excluded (the tails of tailed viruses do not contain DNA), suggesting that while capsid morphologies might be different, the overall size of the viral genome is likely to be similar between the two conditions. That said, the estimations of capsid area made here are imperfect; a transmission electron micrograph is a two-dimensional rendering, and I cannot say with certainty the angle at which a capsid has landed on the grid. Nevertheless, the initial data regarding differences in capsid morphology, in combination with data suggesting that the genome sizes themselves were quite similar despite the divergence of the viral community, provided the impetus for the project described in this chapter. We set up fresh incubations with the hope of tracking this community-level divergence under AQDS treatment over time.

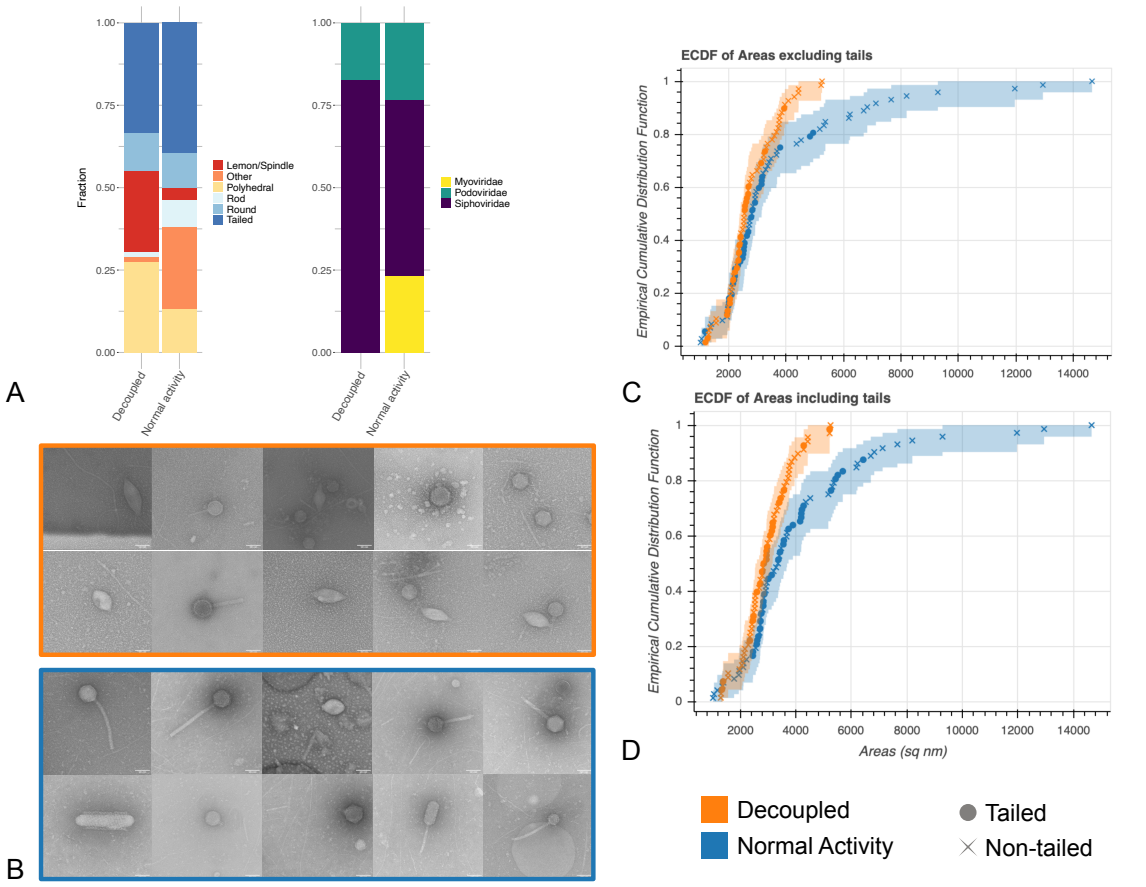

Figure 5.2. Analysis of viral morphologies and sizes between AQDS-treated and sulfate-treated incubations. **A)** Bar graph indicates the relative proportions of the different viral morphologies found in these samples. The graph on the left shows the relative proportions of all the major morphological types, while the graph on the right shows the subset of tailed viruses. The AQDS-decoupled bottles (left bar in each graph) show a higher proportion of lemon/spindle shaped viruses, which have thus far only been found infecting Archaea, while the normal activity bottles (right bar in each graph) show a higher diversity of tailed phages, with representatives from all three major groups. For examples of the different morphological types, see **Appendix I Figure A1.1**. **B)** Example transmission electron micrographs of suspected viruses from AQDS-decoupled (orange) and non-AQDS (blue) samples. Scale bars are 50nm. Microscopy was performed at the National Center for Imaging and Microcopy Research at the University of California, San Diego. **C)** Empirical cumulative distribution function of the areas of the capsids, excluding the tails in the case of tailed viruses. Shaded areas represent 95% confidence intervals. Capsid areas were measured by manually tracing each capsid in imageJ and using their area measurement function. **D)** Empirical cumulative distribution functions of the capsid areas including tails. Shaded areas represent 95% confidence intervals. The Kolmagarov-Smirnov test does not show significant differences in overall sizes of viruses between the bottles when measurements exclude the tails (P=0.084), but do show significance when measurements include the tails (P=0.0149). The measurements excluding tails are the more accurate proxy for genome size. Laboratory incubations from which these data were derived were developed and maintained by Dr. Hang Yu and Yongzhao Guo.

For our experiment, we chose sediments sampled from the same site of active methane seepage in Costa Rica that is described in Chapters 2 and 3, with the goal of creating a more comprehensive picture of the viral community at that site beyond sequence analysis. Though our initial motivation was simply to understand the differences in community composition between AQDS-decoupled and non-decoupled methane oxidizing communities, we realized that this system would allow us to ask questions that went beyond who was present or absent under certain conditions. To that end, we employed the techniques outlined in Chapter 4: viral-bioorthogonal non-canonical amino acid tagging (BONCAT) and viral nanoscale secondary ion mass spectrometry (nanoSIMS). Our goals were as follows:

1. Temporally constrain the timescale on which fixed nitrogen, a crucial nutrient, moves into the viral pool in a methane cold seep ecosystem.
2. Estimate how quickly viruses are produced in methane cold seeps.
3. Understand how the results of the two goals outlined above change if you perturb the keystone taxa in the environment by decoupling ANME and SRB via the addition of the humic acid analog AQDS.

As you will see below, the decoupling using AQDS was not as effective as we might have hoped. With a renewed focus on the canonical methane-oxidizing syntrophy, we implemented both BONCAT and nanoSIMS in a multi-month incubation of sediment treated with isotopically labeled ammonium and the methionine analog necessary for BONCAT, L-homopropargyl glycine. Despite biases in the nanoSIMS dataset towards larger viruses, the broad pattern was that HPG incorporation into viral-like particles appears to occur faster than ^{15}N incorporation, suggesting hierarchical substrate preferences by the hosts, stress-induced changes in the composition of the fixed nitrogen pool, or perhaps entirely distinct populations. Future work will be able to test these hypotheses and more accurately answer the three questions listed above via refined implementation of the techniques that we have now shown to be effective.

MATERIALS AND METHODS

Sample choice and preparation

Because of the large volume of sediment needed for this experiment, we knew we would need to pool sediment from multiple horizons of a sediment pushcore (Appendix II). We examined the sediment cores used in Chapter 3's analysis of Costa Rica viral genomes and used 16S rRNA amplicon sequencing to determine which horizons were similar enough in microbial community

composition to pool (**Figure 5.3**). The samples that best optimized consistency in community composition and had enough sediment for our purposes was the core from Little Knoll, Mound 12. Equal amounts of sediment from the bottom four horizons (sample serial numbers 10140-10143, pushcore depth of 9-21cm below seafloor) were pooled, and sediment from horizon 10138-39 (depth of 5-9cmbsf) was used as a parallel horizon control to help account for effects based on sample choice.

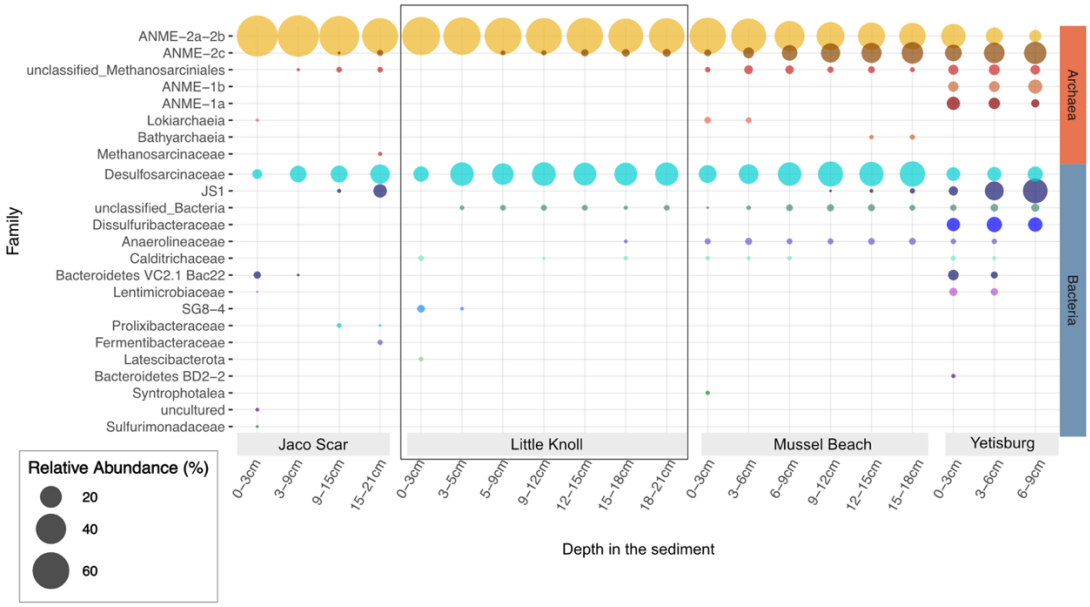

Figure 5.3. 16S rRNA amplicon sequencing of the laboratory incubations of the four sediment pushcores investigated in Chapter 3. Sample names, on the x-axis, are listed with the sediment depth as centimeters below seafloor. The size of the bubbles is proportional to the family-level taxon's relative abundance within the sample, with taxa shown only if they comprise >1% of the total reads in the sample. The Little Knoll core from Mound 12, outlined with a black rectangle, has the most consistent cellular community with increasing sediment depth, and was thus chosen for the experiment outlined below.

AQDS' ability to decouple the ANME/SRB partnership is less effective in the presence of sulfide. It is thought that AQDS reacts with sulfide and produces zero-valent sulfur, which is inhibitory to methane oxidation via an as-yet unknown mechanism of toxicity (Yu et al., 2018, 2021). In order to remove sulfide, which is a product of methane oxidation coupled to sulfate reduction, the sediment was washed with anoxic sulfate-free seawater (See Appendix II). The washing process involved removing the overlying seawater, adding sulfate-free seawater of at least equal volume to the sediment, and mixing vigorously before incubating under a 0.15MPa methane headspace to help oxidize remaining ambient sulfate into sulfide (recall that methane

oxidation coupled to sulfate reduction results in the production of inorganic carbon and sulfide). The sediment was allowed to settle for a week. This washing protocol was repeated three more times before the experiment began.

Incubation setup

The incubations were treated with the following substrates in combinations outlined in **Figure 5.4**:

- ^{15}N-labeled ammonium. The incorporation of heavy isotope-labeled fixed nitrogen to track nitrogen flow through the system.
- ^{13}C-labeled bicarbonate. Bicarbonate is a carbon source for chemosynthetic autotrophs in the deep ocean (Kellermann et al., 2012), and if detectable using our nanoSIMS methods, could give us a picture of the flow of carbon in addition to the nitrogen data above.
- L-homopropargylglycine (HPG): Methionine analog used in BONCAT; a tracer for activity.
- ^{13}C labeled methane: Allows us to track the production of ^{13}C labeled dissolved inorganic carbon (DIC) as a proxy for methane oxidation activity.
- Anthraquinone 2,6-disulfonate (AQDS): alternative electron sink for the process of methane oxidation.

Incubations were set up in 160mL serum vials. To each bottle was added 90mL of sediment slurry, which was one part sediment to two parts seawater, with nitrogen gas bubbling through the sediment to keep it anoxic. The bottles were allowed to settle overnight. The following day, the overlying seawater was removed from the settled sediment and replaced with fresh anoxic seawater. Each bottle treated with AQDS was incubated in low-sulfate (2mM) seawater, while those with normal activity were incubated in seawater containing 10mM sulfate (see recipes in Appendix II, developed by Dr. Hang Yu and Yongzhao Guo). AQDS (disodium anthraquinone-2,6-disulfonate, CAS #853-68-9, Cat. #372382, Tokyo Chemical Industry, Tokyo, Japan) was added in powdered form directly to the relevant incubations. Because labeled ammonium and bicarbonate were used in some incubations, the seawater recipes were made without any ammonium or bicarbonate. Instead, 1M solutions of ^{13}C $NaHCO_3$ (98 at%, CAS #87081-58-1, Cat. 372382, Sigma Aldrich, St. Louis, Missouri, USA) or ^{12}C $NaHCO_3$, and 500mM solutions of ^{15}N NH_4Cl (100 at%, CAS #39466-62-1, Cat. 299251, Sigma Aldrich) or ^{14}N NH_4Cl were made anoxically and added to individual incubations at the concentration required in the

seawater recipes. HPG (Cat. #CCT-1067, Vector Laboratories, Newark, California, USA) was added to the relevant incubations at a final concentration of 50μM using a 50mM anoxic stock solution. For a brief protocol on making anoxic solutions, see Appendix II.

Before the addition of ammonium, bicarbonate, and HPG, the bottles were sealed with butyl rubber stoppers and sparged with 100% methane gas for ten minutes. The relevant amounts of HPG and labeled/unlabeled ammonium and bicarbonate were injected into each bottle using a 23-gauge needle and syringe that were flushed with methane before being used to withdraw reagents from the anoxic stock bottles. Bottles that did not contain ^{13}C labeled methane were pressurized to 0.2MPa ^{12}C methane.

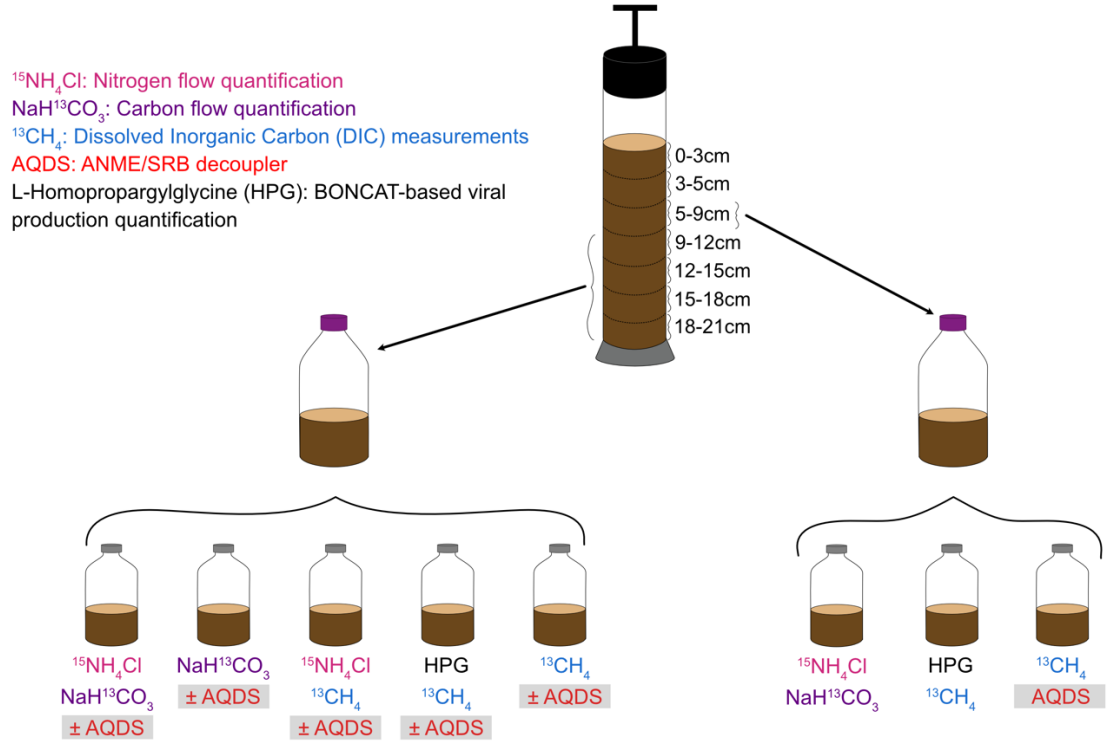

Figure 5.4 Experimental setup. **Left:** Sediment from the bottom four horizons of a pushcore from Costa Rica were incubated with combinations of isotopically labeled fixed nitrogen, isotopically labeled bicarbonate, isotopically labeled methane, and HPG. Each combination was incubated with or without AQDS. Where bottles do not contain isotopically labeled substrates, they are treated with the standard isotope version of the substrate (e.g., ^{14}N or ^{12}C). **Right:** A parallel sediment horizon from the same pushcore was incubated with the same substrates as an internal control for bottle effects. As described in the results and discussion, the AQDS treatment did not work (indicated by grey boxes), and AQDS-treated bottles were not analyzed.

Addition of ^{13}C labeled methane

^{13}C labeled methane (99at%, CAS #6532-48-5, Cat. #490229, Sigma Aldrich) was added to the relevant incubations at a final concentration of 5%. The calculation for the amount of labeled methane to be added is as follows:

1. Calculate the amount of headspace in the bottle. Because we filled 90mL of a 160mL bottle, we had 70mL of headspace left.
2. Multiply the headspace volume by the number of atmospheres to which you are pressurizing your bottles. We used 0.2MPa, which is 2atm. That meant our total headspace volume was 140mL.
3. 5% of 140mL is 7mL of ^{13}C methane.

After injecting 7mL of ^{13}C methane, the remainder of the headspace was filled with ^{12}C methane until it reached the desired final pressure of 0.2MPa.

Sampling protocol

The following types of samples were taken:

1. 1mL slurry for bulk DNA and 16S rRNA amplicon sequencing, flash frozen in liquid N_2 and stored at -80˚C.
2. 1mL slurry for bulk RNA, incubated overnight in RNA Later (Cat. # AM7021, ThermoFisher Scientific) under an anoxic nitrogen headspace. The following day, the sample was spun down and liquid was removed before the sediment pellet was flash frozen in liquid N_2 and stored at -80˚C.
3. 3mL slurry for viral nanoSIMS, viral BONCAT, and transmission electron microscopy; processed as described in Chapter 4 under *Extraction of Viruses from Sediment Samples*.
4. 0.5mL slurry for cellular fluorescence *in situ* hybridization (FISH), fixed for one hour at room temperature in a final concentration of 2% paraformaldehyde. Fixed samples were washed three times with 1X PBS and stored at -20˚C in 1mL of ethanol 50:50 1X PBS.
5. 100µL of overlying seawater to track sulfide production rates (Cline 1969). The sample was rapidly filtered to prevent oxidation and 20µL of filtrate were placed into 400µL of 0.5M zinc acetate solution.
6. 100µL of seawater for dissolved inorganic carbon (DIC), a proxy for methane oxidation to HCO_3^-. The sample was injected into a 12mL Exetainer vial (Cat. #738W, Labco, Lampeter, Wales, UK) containing 100µL of 45% phosphoric acid and flushed with helium gas.
7. 0.5mL of seawater for ion chromatography (IC) analysis, flash frozen in liquid N_2 and stored at -80˚C.

8. 100μL of seawater from the AQDS-containing samples to track the amount of AQDS that is reduced over time.

Sampling occurred weekly (Day 2 after incubation setup, 7, 14, 21, 28, 35, 42, and 56). Additional geochemical samples (sulfide, AQDS, IC, and DIC) and viral samples were taken on days 10 and 17 to monitor activity early in the experiment. No media replacement occurred over the course of the experiment; however, additional methane was added to replace the volume lost from sampling on Days 10, 27, and 42. Final samples were taken on Day 77 and 117, mostly because we were curious as to what might happen after multiple months of incubation with no replacement of substrates.

Transmission Electron Microscopy

TEM was performed as described in the methods of Chapter 2. Electron microscopy was done in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech and at the National Center for Imaging and Microscopy Research at the University of California, San Diego.

16S rRNA gene amplicon sequencing and analysis

All 16S sequencing and data analysis, including the ordination and analysis of variance, was performed as described in Chapter 2. For the ordination plot in which only taxa with greater than 1% abundance were included, the empty cells in the matrix were filled with zeros before Bray-Curtis dissimilarity was calculated.

Analysis of Dissolved Inorganic Carbon

100μL of medium from the incubations was injected into a 12mL Exetainer vial containing 100μL of 45% phosphoric acid and flushed with helium gas prior to sample injection. Headspace measurements of carbon dioxide were measured on a GC-IR-MS GasBench II (Thermo Fisher Scientific, Waltham, MA, USA) using helium as the carrier gas. DIC production was calculated as described in Scheller et al., 2016, including the assumption that isotope effects on methane oxidation were minimal. Briefly, the change in DIC is calculated as

$$\Delta[DIC](t_n) = [DIC](t_0) * (^{13}F(t_n) - {}^{13}F(t_0)) / (^{13}F(CH_4) - {}^{13}F(t_n))$$

where

- [DIC](t_0) = 5mM, the starting amount of sodium bicarbonate in the artificial seawater medium
- $^{13}F(t_n) = R^{13}C/^{12}C/(1+ R^{13}C/^{12}C)$, where $R^{13}C/^{12}C$ is the ratio value measured by the GasBench II at that timepoint
- $^{13}F(t_0) = R^{13}C/^{12}C/(1+ R^{13}C/^{12}C)$, where $R^{13}C/^{12}C$ is the ratio at Day 2 as measured by the GasBench II
- $^{13}F(CH_4) = 0.05$ (the 5% initial concentration of $^{13}CH_4$)

Laboratory BONCAT protocols and NanoSIMS protocols are discussed in Chapter 4.

Analysis of nanoSIMS data

Square image rasters were between 10-15μm on each side, with 8-12 planes taken at a single spot within the sample. Image segmentation of nanoSIMS data was performed manually in FIJI (Schindelin et al., 2012, 2015; Schneider et al., 2012) using the ROI Manager tool on the ^{14}N nanoSIMS images. All ^{14}N images were scaled such that the maximum count value was 1000. ROIs were counted as a virus if they were present in no more than three consecutive planes within an image, thus assuming that any particle persisting over more than three planes was likely too large to be a virus. ROIs were also excluded if the maximum intensity of the particle was on the first or last plane. ROI sizes were limited to no more than 200nm in length. All ROIs were then filled in and converted to a mask using the Mask From ROI(s) ImageJ plugin (Thomas & Trehin, 2021). ^{15}N enrichment was calculated for each ROI as the ^{15}N value divided by the sum of the ^{14}N and ^{15}N values using a pipeline written by Dr. Daniel Utter, a postdoc in the Orphan lab. Nitrogen values for both isotopes were derived solely from the plane on which the ROI was drawn, rather than from the sum value of all planes at that location.

Analysis of BONCAT data

Viral-BONCAT and the subsequent data analysis was performed by Dr. Alon Philosof. Briefly, 1mL of sample was divided into two 500μL aliquots that underwent the BONCAT click reaction in parallel. Viruses were then stained with SYBR-Gold and the sample was filtered onto a Whatman Anodisc aluminum oxide wafer with 0.02μm pore size (Millipore Sigma, Cat. #WHA68096002). The 1mL of sample represents a fraction of the total 14mL obtained from the extraction of VLPs from 1mL of sediment slurry. Between 50-55 images were taken per sample on an Olympus BX51 microscope (Olympus Corporation, Tokyo, Japan) using a 100X

oil objective with a FITC (SYBR-Gold) and CY5 (BONCAT, CalFluor 647) filter cube and an R6 Retiga camera (Teledyne Vision Solutions, Thousand Oaks, CA, USA) with a 200ms exposure time. Viral-like particles were segmented and their pixel intensity was calculated in FIJI. ROI size and circularity thresholds were used for segmentation. For BONCAT-labeled VLPs, particles were between 0.05-2.5px with a circularity of 0.8-1, while SYBR-Gold stained VLPs were between 0.03-3.5px with a circularity of 0.7-1. SYBR-Gold-based segmentation had broader criteria due to the potential for nonspecific binding resulting in slightly larger particle sizes as seen under the microscope.

RESULTS AND DISCUSSION

Methane oxidation activity plateaus after 30 days of incubation

Measurements of dissolved inorganic carbon (DIC) based on the oxidation of isotopically labeled $^{13}CH_4$ indicated that AQDS-decoupled incubations consistently produced less DIC than sulfate-containing incubations, suggesting that methane oxidation activity was substantially lower (**Figure 5.5**); at most, AQDS-treated bottles produced between 36%–65% of the DIC of sulfate-treated bottles. Previous work in the lab has shown that AQDS-decoupled incubations show levels of DIC production at approximately 80% of canonical sulfate-reducing levels (Scheller et al., 2016; Yu et al., 2021). We suspect that the reason for the deficiency in our incubations was insufficient washing of the sediment to remove residual sulfide before the incubations were set up. Because another member of the lab was able to successfully decouple some of his incubations with the same reagent lot number, we do not believe that we had faulty AQDS. Though the failure of the AQDS decoupling process meant we could no longer make progress on the third goal listed in the previous section, our first two questions still addressed an unknown in viral ecology; at this time, there are few studies that attempt to quantify baseline viral turnover and nutrient flow in methane cold seeps, even without the added variable of host metabolic manipulation.

Though showing us the unfortunate failure of the AQDS decoupling, the DIC data help us constrain when methane oxidation activity is at its highest. We find that activity began immediately and continued until approximately Day 30, at which point it asymptotes as methane

and sulfate are presumably consumed. Unfortunately, there were problems with the GC-IR-MS instrument that resulted in contamination during measurement of all samples past Day 56.

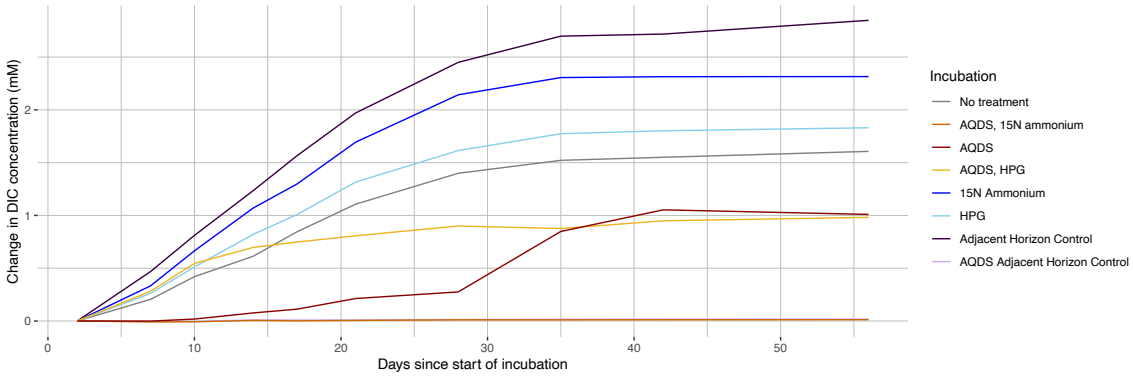

Figure 5.5. Dissolved inorganic carbon (DIC) production over the course of the experiment, a proxy for anaerobic methane oxidation activity; measurements were based on the incorporation of the ^{13}C-labeled methane into the DIC pool. All AQDS-treated incubations produced less DIC than non-AQDS-treated bottles, and some AQDS-treated bottles showed no appreciable DIC production. The bottles without AQDS, which will form the remainder of this chapter, all show rapid onset of activity, which then levels off around Day 30.

The cellular community shifts as substrates are consumed

We performed 16S rRNA amplicon sequencing on multiple timepoints for each incubation to investigate whether the cellular community changed over time and/or by incubation treatment. The community did indeed appear to have shifted (**Figure 5.6a**), and clustering analysis of the community composition at the family level found that incubation time was a significant factor (PERMANOVA, p=0.010), while the substrates added to the bottle were not (p=0.090). Visually at least, the clusters appear to show some separation between Day 21 and Day 35, which matches the time by which DIC production plateaus around Day 28. It may seem unsurprising to see a change in the community as methane oxidation activity, the keystone metabolic process in the seep, ceases. However, doubling times for ANME and SRB consortia are on the order of months (Nauhaus et al., 2007; Orphan et al., 2009; Timmers et al., 2015); while few estimates of doubling times exist for non-ANME/SRB cells in seep sediments, it is reasonable to assume that they are similarly slow. I speculate that the shift in microbial community may partially be a result of viral infection and lysis of cells. Indeed, a recent study examining deep sea sediment communities over 30 days showed that viral activity could shift cellular microbial communities in favor of complex organic matter (COM) degrading organisms (C. Wang et al., 2025), with the

caveat that this study was limited to bacteria and did not consider the archaeal response. By examining the community response to the addition of concentrated phages, the authors hypothesized that the presence of COM may stimulate the production of chronic phages in deep sea sediments, and that these phages themselves may in turn play a role in the breakdown of COM.

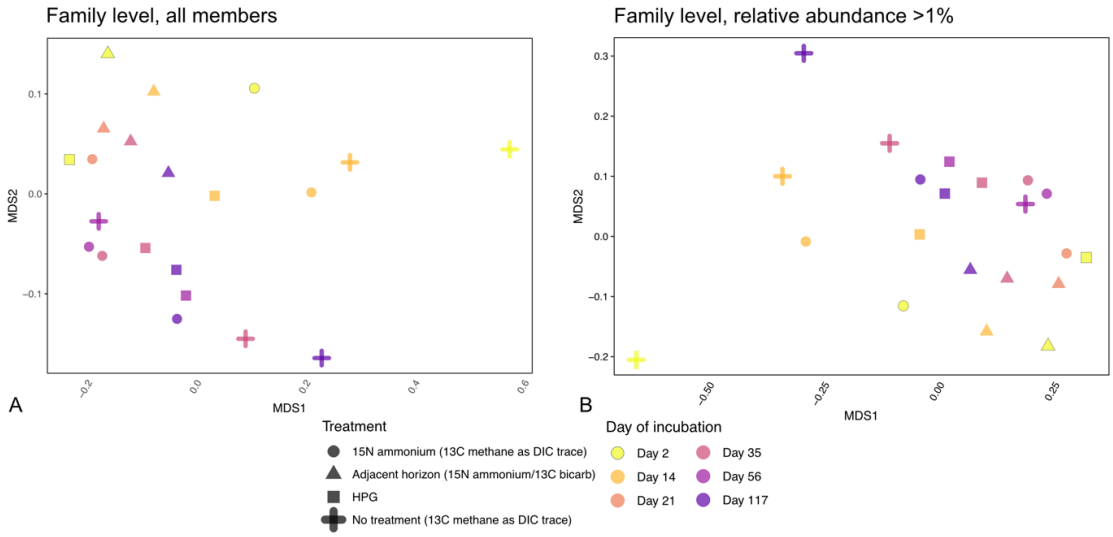

Figure 5.6 Shifts in prokaryotic community composition over the course of the experiment. **A)** Ordination plot, nonmetric multidimensional scaling (NMDS) of 16S rRNA genes at the family taxonomic level. Stress=0.065. ADONIS values indicate that the separation between communities is driven more by the incubation time than by the treatment provided to each incubation. **B)** Ordination plot, NMDS of 16S rRNA genes at the family taxonomic level where only those taxa with at least 1% relative abundance in at least one sample are included. Stress=0.074. ADONIS values indicate that the separation is now driven more by the treatment than the incubation time. However, the general pattern is the same between the two plots, with the layout of points appearing to be mirror images of each other.

Strikingly, when we eliminate all family-level taxa that are below a 1% relative abundance threshold (**Figure 5.6b**), the pattern changes: the incubation time is no longer significant ($p=0.10$), while the treatment is ($p=0.013$), though the latter may be the result of incubation in separate bottles rather than the treatment itself. The temporal shift we see in the complete community (**Figure 5.6a**) may thus be caused by an increase in less abundant members or a decrease in previously abundant members. Because the analysis is based on relative abundance, quantitative 16S rRNA sequencing and/or examining rank abundance data may allow us to understand exactly which taxa drive this change. Preliminarily, when we compare the family-level taxa with >1% relative abundance at the beginning and the end of the incubation

(**Supplementary Figure S5.1**), the consistency in the relative abundance of these taxa over time suggests that the shift is primarily driven by the rare members of the community.

BONCAT and nanoSIMS show different temporal patterns of viral activity

Initial data shows that viral production as measured by BONCAT increased rapidly and began to plateau sometime between Days 28-42, i.e., weeks 4-6 of incubation (**Figure 5.7a**). Interestingly, the shapes of the histogram for both the BONCAT-tagged viruses and the general population stained with SYBR-Gold change dramatically with time, with the SYBR-Gold bulk population becoming more bimodal and the BONCAT-tagged viral distribution becoming more unimodal. This may indicate the presence of multiple distinct viral populations that turned over as substrates were consumed. While the number of VLPs counted varied between timepoints, the percent of BONCAT-tagged viruses plateaued around Day 28 and stayed steady until Day 77 (**Table 5.1**). The total number of viruses, however, dropped between Day 28 and Day 42 before doubling by Day 77.

There are a couple of possible explanations for this pattern. The simplest is that we may simply have a greater degree of background noise in the images from Day 77 that cannot be easily distinguished from real VLPs; this does not appear to be the case (**Supplementary Figure S5.2**). Another, which would be ecologically more interesting, is that this observation is the result of methane oxidizing activity plateauing by Day 28, as shown by the DIC data above. Viral populations began to decay as the hosts directly dependent on AOM became less metabolically active, resulting in a trough in viral numbers by Day 42. By Day 77, a different heterotrophic population was dominating the metabolic landscape. As discussed, the 16S rRNA analysis indicates a shift in the prokaryotic community. The number of BONCAT-active viruses thus increased as a result of higher heterotrophic cellular activity levels. Possibly the percent of HPG-tagged viruses showed no appreciable change (**Table 5.1**) due to the low remaining levels of free HPG in the media, but the signal also did not decrease because a large amount of HPG is released from decaying phage and recently dead cells. It may thus be that 7.6% is the highest labeling percentage possible in this system with the 50μM concentration of HPG in the media. Work in cellular microbes from terrestrial soil samples found a similar unexplained plateau in BONCAT activity, though on the order of days rather than weeks (Couradeau et al., 2019).

Timepoint	*N* SYBR-Gold VLPs	*N* BONCAT VLPs	% BONCAT
Day 2	23260	69	0.3%
Day 14	21371	827	3.9%
Day 21	21027	1019	4.8%
Day 28	23604	1784	7.6%
Day 42*	11687/12856	894/983	7.6%
Day 77	27050	2059	7.6%

Table 5.1 Number of viral-like particles counted per 1mL of viral suspension (i.e., the viruses from 1/14th of 1mL of sediment slurry, see methods). Recall that SYBR-Gold is a nucleic acid stain, and thus represents total VLPs, while BONCAT-active viruses represent those which have incorporated HPG by the given timepoint. *Day 42 has two numbers in each column due to loss of 10% of the sample. In the format [first #]/[second #], the first number is the actual count from the microscopy images and the second number is an estimation of the value had the 10% sample loss not occurred.

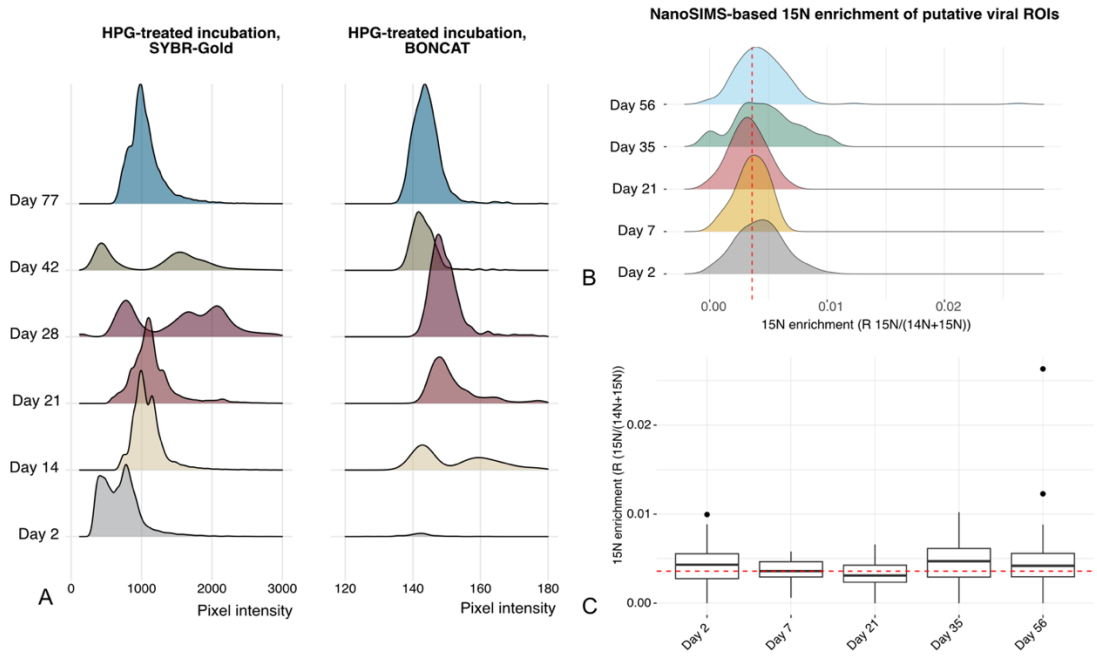

Figure 5.7 Results of viral-BONCAT and viral-nanoSIMS. **A) Left:** Ridge plot showing pixel intensity of all SYBR-Gold events. SYBR Gold is a nucleic acid stain and should reveal all viral-like particles (VLPs). **Right:** Ridge plot showing pixel intensity of BONCAT-positive VLPs. These data were collected from the non-AQDS treated incubation containing only HPG. *N* for VLPs are available in **Table 5.1**. These ridge plots were generated by Dr. Alon Philosof. **B)** Ridge plot of ^{15}N enrichment of all viral ROIs (VLPs) from the ^{15}N ammonium non-AQDS decoupled bottles. The majority of the data is from the bottle containing ^{15}N ammonium and the ^{13}C methane tracer. *N* for VLPs are available in **Table 5.2 C)** Boxplot representation of the same data as in figure B. The drop in enrichment is easier to see with the boxplot, while the overall shape of the distribution is clearer in the ridge plot. In figures B and C, the natural abundance of ^{15}N (0.0036) is plotted as a dashed red line.

In contrast, the nanoSIMS signal has an almost inverse pattern to the BONCAT data (**Figure 5.7b**). Mean enrichment of VLPs was above ^{15}N natural abundance by Day 2 (**Figure 5.7c**), though this value dropped by Day 21. The enrichment on Day 2 may be a genuine signal of

rapid initial uptake of the labeled ammonium or may be due to human error (for example, insufficient washing of the sample or unlikely cross contamination of timepoints during nanoSIMS sample preparation). By Day 35, ^{15}N enrichment increased until it began to drop again by Day 56. The fact that ^{15}N enrichment dropped below natural abundance in the first few weeks of the experiment might be a result of the experimental setup itself. Host cell stress, including but not limited to DNA damage, antibiotic stress, and oxidative stress, is thought to be one mechanism by which lysogenic phages make the switch to a lytic lifestyle (Fang et al., 2017; Levine, 1961; Lwoff, 1953). If the process of washing the sediment and setting up incubations stressed the cells within the sediment, lysogenic phages may have switched to the lytic mode and released a great deal of fixed ^{14}N into the organic matter pool, resulting in increased ^{14}N levels in the surviving cells and the subsequently produced viruses. Though limited data exists on nitrogen isotopic preferences of biological processes, some work on bacterial nitrogen fixation shows that ^{14}N is preferentially used over ^{15}N (Zhang et al., 2014). We may be observing an equivalent preference in the uptake of fixed nitrogen in these samples.

Ideally, we would want to know what had happened on Days 14 and 28 to better understand the pattern we see; unfortunately, due to a freezer malfunction, our samples from those two timepoints were destroyed. An additional consideration is that the nanoSIMS images were too large to resolve viruses smaller than 150nm, an important detail which was not clearly understood until the data had been analyzed. This dataset is therefore biased towards larger viruses.

Timepoint	***N* 14N VLPs**	***N* VLPs with enrichment over natural abundance**	**% enriched over nat. abun.**
Day 2	105	64	61%
Day 7	65	33	51%
Day 21	31	13	42%
Day 35	32	20	62.5%
Day 56	155	97	62.6%

Table 5.2 Percent of viral-like particles out of the total counted that were enriched above ^{15}N natural abundance (0.0036).

There are a few potential explanations for why viruses might show temporally different levels of HPG and ^{15}N incorporation. One is that viruses incorporating HPG and viruses incorporating ^{15}N comprise distinct populations, perhaps ones infecting organisms with preferences for

specific versions of substrates. Another is that had we been able to capture smaller viruses more accurately in the nanoSIMS data, the difference would not have existed at all. It is also possible that rates of uptake for the two substrates is simply different; a complete amino acid like HPG may be more likely to be scavenged in an otherwise energy-limited system, while the corresponding building blocks like ammonium, which require more energy to incorporate into biomass, are scavenged when the more advantageous substrates are consumed. Cellular BONCAT and nanoSIMS data, for which samples have been preserved, would be useful in corroborating (or contradicting) the existing pattern. For now, cellular analysis using samples from Santa Monica Basin by graduate student Rebecca Wipfler and former postdoc Dr. Kriti Sharma hints at agreement with the data above, i.e., that there is variance in ^{15}N and HPG uptake preferences that are possibly dependent on cell identity or metabolic state.

Assessing the benefits and challenges of viral-BONCAT and viral-nanoSIMS in heterogeneous environmental samples

Promisingly, our methods appear to be functional. Embedding the viruses in resin removes an important source of bias compared to previous methods. As discussed in Chapter 4, prior methods involved placing a suspension of viruses on a silicon wafer, imaging them, and rinsing the mounting fluid away (Pasulka et al., 2018). This results in uneven sample loss, which is adds additional uncertainty in heterogeneous environmental samples. An additional advantage of the embedding method is that theoretically, BONCAT could be performed on thin sections before nanoSIMS, allowing us to match BONCAT-positive viruses to ^{15}N incorporation.

However, I believe there are many lessons learned from this trial experiment that will result in more robust future experiments. I list these points below with their relative pros and cons:

- Double purifying viruses by performing density separation on the extracted viruses a second time. **Advantages:** Will remove a greater proportion of cell debris that might be mistaken for a virus in the nanoSIMS image. **Disadvantages:** Will result in additional sample loss, likely with bias.
- Smaller nanoSIMS rasters, e.g., 5x5μm rasters instead of 10x10μm or 15x15μm. **Advantages:** As discussed above, our relatively large raster size has likely resulted in an inability to resolve smaller viruses (a 50-60nm virus would be about 1 pixel). Our dataset is biased towards the activity of larger viruses. **Disadvantages:** Beam drift may be more noticeable between planes.

- Image thin sections with a more powerful microscope, such as the Zeiss LSM 980 with Airyscan2. **Advantages:** Such a microscope can would provide us with higher resolution images and the ability to image in the z-direction. Matching viruses between fluorescence images and nanoSIMS images would be substantially easier—with our current images, it is nearly impossible because we cannot visualize deep enough into the section. In experiments directly combining BONCAT and nanoSIMS analysis, reliable co-registration of viral-like particles between imaging platforms is important. **Disadvantages:** Costly on a per hour basis, unless the lab already owns such a microscope.
- Fluorescently sort viral-like particles after performing BONCAT and before embedding. **Advantages:** Separately embedding BONCAT-positive and BONCAT-negative viral-like particles allows us to easily quantify incorporation of the label in these two distinct populations. Fluorescent sorting will also provide accurate high-throughput counts. **Disadvantages:** May result in additional biases towards certain particle sizes and shapes. A great deal of sorting will also be required to have enough particles for nanoSIMS analysis.

As far as viral-BONCAT is concerned, while we are able to easily visualize viruses under the microscope, the downstream use of fluorescence-activated viral sorting (FAVS) has not been tested on these particular samples. Our preliminary tests with other sediments have shown that viral-like particles can be distinguished using flow cytometry and FAVS, and the next step is the application of this technique to this sample set for high-throughput information on viral counts.

CONCLUSIONS AND FUTURE WORK

Our data allow us to hypothesize that viral production in laboratory sediment microcosms is correlated with metabolic activity and suspected virus-mediated shifts in the cellular community composition, and that the percentage of labeled viruses produced reaches a plateau around the same time as methane oxidizing activity. This may be a result of shifts in dominant metabolic functions within the cell and viral communities as the primary electron donor and acceptor, methane and sulfate, are consumed. Should this hypothesis show merit, we would have a partial answer to our initial question of how viral activity shifts in response to host metabolic changes, even without data from AQDS-decoupled incubations. Future work that stifles methane oxidizing activity, for example by incubating with a nitrogen headspace instead of methane, would help us constrain whether the shifts in viral production truly are correlated with decreased methane oxidation or whether viral production patterns are simply a result of the length of

incubation time. Additional work should focus on constraining decay rates of viruses, which would aid calculations of production rates; our current data can give us approximate timescales on which viruses are most actively produced and decay, but not necessarily at what rates.

We have also generated further questions to test regarding how the uptake rates of various fixed nitrogen sources vary among methane seep microbes and associated viruses, with implications for the cycling of fixed nitrogen in marine sediments. Previous work in the Mound 12, Costa Rica methane seep, which is the same site used in this study, found that nitrogen fixation is predominantly mediated by ANME (Dekas et al., 2009). Viral lysis could be instrumental in returning various fixed nitrogen species to the dissolved organic matter pool; perhaps the size of this pool of bioavailable nitrogen impacts rates of ANME-mediated nitrogen fixation.

We have shown here that the techniques developed in Chapter 4 can be successfully implemented in a complex environmental sample, with lessons for how future scientists can improve upon the existing methods. There is a great deal to be learned simply by continuing to process the existing cellular, genomic, and geochemical samples. In the ideal future experiment, however, the bottles are successfully decoupled with AQDS and incubated with HPG and ^{15}N ammonium, extracted viral-like particles undergo the BONCAT click reaction for fluorophore addition, and viruses are fluorescently sorted. Some of those viruses might undergo single-viral genome sequencing, a technique that has been shown to work in surface ocean systems (Martínez Martínez et al., 2020; Martinez-Hernandez et al., 2017), in order to assign identity to activity. The other BONCAT-positive viral-like particles can undergo nanoSIMS in parallel to those that are BONCAT-negative. The cellular community should also undergo nanoSIMS to compare timescales of nitrogen uptake between cell and virus, and to compare host nitrogen uptake to the amount that is eventually found in the viral pool. The goals would be multifaceted: to quantify differences in viral production rates from AQDS-decoupled bottles and bottles with normal activity, better constrain whether the timescales in which viral activity becomes apparent are different between BONCAT and nanoSIMS, understand the degree to which BONCAT-active and isotope-labeled viruses are distinct and may constitute different populations, and create a more comprehensive and quantitative picture of nitrogen flow.

CONTRIBUTIONS

Dr. Alon Philosof and I jointly planned and set up the experiment and performed all sampling. I performed all electron microscopy, 16S rRNA data analysis, GasBench sample analysis, the majority of the nanoSIMS sample preparation, and analysis of the nanoSIMS data. Dr. Alon Philosof, Dr. Yunbin Guan, and I jointly collected the nanoSIMS data, with Dr. Philosof creating fluorescence microscopy maps of the samples and Dr. Yunbin Guan tuning and operating the instrument. Dr. Alon Philosof performed all BONCAT and the downstream analysis, extracted and amplified DNA for 16S rRNA analysis, and performed much of the nanoSIMS sample preparation. Dr. Stephanie Connon prepared the libraries for the 16S rRNA analysis.

SUPPLEMENTAL DATA

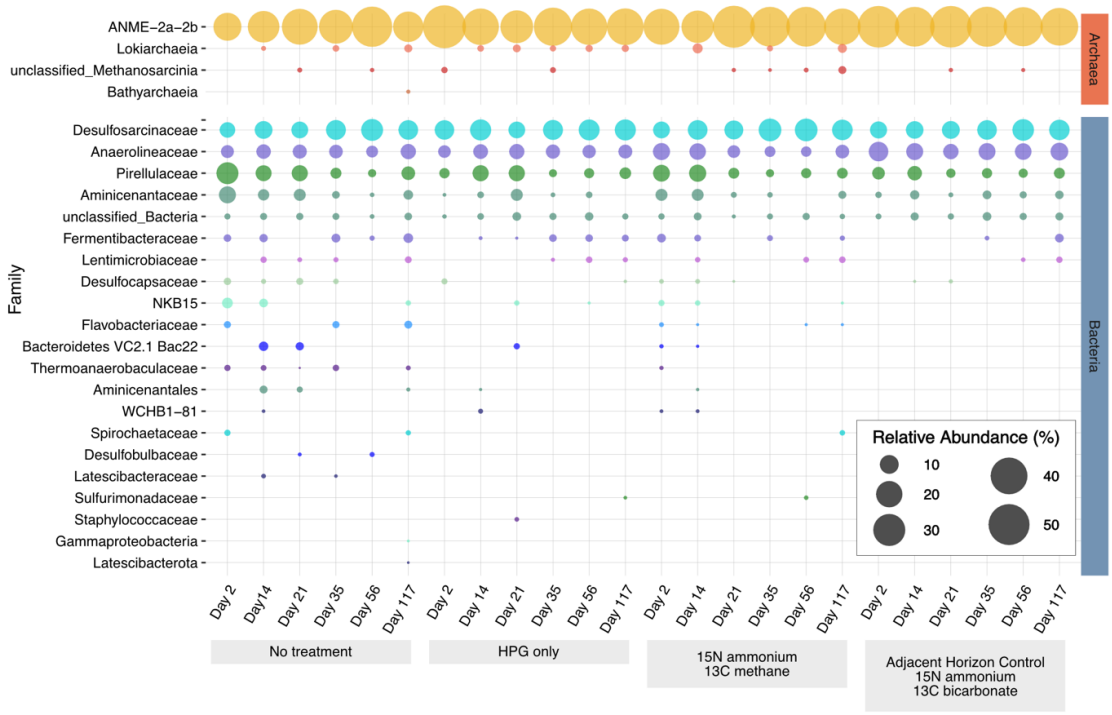

Figure S5.1 16S rRNA amplicon data, shown here at the family level, over the course of the incubation time. Taxa represented comprised >1% of all reads in a given sample.

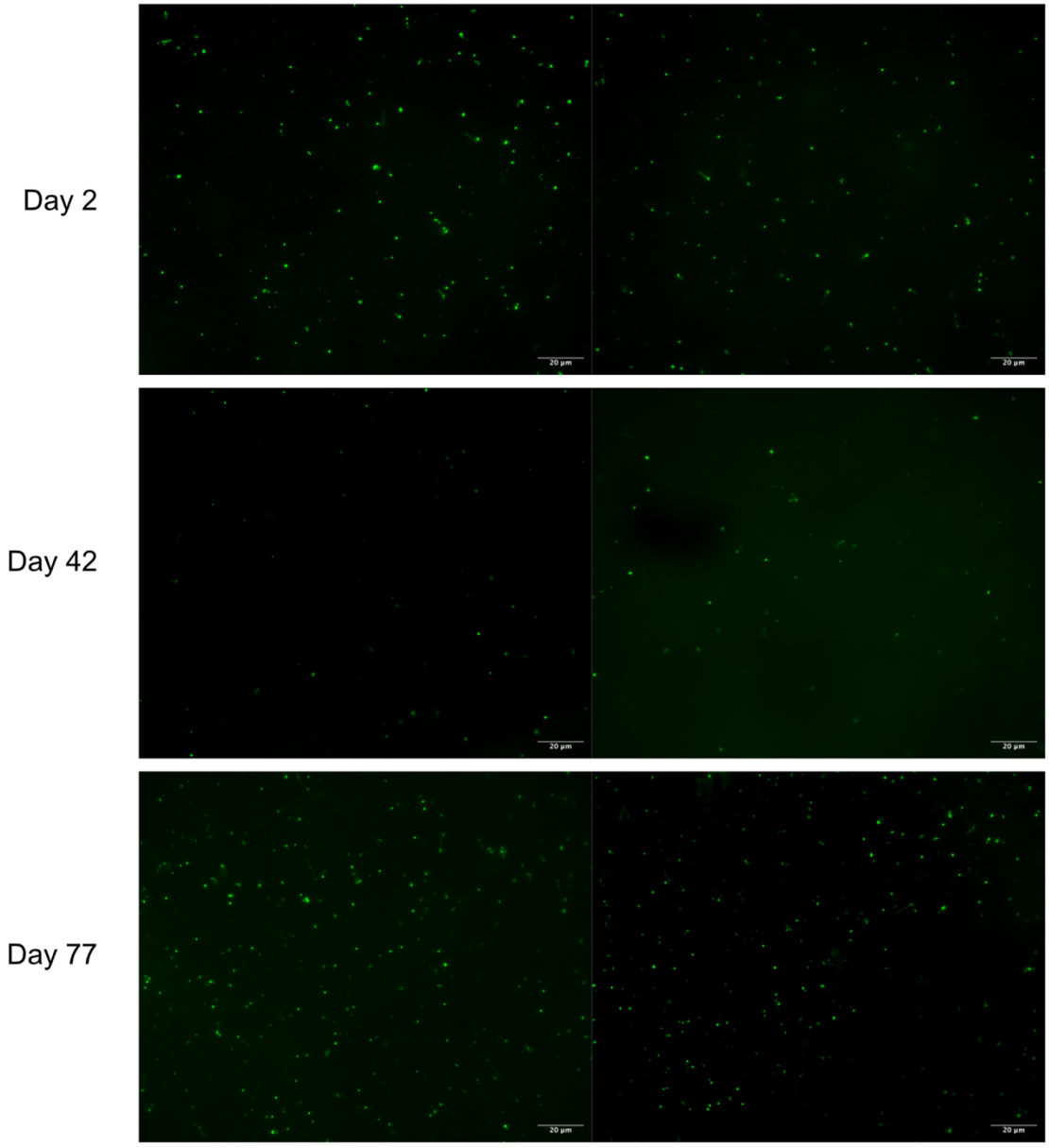

Figure S5.2 Representative SYBR-Gold stained VLP images used in the viral-BONCAT and counts analysis in Figure 5.7. Day 77 does not appear to show a higher amount of background when adjusted according to the analysis pipeline. Day 42 clearly shows a lower level of VLPs and debris, indicating potential viral decay and a decrease in viral production within the incubations. Note that all three timepoints show debris that is too large to be viral; the size thresholds used in the segmentation pipeline will have removed these. Scale bars are 20µm.

Chapter 6

CONCLUSIONS AND LOOKING TO THE FUTURE

You could argue that phages are the ecological elephants in the room, albeit elephants each about 50-250nm wide. They are abundant, active, and often ignored in many environments, including controlled laboratory systems like enrichment cultures. Now, with the door to viral metagenomics thrown wide open and the feasibility of the techniques we developed in Chapter 4, we are well on our way to shining light on an otherwise elusive component of our world.

We started this journey with the intention of answering the following questions:

1. How diverse are the viruses in alkane-oxidizing communities? Who do they infect, and what can we learn about their adaptations and patterns of distribution in relation to various physico-chemical parameters?
2. What are the possible functions of viruses in alkane-oxidizing communities? What kinds of auxiliary metabolic genes (AMGs) do we find?
3. How do nutrients flow through seep microbes into their viral predators, and on what timescales?

We can comfortably say that we have started to answer these questions, at least in part. The viruses of alkane-oxidizing sediments are highly diverse, and we have found hundreds of novel viral families just by directly purifying and sequencing viruses instead of relying on bulk metagenomics. We have also further substantiated existing reports of viral taxa infecting both archaea and bacteria in deep marine sediments (Hwang et al., 2023; L. Wang et al., 2022), and suggested that the complexity of the sediment matrix and the lower energy inputs compared to the surface ocean select for phages with a broad host range. In an environment where the probability of collision with a host is lower than it would be in a flowing water regime, beggars cannot be choosers. This notion may also explain the lack of specificity in the types of auxiliary metabolic genes (AMGs) we find in our dataset. We might expect that AMGs or their abundances might differ between cold and hot environments, or between different geographic locations. We might also expect to find genes specific to alkane oxidation or sulfate reduction, like methyl co-enzyme reductase A (*mcrA*) or dissimilatory sulfite reductase (*dsr*). Neither of these appears to be the case.

So far, I have mostly described a sort of negative "adaptation", or lack of specialization. However, the presence in sediment and soil samples of the *nosD*-like proteins potentially related to surface adhesion suggests that there are some phage genome-level specialization to the subsurface environment, especially when we observe that these proteins have not yet been found in water column viral datasets. In sum, I propose that phage gene content-related adaptations in these alkane-oxidizing subsurface environments are related to the sediment environment in which they are found rather than to host identity and metabolism or other parameters like temperature (recall that our enrichment cultures spanned 10°–60°C. These datasets are a single snapshot in time, and we cannot account for possible loss or gain of AMGs as viruses move from host to host—in other words, absence of evidence is not evidence of absence. However, if it is true that broad host range viruses truly are more common in sediments, and that AMGs can rapidly turn over, we might still expect that at least one virus might contain some gene related to keystone metabolic processes.

That is not to say that temperature, location, and other factors play no role at all in the viral ecology of these sites. We have certainly observed interesting distribution patterns in both datasets. In the sediment-free enrichment cultures, we find that parameters like temperature and sampling location affect the overall community composition, even though they do not appear to impact gene content of individual taxa. In the Costa Rica sediment cores, we find that viral communities are more similar within a core regardless of sediment depth than they are in corresponding sediment depths between cores. We also find that the water column samples taken immediately above two different sampling sites are more similar to each other than they are to the cores taken directly below them. This highlights the importance of studying sediment viruses and the dangers of overreliance on water column samples in understanding marine viral ecology.

Genomics is a revolutionary tool, but it can only take us so far. We have also been able to develop and leverage laboratory techniques to better constrain the rate at which key nutrients like fixed nitrogen are transferred from the cellular biomass to the viral biomass and to positively identify viruses that are produced under different laboratory treatments. These techniques have been successfully implemented in laboratory sediment microcosms and have shown that viral

production is likely responsive to changes in the dominant metabolic processes of the sediment. We find that cellular community composition changes with time in incubations where media and headspace are not replaced and hypothesize that heterotrophic members become more active as methane-oxidizing activity stalls. The plateau in methane-oxidizing activity is correlated with a visible decrease in the number of viruses, possibly due to particle decay, but the viral community then diverges as viral numbers rapidly increase later on. If it is true that heterotrophs are becoming more active as methane is consumed, the viral community may be bouncing back in response, though perhaps with a different composition.

WHAT QUESTIONS SHOULD WE ANSWER NEXT?

The data presented in this thesis is just the beginning of unraveling viral impacts in deep ocean sediments. In looking specifically at alkane-oxidizing taxa, I believe one of the most interesting questions remaining is whether or not the shape of ANME/SRB aggregates is at all influenced by viral infection. Earlier in my PhD, I hoped to run experiments in which I would remove as many viruses as possible from our sediment-free enrichment cultures and examine whether these methane-oxidizing consortia changed in size, aspect ratio, spatial arrangement of ANME/SRB, or thickness of extracellular matrix. If viruses do affect aggregate morphology, I might expect that fewer viruses would result in a thinner extracellular matrix and/or smaller aggregates, as larger aggregates with a thicker matrix might be an adaptation to resist viral attachment. On the other hand, we may also find that viruses play no role at all in consortium morphology, and aggregate structure may be under greater influence by factors such as proximity between syntrophic partners, substrate concentration and diffusion rates, and random variation. That would be an interesting finding too, as it would give us a sense of the relative importance of viruses among the other factors affecting the major taxa in these environments. Helium-ion beam imaging of viral infection, which may be possible through the Kavli Nanoscience Institute at Caltech, is a technique that has a great deal of potential to answer these structure-based questions (Leppänen et al., 2017).

We also do not yet understand how these diverse viruses might affect their hosts physiologically. Biochemical characterization of genes such as the *nosD*-like adhesion-related protein, including expression in a model system, would help clarify their role in infection. There is also vast

potential for qPCR and fluorescence in situ hybridization (FISH) experiments that seek to quantify expression of auxiliary metabolic genes of interest under different conditions. For example, it is hypothesized that direct interspecies electron transfer via large multiheme cytochromes is the mechanism by which ANME and SRB are able to syntrophically couple the methane/sulfate redox reaction (Chadwick et al., 2022; McGlynn et al., 2015; Skennerton et al., 2017; Wegener et al., 2015). When the ANME/SRB partnership is decoupled with the addition of anthraquinone 2,6-disulfonate (as we attempted in Chapter 5), does expression of viral heme-related proteins like those outlined in Chapters 2 and 3 also change? Experiments such as these may also help match host and virus pairs: if we can track which organisms are growing most rapidly and understand which viruses have highly expressed genes at the same time, we may be able to make an educated guess as to which host these viruses are most likely to infect.

There are also additional related questions in the geochemical realm; our preliminary look at nitrogen flow does not constrain other crucial biological elements like carbon, hydrogen, oxygen, phosphorus, and sulfur, as well as limiting nutrients like iron. Additionally, comparing the viral nanoSIMS data to parallel cellular nanoSIMS data would allow us to compare timescales of cell and phage isotope incorporation and provide us with information on the relative fractions of nutrients held in the two biological pools. Lastly, we have not attempted to calculate rates of decay within the viral pool in sediments. As mentioned when describing our protocol to detach viruses from sediment, viral attachment to the sediment matrix is thought to increase the lifespan of the particle. Understanding decay rates and how quickly viruses are returned to the dissolved organic matter pool may further constrain their contribution to carbon and nitrogen cycling in the deep ocean.

Furthermore, incubations such as those in Chapter 5, in which media and headspace were not replaced (except to maintain headspace pressure), also provide the opportunity to examine how ecological succession might occur in the viral community as the main energy source in an ecosystem is depleted. Prior work has shown that in the gradient around a methane seep from high to low/no activity, cellular microbial communities in the inactive areas are more similar to "transition" communities than these transition communities are to areas of high activity (Mullin, 2020). Examining rates of viral production and shifts in viral community composition as a

laboratory microcosm transitions from high to low activity would provide a more holistic view of the seep microbial community under energy limitation. These techniques provide us with a unique opportunity to study host/parasite dynamics in complex, non-culture-based systems.

GOING FURTHER

For the past several years, I have limited myself to investigating alkane-rich sediment environments. The ocean floor, however, is one of the largest biomes on the planet, and it is by no means limited to cold seeps and hydrothermal vents. What happens to the viral community when a dead whale sinks to the ocean floor and is decomposed over the course of years? How do viruses impact carbon sequestration in the shallow sediments of seagrass beds (Serrano et al., 2021)? What about those areas on the ocean floor without such large carbon inputs? Are viruses active in these comparatively desert-like environments? On what timescales? Most critically, how does viral activity on the seafloor impact carbon cycling across the global ocean? The ocean is the planet's largest carbon sink, absorbing far more carbon dioxide than is held in the atmosphere, and burial in the sediment is an excellent way to more permanently sequester that carbon away from the atmosphere (Scott & Lindsey, 2022). Climate change mitigation requires that we deepen our understanding of the organisms that interact with that stored carbon.

The next several years will be a critical period for this kind of research. The planet is warming faster, and the ocean can only buffer our atmospheric CO_2 concentrations for so long. A recent study focusing on ocean acidification estimates that the ocean will begin to lose its ability to absorb carbon dioxide by the year 2100 as acidification continues (Chikamoto et al., 2023). These ecological disasters dovetail unfortunately rather well with the rise of deep-sea mining, oil drilling, and laying of undersea cables for communication networks. The destruction of these habitats bodes poorly for the organisms that live there and actively keep the sequestered carbon where we would prefer it to remain. We have little time left to show the world the magnitude of what will be lost if these kinds of human activities are left unchecked.

The field of ecology is uniquely positioned to push our structures in the right direction. You cannot be an ecologist for very long without feeling humbled by the degree to which all things are interconnected; this very notion underpins the entire field. A parallel idea that I think is also

inherent to ecology is that assigning human moral ideas to individual ecosystem components is not a useful exercise. We see this clearly with viruses—as we have already discussed, viruses may kill photosynthetic cyanobacteria and stop them from fixing carbon, but they also are important in the collapse of toxic blooms in eutrophic waters. The key, as many cultures around the world have known for millennia, is balance. Used in conjunction with that traditional knowledge, the ecological sciences have the tools to show us how to examine and maintain that balance and restore human understanding of our place in the natural world.

Appendix I

STUDYING PHAGE MORPHOLOGICAL DIVERSITY

INTRODUCTION

Phage morphology is an enigma. Though there are several well-known capsid shapes including the classic head-tail virus, the evolutionary relationships between the shapes and how they might confer specific advantages to the phage have not been fully elucidated. In bacteria and archaea, the correlation between structure and function is well-established, in that overall cell shape is often an adaptation to environmental factors such as desiccation risk and nutrient acquisition efficiency. A spherical cell has a much lower ratio of surface area to overall volume compared to a cylindrical cell of equal radius, thus giving the former a much lower risk of losing water through its surface in return for a decreased flux of nutrients through its membrane (Young, 2007). The famous *Haloquadratum walsbyi* is a thin square cell, floating like a sheet of paper on the surface of highly saline aqueous environments (Walsby, 1980). This shape allows the species to maintain buoyancy, and the higher surface area to volume ratio allows it to absorb more nutrients in its typically nutrient-starved habitat and take in as much solar energy as possible for its metabolic needs (Bolhuis et al., 2006).

Correlating structure and function in viruses is less straightforward, not least because they lack the metabolic processes that play such an important role in determining cell shape. Instead, viral shape appears to be determined by other factors related to genome size, environmental conditions, and the ability to bind to and inject their genetic material into a host. Some research indicates that spindle-shaped viruses are descended from rod-shaped viruses, and the expanded middle allows a longer genome to be packed into the capsid (F. Wang et al., 2022). Other more general work has found a clear correlation between internal capsid volume and genome size (Chaudhari et al., 2021), as phage genomes are packed so tightly as to be almost crystalline. It has also been shown that certain viral shapes may lend themselves to binding to specific parts of host outer membranes; for example, the tailless *Turriviridae* has small protrusions that bind to pili-like appendages on the host cell surface (Baquero et al., 2020).

Many of the more exotic viral shapes found in nature, like the spindle shaped viruses (**Figure A1.1**), the droplet-shaped *Guttaviridae*, or the wine bottle-like *Ampullaviridae* have thus far only been observed infecting archaea, while the classic head-tail shape infects both archaea and bacteria (Ackermann & Prangishvili, 2012). It has been suggested that these capsid shapes are the result of the physical influence on the capsid proteins of the extreme conditions in which some of their hosts live (Hartman et al., 2019), as archaea are more likely to be extremophiles. For example, *Ampullaviridae* were found infecting the highly acidophilic and hyperthermophilic genus *Acidianus* (Häring, Rachel, et al., 2005), while both spindle and droplet shaped viruses have been found infecting hyperthermophiles (Mochizuki et al., 2011; Xiang et al., 2005). Some viruses also possess a degree of pleomorphy (Baquero et al., 2024), in some cases changing their shape outside of their host cell in response to high temperatures (Häring, Vestergaard, et al., 2005). However, extreme environments are not exclusively populated by these unusual shapes, nor are archaea exclusively infected by them.

A naively ambitious project during the early stages of my PhD was an attempt to quantitatively assess the morphology of viruses in methane seep sediment and correlate it with cellular community composition, geochemistry, and depth in the sediment in an attempt. At the time, I found a single comprehensive survey looking at surface ocean water from the Tara Oceans Project that found factors like temperature, salinity, and oxygen levels were better correlated with viral morphological abundances than geographical proximity (Brum et al., 2013). I did not get far before realizing that such correlations would not answer my question of what determined phage capsid shape, and that this was the kind of question that research labs spend their lifetimes focusing on. Nonetheless, I did a rather large amount of transmission electron microscopy of viral particles that revealed some interesting trends and truly astonishing shapes (**Figure A1.1**). I started with a broad survey of viral morphologies in incubations of pushcores taken from sites of varying levels of methane activity to see how methane levels might correlate with viral morphology.

METHODS

TEM was performed as described in the methods of Chapter 2. Electron microscopy was done in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech and

at the National Center for Imaging and Microscopy Research at the University of California, San Diego.

RESULTS AND DISCUSSION (of a kind)

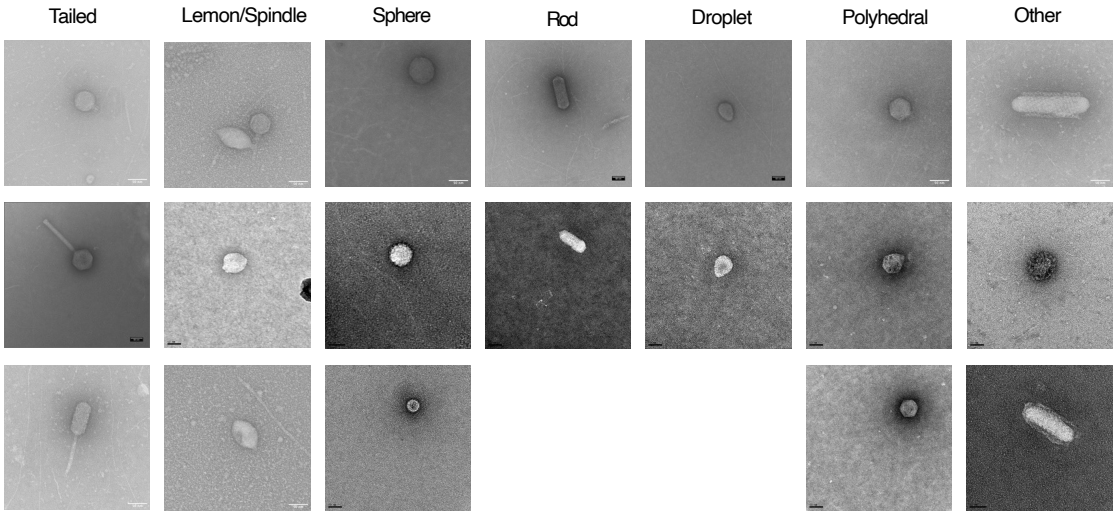

Figure A1.1. Examples of the types of shapes found in various incubations over the course of my time at Caltech.

Do viral morphologies vary with geochemical gradients?

An early question in my exploration of viral morphologies was whether or not certain shapes were more dominant under certain geochemical conditions. As discussed in Chapter 5, decoupling the anerobic methanotrophic archaea/sulfate reducing bacteria syntrophy resulted in vastly different morphological compositions of the viral community compared to samples with normal ANME/SRB activity. However, I began my digging into this question even earlier, looking at samples from a pushcore collected in a transition zone, i.e., an area farther from the center of methane seepage where methane availability is lower but nonzero, and comparing them to a core taken from a site of active seepage. Both pushcores are from the Costa Rica methane seep, the active core being one of those discussed in Chapter 3. Had I not realized that this line of inquiry might be a bit more than I could handle in a PhD, I would have also collected data from an inactive site. Prior work in the lab found that the cellular microbial communities of inactive and transition sites are more similar than the cellular communities of transition and active sites (Mullin, 2020), which, in conjunction with the community level data from Chapters

2 and 3, perhaps allows us to assume that the transition site used here may be a proxy for an inactive site.

In any case, the most striking difference I found between the transition and active cores was the lack of spindle shaped viruses and the lower proportion of tailed viruses in deeper sections of the transition core (**Figure A1.2a**). As mentioned earlier, spindle shaped viruses have thus far been observed exclusively infecting archaea. Perhaps the higher methane fluxes drive higher ANME growth, resulting in more spindle-shaped viruses in the active core than the transition core. As to why there are more tailed and droplet shaped viruses at the surface of the transition core than the deeper sections, I cannot hazard a guess without, at minimum, 16S rRNA amplicon data to give me a sense of which cellular microbes are present. Presumably there are more archaea present if the droplet morphologies are in higher abundance.

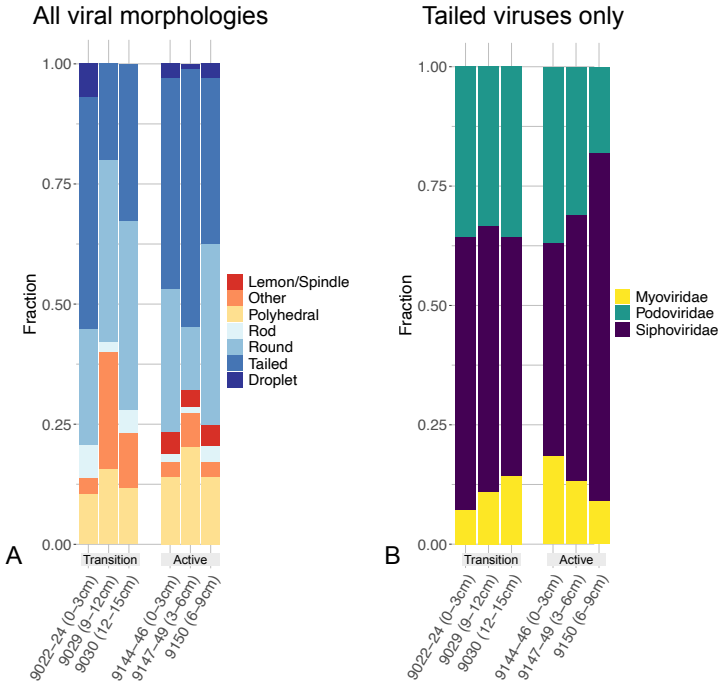

Figure A1.2 Morphological analysis of viral particles based on TEM. **A)** Bar graph of viral morphologies between a pushcore taken from an area of active methane seepage (right) and a pushcore from an transitional area, between zones of inactivity and active seepage. Sample sizes are as follows: 9022-24 (0-3cm)=29, 9029 (9-12cm)=45, 9030 (12-15cm)=43, 9144-46 (0-3cm)=64, 9147-49 (3-6cm)=87, 9150 (6-9cm)=64. **B)** A bar graph examining tailed viruses only. Given that the samples sizes are much smaller, it is more difficult to make comparisons between the two groups. For now, it appears that the potential difference between the two sites is that there is less of a difference in the transition core between the tailed community as sediment depth increases. This may be due to the steeper methane gradient in the active core compared to the transition core.

What is the "hot dog" shaped viral-like particle?

When looking through the viral EM images collected over the last several years, one particular shape caught our attention: the so-called hot dog (**Figure A1.3**). The first time I saw it in a Costa Rica sample, I took a picture of it purely because I thought it had an interesting shape. I regarded it as a one-off, a possible artefact of my sample preparation. But then it showed up again in a different Costa Rica sample—perhaps this was a novel yet Costa Rica specific oddity. The shape is consistently 200-250nm in length, approximately 50-60nm in width, and consists of a solid cylindrical structure partially encased in a crystalline mesh. A possible image of the mesh on its own (**Figure A1.3**, second from the right) and a more intact, though definitely distinct, particle then appeared in a sample from Guaymas Basin. I have not been able to find examples of similar structures in the literature, so the best I can offer are some hypotheses as to what they are:

1. They are members of a cellular taxon that is small enough to fit through the pores of the 0.2μm filter membrane.
2. They are some kind of extracellular vesicle.
3. They are a novel phage or eukaryotic virus.

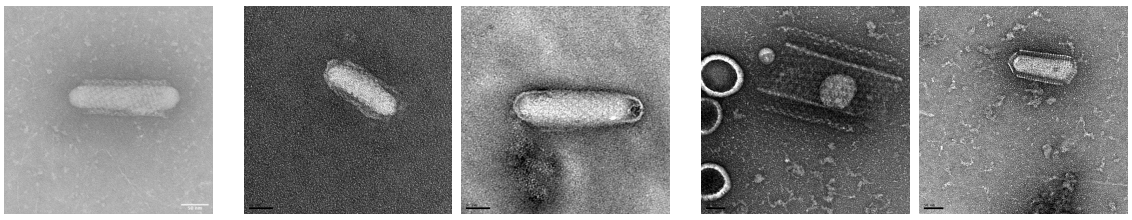

Figure A1.3 Examples of the "hot dog" particle. The three leftmost images are putatively the same type of structure. The fourth is potentially the outer mesh wrap (the hot dog bun, if you will) on its own, though this is much more speculative. The last image may be related but has a very different, though equally bizarre, structure from the first three images. Scale bars are 50nm.

If a sample is known to contain structures such as these in high abundance, I believe it would be worth doing a more detailed examination via cryo-electron tomography. This would give us information on the inner and outer structure of the hot dog particles and allow us to develop better hypotheses as to their identity.

CONCLUSIONS

In hindsight, I wish I had spent more energy thinking about viral origins (though that is a tangled subject in and of itself) and species evolution than about capsid shape alone, as I believe answers regarding the former might shed light on the latter. As discussed in the introduction, some spindle shaped viruses are thought to originate in rod-shaped species, which then expanded in the middle to accommodate a larger genome. Already we see how understanding these evolutionary relationships begins to answer questions of shape and function. Ascribing specific intention to evolutionary processes is of course a tricky business; some shapes might exist through random mutation resulting in particles that were stable enough to survive and infect. But I hope whoever next decides to peer at seep viruses through the screen of an electron microscope feels the excitement I did. The science is almost beside the point[4]—these are specimens from the bottom of the ocean, and we are privileged to be able to see them.

CONTRIBUTIONS

I performed all transmission electron microscopy and downstream analysis in this appendix. Mason Mackey from Prof. Mark Ellisman's group at the National Center for Microscopy and Imaging Research (NCMIR) at UC San Diego trained me on the proper use of an electron microscope.

[4] *Almost.*

Appendix II

ADDITIONAL METHODS AND MEDIA RECIPES

SHIPBOARD SAMPLE COLLECTION AND LABORATORY MAINTENANCE

As the reader has gathered by now, we cannot study the ecosystems of the deep ocean without collecting samples directly from the sites in question. All of the sites used in this study are at pressures that would flatten a human, so we usually send down autonomous vehicles or the DSV Alvin human-occupiable submersible, all of which are rated to withstand high pressure. Typically, samples are collected as pushcores, using a long plastic cylinder that we push into the sediment and retrieve such that we preserve the vertical structure of the sediment layers (**Figure A2.1, A2.2**). Pushcores can vary in length. If we hit rock while pushing into the sediment, our core will be short, sometimes no more than a few centimeters. If we are lucky, we can get a core up to 21cm. We might even sample long cores using tubes up to a meter in length. Once returned to the ship, the sediment is rapidly sectioned into 3cm horizons and sampled for DNA, geochemical information, and cells. Additional sediment is saved in a mylar bag filled with argon, which helps keep oxygen out. Upon return to land and lab, the sediment is moved into glass jars and given an artificial seawater medium or ambient seawater from the sampling site, if available. Empty space in the bottle is filled with methane gas to simulate in situ conditions. The media and headspace are replaced every one to three months, depending on geochemical measures of microbial activity. These incubations are now available for experimental use.

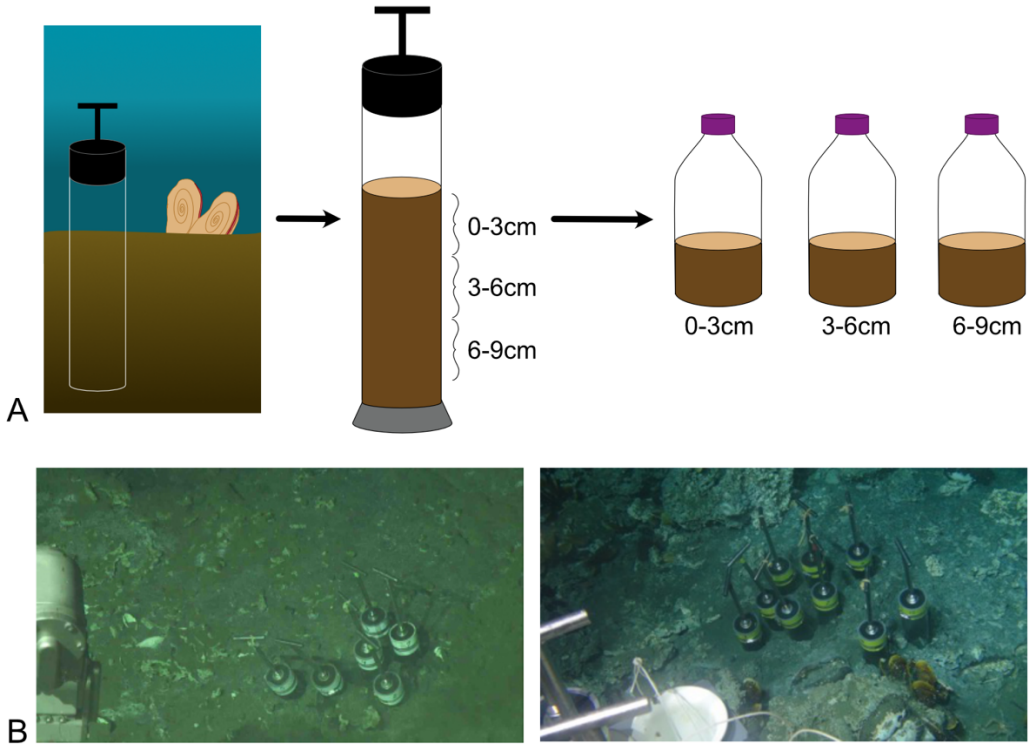

Figure A2.1 Schematic of the standard sediment collection process on the ocean floor, with example images. **A)** Schematic of shipboard sample collection. Sediment cores are taken from the deep sea and brought back to the ship. The cores are then sectioned horizontally, preserving stratification of geochemical gradients and microbial communities at the centimeter scale. The sectioned core is moved into Pyrex bottles and placed under a 100% methane headspace. **B)** Examples of pushcore sampling at the bottom of the ocean. These particular cores were sampled from the Costa Rica seep and are part of the set analyzed in Chapter 3.

ARTIFICIAL SEAWATER RECIPE, CHAPTER 3

Reagent	**Per 2L**	**Final concentration**	**Notes**
Sigma Aldrich Sea Salts	80g		No longer manufactured
NH_4Cl	1.07g	10mM	
K_2HPO_4	0.136g	0.5mM	
$NaHCO_3^-$	2.45g	15mM	
Se/W solution	0.2mL	0.01/0.0067 μM	See Scheller et al, 2016
Trace elements solution	2mL		See Scheller et al, 2016
DSM141 Vitamin solution, 1000X	2mL		

1. Dissolve the Sigma Aldrich Sea Salts in 1.5L ultrapure water, using a magnetic stir bar and a hot plate if needed to encourage dissolution.
2. Add the NH_4Cl, K_2HPO_4, and $NaHCO_3^-$ slowly and sequentially, allowing the solution to clear between additions.
3. Fix pH to approximately 7.5 using NaOH or HCl as needed.
4. Bring volume up to 2L using additional ultrapure water.
5. Filter through a 0.22µm pore size membrane.
6. Add Se/W, trace elements, and vitamin solutions. Store at 4˚C and sparge with nitrogen gas for 30 minutes before use in anoxic cultivation.

ARTIFICIAL SEAWATER RECIPE, CHAPTER 5

Based on the recipe from (Scheller et al., 2016), *Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction*, and modified by Dr. Hang Yu and Yongzhao Guo

1. Begin with the following in 800mL of water, for a final volume of 1L:

Reagent	**Per 1L**	**final mM concentration**	**Notes**
$MgCl_2 \cdot 6H_2O$	5.67g	27.9	
$CaCl_2 \cdot 2H_2O$	0.22g	1.5	
NaCl	26.37g	451.2	
KCl	0.60g	8.0	
Na_2SO_4	1.4g	10	For non-AQDS incubation
Na_2SO_4	0.07g	0.5	For AQDS incubation, and for sulfate-free seawater for washing sediment
Anthraquinone-2,6-disulfonate (AQDS		10	For AQDS incubation
K_2HPO_4	0.17g	1.0	
Se/W solution	0.10mL	0.01/0.0067 uM	See Scheller et al., 2016
Trace elements solution	1.00mL		See Scheller et al., 2016
250mM HEPES buffer *	100.00mL	25.0	
DSM141 Vitamin solution, 1000X	1mL		

2. The pH should be between 7.2-7.45. Filter through a 0.22μm pore size filter to sterilize and flush with nitrogen gas for 20-30 minutes to make the media anoxic. Stopper the media bottle using a butyl rubber stopper and cap.
3. The following solutions should be added using a nitrogen gas-flushed needle and should be prepared anoxically (see the following section for more on preparing anoxic solutions).

Reagent	**g or mL/L**	**final concentration**
1M $NaHCO_3^-$ solution, either ^{12}C or ^{13}C	5mL	5mM
500mM NH_4Cl, either ^{14}N or ^{15}N	4mL	2mM
50mM L-homopropargyl glycine (HPG)	1mL	50μM

*For a stock solution of 250mM HEPES at pH 7.5, combine 32.5g HEPES and 2.7g of NaOH in ultrapure water for a final volume of 500mL. Filter through a 0.22μm pore size membrane.

MAKING ANOXIC SOLUTIONS

This protocol describes how to make anoxic solutions of compounds that were added to the artificial seawater media in Chapter 5. These include solutions of ^{12}C and ^{13}C sodium bicarbonate, ^{14}N and ^{15}N ammonium chloride, and the methionine analog L-homopropargyl glycine (HPG).

Materials

- 10-30mL serum vials, one per solution you plan to make
- Butyl rubber stoppers (e.g., FisherSci Cat #50-143-854)
- Aluminum serum vial seals (e.g., Millipore Sigma Cat #27230-U)
- Aluminum seal crimper (e.g., MedLab Supply, #20001-00-C01A)
- The substrates you wish to make into solutions, measured for the volume of solution you are making

Protocol

1. Place all the above materials into the anaerobic glove box.
2. Autoclave a pyrex bottle of ultrapure water.
3. Remove water from the autoclave; once it is cool enough to set on a plastic surface but still hot, move the water into an anaerobic glove box.
    a. If it is too hot, you may melt the plastic walls of the chamber. If it is allowed to cool too much, you may have atmospheric oxygen re-dissolve into the water.

   b. The vacuum process within the airlock will also help remove any remaining oxygen.
4. Once all supplies are in the glovebox and the oxygen concentration has returned into the low tens, make your solutions by adding the substrates to their respective vials, adding the appropriate amount of water, and sealing with the rubber stopper and aluminum seal.

BIBLIOGRAPHY

Ackermann, H.-W., & Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. *Archives of Virology*, *157*(10), 1843–1849. https://doi.org/10.1007/s00705-012-1383-y

Aiewsakun, P., Adriaenssens, E. M., Lavigne, R., Kropinski, A. M., & Simmonds, P. (2018). Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. *Journal of General Virology*, *99*(9), 1331–1343. https://doi.org/10.1099/jgv.0.001110

Andersson, A. F., & Banfield, J. F. (2008). Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. *Science*, *320*(5879), 1047–1050. https://doi.org/10.1126/science.1157358

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

Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G., & Tratschin, J. D. (1979). Thermal stability and protein structure. *Biochemistry*, *18*(25), 5698–5703. https://doi.org/10.1021/bi00592a028

Atwood, T. B., Witt, A., Mayorga, J., Hammill, E., & Sala, E. (2020). Global Patterns in Marine Sediment Carbon Stocks. *Frontiers in Marine Science*, *7*, 165. https://doi.org/10.3389/fmars.2020.00165

Auslander, N., Gussow, A. B., Benler, S., Wolf, Y. I., & Koonin, E. V. (2020). Seeker: alignment-free identification of bacteriophage genomes by deep learning. *Nucleic Acids Research*, *48*(21), e121–e121. https://doi.org/10.1093/nar/gkaa856

Babich, H., & Stotzky, G. (1980). Reductions in inactivation rates of bacteriophages by clay minerals in lake water. *Water Research*, *14*(2), 185–187. https://doi.org/10.1016/0043-1354(80)90236-5

Bali, S., Lawrence, A. D., Lobo, S. A., Saraiva, L. M., Golding, B. T., Palmer, D. J., et al. (2011). Molecular hijacking of siroheme for the synthesis of heme and d_1 heme. *Proceedings of the National Academy of Sciences*, *108*(45), 18260–18265. https://doi.org/10.1073/pnas.1108228108

Ballard, R. D. (1977). *Notes on a Major Oceanographic Find* (THE DISCOVERY OF HYDROTHERMAL VENTS). Woods Hole Oceanographic Institute.

Baquero, D. P., Liu, Y., Wang, F., Egelman, E. H., Prangishvili, D., & Krupovic, M. (2020). Structure and assembly of archaeal viruses. In *Advances in Virus Research* (Vol. 108, pp. 127–164). Elsevier. https://doi.org/10.1016/bs.aivir.2020.09.004

Baquero, D. P., Bignon, E. A., & Krupovic, M. (2024). Pleomorphic viruses establish stable relationship with marine hyperthermophilic archaea. *The ISME Journal*, *18*(1), wrae008. https://doi.org/10.1093/ismejo/wrae008

Baranyi, U., Klein, R., Lubitz, W., Krüger, D. H., & Witte, A. (2000). The archaeal halophilic virus-encoded Dam-like methyltransferase M.φCh1-I methylates adenine residues and

complements *dam* mutants in the low salt environment of *Escherichia coli*. *Molecular Microbiology*, *35*(5), 1168–1179. https://doi.org/10.1046/j.1365-2958.2000.01786.x

Barry, J. P., Gary Greene, H., Orange, D. L., Baxter, C. H., Robison, B. H., Kochevar, R. E., et al. (1996). Biologic and geologic characteristics of cold seeps in Monterey Bay, California. *Deep Sea Research Part I: Oceanographic Research Papers*, *43*(11–12), 1739–1762. https://doi.org/10.1016/S0967-0637(96)00075-1

Beltrami, E., & Carroll, T. O. (1994). Modeling the role of viral disease in recurrent phytoplankton blooms. *Journal of Mathematical Biology*, *32*(8), 857–863. https://doi.org/10.1007/BF00168802

Benler, S., & Koonin, E. V. (2021). Fishing for phages in metagenomes: what do we catch, what do we miss? *Current Opinion in Virology*, *49*, 142–150. https://doi.org/10.1016/j.coviro.2021.05.008

Biller, S. J., McDaniel, L. D., Breitbart, M., Rogers, E., Paul, J. H., & Chisholm, S. W. (2017). Membrane vesicles in sea water: heterogeneous DNA content and implications for viral abundance estimates. *The ISME Journal*, *11*(2), 394–404. https://doi.org/10.1038/ismej.2016.134

Biswas, A., Staals, R. H. J., Morales, S. E., Fineran, P. C., & Brown, C. M. (2016). CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics*, *17*(1), 356. https://doi.org/10.1186/s12864-016-2627-0

Blanco-Romero, E., Garrido-Sanz, D., Durán, D., Rybtke, M., Tolker-Nielsen, T., Redondo-Nieto, M., et al. (2024). Role of extracellular matrix components in biofilm formation and adaptation of Pseudomonas ogarae F113 to the rhizosphere environment. *Frontiers in Microbiology*, *15*, 1341728. https://doi.org/10.3389/fmicb.2024.1341728

Boetius, A., & Wenzhöfer, F. (2013). Seafloor oxygen consumption fuelled by methane from cold seeps. *Nature Geoscience*, *6*(9), 725–734. https://doi.org/10.1038/ngeo1926

Boetius, A., Ravenschlag, K., Schubert, C. J., Rickert, D., Widdel, F., Gieseke, A., et al. (2000). A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature*, *407*(6804), 623–626. https://doi.org/10.1038/35036572

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B. (2017). vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*. *PeerJ*, *5*, e3243. https://doi.org/10.7717/peerj.3243

Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., et al. (2006). The genome of the square archaeon Haloquadratum walsbyi : life at the limits of water activity. *BMC Genomics*, *7*(1), 169. https://doi.org/10.1186/1471-2164-7-169

Bratbak, G., Egge, J., & Heldal, M. (1993). Viral mortality of the marine alga Emiliania huxleyi (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series*, *93*, 39–48. https://doi.org/10.3354/meps093039

Brum, J. R., Schenck, R. O., & Sullivan, M. B. (2013). Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *The ISME Journal*, *7*(9), 1738–1751. https://doi.org/10.1038/ismej.2013.67

Buchenau, B., Kahnt, J., Heinemann, I. U., Jahn, D., & Thauer, R. K. (2006). Heme Biosynthesis in *Methanosarcina barkeri* via a Pathway Involving Two Methylation Reactions. *Journal of Bacteriology*, *188*(24), 8666–8668. https://doi.org/10.1128/JB.01349-06

Burnim, A. A., Dufault-Thompson, K., & Jiang, X. (2024). The three-sided right-handed β-helix is a versatile fold for glycan interactions. *Glycobiology*, *34*(cwae037). https://doi.org/10.1093/glycob/cwae037

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications*, *7*(1), 10613. https://doi.org/10.1038/ncomms10613

Bushnell, Brian. (n.d.). BBTools. Retrieved from sourceforge.net/projects/bbmap/

Cai, L., Jørgensen, B. B., Suttle, C. A., He, M., Cragg, B. A., Jiao, N., & Zhang, R. (2019). Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years. *The ISME Journal*. https://doi.org/10.1038/s41396-019-0397-9

Cai, W., Wang, H., Tian, Y., Chen, F., & Zheng, T. (2011). Influence of a Bacteriophage on the Population Dynamics of Toxic Dinoflagellates by Lysis of Algicidal Bacteria. *Applied and Environmental Microbiology*, *77*(21), 7837–7840. https://doi.org/10.1128/AEM.05783-11

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

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

Chadwick, G. L., Skennerton, C. T., Laso-Pérez, R., Leu, A. O., Speth, D. R., Yu, H., et al. (2022). Comparative genomics reveals electron transfer and syntrophic mechanisms differentiating methanotrophic and methanogenic archaea. *PLOS Biology*, *20*(1), e3001508. https://doi.org/10.1371/journal.pbio.3001508

Chakraborty, A., Ruff, S. E., Dong, X., Ellefson, E. D., Li, C., Brooks, J. M., et al. (2020). Hydrocarbon seepage in the deep seabed links subsurface and seafloor biospheres. *Proceedings of the National Academy of Sciences*, *117*(20), 11029–11037. https://doi.org/10.1073/pnas.2002289117

Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*, *49*(16), 9077–9096. https://doi.org/10.1093/nar/gkab688

Chao, H., Balcazar, J. L., Wu, Y., Cai, A., Ye, M., Sun, M., & Hu, F. (2023). Phages in vermicomposts enrich functional gene content and facilitate pesticide degradation in soil. *Environment International*, *179*, 108175. https://doi.org/10.1016/j.envint.2023.108175

Chaudhari, H. V., Inamdar, M. M., & Kondabagil, K. (2021). Scaling relation between genome length and particle size of viruses provides insights into viral life history. *iScience*, *24*(5), 102452. https://doi.org/10.1016/j.isci.2021.102452

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*, *38*(23), 5315–5316. https://doi.org/10.1093/bioinformatics/btac672

Chen, L.-X., Méheust, R., Crits-Christoph, A., McMahon, K. D., Nelson, T. C., Slater, G. F., et al. (2020). Large freshwater phages with the potential to augment aerobic methane oxidation. *Nature Microbiology*, *5*(12), 1504–1515. https://doi.org/10.1038/s41564-020-0779-9

Chikamoto, M. O., DiNezio, P., & Lovenduski, N. (2023). Long-Term Slowdown of Ocean Carbon Uptake by Alkalinity Dynamics. *Geophysical Research Letters*, *50*(4), e2022GL101954. https://doi.org/10.1029/2022GL101954

Chivers, C. E., Koner, A. L., Lowe, E. D., & Howarth, M. (2011). How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *The Biochemical Journal*, *435*(1), 55–63. https://doi.org/10.1042/BJ20101593

Ciccarelli, F. D., Copley, R. R., Doerks, T., Russell, R. B., & Bork, P. (2002). CASH – a β-helix domain widespread among carbohydrate-binding proteins. *Trends in Biochemical Sciences*, *27*(2), 59–62. https://doi.org/10.1016/S0968-0004(01)02046-1

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*(6), 066111. https://doi.org/10.1103/PhysRevE.70.066111

Cohen, S. S. (1948). THE SYNTHESIS OF BACTERIAL VIRUSES. *Journal of Biological Chemistry*, *174*(1), 295–303. https://doi.org/10.1016/S0021-9258(18)57398-1

Comeau, A. M., Bertrand, C., Letarov, A., Tétart, F., & Krisch, H. M. (2007). Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery. *Virology*, *362*(2), 384–396. https://doi.org/10.1016/j.virol.2006.12.031

Couradeau, E., Sasse, J., Goudeau, D., Nath, N., Hazen, T. C., Bowen, B. P., et al. (2019). Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nature Communications*, *10*(1), 2770. https://doi.org/10.1038/s41467-019-10542-0

Da Silva, R. A. G., Afonina, I., & Kline, K. A. (2021). Eradicating biofilm infections: an update on current and prospective approaches. *Current Opinion in Microbiology*, *63*, 117–125. https://doi.org/10.1016/j.mib.2021.07.001

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., & Weinbauer, M. (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*, *454*(7208), 1084–1087. https://doi.org/10.1038/nature07268

Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science*, *326*(5951), 422–426. https://doi.org/10.1126/science.1178223

Dekas, Anne E, Connon, S. A., Chadwick, G. L., Trembath-Reichert, E., & Orphan, V. J. (2016). Activity and interactions of methane seep microorganisms assessed by parallel transcription and FISH-NanoSIMS analyses. *The ISME Journal*, *10*(3), 678–692. https://doi.org/10.1038/ismej.2015.145

Dell'Anno, A., & Danovaro, R. (2005). Extracellular DNA Plays a Key Role in Deep-Sea Ecosystem Functioning. *Science*, *309*(5744), 2179–2179. https://doi.org/10.1126/science.1117475

Dell'Anno, A., Corinaldesi, C., Magagnini, M., & Danovaro, R. (2009). Determination of viral production in aquatic sediments using the dilution-based approach. *Nature Protocols*, *4*(7), 1013–1022. https://doi.org/10.1038/nprot.2009.82

Dell'Anno, A., Corinaldesi, C., & Danovaro, R. (2015). Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proceedings of the National Academy of Sciences*, *112*(16), E2014–E2019. https://doi.org/10.1073/pnas.1422234112

Di Giulio, M. (2005). A comparison of proteins from Pyrococcus furiosus and Pyrococcus abyssi: barophily in the physicochemical properties of amino acids and in the genetic code. *Gene*, *346*, 1–6. https://doi.org/10.1016/j.gene.2004.10.008

Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., & Schuman, E. M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences*, *103*(25), 9482–9487. https://doi.org/10.1073/pnas.0601637103

Dombrowski, N., Seitz, K. W., Teske, A. P., & Baker, B. J. (2017). Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome*, *5*(1), 106. https://doi.org/10.1186/s40168-017-0322-2

Eddy, S. R. (2007). HMMER: biosequence analysis using profile hidden Markov models. (Version 3.3.2). Retrieved from hmmer.org

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edwards, R. A., McNair, K., Faust, K., Raes, J., & Dutilh, B. E. (2016). Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews*, *40*(2), 258–272. https://doi.org/10.1093/femsre/fuv048

Engelhardt, T., Kallmeyer, J., Cypionka, H., & Engelen, B. (2014). High virus-to-cell ratios indicate ongoing production of viruses in deep subsurface sediments. *The ISME Journal*, *8*(7), 1503–1509. https://doi.org/10.1038/ismej.2013.245

Fang, H., Kang, J., & Zhang, D. (2017). Microbial production of vitamin B12: a review and future perspectives. *Microbial Cell Factories*, *16*(1), 15. https://doi.org/10.1186/s12934-017-0631-y

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., & Gloor, G. B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS ONE*, *8*(7), e67019. https://doi.org/10.1371/journal.pone.0067019

Fix, I., Heidinger, L., Friedrich, T., & Layer, G. (2023). The Radical SAM Heme Synthase AhbD from Methanosarcina barkeri Contains Two Auxiliary [4Fe-4S] Clusters. *Biomolecules*, *13*(8), 1268. https://doi.org/10.3390/biom13081268

Flemming, H.-C., & Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology*, *8*(9), 623–633. https://doi.org/10.1038/nrmicro2415

Fukuchi, S., & Nishikawa, K. (2001). Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *Journal of Molecular Biology*, *309*(4), 835–843. https://doi.org/10.1006/jmbi.2001.4718

Gambelli, L., Cremers, G., Mesman, R., Guerrero, S., Dutilh, B. E., Jetten, M. S. M., et al. (2016). Ultrastructure and Viral Metagenome of Bacteriophages from an Anaerobic Methane

Oxidizing Methylomirabilis Bioreactor Enrichment Culture. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01740

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, *27*(1), 14–25. https://doi.org/10.1002/pro.3235

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). "Exploring network structure, dynamics, and function using NetworkX" (pp. 11–15). Presented at the SciPy 2008, Pasadena, CA: Proceedings of the 7th Python in Science Conference. Retrieved from http://conference.scipy.org.s3-website-us-east-1.amazonaws.com/proceedings/scipy2008/paper_2/

Hahn, C. J., Laso-Pérez, R., Vulcano, F., Vaziourakis, K.-M., Stokke, R., Steen, I. H., et al. (2020). "Candidatus Ethanoperedens," a Thermophilic Genus of Archaea Mediating the Anaerobic Oxidation of Ethane, *11*(2), 18.

Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R., & Olsen, G. J. (1999). Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proceedings of the National Academy of Sciences*, *96*(7), 3578–3583. https://doi.org/10.1073/pnas.96.7.3578

Häring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R. A., & Prangishvili, D. (2005). Independent virus development outside a host. *Nature*, *436*(7054), 1101–1102. https://doi.org/10.1038/4361101a

Häring, M., Rachel, R., Peng, X., Garrett, R. A., & Prangishvili, D. (2005). Viral Diversity in Hot Springs of Pozzuoli, Italy, and Characterization of a Unique Archaeal Virus, *Acidianus* Bottle-Shaped Virus, from a New Family, the *Ampullaviridae*. *Journal of Virology*, *79*(15), 9904–9911. https://doi.org/10.1128/JVI.79.15.9904-9911.2005

Hartman, R., Munson-McGee, J., Young, M. J., & Lawrence, C. M. (2019). Survey of high-resolution archaeal virus structures. *Current Opinion in Virology*, *36*, 74–83. https://doi.org/10.1016/j.coviro.2019.05.008

Hatfull, G. F., & Hendrix, R. W. (2011). Bacteriophages and their genomes. *Current Opinion in Virology*, *1*(4), 298–303. https://doi.org/10.1016/j.coviro.2011.06.009

Hatzenpichler, R., Connon, S. A., Goudeau, D., Malmstrom, R. R., Woyke, T., & Orphan, V. J. (2016). Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. *Proceedings of the National Academy of Sciences*, *113*(28). https://doi.org/10.1073/pnas.1603757113

Hazard, C., Anantharaman, K., Hillary, L. S., Neri, U., Roux, S., Trubl, G., et al. (2025). Beneath the surface: Unsolved questions in soil virus ecology. *Soil Biology and Biochemistry*, *205*, 109780. https://doi.org/10.1016/j.soilbio.2025.109780

Hellwig, P., Dittrich, A., Heyer, R., Reichl, U., & Benndorf, D. (2024). Detection, isolation and characterization of phage-host complexes using BONCAT and click chemistry. *Frontiers in Microbiology*, *15*, 1434301. https://doi.org/10.3389/fmicb.2024.1434301

Hevroni, G., Flores-Uribe, J., Béjà, O., & Philosof, A. (2020). *Rising through the Ranks: Seasonal and Diel Patterns of Marine Viruses* (preprint). Microbiology. https://doi.org/10.1101/2020.05.07.082883

Heyerhoff, B., Engelen, B., & Bunse, C. (2022). Auxiliary Metabolic Gene Functions in Pelagic and Benthic Viruses of the Baltic Sea. *Frontiers in Microbiology*, *13*, 863620. https://doi.org/10.3389/fmicb.2022.863620

Hobbs, Z., & Abedon, S. T. (2016). Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic.' *FEMS Microbiology Letters*, *363*(7), fnw047. https://doi.org/10.1093/femsle/fnw047

Holler, T., Widdel, F., Knittel, K., Amann, R., Kellermann, M. Y., Hinrichs, K.-U., et al. (2011). Thermophilic anaerobic oxidation of methane by marine microbial consortia. *The ISME Journal*, *5*(12), 1946–1956. https://doi.org/10.1038/ismej.2011.77

Hu, E.-Z., Lan, X.-R., Liu, Z.-L., Gao, J., & Niu, D.-K. (2022). A positive correlation between GC content and growth temperature in prokaryotes. *BMC Genomics*, *23*(1), 110. https://doi.org/10.1186/s12864-022-08353-7

Hurst, L. D., & Merchant, A. R. (2001). High guanine–cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1466), 493–497. https://doi.org/10.1098/rspb.2000.1397

Hwang, Y., Roux, S., Coclet, C., Krause, S. J. E., & Girguis, P. R. (2023). Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. *Nature Microbiology*, *8*(5), 946–957. https://doi.org/10.1038/s41564-023-01347-5

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

Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers*. Milwaukee, Wis: ASQC Quality Press.

Ishida, T., Yu, L., Akutsu, H., Ozawa, K., Kawanishi, S., Seto, A., et al. (1998). A primitive pathway of porphyrin biosynthesis and enzymology in *Desulfovibrio vulgaris*. *Proceedings of the National Academy of Sciences*, *95*(9), 4853–4858. https://doi.org/10.1073/pnas.95.9.4853

Jaffe, A. L., Bardot, C., Le Jeune, A.-H., Liu, J., Colombet, J., Perrière, F., et al. (2023). Variable impact of geochemical gradients on the functional potential of bacteria, archaea, and phages from the permanently stratified Lac Pavin. *Microbiome*, *11*(1), 14. https://doi.org/10.1186/s40168-022-01416-7

Johnson, C., England, A., Munro-Ehrlich, M., Colman, D. R., DuBois, J. L., & Boyd, E. S. (2021). Pathways of Iron and Sulfur Acquisition, Cofactor Assembly, Destination, and Storage in Diverse Archaeal Methanogens and Alkanotrophs. *Journal of Bacteriology*, *203*(17). https://doi.org/10.1128/JB.00117-21

Jørgensen, B. B., & Boetius, A. (2007). Feast and famine — microbial life in the deep-sea bed. *Nature Reviews Microbiology*, *5*(10), 770–781. https://doi.org/10.1038/nrmicro1745

Joye, S. B. (2020). The Geology and Biogeochemistry of Hydrocarbon Seeps. *Annual Review of Earth and Planetary Sciences*, *48*(1), 205–231. https://doi.org/10.1146/annurev-earth-063016-020052

Joye, S. B., Boetius, A., Orcutt, B. N., Montoya, J. P., Schulz, H. N., Erickson, M. J., & Lugo, S. K. (2004). The anaerobic oxidation of methane and sulfate reduction in sediments from Gulf of Mexico cold seeps. *Chemical Geology*, *205*(3–4), 219–238. https://doi.org/10.1016/j.chemgeo.2003.12.019

Judd, A. G. (2004). Natural seabed gas seeps as sources of atmospheric methane. *Environmental Geology*, *46*(8), 988–996. https://doi.org/10.1007/s00254-004-1083-3

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*, e7359. https://doi.org/10.7717/peerj.7359

Karaca, D., Hensen, C., & Wallmann, K. (2010). Controls on authigenic carbonate precipitation at cold seeps along the convergent margin off Costa Rica. *Geochemistry, Geophysics, Geosystems*, *11*(8), 2010GC003062. https://doi.org/10.1029/2010GC003062

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

Kauffman, K. M., Hussain, F. A., Yang, J., Arevalo, P., Brown, J. M., Chang, W. K., et al. (2018). A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature*, *554*(7690), 118–122. https://doi.org/10.1038/nature25474

Kellermann, M. Y., Wegener, G., Elvert, M., Yoshinaga, M. Y., Lin, Y.-S., Holler, T., et al. (2012). Autotrophy as a predominant mode of carbon fixation in anaerobic methane-oxidizing microbial communities. *Proceedings of the National Academy of Sciences*, *109*(47), 19321–19326. https://doi.org/10.1073/pnas.1208795109

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

Kieft, K., Breister, A. M., Huss, P., Linz, A. M., Zanetakos, E., Zhou, Z., et al. (2021). Virus-associated organosulfur metabolism in human and environmental systems. *Cell Reports*, *36*(5), 109471. https://doi.org/10.1016/j.celrep.2021.109471

Knittel, K., Lösekann, T., Boetius, A., Kort, R., & Amann, R. (2005). Diversity and Distribution of Methanotrophic Archaea at Cold Seeps. *Applied and Environmental Microbiology*, *71*(1), 467–479. https://doi.org/10.1128/AEM.71.1.467-479.2005

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

Kosmopoulos, J. C., Klier, K. M., Langwig, M. V., Tran, P. Q., & Anantharaman, K. (2024). Viromes vs. mixed community metagenomes: choice of method dictates interpretation of viral community ecology. *Microbiome*, *12*(1), 195. https://doi.org/10.1186/s40168-024-01905-x

Kreil, D. P. (2001). Identification of thermophilic species by the amino acid compositions deduced from their genomes. *Nucleic Acids Research*, *29*(7), 1608–1615. https://doi.org/10.1093/nar/29.7.1608

Krukenberg, V., Harding, K., Richter, M., Glöckner, F. O., Gruber-Vodicka, H. R., Adam, B., et al. (2016). Candidatus Desulfofervidus auxilii, a hydrogenotrophic sulfate-reducing bacterium involved in the thermophilic anaerobic oxidation of methane: The partner bacterium in thermophilic methanotrophy. *Environmental Microbiology*, *18*(9), 3073–3091. https://doi.org/10.1111/1462-2920.13283

Krukenberg, V., Riedel, D., Gruber-Vodicka, H. R., Buttigieg, P. L., Tegetmeyer, H. E., Boetius, A., & Wegener, G. (2018). Gene expression and ultrastructure of meso- and thermophilic methanotrophic consortia: Interspecies interaction within AOM consortia. *Environmental Microbiology*, *20*(5), 1651–1666. https://doi.org/10.1111/1462-2920.14077

Kühner, M., Haufschildt, K., Neumann, A., Storbeck, S., Streif, J., & Layer, G. (2014). The Alternative Route to Heme in the Methanogenic Archaeon *Methanosarcina barkeri*. *Archaea*, *2014*, 1–13. https://doi.org/10.1155/2014/327637

Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, *24*(5), 719–720. https://doi.org/10.1093/bioinformatics/btm563

Laslett, D. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*(1), 11–16. https://doi.org/10.1093/nar/gkh152

Laso-Pérez, R., Wegener, G., Knittel, K., Widdel, F., Harding, K. J., Krukenberg, V., et al. (2016). Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature*, *539*(7629), 396–401. https://doi.org/10.1038/nature20152

Laso-Pérez, R., Krukenberg, V., Musat, F., & Wegener, G. (2018). Establishing anaerobic hydrocarbon-degrading enrichment cultures of microorganisms under strictly anoxic conditions. *Nature Protocols*, *13*(6), 1310–1330. https://doi.org/10.1038/nprot.2018.030

Laso-Pérez, R., Wu, F., Crémière, A., Speth, D. R., Magyar, J. S., Zhao, K., et al. (2023). Evolutionary diversification of methanotrophic ANME-1 archaea and their expansive virome. *Nature Microbiology*. https://doi.org/10.1038/s41564-022-01297-4

Layer, G. (2021). Heme biosynthesis in prokaryotes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1868*(1), 118861. https://doi.org/10.1016/j.bbamcr.2020.118861

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, *46*(D1), D708–D717. https://doi.org/10.1093/nar/gkx932

Leppänen, M., Sundberg, L.-R., Laanto, E., de Freitas Almeida, G. M., Papponen, P., & Maasilta, I. J. (2017). Imaging Bacterial Colonies and Phage-Bacterium Interaction at Sub-Nanometer

Resolution Using Helium-Ion Microscopy. *Advanced Biosystems*, *1*(8), 1700070. https://doi.org/10.1002/adbi.201700070

Levin, L. A., Orphan, V. J., Rouse, G. W., Rathburn, A. E., Ussler, W., Cook, G. S., et al. (2012). A hydrothermal seep on the Costa Rica margin: middle ground in a continuum of reducing ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2580–2588. https://doi.org/10.1098/rspb.2012.0205

Levin, L. A., Baco, A. R., Bowden, D. A., Colaco, A., Cordes, E. E., Cunha, M. R., et al. (2016). Hydrothermal Vents and Methane Seeps: Rethinking the Sphere of Influence. *Frontiers in Marine Science*, *3*. https://doi.org/10.3389/fmars.2016.00072

Levine, M. (1961). Effect of mitomycin C on interactions between temperate phages and bacteria. *Virology*, *13*(4), 493–499. https://doi.org/10.1016/0042-6822(61)90280-X

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Z., Pan, D., Wei, G., Pi, W., Zhang, C., Wang, J.-H., et al. (2021). Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *The ISME Journal*, *15*(8), 2366–2378. https://doi.org/10.1038/s41396-021-00932-y

Liang, X., Zhang, Y., Wommack, K. E., Wilhelm, S. W., DeBruyn, J. M., Sherfy, A. C., et al. (2020). Lysogenic reproductive strategies of viral communities vary with soil depth and are correlated with bacterial diversity. *Soil Biology and Biochemistry*, *144*, 107767. https://doi.org/10.1016/j.soilbio.2020.107767

Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., & Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proceedings of the National Academy of Sciences*, *101*(30), 11013–11018. https://doi.org/10.1073/pnas.0401526101

Linke, P., Wallmann, K., Suess, E., Hensen, C., & Rehder, G. (2005). In situ benthic fluxes from an intermittently active mud volcano at the Costa Rica convergent margin. *Earth and Planetary Science Letters*, *235*(1–2), 79–95. https://doi.org/10.1016/j.epsl.2005.03.009

Liu, R., Cai, R., Wang, M., Zhang, J., Zhang, H., Li, C., & Sun, C. (2024). Metagenomic insights into Heimdallarchaeia clades from the deep-sea cold seep and hydrothermal vent. *Environmental Microbiome*, *19*(1), 43. https://doi.org/10.1186/s40793-024-00585-2

Lobo, S. A. L., Lawrence, A. D., Romão, C. V., Warren, M. J., Teixeira, M., & Saraiva, L. M. (2014). Characterisation of Desulfovibrio vulgaris haem b synthase, a radical SAM family member. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1844*(7), 1238–1247. https://doi.org/10.1016/j.bbapap.2014.03.016

Łoś, M., & Węgrzyn, G. (2012). Pseudolysogeny. In *Advances in Virus Research* (Vol. 82, pp. 339–349). Elsevier. https://doi.org/10.1016/B978-0-12-394621-8.00019-4

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lovley, D. R. (2017). Syntrophy Goes Electric: Direct Interspecies Electron Transfer. *Annual Review of Microbiology*, *71*(1), 643–664. https://doi.org/10.1146/annurev-micro-030117-020420

Lubofsky, E. (2018, June 11). The Discovery of Hydrothermal Vents. *Oceanus*. Retrieved from https://www.whoi.edu/oceanus/feature/the-discovery-of-hydrothermal-vents/

Luo, X.-Q., Wang, P., Li, J.-L., Ahmad, M., Duan, L., Yin, L.-Z., et al. (2022). Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome*, *10*(1), 190. https://doi.org/10.1186/s40168-022-01384-y

Lwoff, A. (1953). LYSOGENY. *Bacteriological Reviews*, *17*(4), 269–337. https://doi.org/10.1128/br.17.4.269-337.1953

Makarova, K. S., Wolf, Y. I., Shmakov, S. A., Liu, Y., Li, M., & Koonin, E. V. (2020). Unprecedented Diversity of Unique CRISPR-Cas-Related Systems and Cas1 Homologs in Asgard Archaea. *The CRISPR Journal*, *3*(3), 156–163. https://doi.org/10.1089/crispr.2020.0012

Mäntynen, S., Laanto, E., Oksanen, H. M., Poranen, M. M., & Díaz-Muñoz, S. L. (2021). Black box of phage–bacterium interactions: exploring alternative phage infection strategies. *Open Biology*, *11*(9), 210188. https://doi.org/10.1098/rsob.210188

Martínez Martínez, J., Martinez-Hernandez, F., & Martinez-Garcia, M. (2020). Single-virus genomics and beyond. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-020-00444-0

Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M. J., Martínez, J. M., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nature Communications*, *8*, 15892. https://doi.org/10.1038/ncomms15892

McArdle, B. H., & Anderson, M. J. (2001). FITTING MULTIVARIATE MODELS TO COMMUNITY DATA: A COMMENT ON DISTANCE-BASED REDUNDANCY ANALYSIS. *Ecology*, *82*(1), 290–297. https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2

McGlynn, S. E., Chadwick, G. L., Kempes, C. P., & Orphan, V. J. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature*, *526*(7574), 531–535. https://doi.org/10.1038/nature15512

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX : Tools for structure building and analysis. *Protein Science*, *32*(11), e4792. https://doi.org/10.1002/pro.4792

Merkel, A. Y., Huber, J. A., Chernyh, N. A., Bonch-Osmolovskaya, E. A., & Lebedinsky, A. V. (2013). Detection of Putatively Thermophilic Anaerobic Methanotrophs in Diffuse Hydrothermal Vent Fluids. *Applied and Environmental Microbiology*, *79*(3), 915–923. https://doi.org/10.1128/AEM.03034-12

Meyerdierks, A., Kube, M., Lombardot, T., Knittel, K., Bauer, M., Glöckner, F. O., et al. (2005). Insights into the genomes of archaea mediating the anaerobic oxidation of methane. *Environmental Microbiology*, *7*(12), 1937–1951. https://doi.org/10.1111/j.1462-2920.2005.00844.x

Middelboe, M., Riemann, L., Steward, G., Hansen, V., & Nybroe, O. (2003). Virus-induced transfer of organic carbon between marine bacteria in a model community. *Aquatic Microbial Ecology*, *33*, 1–10. https://doi.org/10.3354/ame033001

Middelboe, M., Glud, R., & Filippini, M. (2011). Viral abundance and activity in the deep sub-seafloor biosphere. *Aquatic Microbial Ecology*, *63*(1), 1–8. https://doi.org/10.3354/ame01485

Minot, S., Bryson, A., Chehoud, C., Wu, G. D., Lewis, J. D., & Bushman, F. D. (2013). Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences*, *110*(30), 12450–12455. https://doi.org/10.1073/pnas.1300833110

Mochizuki, T., Sako, Y., & Prangishvili, D. (2011). Provirus Induction in Hyperthermophilic Archaea: Characterization of Aeropyrum pernix Spindle-Shaped Virus 1 and Aeropyrum pernix Ovoid Virus 1. *Journal of Bacteriology*, *193*(19), 5412–5419. https://doi.org/10.1128/JB.05101-11

Mrabet, N. T., Van Den Broeck, A., Van Den Brande, I., Stanssens, P., Laroche, Y., Lambeir, A. M., et al. (1992). Arginine residues as stabilizing elements in proteins. *Biochemistry*, *31*(8), 2239–2253. https://doi.org/10.1021/bi00123a005

Müller, C., Zhang, L., Zipfel, S., Topitsch, A., Lutz, M., Eckert, J., et al. (2022). Molecular interplay of an assembly machinery for nitrous oxide reductase. *Nature*, *608*(7923), 626–631. https://doi.org/10.1038/s41586-022-05015-2

Mullin, S. W. (2020, June 8). *Spatial and Temporal Dynamics of Microorganisms Living Along Steep Energy Gradients and Implications for Ecology and Geologic Preservation in the Deep Biosphere* (Dissertation (Ph.D.)). California Institute of Technology, Pasadena, CA. Retrieved from 10.7907/f3k8-ck13

Muniesa, M., Blanch, A. R., Lucena, F., & Jofre, J. (2005). Bacteriophages May Bias Outcome of Bacterial Enrichment Cultures. *Applied and Environmental Microbiology*, *71*(8), 4269–4275. https://doi.org/10.1128/AEM.71.8.4269-4275.2005

Murali, R., Yu, H., Speth, D. R., Wu, F., Metcalfe, K. S., Crémière, A., et al. (2023). Physiological potential and evolutionary trajectories of syntrophic sulfate-reducing bacterial partners of anaerobic methanotrophic archaea. *PLOS Biology*, *21*(9), e3002292. https://doi.org/10.1371/journal.pbio.3002292

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., & Van Sinderen, D. (2013). Bacteriophage Orphan DNA Methyltransferases: Insights from Their Bacterial Origin, Function, and Occurrence. *Applied and Environmental Microbiology*, *79*(24), 7547–7555. https://doi.org/10.1128/AEM.02229-13

Nakao, R., Senpuku, H., & Watanabe, H. (2006). *Porphyromonas gingivalis galE* Is Involved in Lipopolysaccharide O-Antigen Synthesis and Biofilm Formation. *Infection and Immunity*, *74*(11), 6145–6153. https://doi.org/10.1128/IAI.00261-06

Narayanan, A. K., & Philosof, A. (2020, July 11). Opti-Prep density separation of viruses for metagenomics v1. Protocol, Protocols.io. https://doi.org/10.17504/protocols.io.261geobqwl47/v1

Narayanan, A. K., & Philosof, A. (2024, August 5). SYBR Gold staining of viral-like particles on Anodiscs v1. https://doi.org/10.17504/protocols.io.8epv5rd85g1b/v1

Narayanan, A. K., & Philosof, A. (2025a, January 15). Viral nanoSIMS sample preparation v1. Protocol, Protocols.io. https://doi.org/10.17504/protocols.io.8epv52dp5v1b/v1

Narayanan, A. K., & Philosof, A. (2025b, January 16). Viral extraction from Sediment v1. Protocol, Protocols.io. https://doi.org/10.17504/protocols.io.6qpvr97zbvmk/v1

Nauhaus, K., Albrecht, M., Elvert, M., Boetius, A., & Widdel, F. (2007). *In vitro* cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate. *Environmental Microbiology*, *9*(1), 187–196. https://doi.org/10.1111/j.1462-2920.2006.01127.x

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

Nesper, J., Lauriano, C. M., Klose, K. E., Kapfhammer, D., Kraiß, A., & Reidl, J. (2001). Characterization of *Vibrio cholerae* O1 El Tor *galU* and *galE* Mutants: Influence on Lipopolysaccharide Structure, Colonization, and Biofilm Formation. *Infection and Immunity*, *69*(1), 435–445. https://doi.org/10.1128/IAI.69.1.435-445.2001

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Niou, Y.-K., Wu, W.-L., Lin, L.-C., Yu, M.-S., Shu, H.-Y., Yang, H.-H., & Lin, G.-H. (2009). Role of galE on biofilm formation by Thermus spp. *Biochemical and Biophysical Research Communications*, *390*(2), 313–318. https://doi.org/10.1016/j.bbrc.2009.09.120

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

Nurmemmedov, E., Castelnovo, M., Catalano, C. E., & Evilevitch, A. (2007). Biophysics of viral infectivity: matching genome length with capsid size. *Quarterly Reviews of Biophysics*, *40*(4), 327–356. https://doi.org/10.1017/S0033583508004666

Nweze, J. E., Schweichhart, J. S., & Angel, R. (2024). Viral communities in millipede guts: Insights into the diversity and potential role in modulating the microbiome. *Environmental Microbiology*, *26*(2), e16586. https://doi.org/10.1111/1462-2920.16586

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., et al. (n.d.). vegan: Community Ecology Package (Version 2.6-4). Retrieved from https://vegandevs.github.io/vegan/

Orphan, V. J., House, C. H., Hinrichs, K.-U., McKeegan, K. D., & DeLong, E. F. (2001). Methane-Consuming Archaea Revealed by Directly Coupled Isotopic and Phylogenetic Analysis. *Science*, *293*(5529), 484–487. https://doi.org/10.1126/science.1061338

Orphan, V. J., Turk, K. A., Green, A. M., & House, C. H. (2009). Patterns of^{15} N assimilation and growth of methanotrophic ANME-2 archaea and sulfate-reducing bacteria within structured syntrophic consortia revealed by FISH-SIMS. *Environmental Microbiology*, *11*(7), 1777–1791. https://doi.org/10.1111/j.1462-2920.2009.01903.x

Osorio-Rodriguez, D., Metcalfe, K. S., McGlynn, S. E., Yu, H., Dekas, A. E., Ellisman, M., et al. (2023). Microbially induced precipitation of silica by anaerobic methane-oxidizing consortia and implications for microbial fossil preservation. *Proceedings of the National Academy of Sciences*, *120*(51), e2302156120. https://doi.org/10.1073/pnas.2302156120

Pan, D., Morono, Y., Inagaki, F., & Takai, K. (2019). An Improved Method for Extracting Viruses From Sediment: Detection of Far More Viruses in the Subseafloor Than Previously Reported. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.00878

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996–1004. https://doi.org/10.1038/nbt.4229

Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., & Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology*, *38*(9), 1079–1086. https://doi.org/10.1038/s41587-020-0501-8

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1), D785–D794. https://doi.org/10.1093/nar/gkab776

Pasulka, A. L., Thamatrakoln, K., Kopf, S. H., Guan, Y., Poulos, B., Moradian, A., et al. (2018). Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environmental Microbiology*, *20*(2), 671–692. https://doi.org/10.1111/1462-2920.13996

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Paul, B. G., Bagby, S. C., Czornyj, E., Arambula, D., Handa, S., Sczyrba, A., et al. (2015). Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nature Communications*, *6*(1). https://doi.org/10.1038/ncomms7585

Paull, C. K., Hecker, B., Commeau, R., Freeman-Lynde, R. P., Neumann, C., Corso, W. P., et al. (1984). Biological Communities at the Florida Escarpment Resemble Hydrothermal Vent Taxa. *Science*, *226*(4677), 965–967. https://doi.org/10.1126/science.226.4677.965

Peña-Salinas, M. E., Speth, D. R., Utter, D. R., Spelz, R. M., Lim, S., Zierenberg, R., et al. (2024). Thermotogota diversity and distribution patterns revealed in *Auka* and *JaichMaa 'ja 'ag* hydrothermal vent fields in the Pescadero Basin, Gulf of California. *PeerJ*, *12*, e17724. https://doi.org/10.7717/peerj.17724

Peng, Y., Lu, Z., Pan, D., Shi, L.-D., Zhao, Z., Liu, Q., et al. (2023). Viruses in deep-sea cold seep sediments harbor diverse survival mechanisms and remain genetically conserved within

species. *The ISME Journal*, *17*(10), 1774–1784. https://doi.org/10.1038/s41396-023-01491-0

Perez Sepulveda, B., Redgwell, T., Rihtman, B., Pitt, F., Scanlan, D. J., & Millard, A. (2016). Marine phage genomics: the tip of the iceberg. *FEMS Microbiology Letters*, *363*(15), fnw158. https://doi.org/10.1093/femsle/fnw158

Perez-Molphe-Montoya, E., Küsel, K., & Overholt, W. A. (2022). Redefining the phylogenetic and metabolic diversity of phylum Omnitrophota. *Environmental Microbiology*, *24*(11), 5437–5449. https://doi.org/10.1111/1462-2920.16170

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF CHIMERAX : Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Philosof, A., Narayanan, A. K., Murali, R., & Orphan, V. J. (n.d.). Phage infection networks spanning distant taxonomic groups are a defining feature of deep marine methane seeps. *In Prep*.

Poorvin, L., Rinta-Kanto, J. M., Hutchins, D. A., & Wilhelm, S. W. (2004). Viral release of iron and its bioavailability to marine plankton. *Limnology and Oceanography*, *49*(5), 1734–1741. https://doi.org/10.4319/lo.2004.49.5.1734

Poulos, B. (2012, October 17). DNA Extraction of Viruses Using Wizard Columns Version 1. Tucson Marine Phage Lab. Retrieved from https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/e/20087/files/2015/08/Wizard_Column_Viral_DNA_Extraction_v1-2dqudis.pdf

Poulos, B. T., John, S. G., & Sullivan, M. B. (2018). Iron Chloride Flocculation of Bacteriophages from Seawater. In M. R. J. Clokie, A. M. Kropinski, & R. Lavigne (Eds.), *Bacteriophages* (Vol. 1681, pp. 49–57). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-7343-9_4

Pride, D. T., Sun, C. L., Salzman, J., Rao, N., Loomer, P., Armitage, G. C., et al. (2011). Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Research*, *21*(1), 126–136. https://doi.org/10.1101/gr.111732.110

Puxty, R. J., Millard, A. D., Evans, D. J., & Scanlan, D. J. (2016). Viruses Inhibit CO 2 Fixation in the Most Abundant Phototrophs on Earth. *Current Biology*, *26*(12), 1585–1589. https://doi.org/10.1016/j.cub.2016.04.036

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Quinn, T. P., Erb, I., Richardson, M. F., & Crowley, T. M. (2018). Understanding sequencing data as compositions: an outlook and review. *Bioinformatics*, *34*(16), 2870–2878. https://doi.org/10.1093/bioinformatics/bty175

Reed, C. J., Lewis, H., Trejo, E., Winston, V., & Evilia, C. (2013). Protein Adaptations in Archaeal Extremophiles. *Archaea*, *2013*, 1–14. https://doi.org/10.1155/2013/373275

Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., & Hatzenpichler, R. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate

amendment. *The ISME Journal*, *14*(11), 2851–2861. https://doi.org/10.1038/s41396-020-00749-1

Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davín, A. A., Waite, D. W., et al. (2021). A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nature Microbiology*, *6*(7), 946–959. https://doi.org/10.1038/s41564-021-00918-8

Ripp, S., & Miller, R. V. (1995). Effects of Suspended Particulates on the Frequency of Transduction among Pseudomonas aeruginosa in a Freshwater Environment. *Applied and Environmental Microbiology*, *61*(4), 1214–1219. https://doi.org/10.1128/aem.61.4.1214-1219.1995

Romig, W. R., & Brodetsky, A. M. (1961). ISOLATION AND PRELIMINARY CHARACTERIZATION OF BACTERIOPHAGES FOR BACILLUS SUBTILIS. *Journal of Bacteriology*, *82*(1), 135–141. https://doi.org/10.1128/jb.82.1.135-141.1961

Römling, U., & Balsalobre, C. (2012). Biofilm infections, their resilience to therapy and innovative treatment strategies. *Journal of Internal Medicine*, *272*(6), 541–561. https://doi.org/10.1111/joim.12004

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, *537*(7622), 689–693. https://doi.org/10.1038/nature19366

Ruff, S. E., Biddle, J. F., Teske, A. P., Knittel, K., Boetius, A., & Ramette, A. (2015). Global dispersion and local diversification of the methane seep microbiome. *Proceedings of the National Academy of Sciences*, *112*(13), 4015–4020. https://doi.org/10.1073/pnas.1421865112

Ruff, S. E., Kuhfuss, H., Wegener, G., Lott, C., Ramette, A., Wiedling, J., et al. (2016). Methane Seep in Shallow-Water Permeable Sediment Harbors High Diversity of Anaerobic Methanotrophic Communities, Elba, Italy. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00374

Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A., & Sørensen, S. J. (2020). CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *The CRISPR Journal*, *3*(6), 462–469. https://doi.org/10.1089/crispr.2020.0059

Santos-Medellin, C., Zinke, L. A., Ter Horst, A. M., Gelardi, D. L., Parikh, S. J., & Emerson, J. B. (2021). Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *The ISME Journal*, *15*(7), 1956–1970. https://doi.org/10.1038/s41396-021-00897-y

Santos-Medellín, C., Blazewicz, S. J., Pett-Ridge, J., Firestone, M. K., & Emerson, J. B. (2023). Viral but not bacterial community successional patterns reflect extreme turnover shortly after rewetting dry soils. *Nature Ecology & Evolution*, *7*(11), 1809–1822. https://doi.org/10.1038/s41559-023-02207-5

Savvichev, A. S., Rusanov, I. I., Kadnikov, V. V., Beletsky, A. V., Zakcharova, E. E., Samylina, O. S., et al. (2023). Biogeochemical Activity of Methane-Related Microbial Communities in Bottom Sediments of Cold Seeps of the Laptev Sea. *Microorganisms*, *11*(2), 250. https://doi.org/10.3390/microorganisms11020250

Scheller, S., Yu, H., Chadwick, G. L., McGlynn, S. E., & Orphan, V. J. (2016). Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science*, *351*(6274), 703–707. https://doi.org/10.1126/science.aad7154

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, *82*(7–8), 518–529. https://doi.org/10.1002/mrd.22489

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Schoell, M. (1988). Multiple origins of methane in the Earth. *Chemical Geology*, *71*(1–3), 1–10. https://doi.org/10.1016/0009-2541(88)90101-5

Schouten, S., Wakeham, S. G., Hopmans, E. C., & Sinninghe Damsté, J. S. (2003). Biogeochemical Evidence that Thermophilic Archaea Mediate the Anaerobic Oxidation of Methane. *Applied and Environmental Microbiology*, *69*(3), 1680–1686. https://doi.org/10.1128/AEM.69.3.1680-1686.2003

Scott, M., & Lindsey, R. (2022, September 29). Understanding blue carbon [Government]. Retrieved March 3, 2025, from https://www.climate.gov/news-features/understanding-climate/understanding-blue-carbon

Serrano, O., Gómez-López, D. I., Sánchez-Valencia, L., Acosta-Chaparro, A., Navas-Camacho, R., González-Corredor, J., et al. (2021). Seagrass blue carbon stocks and sequestration rates in the Colombian Caribbean. *Scientific Reports*, *11*(1), 11067. https://doi.org/10.1038/s41598-021-90544-5

Seymour, C. O., Palmer, M., Becraft, E. D., Stepanauskas, R., Friel, A. D., Schulz, F., et al. (2023). Hyperactive nanobacteria with host-dependent traits pervade Omnitrophota. *Nature Microbiology*, *8*(4), 727–744. https://doi.org/10.1038/s41564-022-01319-1

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sharp, P. M., Emery, L. R., & Zeng, K. (2010). Forces that influence the evolution of codon bias. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1544), 1203–1212. https://doi.org/10.1098/rstb.2009.0305

Shelford, E., Middelboe, M., Møller, E., & Suttle, C. (2012). Virus-driven nitrogen cycling enhances phytoplankton growth. *Aquatic Microbial Ecology*, *66*(1), 41–46. https://doi.org/10.3354/ame01553

Shieh, P., Dien, V. T., Beahm, B. J., Castellano, J. M., Wyss-Coray, T., & Bertozzi, C. R. (2015). CalFluors: A Universal Motif for Fluorogenic Azide Probes across the Visible Spectrum. *Journal of the American Chemical Society*, *137*(22), 7145–7151. https://doi.org/10.1021/jacs.5b02383

Singer, G. A. C., & Hickey, D. A. (2003). Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. *Gene*, *317*, 39–47. https://doi.org/10.1016/S0378-1119(03)00660-7

Skennerton, C. T., Chourey, K., Iyer, R., Hettich, R. L., Tyson, G. W., & Orphan, V. J. (2017). Methane-Fueled Syntrophy through Extracellular Electron Transfer: Uncovering the Genomic Traits Conserved within Diverse Bacterial Partners of Anaerobic Methanotrophic Archaea. *mBio*, *8*(4), e00530-17. https://doi.org/10.1128/mBio.00530-17

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.1

Speth, D. R., Yu, F. B., Connon, S. A., Lim, S., Magyar, J. S., Peña-Salinas, M. E., et al. (2022). Microbial communities of Auka hydrothermal sediments shed light on vent biogeography and the evolutionary history of thermophily. *The ISME Journal*, *16*(7), 1750–1764. https://doi.org/10.1038/s41396-022-01222-x

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Steward, G. F., Montiel, J. L., & Azam, F. (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography*, *45*(8), 1697–1706. https://doi.org/10.4319/lo.2000.45.8.1697

Steward, K. F., Eilers, B., Tripet, B., Fuchs, A., Dorle, M., Rawle, R., et al. (2020). Metabolic Implications of Using BioOrthogonal Non-Canonical Amino Acid Tagging (BONCAT) for Tracking Protein Synthesis. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00197

Sullivan, M. (2015, June 19). Phage Buffer v1. Protocol, Protocols.io. https://doi.org/10.17504/protocols.io.c5ey3d

Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P., & Chisholm, S. W. (2006). Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts. *PLoS Biology*, *4*(8), e234. https://doi.org/10.1371/journal.pbio.0040234

Sutherland, I. W., Hughes, K. A., Skillman, L. C., & Tait, K. (2004). The interaction of phage and biofilms. *FEMS Microbiology Letters*, *232*(1), 1–6. https://doi.org/10.1016/S0378-1097(04)00041-2

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

Takeuchi, M., Yoshioka, H., Seo, Y., Tanabe, S., Tamaki, H., Kamagata, Y., et al. (2011). A distinct freshwater-adapted subgroup of ANME-1 dominates active archaeal communities in terrestrial subsurfaces in Japan: ANME-1 archaea in terrestrial freshwater subsurfaces. *Environmental Microbiology*, *13*(12), 3206–3218. https://doi.org/10.1111/j.1462-2920.2011.02517.x

Teske, A., Hinrichs, K.-U., Edgcomb, V., de Vera Gomez, A., Kysela, D., Sylva, S. P., et al. (2002). Microbial Diversity of Hydrothermal Sediments in the Guaymas Basin: Evidence for

Anaerobic Methanotrophic Communities. *Applied and Environmental Microbiology*, *68*(4), 1994–2007. https://doi.org/10.1128/AEM.68.4.1994-2007.2002

Thingstad, T., & Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, *13*, 19–27. https://doi.org/10.3354/ame013019

Thomas, L., & Trehin, P. (2021, July 22). LauLauThom/MaskFromRois-Fiji: Masks from ROIs plugins for Fiji - initial release (Version 1.0.0). Zenodo. https://doi.org/10.5281/ZENODO.5121890

Timmers, P. H. A., Gieteling, J., Widjaja-Greefkes, H. C. A., Plugge, C. M., Stams, A. J. M., Lens, P. N. L., & Meulepas, R. J. W. (2015). Growth of Anaerobic Methane-Oxidizing Archaea and Sulfate-Reducing Bacteria in a High-Pressure Membrane Capsule Bioreactor. *Applied and Environmental Microbiology*, *81*(4), 1286–1296. https://doi.org/10.1128/AEM.03255-14

Timmers, P. H. A., Welte, C. U., Koehorst, J. J., Plugge, C. M., Jetten, M. S. M., & Stams, A. J. M. (2017). Reverse Methanogenesis and Respiration in Methanotrophic Archaea. *Archaea*, *2017*, 1–22. https://doi.org/10.1155/2017/1654237

Trubl, G., Jang, H. B., Roux, S., Emerson, J. B., Solonenko, N., Vik, D. R., et al. (2018). Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems*, *3*(5), 10.1128/msystems.00076-18. https://doi.org/10.1128/msystems.00076-18

Tyson, G. W., & Banfield, J. F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*, *10*(1), 200–207. https://doi.org/10.1111/j.1462-2920.2007.01444.x

Valentini, T. D., Lucas, S. K., Binder, K. A., Cameron, L. C., Motl, J. A., Dunitz, J. M., & Hunter, R. C. (2020). Bioorthogonal non-canonical amino acid tagging reveals translationally active subpopulations of the cystic fibrosis lung microbiota. *Nature Communications*, *11*(1), 2287. https://doi.org/10.1038/s41467-020-16163-2

Vulcano, F., Hahn, C. J., Roerdink, D., Dahle, H., Reeves, E. P., Wegener, G., et al. (2022). Phylogenetic and functional diverse ANME-1 thrive in Arctic hydrothermal vents. *FEMS Microbiology Ecology*, *98*(11), fiac117. https://doi.org/10.1093/femsec/fiac117

Waldbauer, J. R., Coleman, M. L., Rizzo, A. I., Campbell, K. L., Lotus, J., & Zhang, L. (2019). Nitrogen sourcing during viral infection of marine cyanobacteria. *Proceedings of the National Academy of Sciences*, *116*(31), 15590–15595. https://doi.org/10.1073/pnas.1901856116

Walsby, A. E. (1980). A square bacterium. *Nature*, *283*(5742), 69–71. https://doi.org/10.1038/283069a0

Wang, C., Zheng, R., & Sun, C. (2025, March 22). Characterization of deep-sea viruses reveals their unexpected diversity and role in facilitating host metabolism of complex organic matter. https://doi.org/10.1101/2025.03.21.644689

Wang, F., Cvirkaite-Krupovic, V., Vos, M., Beltran, L. C., Kreutzberger, M. A. B., Winter, J.-M., et al. (2022). Spindle-shaped archaeal viruses evolved from rod-shaped ancestors to package a larger genome. *Cell*, *185*, 1–11. https://doi.org/10.1016/j.cell.2022.02.019

Wang, L., Wang, Y., Huang, X., Ma, R., Li, J., Wang, F., et al. (2022). Potential metabolic and genetic interaction among viruses, methanogen and methanotrophic archaea, and their syntrophic partners. *ISME Communications*, *2*(1), 50. https://doi.org/10.1038/s43705-022-00135-2

Wankel, S. D., Adams, M. M., Johnston, D. T., Hansel, C. M., Joye, S. B., & Girguis, P. R. (2012). Anaerobic methane oxidation in metalliferous hydrothermal sediments: influence on carbon flux and decoupling from sulfate reduction. *Environmental Microbiology*, *14*(10), 2726–2740. https://doi.org/10.1111/j.1462-2920.2012.02825.x

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E., & Boetius, A. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature*, *526*(7574), 587–590. https://doi.org/10.1038/nature15733

Wegener, G., Laso-Pérez, R., Orphan, V. J., & Boetius, A. (2022). Anaerobic Degradation of Alkanes by Marine Archaea. *Annual Review of Microbiology*, *76*(1), 553–577. https://doi.org/10.1146/annurev-micro-111021-045911

Wilhelm, S. W., & Suttle, C. A. (1999). Viruses and Nutrient Cycles in the Sea. *BioScience*, *49*(10), 781–788. https://doi.org/10.2307/1313569

Williamson, S. J., Cary, S. C., Williamson, K. E., Helton, R. R., Bench, S. R., Winget, D., & Wommack, K. E. (2008). Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *The ISME Journal*, *2*(11), 1112–1121. https://doi.org/10.1038/ismej.2008.73

Xiang, X., Chen, L., Huang, X., Luo, Y., She, Q., & Huang, L. (2005). *Sulfolobus tengchongensis* Spindle-Shaped Virus STSV1: Virus-Host Interactions and Genomic Features. *Journal of Virology*, *79*(14), 8677–8686. https://doi.org/10.1128/JVI.79.14.8677-8686.2005

Yan, M., Pratama, A. A., Somasundaram, S., Li, Z., Jiang, Y., Sullivan, M. B., & Yu, Z. (2023). Interrogating the viral dark matter of the rumen ecosystem with a global virome database. *Nature Communications*, *14*(1), 5254. https://doi.org/10.1038/s41467-023-41075-2

Yi, Y., Liu, S., Hao, Y., Sun, Q., Lei, X., Wang, Y., et al. (2023). A systematic analysis of marine lysogens and proviruses. *Nature Communications*, *14*(1), 6013. https://doi.org/10.1038/s41467-023-41699-4

Young, K. D. (2007). Bacterial morphology: why have different shapes? *Current Opinion in Microbiology*, *10*(6), 596–600. https://doi.org/10.1016/j.mib.2007.09.009

Yu, H., Susanti, D., McGlynn, S. E., Skennerton, C. T., Chourey, K., Iyer, R., et al. (2018). Comparative Genomics and Proteomic Analysis of Assimilatory Sulfate Reduction Pathways in Anaerobic Methanotrophic Archaea. *Frontiers in Microbiology*, *9*, 2917. https://doi.org/10.3389/fmicb.2018.02917

Yu, H., Skennerton, C. T., Chadwick, G. L., Leu, A. O., Aoki, M., Tyson, G. W., & Orphan, V. J. (2021). Sulfate differentially stimulates but is not respired by diverse anaerobic methanotrophic archaea. *The ISME Journal*. https://doi.org/10.1038/s41396-021-01047-0

Yu, H., Speth, D. R., Connon, S. A., Goudeau, D., Malmstrom, R. R., Woyke, T., & Orphan, V. J. (2022). Community Structure and Microbial Associations in Sediment-Free Methanotrophic Enrichment Cultures from a Marine Methane Seep. *Applied and Environmental Microbiology*, e02109-21. https://doi.org/10.1128/aem.02109-21

Zhang, R., Mirdita, M., Levy Karin, E., Norroy, C., Galiez, C., & Söding, J. (2021). SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics*, *37*(19), 3364–3366. https://doi.org/10.1093/bioinformatics/btab222

Zhang, Xinning, Sigman, D. M., Morel, F. M. M., & Kraepiel, A. M. L. (2014). Nitrogen isotope fractionation by alternative nitrogenases and past ocean anoxia. *Proceedings of the National Academy of Sciences*, *111*(13), 4782–4787. https://doi.org/10.1073/pnas.1402976111

Zhang, Xueqin, Joyce, G. H., Leu, A. O., Zhao, J., Rabiee, H., Virdis, B., et al. (2023). Multi-heme cytochrome-mediated extracellular electron transfer by the anaerobic methanotroph 'Candidatus Methanoperedens nitroreducens.' *Nature Communications*, *14*(1), 6118. https://doi.org/10.1038/s41467-023-41847-w

Zhong, Z.-P., Du, J., Köstlbacher, S., Pjevac, P., Orlić, S., & Sullivan, M. B. (2024). Viral potential to modulate microbial methane metabolism varies by habitat. *Nature Communications*.

Zulko. (2016). DNA Features Viewer. Python, Edinburgh Genome Foundry. Retrieved from https://edinburgh-genome-foundry.github.io/DnaFeaturesViewer/

Zumft, W. G., & Kroneck, P. M. H. (2006). Respiratory Transformation of Nitrous Oxide (N2O) to Dinitrogen by Bacteria and Archaea. In *Advances in Microbial Physiology* (Vol. 52, pp. 107–227). Elsevier. https://doi.org/10.1016/S0065-2911(06)52003-X