ANNUAL REVIEWS

Annual Review of Virology

Viruses in Subsurface Environments

Jennifer Wirth and Mark Young

Department of Plant Science and Plant Pathology and Thermal Biology Institute, Montana State University, Bozeman, Montana, USA; email: myoung@montana.edu

Annu. Rev. Virol. 2022. 9:99–119

The *Annual Review of Virology* is online at virology.annualreviews.org

https://doi.org/10.1146/annurev-virology-093020-015957

Keywords

subsurface viruses, subsurface metagenomics, virus abundance, virus ecology

Abstract

Over the past 20 years, our knowledge of virus diversity and abundance in subsurface environments has expanded dramatically through application of quantitative metagenomic approaches. In most subsurface environments, viral diversity and abundance rival viral diversity and abundance observed in surface environments. Most of these viruses are uncharacterized in terms of their hosts and replication cycles. Analysis of accessory metabolic genes encoded by subsurface viruses indicates that they evolved to replicate within the unique features of their environments. The key question remains: What role do these viruses play in the ecology and evolution of the environments in which they replicate? Undoubtedly, as more virologists examine the role of viruses in subsurface environments, new insights will emerge.

Michael Crichton wrote, "Life will find a way."

We modify this to, "Viruses will find a way."

OVERVIEW

Viruses in subsurface environments are important players in influencing the ecology and evolution of the microbial communities they infect. Beginning with the pioneering work of Bergh et al. (1) and Proctor & Fuhrman (2) in the late 1980s and 1990, it has long been appreciated that ocean surface waters and terrestrial soil environments harbor on the order of $10^{30–31}$ virus particles, making viruses the most abundant life-like agents on Earth (3–7). Recent estimates suggest that there are roughly an equivalent number of microbes living at and below the seafloor and in the continental subsurface terrestrial surface environments [$\sim10^{29}$ cells in each (6, 8–10)] as in the surface oceans. With estimates of 10–100 virus particles per cell, subsurface environments may contain an equivalent or even greater number of virus particles than surface environments (11). It is likely that these numbers are an underestimation of the true viral abundance. High virus abundance combined with corresponding high estimates for virus infection rates of more than 10^{23} infections/s (4, 12) directly implicates host-virus interactions as key players in the ecology and evolution of life wherever it is found. This review focuses on recent advances in our understanding of the diversity, abundance, and ecological roles of viruses in Earth's subsurface environments.

It is difficult to precisely define a subsurface environment. In general, Earth habitats can be divided into three broad categories: surface terrestrial, surface marine, and subsurface environments. Subsurface environments can be both marine and terrestrial but represent environments deep (50 m to thousands of meters) below the surface. Here we use the definition of subsurface environments as terrestrial habitats below 8 m, oceans below 200 m, and their sediments (5), or as habitats where direct influences of photosynthesis and interactions with Earth's atmosphere are minimal. Subsurface environments can be further divided into two categories: either productive or detrital depending on whether organic carbon or inorganic carbon is the main carbon source for the microbial community (13). Because no or little light is present in subsurface environments, alternative electron donors to water are utilized, typically hydrogen (H_2) or sulfide (H_2S). In these environments, diverse microbial communities and their viruses dominate. It is estimated that microbial communities in subsurface environments contain 13% of life's carbon (70 billion tC) (6) and account for ~60% of all microorganisms on Earth (10, 14). Subsurface environments represent an extensive and diverse collection of microbial habitats that span a wide spatial range of environments from terrestrial caves, hot and cold deep-sea vents, microfractures in terrestrial and marine sediments, and deep within ice sheets to subsurface ocean environments (**Table 1**). While there are interactions between surface and subsurface environments, the temporal scale of these interactions can vary widely from months to millions of years depending on the environment and the unique physical, geochemical, and biogeochemical conditions driving the ecology of the microbial communities and their viruses in subsurface environments (9, 11, 15). As discussed below, viruses in subsurface environments have evolved along with their hosts to thrive in these unusual environments. This review attempts to update our current knowledge of viral communities in subsurface environments and to look forward to the enormous potential that future studies of subsurface viral communities have to contribute to our appreciation of viral diversity on Earth and the role viruses play in the ecology and evolution of life.

When we consider viruses in subsurface environments, it is important to remember that the parasitic nature of all viruses demands that they have the same requirements to carry out their replication as the minimum requirements for cellular life (16, 17) as well as a few additional unique requirements. Both viruses and cells require access to liquid water, the ability to take advantage

Table 1 Important findings from representative investigations of subsurface environments

Environment	Key findings	Estimated viral/host community size	Cell/mL	VLP/mL	Reference(s)
Deep ocean (hadal environment)	There is little overlap in OTUs in sediments versus seawater and much overlap between hadal and nonhadal seawater. Most OTUs are novel. Lysogeny is more common in deep ocean than in surface waters.	12,710 vOTUs	~1×10^6 (sediments, decreases with depth)	~10 (sediments, decreases with depth)	5, 107
Deep ocean vents	Virus communities found across sites persist through time (~3 years). There is high richness, and there are no dominant strains and narrow host ranges.	28 virus families	1×10^5–9×10^9	2.7×10^4–4.5×10^9	97, 98
Deep microfractures in terrestrial sediments (fracking)	There is little overlap between viruses in different hydraulically fractured ecosystems. Evidence exists of high levels of viral predation. Lysogeny is enriched in established versus input viral communities.	<10 host OTUs, 1,838 vOTUs	Not available	Not available	126, 137
Global Ocean Virome	Many novel viruses have been identified. There are few shared genes between surface and deep ocean. Some viruses are abundant and widespread in oceans. Hosts for most abundant viruses have been identified.	488,130 viral populations (viral contigs >5 kb with >70% of shared genes at ≥95% average nucleotide identity)	Not available	Not available	72, 77, 117
High-temperature acidic hot spring (deep subsurface analog)	Many new viral families have been identified. Greater than 60% of cells contain at least one virus; many contain more than one. There is indication of a broad host range.	~10 hosts and 100–200 viral types	<1×10^6	1×10^3–1×10^5	141

Abbreviations: OTU, operational taxonomic unit; VLP, virus-like particle; vOTU, virus operational taxonomic unit.

of energy-yielding redox reactions, temperatures below 121°C, and sufficient time to evolve. In addition, all viruses require metabolically active cellular hosts to complete their replication cycles as well as a mechanism to infect new hosts, either by horizontal transmission—typically by extracellular release of infectious progeny virus particles—or by vertical transmission, where infectious viral genomes are passed on to progeny cells during host cell division. Many subsurface environments meet these minimal physical and biological requirements, and we therefore should expect viruses to be present in diverse subsurfaces where these conditions are met.

It has been proposed and it is reasonable to believe that viruses and virus-like agents have coevolved since life emerged on Earth (18–25). In fact, deep-sea vents have been proposed to be a site of early life and virus emergence on early Earth (26). It is likely that from the time of the emergence of life on Earth and continuing to the present, cellular and virus evolution is, in part, dependent on their constant interactions. To our knowledge, there is no known cellular life, either extant or in the past, that has not interacted with viruses, and these interactions have been critical for the evolution of both cells and viruses. Virus-infected cells take on a distinct physiological state, termed the virocell state, as compared to uninfected cells (27, 28). It is estimated that 20–40% of all prokaryotic cells in aquatic environments are in the virocell state (12, 29), and this may well be an underestimation. This high level of host-virus interaction contributes to the fluid exchange of genes between cells and viruses and vice versa that is well documented (30–34). It is estimated that viruses transfer $\sim 10^{29}$ genes per day globally (34). Viral genomes can acquire and modify host genes through recombination and other mutation mechanisms. Likewise, viruses can alter the genetic content of their hosts through the process of transduction as well as by the direct addition of viral genetic material to the host's genetic capacity. Transduction can lead to both the hosts and viruses acquiring new functional genes that can be beneficial for the host and/or viral fitness (35–37). The two-way exchange of host and viral genetic material has had a strong influence on creating most, if not all, present-day cellular and viral genomes. Host-lytic virus interactions in microbial communities can play a key role in regulating host and virus abundance and community structure. A second major ecological influence of viruses in many microbial environments is their role in creating a viral shunt. The viral shunt is created by viral lysis of infected hosts, which releases organic biomatter that can be either partitioned back into biomatter in the form of new cells or released into the dissolved organic carbon fraction that can be removed from the habitat by various transport mechanisms (4, 38–41). Through the viral shunt process, viruses are thought to significantly contribute to Earth's carbon, nitrogen, and phosphorus biogeochemical cycles (4, 41–44). It is estimated that the viral shunt releases nearly 10 billion tons of carbon per day and is a key component in nutrient cycling in the oceans (12, 35, 38, 39, 42). There is no reason we should not expect that host-virus interactions in subsurface environments also play similar critical roles in the ecology and evolution of the environments in which they exist.

In recent years the view of host-virus interactions has expanded (45), with many identified in subsurface environments. Lytic and temperate viruses with a lytic phase dominate our thinking and modeling of host-virus interactions. In many bacteria-dominated microbial communities, lytic outcomes can predominate, which is important for controlling host and virus diversity and the microbial community structure. In contrast, nonlytic viruses or viruses that can convert a host into a lysogen are also quite common. Lysogenic states include the provirus state where viral genomes are integrated into the host genome, pseudolysogens that produce high levels of virus progeny without killing their hosts, and polylysogens in which there is infection by multiple viruses within the same cell. Both lytic and lysogenic states are present in all domains of life: eukarya, bacteria, and archaea. Lysogeny appears to be a particularly successful replication strategy in low cell density (approximately fewer than 10^6 cells/mL), low productivity environments like those found in many subsurface environments (38, 46). Lysogens often take advantage of both a horizontal transmission

pathway through budding of progeny virions and a vertical transmission pathway through host cell division. It has been proposed that a large pool of proviruses in a microbial community provides an important reservoir of virus-encoded genes (upon activation) for responding to environmental stress, such as changes in nutrient availability (46, 47) or exposure to antimicrobial agents (48). In many subsurface microbial communities—for example, deep-sea hydrothermal vents—lysogenic states appear to be over-represented as compared to in surface environments (49, 50).

It is widely recognized that host-virus interactions can drive the diversity and abundance of both viruses and their hosts, typically through a kill the winner model (Lotka-Volterra-type dynamics) (51). However, this is likely an oversimplification of host-virus interactions found in nature, including in subsurface environments. This is because kill the winner models do not take into full account changes in the environment, the presence of multiple viruses competing for the same host, the ability of some viruses to infect more than one host, that many viruses have replication cycles that do not require cell lysis, and that some virus infections can actually increase their host fitness (52, 53). A modification of the kill the winner model, the piggyback the winner model, was recently proposed (54, 55). This revised model takes into account the observation that in some environments there is a reduction in the virus-to-microbe ratio at high host abundance (55–58). This decrease is not a function of increased host resistance. The piggyback the winner model postulates that high host abundance favors a switch to lysogenic infections as compared to lytic infections, and this has been observed in the ocean water column (59). In summary, while there is a general appreciation that host-virus interactions are important in microbial ecology and evolution, much remains to be understood about these host-virus interactions, especially in subsurface environments.

Virologists, and scientists in general, have a tendency to view viruses through a lens of their role as disease-causing agents. While there is no doubt that viruses are responsible for important human, animal, and plant diseases, this myopic view downplays the potential benefits viruses can provide their hosts. Increasingly, virus infections are being demonstrated to provide a fitness advantage to their hosts (60–62). While some virus infections lead to pathogenesis in their hosts, many do not. Given the sheer number of host-virus interactions, it seems likely that viral pathogenesis may not be as common as viral commensalism and mutualism. An expanded view of virus genome architecture has emerged over the past few years. This view holds that viral genomes can be composed of up to three classes of genes: (*a*) genes involved directly in completing the virus replication cycle (e.g., viral polymerases, virion structural proteins, lysis proteins), (*b*) auxiliary metabolic genes (AMGs) that can provide niche-specific metabolic functions to their hosts, often enhancing viral replication, and (*c*) viral genes that function to counter host antiviral defense systems (**Figure 1**).

Knowledge of viral AMGs found in surface and subsurface environments is expanding (36, 37, 63–69). Most AMGs are thought to have originated from cellular genomes and to have been co-opted by viruses to increase virus and/or host fitness under changing environmental conditions. Two contrasting outcomes of AMGs have been noted (63, 70). In the first, AMGs optimize virus propagation by altering host metabolism to favor virus replication. In the second, AMGs increase host fitness by increasing substrate uptake and by altering interactions with other microbial community members, including other viruses. Two classes of AMGs are recognized (63, 71, 72): those that have metabolic functions that can be assigned in the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways (Class I) and those that are absent from KEGG and have a more peripheral role in metabolism (Class II). One of the first AMGs discovered was in marine cyanophage (37, 65, 73–75). Many of the marine cyanophage that infect cyanobacterial *Prochloroccus* and *Synechococcus* hosts carry the *psbA* gene in their genome that codes for a protein of the core photosystem II reaction center D1. Viral expression of this protein helps maintain the

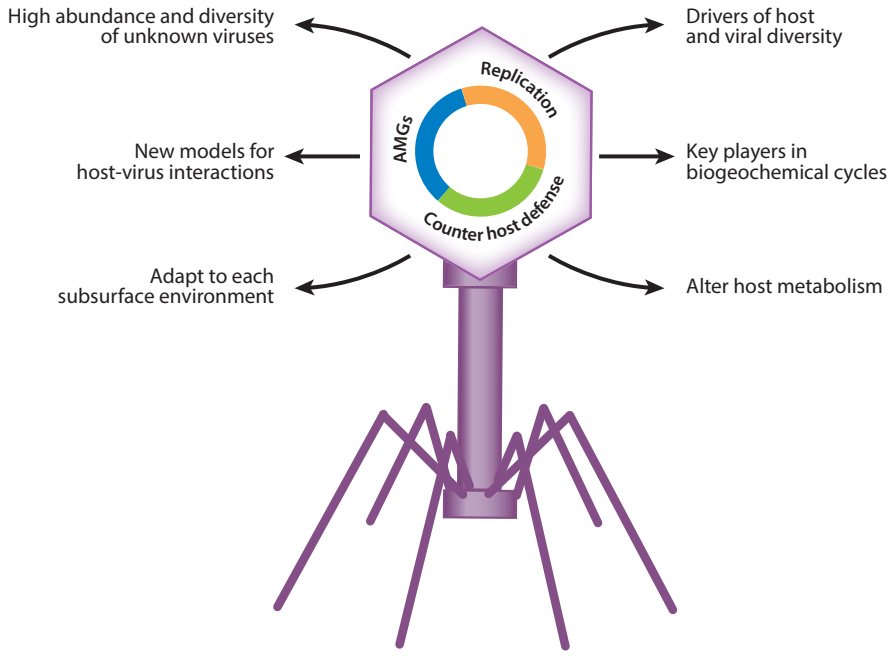

Figure 1

Roles and attributes of viruses in subsurface environments: a conceptual framework for the three classes of viral genes present in many subsurface viral genomes. Abbreviation: AMGs, auxiliary metabolic genes.

cell's photosynthetic activity, providing cellular energy needed for viral replication. Analysis of marine viral metagenomic data sets from both surface and deep ocean waters has found AMGs in aggregate that cover nearly all of the genes in central carbon metabolism (36, 76). It is likely that these AMGs allow virus infection to reprogram host carbon metabolism. It has been proposed that these AMGs mimic a carbon-limited metabolic state in the host, often inducing a state of starvation that promotes production of host resources such as increased dNTP biosynthesis needed for virus propagation (36, 63). AMGs encoded in viruses from subsurface environments have been found to directly provide fitness advantages to their hosts. For example, deep-sea viruses carry AMGs enhancing sulfur oxidation as a means of increasing chemosynthesis in deep ocean environments (69, 77). Anantharaman et al. (77) showed that these viruses infect the sulfur-oxidizing bacteria SUP05. Gammaproteobacteria encode AMGs for the alpha (*rdsrA*) and gamma (*rdsrC*) subunits of the reverse-acting dissimilatory sulfite reductase complex (*Rdsr*) for sulfur oxidation. In addition, Roux et al. (67) found AMGs for dissimilatory sulfite reductase subunit C (*dsrC*) by analysis of single-cell genomes from cells collected from marine oxygen minimum zones. AMGs associated with viruses found at deep marine hydrothermal vent sites are thought to create branched metabolic pathways for pyrimidine, alanine, aspartate, glutamate, nitrogen, amino, and nucleotide sugar metabolism, likely enhancing the host's metabolic capacity and leading to increased host fitness (78).

Viral genomes can encode for genes that counteract host antiviral systems. In recent years understanding of the diversity and function of host viral defense systems has expanded (79). We now have a more mechanistic understanding of how many of these defense systems function to prevent or subvert viral infection. Common host defense systems include abortive infection systems, restriction-modification systems, toxin-antitoxin systems, CRISPR-Cas systems, argonautes, and

cyclic nucleotide-based systems (reviewed in 80–82). Not surprisingly, viruses, including viruses present in subsurface environments, have evolved systems to overcome host defense systems. The anti-CRISPR (Acr) proteins are recently discovered anti-host defense proteins (reviewed in 83). CRISPR-Cas systems are a common defense mechanism in prokaryotes and are estimated to exist in 50% of bacteria and 90% of archaea (84, 85). It has been observed that the CRISPR-Cas system is under tight regulation and in some cases is not activated until cell density is high, when microbial populations are more susceptible to virus infection. To counteract cellular CRISPR-Cas systems, viruses have evolved Acr proteins (53, 86–90). Acr proteins tend to be small (53–333 aa) and usually directly interact with CRISPR-Cas proteins to inactivate them. Enzymatic activities of Acrs have also been discovered; they can act by cleaving CRISPR RNA (87, 91, 92), acetylating protospacer-adjacent motif recognition sites (93), or reducing nuclease activity that rapidly degrades CRISPR cyclic signaling molecules (94). It has been shown that the multiplicity of infection (MOI) influences the efficacy of Acrs (86, 95). At low MOI, Acrs are less effective, presumably due to the rapid kinetics of the CRISPR-Cas systems. Models suggest that the MOI threshold for Acr effectiveness is an indication of virus cooperation in which high Acr expression induces a state of host immunosuppression benefiting the subsequent viral infections. *Acr* genes have been discovered in the genomes of viruses found in subsurface environments. For example, in the archaeal *Sulfolobus islandicus* rod-shaped virus (SIRV), 12 *acrID1* paralogs were found. The acrlD1 protein interacts with the host's (*Sulfolobus*) Cas10d protein, thereby inhibiting CRISPR-Cas subtype I-D immunity (96). It is unknown if these 12 Acrs function independently or act cooperatively. As He and colleagues (96) pointed out, multiple copies of *acr* genes may result in higher Acr protein levels that can inhibit the *Sulfolobus* CRISPR-Cas system in a timelier manner to the benefit of viral replication. To our knowledge, there has not been a systematic search of surface and subsurface metagenomic data sets for Acrs, which would be insightful for understanding the diversity and ecological role for ACRs in subsurface environments.

CASE STUDIES

Our knowledge of viruses in subsurface environments is limited by access, difficulties in sampling, costs, and the ability to recover virus particles across broad size and composition types. To date, viruses have been examined to some extent in multiple subsurface environments but most extensively in marine environments (**Table 1**). These include deep-sea hydrothermal vent fluids (97, 98), deep-sea cold seep sediments (99, 100), other deep-sea sediments (15, 101–110), surface and deep seawaters (59, 72, 111–121), permafrost (122, 123), arctic lakes (124), glaciers (125), and terrestrial subsurfaces disturbed by hydraulic fracking (126). While there has been an increased examination of viruses present in subsurface environments, there are still many unexplored environments. A number of subsurface microbial communities have been described to some extent, but their viral communities have yet to be explored, including karst cave systems (127, 128), volcanic ice caves (129), deep mines (130, 131), and deep ice environments (132).

While it is beyond the scope of this review to discuss all studies examining viruses inhabiting subsurface environments, we highlight a few of them in the following case studies (**Figure 2**). In general, most studies looking at viruses in subsurface environments are focused on defining the virus diversity present, their relative abundance, and to a lesser extent their lifestyle (temperate or lytic), metabolic capacity, and possible ecological role in the environments in which they are located. Metagenomics, the creation and analysis of the resulting genetic sequences, has greatly expanded our knowledge of microbial and viral diversity and has become the dominant approach to investigate microbial communities present in subsurface environments. The availability of ever-better DNA sequencing technologies and tools for quantitative analysis of metagenomic data

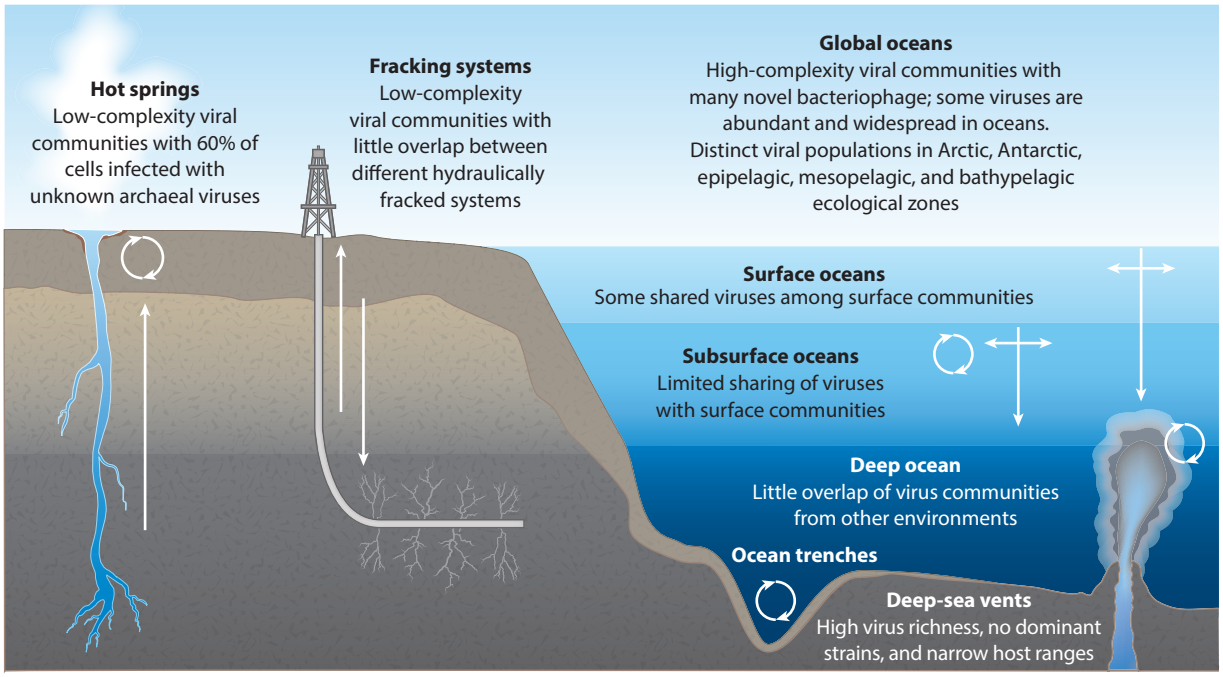

Figure 2

Viruses in representative subsurface environments and their general features. Arrows indicate the potential for virus movement between and within the subsurface environments.

sets has allowed for deeper DNA sequencing and analysis of microbial and viral genetic material recovered directly from subsurface environmental samples. For example, there have been multiple metagenomic studies on viruses present at deep-sea hydrothermal vents (47, 98, 133). From the early studies it was suggested that deep-sea vent sites support viral communities that are diverse and abundant and that infect a broad range of microbial hosts. Furthermore, temperate viruses were found to be quite common, suggesting that this viral lifestyle is particularly well suited to deep-sea vent environments. Thomas et al. (98) used metagenomic analysis to compare the virus communities across three geochemically and geographically distinct deep-sea hydrothermal vent fluids. They found that viruses near hydrothermal vents are active, abundant, and diverse, but not broadly shared across different vent sites. In this study (98), 31 microbial metagenomic data sets were generated and bioinformatically examined for the presence of viral sequences using VirFinder (134). They looked at low pH and high pH vents located in close proximity to each other (20 km) on the Mid-Cayman Rise in the Caribbean Sea and a low pH, high metal content submarine volcano vent located on the Juan de Fuca Ridge in the Pacific Ocean at a 1,520-m depth. Samples were collected both at the vent orifice and in vent plume waters. Across the three vent areas, 28 different viral families were detected, 13 of which were detected at all three vent areas, and 14 viral families were found in both Caribbean vent sites. While VirFinder likely underdetects viral sequences in metagenomic data sets, it does allow for the relative abundances of the detected viral sequences across data sets to be compared with high confidence. Both archaeal and bacterial virus families were detected. Interestingly, archaeal virus families originally found in terrestrial hot springs (135, 136), including the *Fuselloviridae*, *Bicaudaviridae*, and *Guttaviridae*, were detected. In the highest temperature and lowest pH vent samples, archaeal viruses were

more common. Outside of these vent sites, family members of the order *Caudovirales* (*Myoviridae*, *Podoviridae*, *Siphoviridae*) were most common. However, differences in viral taxonomy were found between the vent sites. For example, Hurwitz & Sullivan (113) found that *Myoviridae* had the highest relative abundance in 12 of the 16 Pacific vent samples, while they were much lower in the Caribbean vent sites that had higher relative abundance of *Podoviridae* and *Guttaviridae* virus families. At least one vent site showed evidence that the viral community can persist over a three-year sampling interval. Overall, the viral communities in the three vent sites had high richness and were not dominated by specific viral strains. However, rare fraction analysis indicated that the total diversity of viruses was not captured in this study. Network analysis suggested that few viruses infected multiple microbial linages, indicating that few viruses were generalists. Interestingly, Thomas et al. (98) found that the vent viral communities were endemic to individual vent sites, suggesting that there is limited dispersal of viruses between vent fields and/or that there is rapid local adaptation to each particular vent site.

Jian and colleagues (107) examined the diversity and distribution of viruses in the hadal biosphere, defined as the deep ocean environments below 6,000 m. In this study, the microbial metagenomes from seawater and sediment samples from three geographically distinct oceanic trench sites were examined. Overall, 12,710 virus operational taxonomic units (vOTUs), which roughly correspond to virus types or species, were defined using network clustering of data sets from 19 publicly available metagenomes. In the seawater, 6,011 vOTUs were defined, while 6,573 vOTUs were defined in the sediments. Surprisingly, only 6 vOTUs were found in both environments. In contrast, 1,027 vOTUs were present in both the hadal metagenomes and the controls of nonhadal marine environments, indicating a connection between these two environments. Most of the vOTUs in the hadal data sets were novel. Greater than 99% of the vOTUs in the hadal biosphere had no close homologs to viruses in the National Center for Biotechnology Information Viral RefSeq, Global Oceans Viromes (GOV) 2.0, and IMG/VR viral databases. The vast majority of the hadal biosphere vOTUs (85%) could not be taxonomically classified. Of the 15% that could be classified, 13.5% fell into the *Caudovirales* order of double-stranded DNA (dsDNA) viruses (*Myoviridae*, *Siphoviridae*, and *Podoviridae* viral families), and interestingly 0.6% were classified as members of the nucleocytoplasmic large DNA viruses group. Among the predicted bacterial hosts, *Gammaproteobacteria*, *Alphaproteobacteria*, *Bacteroidia*, and *Acitinobacteria* were the most common, while *Thaumarchaeota* was the most common archaeal phylum present. Viral hosts could be assigned to 1,229 vOTUs (9.7% of total vOTUs), including 27 vOTUs that were assigned as *Thaumarchaeota* viruses. Jian et al. (107) also noted that a significant number of vOTUs co-occurred across the three trench sites, suggesting the possibility of exchange of viral populations. There were 815 vOTUs present in two or more of the hadal trench sites. Both intertrench and intratrench exchanges of viral communities were detected, which the authors speculate were likely a result of current flow and deep ocean water circulation, upwelling of deep waters, and/or the downward movement of particulate organic matter. The availability of nutrients and viral lifestyles of viruses in the hadal biosphere was examined. It was found that the availability of nutrients, especially nitrogen and carbon, influenced niche-dependent distribution of viral populations. It was also found that the lysogenic lifestyle was more common in the deep ocean than in the upper ocean. Interestingly, niche-specific AMGs were found in the hadal viruses. One group of AMGs was associated with nucleotide, amino acid, vitamin, cofactor, and metabolism-related protein families. A second group of AMGs was associated with genetic information processing and cellular signaling processes. These results suggest that viruses in the hadal biosphere contribute to reprogramming the nitrogen and carbon metabolism of their hosts. Overall, the presence of diverse AMGs indicates that virus infections are contributing to host metabolism, affecting ocean biogeochemical cycles, and performing important ecological functions in their environments.

Studies of the viral communities established in fracture shale ecosystems provide insights into the rapid selection and establishment of viral communities when they are artificially introduced from surface solutions (126, 137). These studies are particularly interesting because they temporally track host and virus community dynamics during the introduction of surface solutions into the previously closed environments in the deep subsurface. Hydraulic fracturing uses high-pressure injection of chemicals and surface water into a shale formation, generating fracture networks that release gas and oil. The depth of injection is typically greater than 1,000 m, creating a fractured network environment that is characterized by high pressures, temperatures, and salinities. Input solutions and the resulting fluids that have passed through the shale were collected at wellheads and in the oil-gas-water separators used in the extraction process. Temporal samples were collected, and metagenomic analysis was used to monitor the development of the microbial and viral communities. Changes in the host and virus community structures were examined for up to ~300 days after the start of the fracturing process. Daly and colleagues (126) monitored viral and host community structures in multiple geographically separated hydraulically fractured wells. In general, they found over time that a low-diversity host community of less than 10 microbial operational taxonomic units was established that are dominated by anerobic *Halanaerobium spp*. Interestingly, this low-diversity host environment was characterized by a high-diversity virus community. Through a combination of the data from five well sites, 1,838 vOTUs at the species-level taxonomy were identified, which is comparable to the estimated viral diversity seen in soil ecosystems that harbor high microbial diversity (122, 138, 139). Of these 1,838 vOTUs, 34.8% could be taxonomically assigned to the order *Caudovirales*, while 46% could not be assigned to taxonomic groups. However, this 46% of vOTUs could be network clustered into 156 groups that likely represent new candidate virus genus-level groups. The high level of virus diversity in hydraulic fracture ecosystems is on par with the 658 previously described candidate genera from 104 viral-targeted samples collected for the GOV data sets (117). Unlike what was observed in the hadal habitats, there is little detected overlap between the viral communities in the different sampled hydraulically fractured shale ecosystems. This is likely a result of the differences between the two subsurface environments, one being a closed system where a foreign microbial community was recently and rapidly introduced by human intervention (hydraulically fractured shale ecosystem), while the other is a long-standing, more open system that has not experienced the rapid introduction of new/foreign microbial communities. In the case of the hydraulically fractured shale ecosystem, Jian et al. (107) suggest that either strong founder effects and/or very rapid diversification processes are responsible for the unique viral communities in each site. An active lysogenic lifestyle was found to be enriched in the established viral communities as compared to the input viral community as indicated by the strong correlation between viral and host relative abundances. It is interesting that hosts are surviving despite high predation from free viruses and induction of prophage. Furthermore, the large temporal fluctuations in host relative abundance, in the absence of other predators, suggest that the changes in host strains are exclusively due to viral predation. Overall, these studies (126, 137) show that viral lysis of hosts in hydraulically fractured shale ecosystems both influences viral population structures and controls the release of compounds from the lysed cells that can likely be utilized by other microbes.

The GOV 2.0 data sets have provided some of the most comprehensive surveys of ocean ds-DNA viruses (117, 118). The GOV data set is composed of 145 individual data sets with 3.95 Tb of metagenomic sequencing data from virus communities, principally in epipelagic and mesopelagic ocean waters from the world's oceans. Through use of a combination of quantitative metagenomic approaches including deeper DNA sequencing, improved assembly, and binning methods [STAR methods (118)], a more robust picture of the ocean's viral populations has emerged from these studies. A total of 488,130 viral populations were identified where a viral population consists of

viral contigs greater than 5 kb where greater than 70% of the shared genes have greater than or equal to 95% average nucleotide identity across the shared contigs. Of the viral populations, 90% could not be taxonomically classified to existing viral families. The remaining 10% that could be classified were mostly dsDNA viruses belonging to the *Caudovirales* order (118). The distribution of the viral populations fell within five meta-community zones defined by Bray-Curtis dissimilarity distances that represented five ecological zones: the Artic, Antarctic, epipelagic (0–150 m BSL), mesopelagic (150–1,000 m BSL), and bathypelagic (>2,000 m BSL). The distribution of viral populations mirrored the distribution of microbial populations, suggesting that ocean physiochemical properties structure both. In support of this, temperature was found to be the major driver structuring both viral and microbial populations (118). Within each sample site within the five ecological zones there was evidence for high microdiversity, an indication that there is local adaptation for the resident viral populations. In a separate previous analysis using a smaller number of metagenomic data sets [61 data sets (36, 72)], only a limited number of viral genes were shared between the surface (phototropic zone) and subsurface (aphotic) zones. Low numbers of protein clusters (24 clusters) were shared among all surface and deeper subsurface sites, indicating that the deeper ocean viral communities are distinct from those of the upper ocean (72). Likewise, a separate analysis of 78 marine viromes resulted in the assembly of 27,346 marine viral contigs. When combined with 43 *Tara* Oceans data sets (72), an average of 66% (Standard Deviation 19) of contigs could not be assigned to known viruses and many viral contigs form uncharacterized viral lineages (119). Coutinho et al. (119) also demonstrated that there are distinct viral communities in the deeper oceans as compared to the surface waters. A more recent study of surface and deep-sea viral communities in the South China Sea also found distinct viral communities in the deep ocean as compared to the surface ocean (120). The deep-sea viral communities were less diverse than their surface viral community counterparts. Of the vOTUs, 5,859 were found exclusively in the deep-sea samples and 16,191 were found exclusively in surface waters, while 7,918 were found in both. It was speculated that some of the shared vOTUs may have originated in the surface waters that had been vertically transported into the deep ocean. Few (10 vOTUs) were found in all 10 sampling sites, and only 74 of the 29,967 vOTUs could be taxonomically matched to cultivated viruses (120). Likewise, analysis of the GOV data sets showed that most virus groups were unique to the GOV (658 candidate genera) or shared at least one GOV sequence (209 groups). Importantly, 38 of these 867 GOV clusters represented 80% of the sample diversity in two or more of the sampling stations, indicating that these virus types are abundant and widespread in the oceans. Hosts were identified at the phylum level for 392 of the 867 viral clusters. Hosts for the 38 most abundant virus types were identified; most were restricted to eight microbial host phyla present in the epipelagic zone: seven bacterial and one archaeal phylum. In general, Roux et al. (117) found that among these 38 globally abundant viral clusters, there was a positive correlation with the global richness of the host rather than the host's relative abundance.This suggested to the authors that globally distributed and abundant hosts are minimally diverse and provide fewer viral niches as compared to the more diverse host groups at lower abundance, which may provide more opportunities for virus niche differentiation (117).

A total of 243 likely AMGs were extracted from the GOV data sets, of which 95 were known AMGs (117). Four of these AMGs associated with sulfur and nitrogen metabolism (*dsrC*, *soxYZ*, *P-II*, and *amoC*) were further investigated. This subset of four AMGs included 12 gene lineage subtypes. The presence of sulfur and nitrogen metabolism AMGs suggests that ocean viruses are contributing to global nitrogen and sulfur cycling. AMG genes for sulfur reduction [dissimilatory sulfur reductases (*Dsr*-like)] and sulfur oxidation (*Sox*-like) genes were detected. *Dsr* genes are known from previous studies of viruses from deep-sea anoxic environments (77), and it is noteworthy to see them detected in viruses from epipelagic zones. AMGs associated with

nitrogen metabolism included *P-II*-related genes involved in nitrogen metabolism regulation in both archaea and bacteria. A single GOV contig contained an *amoC* gene encoding the C subunit of ammonia oxidation that was closely related to the *AmoC* gene in the cellular archaeal phylum *Thaumarchaeota*. It was noted that seven of the examined AMGs were geographically limited, mostly to southern oceans, while five were widespread across multiple epipelagic sampling sites and one was widespread across mesopelagic sampling sites. The nutrient status of the sampling site appeared to be indicative of the AMGs present. Low-nutrient sites tended to be locations where *dsrC-5* and *soxYZ* AMGs were detected, while *P-II-1* was detected in high-nutrient sites. This suggested to the authors that *DSrC-5*- and *SoxYZ*-containing viruses function to boost sulfur oxidation in infected hosts while infection with *P-II*-expressing viruses in high-nutrient sites might induce a state of nitrogen starvation that would induce alternative cellular nitrogen producing pathways favoring nucleotide production for virus replication needs (117). In a separate analysis, Coutinho et al. (119) also concluded that viral community structure adapts to the host community to better exploit the host community. This was observed in AMGs associated with purine/pyrimidine metabolism, nucleic acid biosynthesis, ATP-binding cassette transporters, and transcriptional regulators between surface and deep ocean waters. Like the finding of exchange of viral genes between geographically distinct hadal environments, there is evidence for passive transport of viral communities on oceanic currents and that host and viral communities are locally structured by environmental conditions. Other studies have similarly found that surface ocean virus communities have high local diversity and low global diversity (140). These observations support the previously proposed seed-bank model whereby viral community structures in the ocean surface waters draw genetic variation from a common and relatively limited global viral gene pool, and that local environmental conditions drive high local viral diversity (119, 140). It remains to be determined if this model will also apply to deep ocean environments.

One factor limiting our understanding of the role of viruses in ecosystems is the fact that most environments have large host and virus community structures, which limits our ability to comprehensively examine all host-virus interactions simultaneously. However, high-temperature (>75°C) acidic (pH < 4) hot springs found in Yellowstone National Park (YNP) provide an opportunity to define and monitor the near-complete host and virus communities due to the highly simplified microbial communities present in these hot springs (141, 142). These hot springs can be viewed as subsurface environments because they harbor chemolithic microbial communities that obtain energy by the oxidation of electron donors in their environments, and a component of the waters that feed these hot springs originates in the deep subsurface. These molecules can be organic (chemoorganotrophs) or inorganic (chemolithotrophs). Photosynthetic organisms are not present in these hot springs. The microbial communities in these environments are almost exclusively hyperthermophilic archaea and their viruses. Few bacteria and no eukaryotes are present. The cells in these hot springs are under chronic energy stress and are found at relatively low microbial cell densities (typically 10^3–10^5 cells/mL) and at moderate virus particle-to-cell ratios [typically 0.1–1 virus-like particle-to-cell ratio). One appeal of examining microbial communities and viruses in these hot springs as an analog to deep subsurface environments is the relative ease of sampling, allowing for both short- and long-term temporal studies of viruses and their hosts. More than 15 years of continuous monitoring of YNP hot springs using both culture-dependent and culture-independent methods (metagenomics) has shown that these hot springs are relatively simple microbial communities consisting of ~10 archaeal and 1–2 bacterial members that support a virus community of ~200 members (141–143) (**Figure 3**). Many of the archaeal viruses discovered in these hot springs have unique virion morphology and gene content, often forming new virus families (136, 144–147) (**Figure 4**). Most of the hosts are members of the *Crenarchaeota* phylum of hyperthermophilic archaea. The bacterial hosts are usually species of

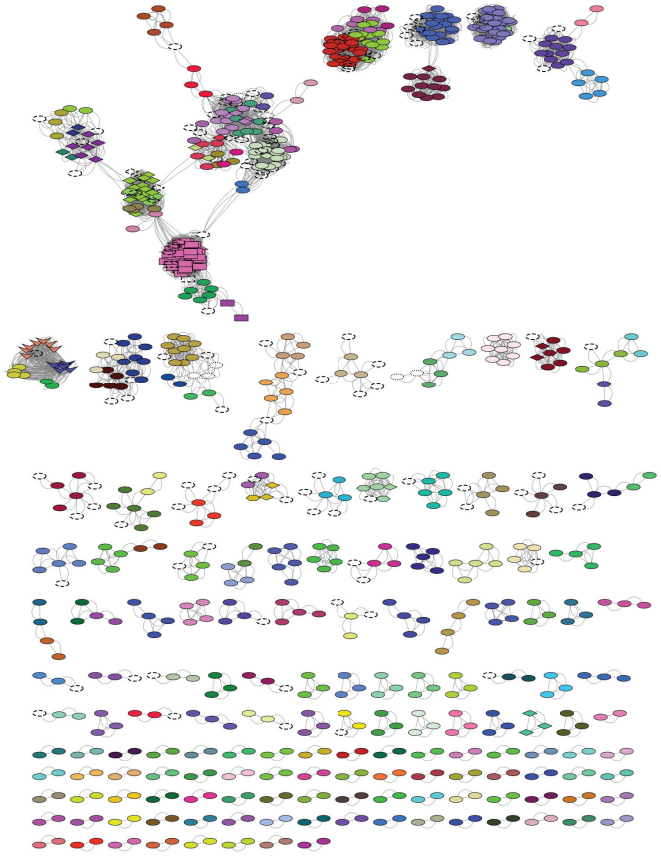

Figure 3

Network analysis of virus populations present in acidic hot springs found in Yellowstone National Park, USA. The 198 identified virus operational taxonomic units represent the majority of virus types present in these hot springs, of which 179 have no characterized representatives. Dots represent individual viral contigs and are colored by the viral cluster they are assigned to, and edges indicate genomic similarity between contigs. Different colors represent different hot springs. Analysis performed by Jacob Munson-McGee.

Hydrogenobacterium. Long-term monitoring over years has demonstrated that the overall host and viral community structures have remained relatively stable. However, there can be large changes in the relative abundance of both hosts and viruses, which often fluctuate with seasonal variations. Several new archaeal viruses have been characterized from these hot springs, most of which have been founding members of new virus families. Many of these viruses possess unusual virion morphologies and gene content. These new morphologies include spindle-shaped and two-tailed virions. Likewise, the gene content of these viruses can be unusual, containing genes with few homologs to known genes of viruses from bacterial or eukaryotic hosts. Many of these archaeal genomes, representing different taxonomic families of archaeal viruses such as *Sulfolobus* spindle-shaped virus, *Sulfolobus* turreted icosahedral virus, and SIRV, encode for suspected AMGs and genes to counter host CRISPR-Cas defense systems (94). These AMGs include suspected glycosidases and transcriptional regulators while anti-CRISPR-Cas proteins directly target and inactivate host Cas proteins (Cas 10D) or enzymatically degrade CRISPR-Cas cyclic nucleotide signaling molecules (96).

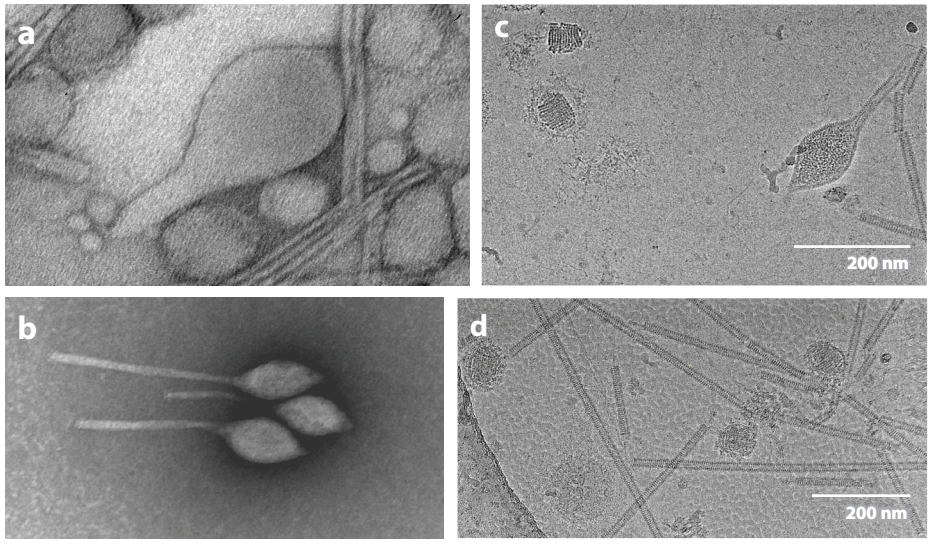

Figure 4

Negative stain (*a*,*b*) and cryo- (*c*,*d*) transmission electron microscopy of unusual virion morphologies found in environmental samples of acidic hot springs present in Yellowstone National Park, USA. Spindle, spherical, and rod-shaped morphologies are detected. Photos *a* and *b* courtesy of Sue Brumfield. Photos *c* and *d* courtesy of Colin Gauvin.

Viral metagenomic sequence analysis from several different YNP hot springs, viral network analysis, and cellular DNA and RNA sequencing have been carried out to develop a more complete understanding of the virus community structure and activity (141, 142). We (M. Young, unpublished data) defined 198 vOTUs that represent the majority of virus types present in these hot springs, of which 179 have no characterized representatives. Most of these virus groups are likely new archaeal viruses that are widely distributed among the sampled hot springs. A surprisingly high percentage of sequence reads in both the cellular DNA and RNA data sets map to the virus network, indicating that most if not all cells are interacting with and replicating viruses. In support of this finding, single-cell sequencing was combined with metagenomics to further characterize the structure of host-virus associations. Through a combination of hexanucleotide analysis, single-cell read mapping, network-based analytics, and CRISPR-based inference, it was estimated that at least 60% of cells contain at least one virus type and many cells contain two or more virus types. More than half of the detected viruses were found in more than two cellular clades, indicating a broad host range. Overall, this work revealed a network of host-virus interactions in hot spring environments.

SUMMARY AND FUTURE PERSPECTIVES

As this review has tried to highlight, over the past 10 years there have been significant advancements in our understanding of viral communities in diverse subsurface environments. There are at least four major findings we can take away from these efforts: (*a*) diverse and abundant viral communities exist in subsurface environments that rival or possibly exceed the viral diversity and abundance present in surface environments, (*b*) most of the viruses in subsurface environments are new to science and remain uncharacterized, (*c*) viral replication at higher host densities where hosts are thriving favors the lysogenic viral lifestyle, and (*d*) the detailed characterization of these

new viruses, particularly of their AMGs and antihost defense genes, will likely lead to new insights into how these viruses influence the ecology and evolution of life in subsurface environments.

Future studies of viral communities in subsurface environments will likely shift focus to a more detailed understanding of the host-virus interactions and how those interactions control the ecology and evolution of microbial communities and impact biogeochemical cycles. As the examination of viral communities in subsurface environments moves forward, it will become increasingly important to study viruses in the context of the temporal and spatial dynamics of the viral and microbial communities in which they exist. Further advances in DNA sequencing technology (148) as well as improved methods for detection of viral sequences in metagenomic data sets (149, 150) will help establish viral host ranges and expand our knowledge of viral diversity in subsurface environments. Furthermore, development of culture-based-like approaches that can maintain host and viral community structures under laboratory settings will allow for detailed, controlled, mechanistic studies of host-virus interactions. The past 20 years of subsurface environmental virology have opened our eyes to a new world of virology. The next 20 years of exploration will undoubtedly lead to new discoveries on the role viruses play in subsurface environments.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like the acknowledge Sue Brumfield, Martin Lawrence, Colin Gauvin, and Jacob Munson-McGee for their contributions of unpublished data to this review. This work was supported by the National Science Foundation (grant DEB-1342876) and The Moore Foundation (grant GBMF9195).

LITERATURE CITED

1. Bergh O, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature* 340:467–68
2. Proctor LM, Fuhrman JA. 1990. Viral mortality of marine bacteria and cyanobacteria. *Nature* 343:60–62
3. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *PNAS* 96:2192–97
4. Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–61
5. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. *PNAS* 95:6578–83
6. Bar-On YM, Phillips R, Milo R. 2018. The biomass distribution on Earth. *PNAS* 115:6506–11
7. Mushegian AR. 2020. Are there 10^{31} virus particles on Earth, or more, or fewer? *J. Bacteriol.* 202(9):e00052-20
8. Parkes RJ, Cragg BA, Bale SJ, Getliff JM, Goodman K, et al. 1994. Deep bacterial biosphere in Pacific Ocean sediments. *Nature* 371:410–13
9. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. *PNAS* 109:16213–16
10. Magnabosco C, Lin LH, Dong H, Bomberg M, Ghiorse W, et al. 2018. The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* 11:707–17
11. Orsi WD. 2018. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat. Rev. Microbiol.* 16:671–83
12. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5:801–12
13. Stevens T. 1997. Lithoautotrophy in the subsurface. *FEMS Microbiol. Rev.* 20:327–37
14. Flemming H-C, Wuertz S. 2019. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* 17:247–60

15. Cai L, Jorgensen BB, Suttle CA, He M, Cragg BA, et al. 2019. Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years. *ISME J.* 13:1857–64
16. McKay CP. 2004. What is life—and how do we search for it in other worlds? *PLOS Biol.* 2:1260–63
17. McKay CP. 2014. Requirements and limits for life in the context of exoplanets. *PNAS* 111:12628–33
18. Koonin EV, Senkevich TG, Dolja VV. 2006. The ancient Virus World and evolution of cells. *Biol. Direct* 1:29
19. Krupovic M, Dolja VV, Koonin EV. 2019. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* 17:449–58
20. Krupovic M, Dolja VV, Koonin EV. 2020. The LUCA and its complex virome. *Nat. Rev. Microbiol.* 18:661–70
21. Koonin EV, Senkevich TG, Dolja VV. 2009. Compelling reasons why viruses are relevant for the origin of cells. *Nat. Rev. Microbiol.* 7:615
22. Moelling K. 2009. Evolution of viruses and antiviral defense. *Retrovirology* 6:P59
23. Moelling K. 2012. Are viruses our oldest ancestors? *EMBO Rep.* 13:1033
24. Moelling K, Broecker F. 2019. Viruses and evolution—viruses first? A personal perspective. *Front. Microbiol.* 10:523
25. Harris HMB, Hill C. 2021. A place for viruses on the tree of life. *Front. Microbiol.* 11:3449
26. Martin W, Baross J, Kelley D, Russell MJ. 2008. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* 6:805–14
27. Forterre P. 2013. The virocell concept and environmental microbiology. *ISME J.* 7:233–36
28. Howard-Varona C, Lindback MM, Bastien GE, Solonenko N, Zayed AA, et al. 2020. Phage-specific metabolic reprogramming of virocells. *ISME J.* 14:881–95
29. Heldal M, Bratbak G. 1991. Production and decay of viruses in aquatic environments. *Mar. Ecol. Prog. Ser.* 72:205–12
30. Jiang SC, Paul JH. 1998. Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* 64:2780–87
31. Clokie MR, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. *Bacteriophage* 1:31–45
32. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol.* 8:499–500
33. Hatfull GF, Hendrix RW. 2011. Bacteriophages and their genomes. *Curr. Opin. Virol.* 1:298–303
34. Paul JH. 1999. Microbial gene transfer: an ecological perspective. *J. Mol. Microbiol. Biotechnol.* 1:45–50
35. Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, et al. 2008. Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J.* 2:575–78
36. Hurwitz BL, Hallam SJ, Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* 14:R123
37. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, et al. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *PNAS* 108:E757–64
38. Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64:69–114
39. Breitbart M. 2012. Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* 4:425–48
40. Mojica KDA, Brussaard CPD. 2014. Factors affecting virus dynamics and microbial host–virus interactions in marine environments. *FEMS Microbiol. Ecol.* 89:495–515
41. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, et al. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532:465–70
42. Wilhelm SW, Suttle CA. 1999. Viruses and nutrient cycles in the sea—Viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* 49:781–88
43. Jiao N, Herndl GJ, Hansell DA, Benner R, Kattner G, et al. 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* 8:593–99
44. Endo H, Blanc-Mathieu R, Li Y, Salazar G, Henry N, et al. 2020. Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat. Ecol. Evol.* 4:1639–49
45. Weitz JS, Li G, Gulbudak H, Cortez MH, Whitaker RJ. 2019. Viral invasion fitness across a continuum from lysis to latency. *Virus Evol.* 5:vez006

46. Paul JH. 2008. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.* 2:579–89
47. Anderson RE, Brazelton WJ, Baross JA. 2011. Is the genetic landscape of the deep subsurface biosphere affected by viruses? *Front. Microbiol.* 2:219
48. Colombo S, Arioli S, Neri E, Della Scala G, Gargari G, Mora D. 2017. Viromes as genetic reservoir for the microbial communities in aquatic environments: a focus on antimicrobial-resistance genes. *Front. Microbiol.* 8:1095
49. Labonte JM, Pachiadaki M, Fergusson E, McNichol J, Grosche A, et al. 2019. Single cell genomics-based analysis of gene content and expression of prophages in a diffuse-flow deep-sea hydrothermal system. *Front. Microbiol.* 10:1262
50. Williamson SJ, Cary SC, Williamson KE, Helton RR, Bench SR, et al. 2008. Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME J.* 2:1112–21
51. Winter C, Bouvier T, Weinbauer MG, Thingstad TF. 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. *Microbiol. Mol. Biol. Rev.* 74:42–57
52. DeWerff SJ, Bautista MA, Pauly M, Zhang C, Whitaker RJ. 2020. Killer archaea: Virus-mediated antagonism to CRISPR-immune populations results in emergent virus-host mutualism. *mBio* 11:e00404-20
53. Bondy-Denomy J, Davidson AR. 2014. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J. Microbiol.* 52:235–42
54. Silveira CB, Rohwer FL. 2016. Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms Microbiomes* 2:16010
55. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, et al. 2016. Lytic to temperate switching of viral communities. *Nature* 531:466–70
56. Danovaro R, Manini E, Dell'Anno A. 2002. Higher abundance of bacteria than of viruses in deep Mediterranean sediments. *Appl. Environ. Microbiol.* 68:1468–72
57. Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF, et al. 2016. Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.* 1:15024
58. Wei M, Xu K. 2020. New insights into the virus-to-prokaryote ratio (VPR) in marine sediments. *Front. Microbiol.* 11:1102
59. Luo E, Eppley JM, Romano AE, Mende DR, DeLong EF. 2020. Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J.* 14:1304–15
60. Roossinck MJ. 2011. The good viruses: viral mutualistic symbioses. *Nat. Rev. Microbiol.* 9:99–108
61. Roossinck MJ. 2015. Move over, bacteria! Viruses make their mark as mutualistic microbial symbionts. *J. Virol.* 89:6532–35
62. Moelling K. 2020. Viruses more friends than foes. *Electroanalysis* 32:669–73
63. Hurwitz BL, U'Ren JM. 2016. Viral metabolic reprogramming in marine ecosystems. *Curr. Opin. Microbiol.* 31:161–68
64. Chisholm SW, Bragg JG. 2008. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLOS ONE* 3(10):e3550
65. Mann NH, Cook A, Millard A, Bailey S, Clokie M. 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* 424:741
66. Breitbart M, Thompson LR, Suttle CA, Sullivan MB. 2007. Exploring the vast diversity of marine viruses. *Oceanography* 20:135–39
67. Roux S, Hawley AK, Beltran MT, Scofield M, Schwientek P, et al. 2014. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* 3:e03125
68. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3:754–66
69. Kieft K, Zhou Z, Anderson RE, Buchan A, Campbell BJ, et al. 2021. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun.* 12:3503
70. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virol. J.* 16:15
71. Hurwitz BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J.* 9:472–84

72. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, et al. 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498
73. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLOS Biol.* 4:1344–57
74. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLOS Biol.* 3:790–806
75. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. 2015. Shedding new light on viral photosynthesis. *Photosynth. Res.* 126:71–97
76. Gazitua MC, Vik DR, Roux S, Gregory AC, Bolduc B, et al. 2021. Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J.* 15:981–98
77. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. 2014. Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344:757–60
78. He T, Li H, Zhang X. 2017. Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. *mBio* 8(4):e00893-17
79. Bernheim A, Sorek R. 2020. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* 18:113–19
80. Isaev AB, Musharova OS, Severinov KV. 2021. Microbial arsenal of antiviral defenses. Part II. *Biochem. (Moscow)* 86:449–70
81. Isaev AB, Musharova OS, Severinov KV. 2021. Microbial arsenal of antiviral defenses—part I. *Biochem. (Moscow)* 86:319–37
82. Hille F, Richter H, Wong SP, Bratovic M, Ressel S, Charpentier E. 2018. The biology of CRISPR-Cas: backward and forward. *Cell* 172:1239–59
83. Li Y, Bondy-Denomy J. 2021. Anti-CRISPRs go viral: the infection biology of CRISPR-Cas inhibitors. *Cell Host Microbe* 29:704–14
84. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, et al. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13:722–36
85. Makarova KS, Koonin EV. 2015. Annotation and classification of CRISPR-Cas systems. *CRISPR* 1311:47–75
86. Borges AL, Zhang JY, Rollins MF, Osuna BA, Wiedenheft B, Bondy-Denomy J. 2018. Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity. *Cell* 174:917–25
87. Marino ND, Zhang JY, Borges AL, Sousa AA, Leon LM, et al. 2018. Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science* 362:240–42
88. Davidson AR, Lu W-T, Stanley SY, Wang J, Mejdani M, et al. 2020. Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. *Annu. Rev. Biochem.* 89:309–32
89. Wiegand T, Karambelkar S, Bondy-Denomy J, Wiedenheft B. 2020. Structures and strategies of anti-CRISPR-mediated immune suppression. *Annu. Rev. Microbiol.* 74:21–37
90. Hampton HG, Watson BNJ, Fineran PC. 2020. The arms race between bacteria and their phage foes. *Nature* 577:327–36
91. Watters KE, Fellmann C, Bai HB, Ren SM, Doudna JA. 2018. Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science* 362:236–39
92. Watters KE, Shivram H, Fellmann C, Lew RJ, McMahon B, Doudna JA. 2020. Potent CRISPR-Cas9 inhibitors from *Staphylococcus* genomes. *PNAS* 117:6531–39
93. Dong L, Guan X, Li N, Zhang F, Zhu Y, et al. 2019. An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat. Struct. Mol. Biol.* 26:308–14
94. Athukoralage JS, McMahon SA, Zhang C, Grueschow S, Graham S, et al. 2020. An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* 577:572–75
95. Landsberger M, Gandon S, Meaden S, Rollie C, Chevallereau A, et al. 2018. Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. *Cell* 174:908–16
96. He F, Bhoobalan-Chitty Y, Van LB, Kjeldsen AL, Dedola M, et al. 2018. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat. Microbiol.* 3:461–69

97. Yoshida-Takashima Y, Nunoura T, Kazama H, Noguchi T, Inoue K, et al. 2012. Spatial distribution of viruses associated with planktonic and attached microbial communities in hydrothermal environments. *Appl. Environ. Microbiol.* 78:1311–20
98. Thomas E, Anderson RE, Li V, Rogan LJ, Huber JA. 2021. Diverse viruses in deep-sea hydrothermal vent fluids have restricted dispersal across ocean basins. *mSystems* 6:e00068-21
99. Joye SB. 2020. The geology and biogeochemistry of hydrocarbon seeps. *Annu. Rev. Earth Planet. Sci.* 48:205–31
100. Li Z, Pan D, Wei G, Pi W, Zhang C, et al. 2021. Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *ISME J.* 15:2366–78
101. Yoshida M, Takaki Y, Eitoku M, Nunoura T, Takai K. 2013. Metagenomic analysis of viral communities in (hado)pelagic sediments. *PLOS ONE* 8:e57271
102. 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. *Front. Microbiol.* 10:878
103. Backstrom D, Yutin N, Jorgensen SL, Dharamshi J, Homa F, et al. 2019. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *mBio* 10:e02497-18
104. Middelboe M, Glud RN, Wenzhofer F, Oguri K, Kitazato H. 2006. Spatial distribution and activity of viruses in the deep-sea sediments of Sagami Bay, Japan. *Deep Sea Res. Part I* 53:1–13
105. Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, et al. 2008. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* 454:1084–87
106. Corinaldesi C, Dell'Anno A, Danovaro R. 2007. Viral infection plays a key role in extracellular DNA dynamics in marine anoxic systems. *Limnol. Oceanogr.* 52:508–16
107. Jian HH, Yi Y, Wang JH, Hao YL, Zhang MJ, et al. 2021. Diversity and distribution of viruses inhabiting the deepest ocean on Earth. *ISME J.* 15:3094–110
108. Manea E, Dell'Anno A, Rastelli E, Tangherlini M, Nunoura T, et al. 2019. Viral infections boost prokaryotic biomass production and organic C cycling in hadal trench sediments. *Front. Microbiol.* 10:1952
109. Zheng X, Liu W, Dai X, Zhu Y, Wang J, et al. 2021. Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation. *Environ. Microbiol.* 23:728–43
110. Zhao J, Jing HM, Wang Z, Wang L, Jian H, et al. 2022. Novel viral communities potentially assisting in carbon, nitrogen, and sulfur metabolism in the upper slope sediments of Mariana Trench. *mSystems* 7:e01358-21
111. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, et al. 2006. The marine viromes of four oceanic regions. *PLOS Biol.* 4:2121–31
112. Culley AI, Lang AS, Suttle CA. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1795–98
113. Hurwitz BL, Sullivan MB. 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLOS ONE* 8(2):e57355
114. Winter C, Garcia JAL, Weinbauer MG, DuBow MS, Herndl GJ. 2014. Comparison of deep-water viromes from the Atlantic Ocean and the Mediterranean Sea. *PLOS ONE* 9(6):e100600
115. Martinez JM, Swan BK, Wilson WH. 2014. Marine viruses, a genetic reservoir revealed by targeted viromics. *ISME J.* 8:1079–88
116. Mizuno CM, Ghai R, Saghai A, Lopez-Garcia P, Rodriguez-Valera F. 2016. Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7(4):e00805-16
117. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, et al. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–93
118. Gregory AC, Zayed AA, Conceicao-Neto N, Temperton B, Bolduc B, et al. 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177:1109–23
119. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, et al. 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* 8:15955
120. Liang Y, Wang L, Wang Z, Zhao J, Yang Q, et al. 2019. Metagenomic analysis of the diversity of DNA viruses in the surface and deep sea of the South China Sea. *Front. Microbiol.* 10:1951
121. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, et al. 2017. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* 2:e00359-16

122. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, et al. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* 3:870–80
123. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, et al. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *PNAS* 111:4274–79
124. Labbe M, Girard C, Vincent WF, Culley AI. 2020. Extreme viral partitioning in a marine-derived high arctic lake. *mSphere* 5(3):e00334-20
125. Zhong Z-P, Tian F, Roux S, Gazitua MC, Solonenko NE, et al. 2021. Glacier ice archives nearly 15,000-year-old microbes and phages. *Microbiome* 9:160
126. Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, et al. 2019. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat. Microbiol.* 4:352–61
127. Zhu HZ, Zhang ZF, Zhou N, Jiang CY, Wang BJ, et al. 2019. Diversity, distribution and co-occurrence patterns of bacterial communities in a karst cave system. *Front. Microbiol.* 10:1726
128. Hershey OS, Kallmeyer J, Wallace A, Barton MD, Barton HA. 2018. High microbial diversity despite extremely low biomass in a deep karst aquifer. *Front. Microbiol.* 9:2823
129. Tebo BM, Davis RE, Anitori RP, Conne LB, Schiffman P, Staudigel H. 2015. Microbial communities in dark oligotrophic volcanic ice cave ecosystems of Mt. Erebus, Antarctica. *Front. Microbiol.* 6:179
130. Takai K, Komatsu T, Inagaki F, Horikoshi K. 2001. Distribution of archaea in a black smoker chimney structure. *Appl. Environ. Microbiol.* 67:3618–29
131. Osbum MR, Kruger B, Masterson AL, Casar CP, Amend JP. 2019. Establishment of the Deep Mine Microbial Observatory (DeMMO), South Dakota, USA, a geochemically stable portal into the deep subsurface. *Front. Earth Sci.* 7:196
132. Anesio AM, Lutz S, Chrismas NAM, Benning LG. 2017. The microbiome of glaciers and ice sheets. *NPJ Biofilms Microbiomes* 3:10
133. Ortmann AC, Suttle CA. 2005. High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. *Deep Sea Res. Part I* 52:1515–27
134. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel *k*-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69
135. Dellas N, Snyder JC, Bolduc B, Young MJ. 2014. Archaeal viruses: diversity, replication, and structure. *Annu. Rev. Virol.* 1:399–426
136. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV. 2018. Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Virus Res.* 244:181–93
137. Danczak RE, Daly RA, Borton MA, Stegen JC, Roux S, et al. 2020. Ecological assembly processes are coordinated between bacterial and viral communities in fractured shale ecosystems. *mSystems* 5:e00098-20
138. Williamson KE, Radosevich M, Wommack KE. 2005. Abundance and diversity of viruses in six Delaware soils. *Appl. Environ. Microbiol.* 71:3119–25
139. Swanson MM, Fraser G, Daniell TJ, Torrance L, Gregory PJ, Taliansky M. 2009. Viruses in soils: morphological diversity and abundance in the rhizosphere. *Ann. Appl. Biol.* 155:51–60
140. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 13:278–84
141. Munson-McGee JH, Peng SY, Dewerff S, Stepanauskas R, Whitaker RJ, et al. 2018. A virus or more in (nearly) every cell: ubiquitous networks of virus-host interactions in extreme environments. *ISME J.* 12:1706–14
142. Bolduc B, Wirth JF, Mazurie A, Young MJ. 2015. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *ISME J.* 9:2162–77
143. Campbell KM, Kouris A, England W, Anderson RE, McCleskey RB, et al. 2017. *Sulfolobus islandicus* meta-populations in Yellowstone National Park hot springs. *Environ. Microbiol.* 19:2334–47
144. Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, et al. 2009. Structural and functional studies of archaeal viruses. *J. Biol. Chem.* 284:12599–603
145. Snyder JC, Bolduc B, Young MJ. 2015. 40 Years of archaeal virology: expanding viral diversity. *Virology* 479:369–78

146. Wagner C, Reddy V, Asturias F, Khoshouei M, Johnson JE, et al. 2017. Isolation and characterization of Metallosphaera turreted icosahedral virus, a founding member of a new family of archaeal viruses. *J. Virol.* 91:e00925-17
147. Wirth J, Young M. 2020. The intriguing world of archaeal viruses. *PLOS Pathog.* 16(8):e1008574
148. Gamaarachchi H, Samarakoon H, Jenner SP, Ferguson JM, Amos TG, et al. 2022. Fast nanopore sequencing data analysis with SLOW5. *Nat. Biotechnol.* 2022. **https://doi.org/10.1038/s41587-021-01147-4**
149. Wang W, Ren J, Tang K, Dart E, Ignacio-Espinoza JC, et al. 2020. A network-based integrated framework for predicting virus-prokaryote interactions. *NAR Genom. Bioinform.* 2:lqaa044
150. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, et al. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* 37:632–39