Check for updates

Diverse Viruses in Deep-Sea Hydrothermal Vent Fluids Have Restricted Dispersal across Ocean Basins

Elaina Thomas,[a,b] Rika E. Anderson,[a] Viola Li,[a] L. Jenni Rogan,[a] Julie A. Huber[c]

[a]Biology Department, Carleton College, Northfield, Minnesota, USA
[b]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[c]Marine Chemistry & Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

Elaina Thomas and Rika E. Anderson contributed equally to this work. Author order was determined in order of increasing seniority.

ABSTRACT In the ocean, viruses impact microbial mortality, regulate biogeochemical cycling, and alter the metabolic potential of microbial lineages. At deep-sea hydrothermal vents, abundant viruses infect a wide range of hosts among the archaea and bacteria that inhabit these dynamic habitats. However, little is known about viral diversity, host range, and biogeography across different vent ecosystems, which has important implications for how viruses manipulate microbial function and evolution. Here, we examined viral diversity, viral and host distribution, and virus-host interactions in microbial metagenomes generated from venting fluids from several vent sites within three different geochemically and geographically distinct hydrothermal systems: Piccard and Von Damm vent fields at the Mid-Cayman Rise in the Caribbean Sea, and at several vent sites within Axial Seamount in the Pacific Ocean. Analysis of viral sequences and clustered regularly interspaced short palindromic repeat (CRISPR) spacers revealed highly diverse viral assemblages and evidence of active infection. Network analysis revealed that viral host range was relatively narrow, with very few viruses infecting multiple microbial lineages. Viruses were largely endemic to individual vent sites, indicating restricted dispersal, and in some cases, viral assemblages persisted over time. Thus, we show that hydrothermal vent fluids are home to novel, diverse viral assemblages that are highly localized to specific regions and taxa.

IMPORTANCE Viruses play important roles in manipulating microbial communities and their evolution in the ocean, yet not much is known about viruses in deep-sea hydrothermal vents. However, viral ecology and evolution are of particular interest in hydrothermal vent habitats because of their unique nature: previous studies have indicated that most viruses in hydrothermal vents are temperate rather than lytic, and it has been established that rates of horizontal gene transfer (HGT) are particularly high among thermophilic vent microbes, and viruses are common vectors for HGT. If viruses have broad host range or are widespread across vent sites, they have increased potential to act as gene-sharing "highways" between vent sites. By examining viral diversity, distribution, and infection networks across disparate vent sites, this study provides the opportunity to better characterize and constrain the viral impact on hydrothermal vent microbial communities. We show that viruses in hydrothermal vents are diverse and apparently active, but most have restricted host range and are not widely distributed among vent sites. Thus, the impacts of viral infection are likely to be highly localized and constrained to specific taxa in these habitats.

KEYWORDS hydrothermal vent, viral ecology

Deep-sea hydrothermal vents are regions on the seafloor where high-temperature hydrothermal vent fluid is created from water-rock reactions deep within the oceanic crust, mixing with seawater beneath and at the seafloor to create a dynamic,

gradient-dominated habitat that supports diverse microbial communities. These low-temperature diffuse vent fluids are hot spots of primary production in the deep ocean, dominated by chemolithoautotrophic bacteria and archaea carrying out a variety of metabolisms utilizing hydrogen, sulfur compounds, nitrate, and methane (1–7). An important, though understudied, driver of microbial diversity and evolution in hydrothermal systems is viruses. Viruses are major sources of microbial mortality in marine systems and play key roles in mediating biogeochemical cycling and shaping microbial community structure (8–11). At deep-sea hydrothermal vents, viruses are abundant (12) and infect a wide range of microbial hosts (13). Temperate viruses, which often integrate into the genomes of their microbial hosts, are particularly abundant in hydrothermal fluids compared to other marine habitats (14, 15). The high abundance of temperate viruses suggests that there are unique attributes to the vent environment that influence viral infection strategies. Viruses can "metabolically reprogram" their microbial hosts via the introduction and expression of auxiliary metabolic genes (AMGs) (16), which has profound impacts on the ecology and evolution of microbial populations. Like horizontally transferred genes, integrated prophage with AMGs can alter the functional potential of a given organism, allowing it to adapt to changing conditions or expand to new ecological niches. Viruses in hydrothermal habitats have been found to encode AMGs (17–19), and thus can manipulate the metabolic potential of microbial populations in hydrothermal vents.

Viruses also function as vectors of horizontal gene transfer via transduction. Previous work has suggested that horizontal gene transfer is particularly prevalent among microbes inhabiting high-temperature environments (20–22). Transposases, which catalyze the movement of mobile genetic elements within and between genomes (23, 24), are abundant in hydrothermal vent sites (18, 25). Given the abundance of viruses and the inferred high rates of transduction at deep-sea vents, it is likely that viruses are an important vector for horizontal gene transfer in these dynamic systems. Therefore, a clear understanding of which viruses infect which hosts, and what genes those viruses carry, can provide insight into highways of gene sharing in the deep sea.

One large gap in our understanding of ecoevolutionary dynamics in deep-sea hydrothermal vents is how broadly viruses are distributed within and between vent systems and how viral interactions shift across hydrothermal vent types. Previous work has shown that microbial communities exhibit high endemism locally (4, 7, 26, 27), but some cosmopolitan species are distributed globally (28–30). However, less is known about the geographic distribution of viral populations. If viruses exhibit a restricted distribution, this would limit their role as vectors of gene flow across and between hydrothermal systems. Microbial populations and their viruses are also limited by environmental conditions: hydrothermal systems hosted in basalt rocks are characterized by metal-enriched, low pH fluids up to 400°C (31). In contrast, hydrothermal systems hosted in peridotite are influenced by serpentinization and feature organic carbon-enriched, high pH fluids with slightly lower temperature (32). Microbial populations in basalt-hosted and peridotite-hosted vents exhibit distinct patterns of genomic variation (33), suggesting that microbes are subject to different selection pressures depending on the vent type, but we know little about the role viruses play in driving those differences, nor how viral diversity varies across hydrothermal systems.

To gain further insight into viral diversity, distribution, and host range across hydrothermal systems, we compared viral sequences recovered from microbial metagenomes collected from two hydrothermal regions in two distinct ocean basins: Axial Seamount, a submarine volcano located on the Juan de Fuca Ridge in the Pacific Ocean at ~1,520-m depth, and the Mid-Cayman Rise, an ultraslow spreading ridge in the Caribbean Sea. Axial Seamount is a basalt-hosted, magma-driven system with fluids that are low in pH and high in carbon dioxide and hydrogen sulfide (34). Microbial communities at Axial are spatially restricted but temporally stable at individual vent sites over time (4, 27). In contrast, the Mid-Cayman Rise hosts two geologically and geochemically distinct hydrothermal vent fields in close proximity to each other:

Piccard hydrothermal field, located in basalt rocks along the ridge axis, is the deepest hydrothermal vent field discovered thus far (~4,950-m depth), and is characterized by fluids that are acidic with high hydrogen and hydrogen sulfide content, whereas the Von Damm vent field, located approximately 20 km away on a nearby massif at ~2,350-m depth, is influenced by serpentinization and is characterized by fluids that are high in hydrogen, methane, and small carbon compounds (35–38). Previous work found distinct community composition but similar functional potential in the microbial communities at Piccard and Von Damm (7, 27, 33, 39, 40). Here, we compare viral sequences in venting fluids from Von Damm vent field, Piccard vent field, and Axial Seamount in a comparative survey of viral diversity and gene content across geographically and geochemically distinct hydrothermal vent sites. By examining viruses from the microbial fraction (0.22 μm), we are closely examining the relationship between prophage, actively infecting viruses, and their hosts. We examined viral sequences identified in the metagenomes as well as clustered regularly interspaced short palindromic repeat (CRISPR) sequences, which provide context for historical viral infections within each of the communities, and provide a means to match viruses with putative hosts. We show that viral populations have high diversity but restricted distribution and host range across hydrothermal systems, limiting the capacity of viruses to act as vectors of gene flow between disparate hosts and vent locations.

RESULTS

Identification of putative viral contigs. We used VirFinder to identify putative viral contigs because it is a k-mer frequency-based method that has higher potential to identify novel viruses (41). However, this method allows for the possibility of false-positive results. VirFinder assigns scores between 0 and 1 to indicate the likelihood that a contig is viral, with higher values reflecting a higher likelihood that the sequence is viral. Almost all VirFinder sequences from Axial Seamount (99.96%) and Piccard (98.54%) and Von Damm (98.96%) vent fields were assigned scores greater than 0.7. In addition, 47.35% of Axial Seamount, 44.21% of Piccard, and 37.92% of Von Damm VirFinder sequences had scores greater than 0.9. The mean scores of VirFinder sequences were 0.89, 0.88, and 0.87 for Axial Seamount, Piccard, and Von Damm, respectively. Figure S1a in the supplemental material shows the distribution of scores for the contigs identified with VirFinder at the Axial Seamount, Piccard, and Von Damm vent fields. Similarly, the majority of contigs identified as putatively viral were short, with shorter average contigs at Axial Seamount (625 bp) than those at Piccard (2,172 bp) or Von Damm (2,037 bp). The VirFinder score did not increase with sequence length (Fig. S1b).

Taxonomy of viruses and hosts. All analyses were carried out from diffuse flow fluids sampled directly from the vent orifice, as well as vent plume waters, which were sampled up to 100 m above the vent orifice and had much lower temperatures (see Table S1 in the supplemental material). It is possible to examine viruses in 0.22-μm-filtered fluids because this fraction includes integrated prophage, lytic infections in progress, and free viral particles captured on filters. However, our analysis misses free viral particles that were not retained on the filters, and we are only capturing the viral diversity and variation across sites based on the viral sequences identified in the metagenomes.

The viral sequences identified at Axial Seamount, Von Damm, and Piccard vents represent 28 families of viruses (Fig. 1). Thirteen of the viral families were present at all three vent fields. Fourteen viral families were present at both of the Mid-Cayman Rise vent fields, Von Damm and Piccard, while Axial Seamount shared 13 viral families with each of the Mid-Cayman Rise vent fields. We observed a distinct difference in viral taxonomic groups between the vent fields: whereas the *Myoviridae* viral family had the highest relative abundance in 12 of the 16 Axial Seamount samples, the *Myoviridae* were in much lower abundance at Von Damm and Piccard, which had higher abundances of reads matching the *Podoviridae* family of viruses and the *Guttaviridae*.

Relative abundance of viral and microbial reads. In order to examine the distribution and abundance of viruses across individual vent sites, vent fields, and regions,
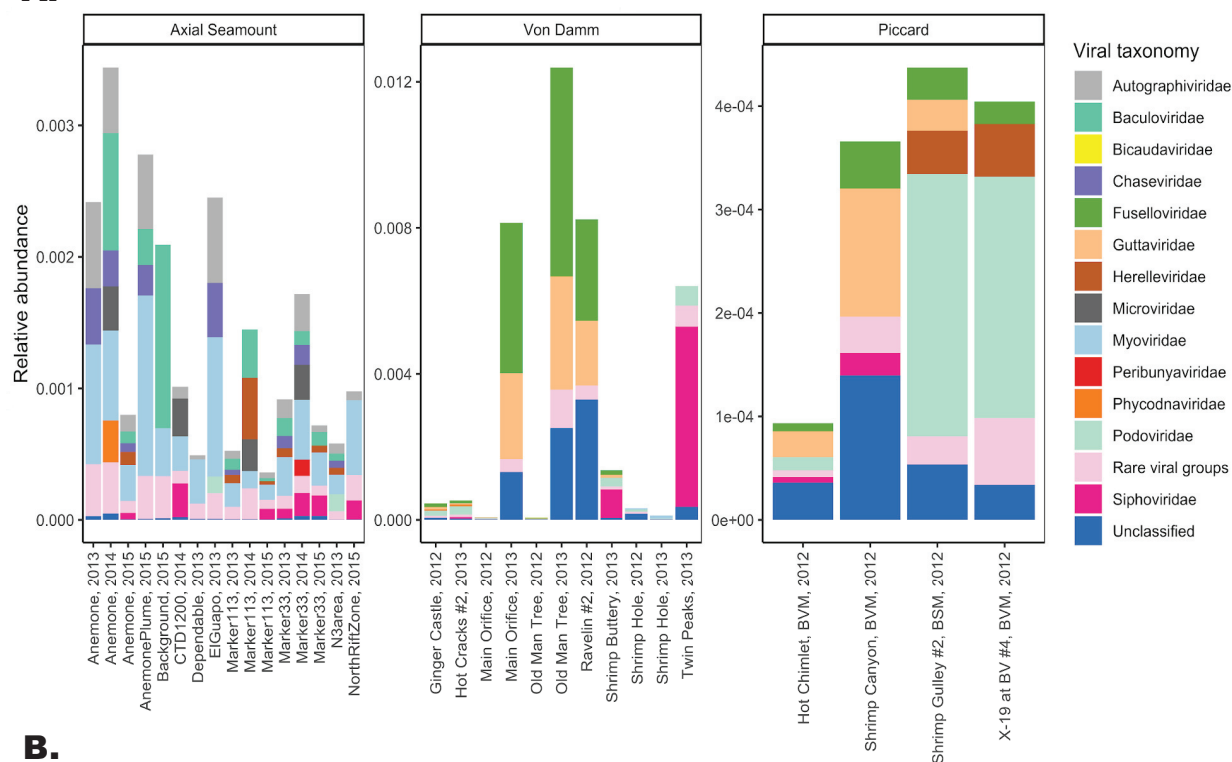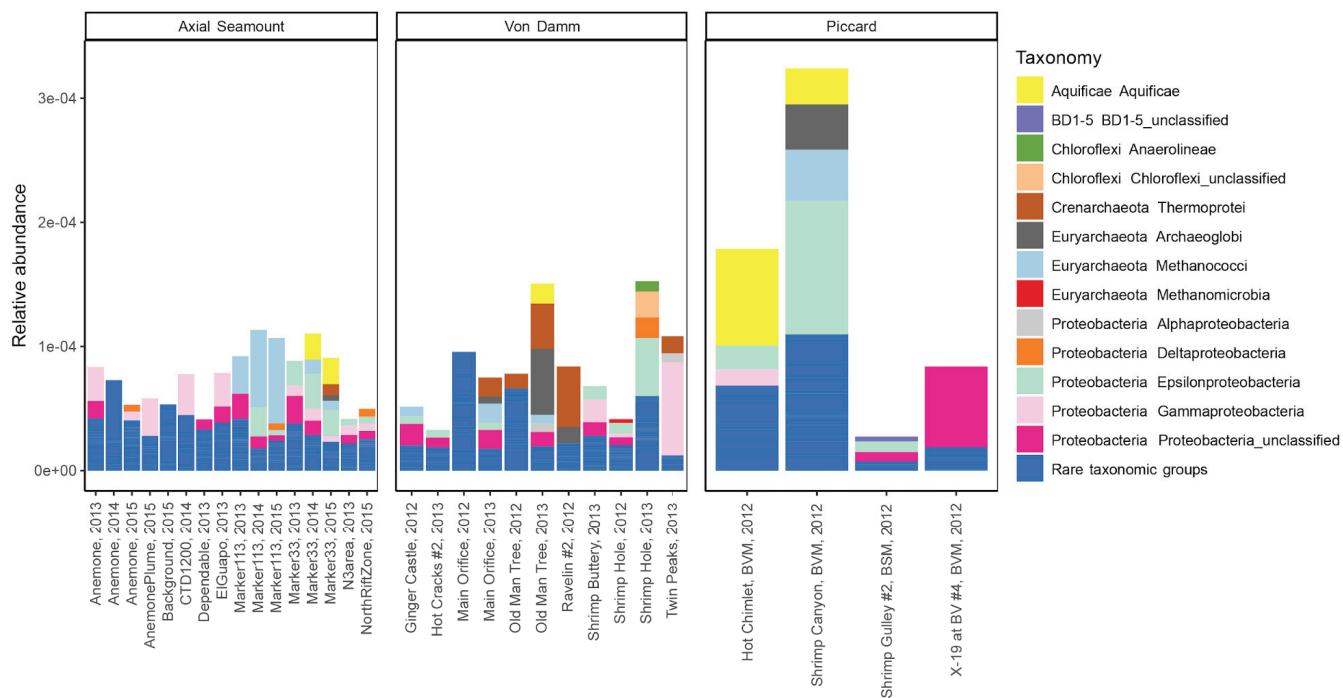
FIG 1 Taxonomy of viral (A) and microbial (B) reads in the Von Damm, Piccard, and Axial Seamount vent fields. In panel A, the y axis represents the number of reads that mapped to VirFinder contigs of each viral taxonomy normalized by the number of reads in each sample. Rare viral groups represent viral taxonomies that comprised less than 5% of reads that mapped to viral contigs in a given sample. In panel B, the y axis represents the number of reads that mapped to 16S rRNA gene sequences of each host normalized by the number of reads in each sample. Reads that mapped to 16S gene rRNA sequences that were unclassified or classified as eukaryotes were excluded. Rare taxonomic groups represent taxonomies that comprised less than 1% of reads that mapped to 16S rRNA gene sequences in a given sample.
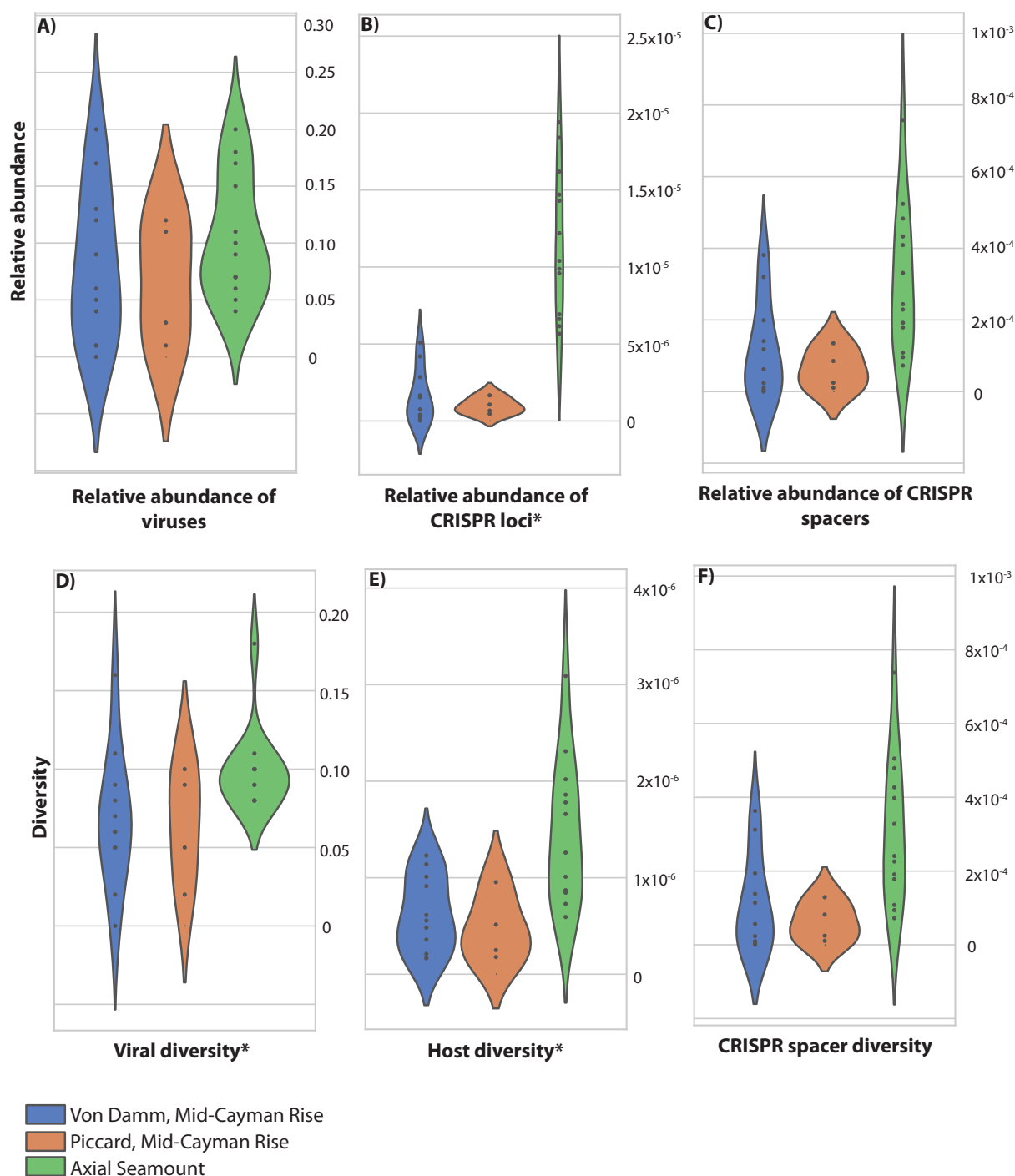
FIG 2 Abundance and diversity of viral sequences, CRISPR spacers, CRISPR loci, and microbes in diffuse flow samples from the Von Damm vent field (blue), Piccard vent field (orange), and Axial Seamount (green). Relative abundance reflects the relative number of reads mapping to viral contigs within each region, normalized by read abundance (see Materials and Methods). Diversity reflects the number of clusters per read, normalized by read abundance (see Materials and Methods). Note that the *y*-axis values are different for each variable. Values for individual samples are indicated with black dots. Violins represent the kernel density estimation of the underlying data distribution. Variables with significant differences between Von Damm, Piccard, and Axial Seamount are indicated with asterisks (based on analysis of variance [ANOVA] test). For these comparisons, samples from background seawater and plumes were excluded.

we quantified the relative abundance of viral sequences and CRISPR loci and spacers in each of the metagenomes (Fig. 2, samples from background seawater and plumes are excluded; Table S2). CRISPR loci are a microbial immune system found in archaeal and bacterial genomes, consisting of direct repeats interspersed by "spacers," which match

foreign DNA (predominantly viruses, but also including plasmids and other forms of foreign DNA) that the cell has been exposed to previously. The relative abundance of CRISPR loci serves as an indication of how many microbial lineages use CRISPR as a mechanism for viral immunity. It is also important to note that CRISPR loci vary in number across microbial genomes, and we did not distinguish between CRISPR loci with the same direct repeat type. Therefore, our measure of CRISPR relative abundance is not a direct proxy for the abundance of viruses. Instead, it gives an indication of how commonly CRISPR loci are used as an antiviral mechanism within the community. Moreover, CRISPRs serve as a record of past infections, and thus while viral diversity reflects the diversity of viral particles sequenced at the time of sampling, CRISPR spacer diversity reflects the diversity of past viral infections.

Our results indicate that the relative abundance of viral sequences (Fig. 2A) was similar both across and within each of the hydrothermal vent regions we studied ($P = 0.42$, t test). There was a higher relative abundance of CRISPR loci in samples collected at Axial Seamount compared to both Piccard and Von Damm vent fields (Fig. 2B; $P = 1.49e−05$, t test). However, there was no difference in the relative abundance of spacers within CRISPR loci per read (Fig. 2C; $P = 0.10$, t test). Within the Mid-Cayman Rise, we compared samples from mafic-hosted (Piccard) versus ultramafic-hosted (Von Damm) hydrothermal systems. We did not observe significant differences in the relative abundance of viral sequences, CRISPR loci, or CRISPR spacers between Piccard and Von Damm (viral sequences, $P = 0.66$, t test; CRISPR loci, $P = 0.40$, t test; spacers, $P = 0.37$, t test) (Fig. 2A to C). Similar results emerged from comparisons among vent fields within Axial Seamount: we did not observe significant differences in the relative abundance of viral sequences, CRISPR loci, or CRISPR spacers between vent fields at Axial Seamount (viral sequences, $P = 0.17$, t test; CRISPR loci, $P = 0.41$, t test; spacers, $P = 0.11$, t test). Finally, within Axial Seamount, we also compared the relative abundance of viruses in samples taken from plume and diffuse flow hydrothermal fluid at Anemone vent. The Anemone diffuse flow samples had a higher relative abundance of CRISPR spacers and CRISPR loci than the Anemone plume sample, which was sampled above the vent. However, the relative abundance of viral sequences did not differ between the Anemone plume and diffuse flow samples (Table S2).

Diversity of viral assemblage and microbial community. Overall, viral diversity analyses revealed that the viral assemblages within these vents had high richness and were not dominated by specific viral strains. With the exception of a few samples, the rarefaction curves for the viruses and microbes did not reach saturation (Fig. S2), indicating that the viral operational taxonomic units (vOTUs) recovered in this study did not capture the total diversity in the samples. In each sample, there were no dominant viral or spacer clusters, and each of the viral and spacer clusters present was relatively rare (Fig. S3 and S4). Moreover, the viral and spacer clusters with the highest coverage did not correspond across samples: in both Piccard and Axial Seamount, one of the most common spacer clusters matched with one of the most common vOTUs by BLAST, while none of the most common spacer clusters and vOTUs aligned in Von Damm. Thus, there were no dominant viral sequences that were consistently found across all samples.

We observed a more diverse viral assemblage at Axial Seamount compared to the Mid-Cayman Rise. We observed higher richness of both viruses (Fig. 2D; $P = 0.015$, t test) and their microbial hosts (Fig. 2E; $P = 0.0056$, t test) in samples from vents at Axial Seamount compared to samples from Von Damm and Piccard vent fields. However, we did not observe meaningful correlations between vOTU, spacer cluster, or host diversity. Moreover, we did not observe significant differences in either viral or host diversity between the Piccard and Von Damm vent fields at the Mid-Cayman Rise (viral diversity, $P = 0.053$, t test; host diversity, $P = 0.66$, t test). We also examined the relative diversity of CRISPR spacers, which provides an indication of past viral infections rather than the current

TABLE 1 Distribution of vOTUs, CRISPR spacer clusters, and microbial hosts[a]

Vent field and variable	% in more than one vent site	% in all vent sites in region	% found in all three regions (Axial, Von Damm, and Piccard)
Von Damm			
vOTUs	13.71	0	
Spacer clusters	5.1	0	
16S rRNA host	46.15	1.44	
Piccard			
vOTUs	10.78	0	
Spacer clusters	9.13	0	
16S rRNA host	46.09	11.72	
Axial Seamount			
vOTUs	8.88	0.05	
Spacer clusters	1.01	0	
16S rRNA host	54.55	2.36	
Viral OTUs			0.02
Spacer clusters			0
16S rRNA host			16.62

[a]Distribution of vOTUs, CRISPR spacer clusters, and microbial hosts (classified at the lowest taxonomic level available based on 16S rRNA genes) across vent sites and vent fields at Piccard, Von Damm, and Axial Seamount. The background and CTD1200 samples were excluded. Reads that mapped to 16S rRNA sequences that were classified as eukaryotes were excluded.

virus pool. We did not observe a significant difference in the diversity of CRISPR spacers between samples from Piccard, Von Damm, and Axial Seamount (Fig. 2F, $P = 0.093$, t test).

We observed significant differences in viral diversity between vent sites within Axial Seamount ($P = 0.022$, t test). However, no significant differences emerged in terms of the diversity of microbial hosts and CRISPR spacers (host diversity, $P = 0.74$, t test; spacer diversity, $P = 0.11$, t test). At Anemone vent, the diffuse flow samples had higher CRISPR spacer and microbial diversity than the plume sample, but viral diversity did not differ between these samples (Table S2).

Viral distribution across vent sites. In order to characterize viral distribution and community similarity across hydrothermal vent fluids, we evaluated the extent to which viral sequences and CRISPR spacers were distributed across samples and then compared these results to the host microbial community. As before, we clustered sequences based on similarity to compare across all diffuse flow samples, excluding background seawater and plume samples. On the whole, viral sequences and CRISPR spacers had fairly limited distributions. Only a few vOTUs (0.02%) were present at all vent sites (Von Damm, Piccard, and Axial Seamount), and no CRISPR spacer clusters were present in all three regions (Table 1). The most cosmopolitan viruses and CRISPR spacers did not match each other: at Von Damm, 2 of the 100 most widely distributed vOTUs matched with 3 of the 100 most widely distributed CRISPR spacer clusters according to BLAST. At Piccard, one of the most widely distributed vOTUs aligned with one of the most widely distributed spacer clusters. At Axial Seamount, 1 of the 100 most widely distributed vOTUs matched with 2 of the 100 most widely distributed CRISPR spacer clusters. In contrast, microbial taxa were much more cosmopolitan, with ~17% of taxa shared between Von Damm, Piccard, and Axial Seamount according to 16S rRNA gene classification at the lowest taxonomic level available (Table 1). Viral and CRISPR spacer clusters were shared more widely among vent sites within Von Damm and Piccard compared to Axial Seamount, but microbial lineages were shared more widely among vent sites at Axial Seamount (Table 1).

We created hierarchical dendrograms to assess the similarity of samples based on their viral content. Viral assemblages in samples from the Mid-Cayman Rise and Axial Seamount grouped separately (Fig. 3A). Within Axial Seamount, samples taken in successive years from the same vent tended to have similar viral assemblage compositions

FIG 3 Hierarchical clustering of Von Damm (black), Piccard (blue), and Axial Seamount (red) diffuse flow samples based on viral assemblage (A), microbial host composition (B), and CRISPR spacer

(Continued on next page)

(Fig. 3A). In contrast, at the Mid-Cayman Rise, samples taken from the same site in two different years did not cluster together, and we observed weak clustering of samples by location (Fig. 3A). Samples from Piccard vent field and Von Damm vent field did not cluster separately. Grouping of microbial communities based on classification of 16S rRNA reads in the metagenomes showed similar patterns to the viral assemblages. Samples collected from vents at Axial and the Mid-Cayman Rise grouped separately from each other, with only some clustering of samples at Piccard and Von Damm. Similarly, we observed stronger temporal clustering among samples at Axial Seamount than at Von Damm (Fig. 3B).

In contrast to the viral and microbial assemblages, hierarchical dendrograms based on CRISPR spacer compositions showed little clustering by location (Fig. 3C). Based on spacer assemblages, samples from either Axial Seamount or the Mid-Cayman Rise did not cluster together, and samples taken from the same vent sites in different years did not cluster together. Very few CRISPR spacer clusters were found at multiple vent sites (Table 1).

Networks of viral infection. Networks of viral infection generated from viral sequences, CRISPRs, and metagenome assembled genomes (MAGs) were used to examine the distribution and host specificity of vent viruses. These virus networks show connections between CRISPR spacers, which represent a record of previous viral infection in the host, and viral sequences recovered from the metagenome, which represent viral sequences present in the community at the time of sampling. Virus-host networks made for Axial Seamount (Fig. 4A) and Von Damm (Fig. 4B) and Piccard (Fig. 4C) vent fields indicate that vent viruses are restricted in terms of both host range and geographic distribution. We did not observe any virus-host connections shared between Piccard, Von Damm, or Axial Seamount. Within Axial Seamount, there were 89 vOTUs linked to 18 MAGs, and 105 connections between distinct pairs of vOTUs and MAGs (Fig. 4A). We did not observe any connections between MAGs and vOTUs in any of the nondiffuse flow samples (background, CTD1200, and the Anemone plume sample). The number of viral connections appeared to be related to, but was not significantly correlated with, the microbial hosts' relative abundance (Table S3): for example, a *Clostridia* MAG in the 2015 North Rift Zone sample (Clostridia_35) had high normalized coverage and was linked to nine vOTUs. We observed a high number of viruses shared among *Aquificae* MAGs sampled from the Anemone vent in 2013 and 2014 (Fig. 4A). Finally, we observed only a single vOTU that was linked to MAGs of different taxonomic classes from different vent sites, with a vOTU connected to an *Ignavibacteria* MAG (Ignavibacteria_15) in the 2014 Marker 33 sample and an *Aquificae* MAG (Aquificae_43) in the 2013 Anemone sample (Fig. 4A).

The virus infection networks reflected the geographic separation observed in the cluster dendrograms and also revealed fairly narrow virus-host ranges for the viral sequences observed in our data sets. Within the Von Damm vent field in the Mid-Cayman Rise, there were 62 vOTUs linked to nine MAGs, with 66 connections between distinct pairs of vOTUs and MAGs (Fig. 4B). The size of the viral infection network was similar in Piccard, where 48 vOTUs were linked to seven MAGs and there were 64 connections between distinct pairs of vOTUs and MAGs (Fig. 4C). All of the vOTUs in the

FIG 3 Legend (Continued)

composition (C). For viral composition, the number of reads in each sample that mapped to each vOTU was calculated. For host composition, the number of reads that mapped to 16S rRNA gene sequences of each host (classified at the lowest taxonomic level available) in each sample was calculated. Reads that mapped to 16S rRNA gene sequences that were either unclassified or classified as eukaryotes were excluded. For CRISPR composition, the number of CRISPR spacers in each sample in each spacer cluster was calculated. For all samples, the zero counts were replaced with estimates, and a centered log ratio (clr) transformation was applied. The y axes indicate distance between samples as calculated by the ward.D2 method based on the transformed counts. Gray values at each edge are P values (as a percentage). Red values are approximately unbiased (AU) P values, computed by multiscale bootstrap resampling, and green values are bootstrap probability (BP) P values, which are computed by normal bootstrap resampling. All uncertainty metrics in the dendrogram were calculated using the R package pvclust.
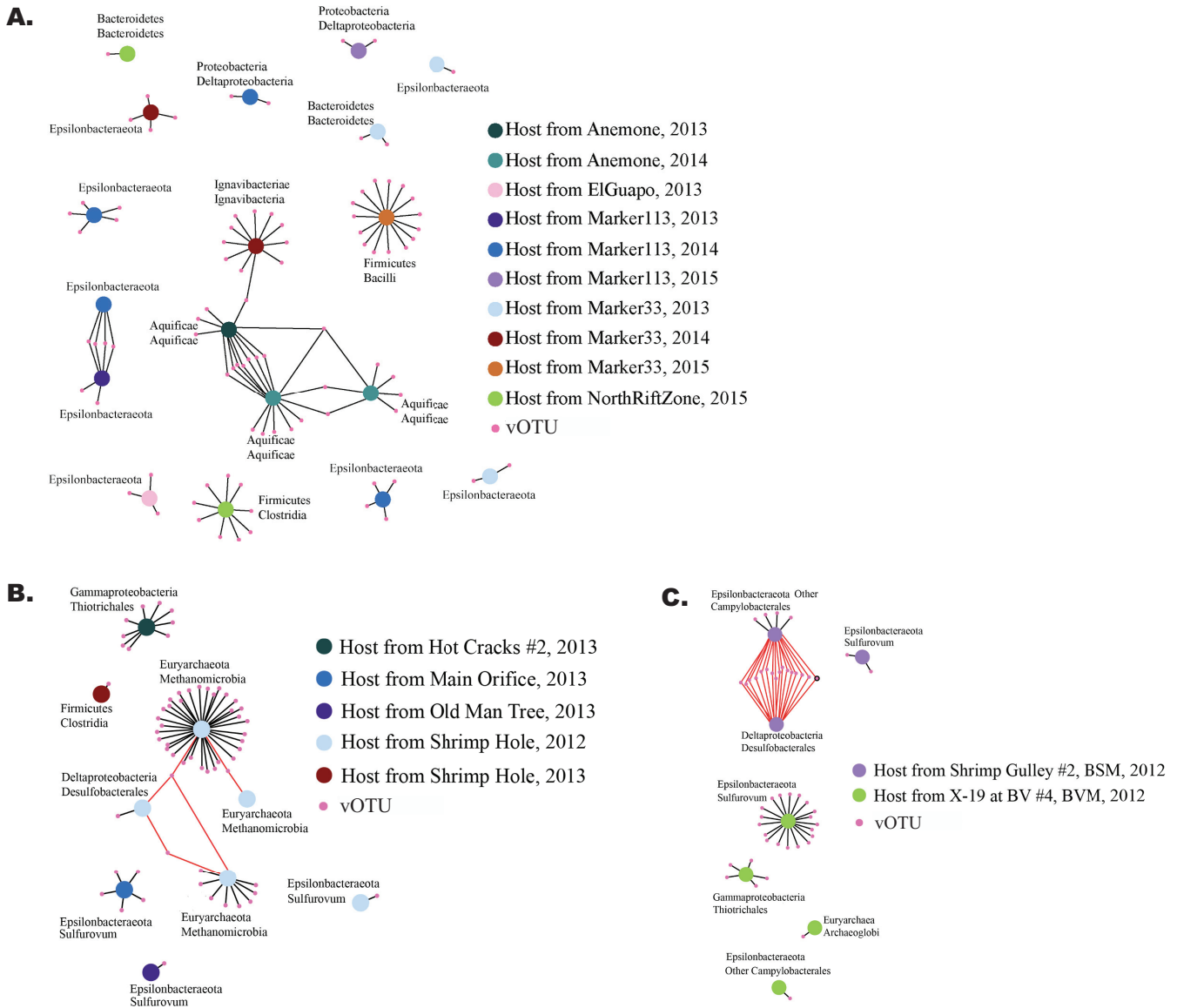
FIG 4 Infection network showing the links between viral clusters (vOTUs) and MAGs at Axial Seamount (A), Von Damm (B), and Piccard (C). vOTUs were linked to MAGs via the spacers and direct repeats in CRISPR loci. Edges are colored red when a vOTU is linked to multiple MAGs in the same sample through the same direct repeat type. Due to the Crass algorithm, when the same direct repeat type is found in multiple MAGs in the same sample, it cannot be determined which MAG the spacers associated with the direct repeat type came from. Therefore, the red edges are links between vOTUs and MAGs that could not be definitely proven. vOTUs with high relative abundance (top six most abundant) in at least one sample are outlined in black.

Von Damm and Piccard networks represented relatively rare viruses with the exception of one. This vOTU was present in seven of the Von Damm vents and all of the Piccard vents, was among the top six most abundant viruses in both the X-19 and Shrimp Gully 2 vents within Piccard, and was linked to a *Campylobacterales* and a *Desulfobacterales* MAG, both from Shrimp Gully 2. This was the only vOTU with high relative abundance in at least one sample that was present in the Axial Seamount, Von Damm, and Piccard viral infection networks (Fig. 4). Some MAGs had more viral connections than others: for example, a *Methanomicrobia* MAG (Methanomicrobia_41) from the 2012 sample from Shrimp Hole within Von Damm was linked to 35 vOTUs (Fig. 4B). However, this is not a direct indication of the number of different viruses infecting a specific strain because some MAGs had more CRISPR spacers than others, increasing the possibility of finding a viral connection. As with Axial Seamount, there were instances of highly abundant MAGs with many viral connections (Table S3): a *Sulfurovum* MAG (Sulfurovum_99) in the 2012

sample from the X-19 vent within Piccard had the highest normalized coverage across Von Damm and Piccard and was linked to 18 vOTUs, making it the MAG with the third highest number of viral connections within either network. Another *Sulfurovum* MAG (Sulfurovum_37) had a high relative abundance in the 2012 sample from Shrimp Gulley #2 within Piccard and had two viral connections. Within Von Damm, Piccard, and Axial Seamount, we observed several cases in which viruses were connected to multiple microbial hosts within the network (Fig. 4). However, in some cases, the MAGs were linked to these shared vOTUs through the same CRISPR direct repeat type and were in the same sample, suggesting that these specific connections may have been due to matching CRISPR direct repeat types rather than true cross-infection. These have been indicated in Fig. 4 (red lines).

Finally, an additional analysis of virus-host interaction was conducted by identifying prophages and auxiliary metabolic genes (AMGs) in MAGs. However, the fragmented nature of the assemblies and the absence of clear viral hallmark genes next to putative AMGs did not allow us to reach strong conclusions regarding the prevalence of lysogeny and AMGs in these hydrothermal vent samples. These analyses are described in Text S1 in the supplemental material.

DISCUSSION

Viruses are important drivers of microbial mortality, ecology, and evolution in the ocean, but studies of their distribution and impact in the deep sea are lacking. Our results indicate that viral diversity is high in venting fluids from hydrothermal systems, and the viruses we analyzed have restricted geographic distributions and host ranges. This implies that viruses do not spread widely between vent sites and that the viral role in mediating horizontal gene transfer across taxa and between vent sites is relatively restricted. However, the high abundance and diversity of viral sequences as well as the large number of virus-host CRISPR connections implies rapid viral mutation and ongoing viral infection, indicating active and ongoing interactions between viruses and their microbial hosts in venting fluids from the hydrothermal vents examined here.

Viral assemblages are taxonomically distinct and spatially restricted across vent sites. Previous studies of viruses in marine systems have observed viral populations to be commonly found across multiple samples (42), whereas others have found that most viruses are biogeographically restricted, with only a few cosmopolitan groups (43, 44). In hydrothermal vents at Von Damm and Piccard vent fields and Axial Seamount, we found high viral diversity with a limited distribution, potentially indicating rapid diversification in vents. Most vOTUs we observed were found only at individual vent sites, with very few cosmopolitan viruses. Previous work examining microbial distribution in hydrothermal systems through fine-scale 16S rRNA gene analyses indicates that while some microbial lineages are endemic to individual vent sites, others are widespread across vent fields (28, 40). Our work revealed that viral biogeographic patterns roughly corresponded to those of their microbial hosts. Taxonomic analysis revealed that there are distinct differences in viral taxonomy between vent sites—most notably, we observed a high abundance of *Myoviridae* at Axial Seamount that was not found at either Piccard or Von Damm. These observations indicate either restricted gene flow or strong environmental selection for specific groups within each site. Our analysis of viral sequence similarity using hierarchical clustering revealed similar results: there were some similarities between the viral assemblages in vents at Piccard and Von Damm, which are located approximately 20 km apart in the Caribbean Sea, but the viral assemblages from Axial Seamount in the Pacific were very different from those at Piccard and Von Damm, despite the fact that vents at Piccard and Axial are both hosted in basalt and have metal-enriched, low-pH fluids. Previous work on microbial community similarity and distribution in hydrothermal systems has shown that vents in close proximity are often more similar to one another in terms of microbial community structure than to geographically distant vents (4, 27, 40, 45). This may result from subseafloor plumbing that restricts fluid flux between sites, creating "islands" of microbial diversity that are distinct from one vent site to the next (26).

Here, we show that these barriers to dispersal apply to viruses as well and that viruses may be even more spatially restricted than their microbial hosts. Microbial lineages that spread between vent sites may face infection from novel viral strains not found in other vent sites. These endemic viral populations thus further shape distinct microbial community structure at individual vent sites.

Although the viral assemblages at hydrothermal vents had high diversity and restricted dispersal between vent sites and vent fields, our results show that the viral assemblages in vents persist over time, particularly at Axial Seamount where the same vents were sampled over a 3-year period. Viral assemblages from samples from the same vent sites at Axial across 3 years clustered together in the hierarchical dendrograms (Fig. 3A), and our viral infection networks revealed that a number of viruses at Axial Seamount were linked to specific microbial hosts at the same vent over multiple years (Fig. 4A). These patterns match those observed in the microbial communities (Fig. 3B) and are consistent with previous work showing spatially restricted but temporally stable microbial communities over time at Axial Seamount (4, 27). Our work extends this to the viral world at Axial Seamount, indicating that viral assemblages follow the same temporal patterns as their microbial hosts and that virus-host relationships persist over time. However, temporal stability in the viral and microbial community was generally not preserved at the Von Damm vent field for sites sampled in both 2012 and 2013. There were fewer sites and time points sampled at Von Damm compared to Axial Seamount, so it is unclear whether this results from a true biological signal or insufficient data.

Vents host diverse and active assemblages of viruses with restricted host range. The viral infection networks show diverse microbial lineages across several samples infected by many different viruses at the time of sampling. Although individual microbial lineages could be targeted by multiple viral strains, the viruses we identified in these hydrothermal habitats had fairly restricted host ranges, infecting specific individual microbial strains. The only clear example of viral infection across microbial taxa emerged at Axial Seamount, where viral sequences associated with *Ignavibacteria* were also linked to *Aquificales* (Fig. 4A). All other examples of viral infection across microbial taxa may either be true examples of virus-host cross-infection or may instead result from shared direct repeats. The limited host range of viruses from hydrothermal systems is consistent with previous work indicating that viruses tend to be host specific (44), and we found very little evidence for generalist viruses, despite the fact that this has been reported previously (44, 46). Furthermore, we identified many viral connections to Epsilonbacteraeota, confirming that these highly successful microbial groups in hydrothermal vent diffuse flow fluids are susceptible to viral infection (3, 39, 47–49).

Matches between CRISPR loci and viruses provide an indication of which viruses are being targeted by the CRISPR immune response. We found that while some CRISPR spacers target relatively abundant viruses, most CRISPRs target relatively rare viruses in these systems. This is consistent with previous observations in an archaea-dominated hypersaline lake (50), where the vast majority of CRISPRs were found to target viruses with populations too small to allow for the assembly of contigs. The relative scarcity of viruses targeted by CRISPRs may result from an evolutionary arms race: CRISPRs limit the abundance of the viral populations they target, while concurrently, viruses undergo mutations, limiting the ability of CRISPR spacers to target them. Alternatively, these observations could result from an abundance of inactive spacers inherited over multiple generations. However, we do not believe this to be the case because CRISPR spacer clusters were infrequently present across multiple samples, suggesting that spacers were integrated on subgenerational timescales.

Patterns recorded in microbial CRISPR loci do not reflect the contemporary viral assemblage. In contrast to the vOTUs, the CRISPR spacers did not demonstrate any clear biogeographic patterns. We expected CRISPR spacers to be more widespread than viruses since viral composition represents the virus community at the time of sampling, while spacers represent a history of viral infection. In contrast to our predictions, while both had limited distributions, we found viruses to be more widespread

than CRISPR spacers at all scales examined (Table 1). These results are in contrast to previous studies of CRISPR spacer biogeography in terrestrial hot springs, where both viruses and CRISPR spacers showed clear biogeographic structure (51). This suggests that there is selective pressure for CRISPR spacer composition to evolve more rapidly than viral sequences via the loss or mutation of CRISPR spacers. However, it is also possible that we found viruses to be more widespread because of undersampling of CRISPR spacers, or our use of different clustering algorithms for viruses and spacers: spacers may have been clustered at a finer resolution, resulting in a narrower distribution for each spacer.

Given that CRISPR spacers provide a history of infection, comparing the record of past viral infections via CRISPR arrays with viral sequences in the metagenomes can provide insights into whether CRISPR arrays provide an accurate representation of the viral assemblage at the time of sampling, as well as the rate at which CRISPR spacers are accumulated. We did not find that the most common or cosmopolitan viruses and CRISPR spacers matched each other. This may arise from a temporal mismatch: it takes time for spacers to be incorporated into and lost from CRISPR loci as virus abundances shift and viruses evolve. Additionally, just one single nucleotide polymorphism (SNP) (which we allowed for when aligning spacers to viruses) can prevent a CRISPR spacer from providing resistance against a virus (52). This could explain the discrepancy between viruses and spacers: once resistance of a spacer to a particular virus is suppressed, the population of the virus is freed to shift independently from the spacer in the host population. The discrepancy we observed between viruses and spacers is important to note when attempting to use CRISPR spacers to study viral populations or vice versa.

Examination of the prophage abundance in MAGs revealed a relatively high abundance of prophage encoded in each MAG, confirming previous results indicating that lysogeny is a common lifestyle in hydrothermal systems (14, 15). However, while examination of the gene content of viral contigs revealed several open reading frames (ORFs) encoding a range of functions, including cell membrane function and energy production, it was difficult to prove conclusively that these genes were virus-encoded AMGs rather than potential cellular contamination, and thus, no concrete conclusions were provided here (see Text S1 in the supplemental material). Previous research has suggested that vent viruses encode AMGs (18, 19), and thus, it is likely that many of these genes represent virus-encoded AMGs, but further research is necessary to determine AMG diversity and prevalence across vent systems.

Taken together, our results show that hydrothermal vent viruses are active, abundant, and diverse. These viruses are restricted in their host range and biogeographic extent, but their interactions with hosts persist over time. Thus, while viruses in venting fluids from deep-sea hydrothermal systems have the capacity to play an important role in driving the evolution and ecology of microbial communities, their influence appears to be highly localized to specific regions and taxa. Future work examining viral diversity and distribution across higher-resolution transects over space and time should reveal further insights into the extent of the viral influence in deep-sea hydrothermal vents.

MATERIALS AND METHODS

Sample collection and DNA preparation and sequencing. Low-temperature diffuse flow fluid samples were collected from the vent fields Piccard and Von Damm at the Mid-Cayman Rise in January 2012 and June 2013 during research cruises on the R/V *Atlantis* and R/V *Falkor*, respectively. We analyzed a total of 11 metagenomes from eight different vents at Von Damm vent field and 4 metagenomes from four different vents at Piccard vent field. Sample locations, depth, and metagenomic data are given in Table S1 in the supplemental material. The ROV *Jason II* and Mat Sampler were used to collect the 2012 Mid-Cayman Rise samples, as previously described (53). The 2013 Mid-Cayman Rise samples were collected using the SUPR version 2 sampler and HROV *Nereus* (54). For microbial DNA collection, approximately 3 to 6 liters of diffuse flow fluid were pumped through 0.22-μm Sterivex filters (Millipore). Shipboard, the filters were flooded with RNALater (Ambion), sealed with Luer caps, stored in sterile Falcon tubes, and frozen at −80°C. Sample collection and preservation are further described by Anderson et al. (33) and Reveillaud et al. (40). Genomic DNA was extracted and metagenomic libraries

were constructed as described by Anderson et al. (33). Sequencing was done on an Illumina HiSeq 1000 at the W. M. Keck Facility in the Josephine Bay Paul Center at the Marine Biological Laboratory.

Diffuse flow fluid samples were collected from Axial Seamount in September 2013, August 2014, and August 2015 (approximately 5 months after the eruption of Axial Seamount) during research cruises on the R/V *Falkor* and R/V *Thompson* in 2013, R/V *Brown* in 2014, and R/V *Thompson* in 2015 (Table S1). Diffuse flow samples were collected from four different vent fields within Axial Seamount (ASHES, International District, North Rift Zone, and Dependable) using the ROVs *ROPOS* and *Jason*. We analyzed a total of 16 metagenomes from 10 different vents, 1 plume, and 2 deep seawater samples at Axial Seamount. For collection of microbial DNA, 3 liters of diffuse fluid was pumped through 0.22-μm, 47-mm GWSP filters (Millipore), and the filters were flooded with RNALater (Ambion) on the seafloor as described by Fortunato et al. (4). Fluids from a hydrothermal plume above Anemone and background seawater were collected using a Seabird SBE911 CTD and 10-liter Niskin bottles, and 3 liters of the plume and seawater fluid was filtered through 0.22-μm Sterivex filters (Millipore). DNA was extracted from the filters using a phenol-chloroform method adapted from Crump et al. (55) and Zhou et al. (56). The Ovation Ultralow Library DR multiplex system (Nugen) was used to prepare metagenomic libraries following the manufacturer's instructions. DNA extraction and metagenomic library construction are further described in Fortunato and Huber (57). The 2013 and 2014 samples were sequenced on an Illumina HiSeq 1000, and the 2015 samples were sequenced on a NextSeq 500. Sequencing was done at the W. M. Keck sequencing facility at the Marine Biological Laboratory.

For all metagenomes, paired-end partially overlapping reads were merged and quality filtered using the illumina-utils package (58) using the iu-quality-filter-minoche flag, then assembled using idba-ud (59) v1.1.2 with default settings. All data from both sites are available in the European Nucleotide Archives under study accession number PRJEB15541 for the Mid-Cayman Rise and under study accession numbers PRJEB7866, PRJEB12000, and PRJEB19456 for Axial samples in the years 2013, 2014, and 2015, respectively (Table S1).

Identification of CRISPR loci, viral populations, and spacer assemblages. Crass (60) v1.0.1 was used to identify CRISPR loci (i.e., unique direct repeat types) and CRISPR spacers in the metagenomic reads. Viral contigs in the metagenomic assemblies were identified with VirFinder (41, 60) v1.1 using a *P* value threshold of 0.05. Given that viruses from hydrothermal systems are not well represented in sequence databases, we elected to use VirFinder for identification of viral contigs for diversity and abundance analysis because VirFinder is a k-mer frequency-based method that avoids gene-based similarity searches, and thus has higher potential to identify novel viruses. We used ClusterGenomes (https://bitbucket.org/MAVERICLab/stampede-clustergenomes) v1.1.3 (95% identity, 80% coverage) to identify viral clusters (or viral operational taxonomic units [vOTUs]) of viral contigs. To cluster the CRISPR spacers, we performed an all-versus-all BLAST using BLASTn v2.5.0 (E-value threshold of 10^{-8}) and then clustered using Markov cluster algorithm (MCL) (61) v14-137 (inflation 1.2 and scheme 7) based on bitscore.

Recovery of MAGs. All metagenome assembled genomes (MAGs) were recovered from metagenomic assemblies using anvi'o (62). Supervised clustering was used to recover bins from Piccard and Von Damm contigs using anvi'o v2.1.0 (33). MAGs from Axial Seamount assemblies were recovered using unsupervised binning with CONCOCT (61, 63, 64) within the anvi'o v4.0 pipeline, followed by manual refinement within anvi'o (39). For all analyses, only bins with ≥70% completeness and ≤10% redundancy were retained as MAGs for this analysis. The MAGs were assigned taxonomies using PhyloSift (61, 63) v1.0.1 as described in Anderson et al. (33). Only MAGs for which taxonomies could be identified were used in analyses. MAG coverage was normalized by the number of merged reads in the metagenome. All of the Mid-Cayman Rise MAGs were previously described in Anderson et al. (33).

Taxonomic identification of microbial and viral reads. The reads from all metagenomes were mapped to the Silva small subunit (SSU) and large subunit (LSU) Parc databases (65, 66) (release 132) with bowtie2 (67) v2.2.9 using default settings and local alignment. Mapped reads were assigned taxonomies using the classify.seqs function in mothur (68, 69) v1.38.1 with the Silva 16S rRNA database (release 119) and a cutoff of 50. Reads mapping to 16S rRNA gene sequences that were classified as eukaryotes were excluded from analyses.

In order to identify the viral taxa at Von Damm and Piccard vent fields and Axial Seamount, we conducted a BLAST search of the representative sequence of each vOTU against all of the viral genomic sequences in RefSeq (downloaded on 16 September 2020; E value ≤ 10^{-5}). The best match for each vOTU representative sequence was selected based on E value and percent identity. The taxonomy of the RefSeq viral sequence with the best alignment was assigned to the given vOTU. To quantify the relative abundance of viral families in each sample, the reads that mapped to vOTUs of a viral family were divided by the total number of reads in the sample.

Viral, spacer, and host diversity. Rarefaction curves for vOTUs, CRISPR spacer clusters, and reads matching 16S rRNA genes categorized at the class level were created using the Vegan R package (70) v2.4-5. The number of vOTUs per contig and spacer clusters per read (paired reads) were used as proxies for viral diversity, and the number of different taxa at the class level matching 16S rRNA genes reflected microbial diversity, calculated on a per read basis (merged reads).

Viral, spacer, and CRISPR relative abundance. To calculate relative viral abundance, the reads in each of the metagenomes were mapped against all of the viral contigs from the corresponding geographic region using bowtie2 (67) v2.2.9. The reads from each sample were mapped against all of the viral contigs from the corresponding geographic region rather than solely the viral contigs in the sample because there were viral reads in samples that did not assemble into viral contigs. This method therefore allowed the identification of more viral sequences. The number of reads that mapped to viral contigs was normalized by the number of merged reads as a measure of relative viral abundance. This measure

of relative viral abundance reflects only the proportion of viruses that were retained on the filter as viral capsids or prophages. We used the number of spacers per read and CRISPR direct repeat types per read as measures of spacer and CRISPR relative abundance, respectively. Paired rather than merged reads were used for these analyses. Relative abundance and diversity of viruses, microbes, CRISPR spacers, and CRISPR loci were visualized using the Seaborn library within Python (71).

Relative compositions and abundances of viral populations, spacer assemblages, and hosts. The relative compositions of the microbial community, viral assemblage, and CRISPR spacers were compared between vent sites within Von Damm, Piccard, and Axial Seamount. To calculate the relative abundance of each vOTU in each sample, the number of reads in the sample that mapped to the viral contigs in the vOTU was determined using bowtie2 (67) v2.2.9. The number of reads in each metagenome that mapped to the vOTU was normalized by the total length of the viral contigs in the vOTU and the number of merged reads in the metagenome. We defined the most common vOTUs as the six clusters with the highest relative abundance in each sample. The relative abundance of each CRISPR spacer cluster in each sample was calculated as the percentage of spacers in the sample that were part of the spacer cluster. For spacer clusters, we defined the most common as the three clusters with the highest relative abundance in each sample. To compare the compositions of viral contigs and CRISPR spacers, all of the spacers were blasted against all of the viral contigs using blastn (E value of $\leq 10^{-5}$ and ≤ 1 mismatch per Emerson et al. [72]). The relative abundance of each microbial host was measured as the number of reads that mapped to 16S rRNA gene sequences of the given taxon, normalized by the number of 16S rRNA gene reads in the sample. Analyses of microbial taxa were done at either the class level or the lowest taxonomic level available, depending on the analysis.

Microbiome data sets are compositional in nature because sequencing instruments impose an arbitrary total (73). Therefore, to conduct hierarchical clustering of samples based on viral, spacer, and host composition, we used the protocol outlined by Gloor et al. (73) and Gloor and Reid (74) for computing distances between samples containing compositional data. For hierarchical clustering, we did not normalize; we performed analyses on the number of reads that mapped to each vOTU, the number of CRISPR spacers in each spacer cluster, and the number of reads that mapped to 16S rRNA gene sequences of each microbial host in each sample; for microbial hosts, we did not include reads that mapped to unclassified sequences or sequences classified as eukaryotes. We replaced zero counts with estimates using the count zero multiplicative method for vOTUs and hosts and the Bayes-Laplace Bayesian multiplicative method for spacer clusters via the zCompositions R package (75) v1.2.0. Using the CoDaSeq R package (https://github.com/ggloor/CoDaSeq) (73) v0.99.3, we applied a centered log ratio (clr) transformation to the count data (that lacked zero counts), thereby capturing the ratios between parts. To calculate distances between samples for hierarchical clustering, we used the ward.D2 method on the transformed counts.

Networks of viral infection. Infection networks of virus-host interactions were created using CRISPR sequences to connect microbial hosts with clusters of viral contigs, adapted from the methods used by Daly et al. (76) and Emerson et al. (72). First, MAGs were connected to CRISPR direct repeat types within each sample using BLASTn v2.5.0 (E value $\leq 10^{-10}$, 100% nucleotide identity, as per Emerson et al. (72). Then, each CRISPR direct repeat type was matched to a set of CRISPR spacers as identified by Crass (60) v1.0.1. Finally, the CRISPR spacers in each sample were matched to viral contigs in the corresponding region using BLASTn v2.5.0 with an E-value cutoff of 10^{-5} and a maximum of one mismatch, as per Emerson et al. (72). Only one mismatch was allowed because resistance has been found to be lost by single nucleotide differences between bacterial spacers and target phage sequences (52). In contrast, resistance by archaeal CRISPR systems can still be provided when there are up to three mismatches between spacers and target phage sequences (77, 78). We did not allow for more mismatches between archaeal spacers and phage sequences because resistance is weakened by more mismatches (78). Each direct repeat type provided by Crass v1.0.1 does not necessarily represent an individual CRISPR locus (60).

Data analysis software. The majority of analyses were conducted in RStudio (79) v1.0.136. R packages used were readr (80) v1.1.1, readxl (81) v1.0.0, tidyr (82) v0.7.2, stringr (83) v1.3.0, dplyr (84) v0.7.4, ggnetwork (85) v0.5.1, statnet (86) v2016.9, ggpubr (87) v0.1.7, ggplot2 (88) v2.2.1, and svglite (89) v1.2.1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.03 MB.
FIG S1, PDF file, 0.8 MB.
FIG S2, PDF file, 0.9 MB.
FIG S3, PDF file, 0.7 MB.
FIG S4, PDF file, 1 MB.
FIG S5, PDF file, 0.5 MB.
TABLE S1, XLSX file, 0.01 MB.
TABLE S2, XLSX file, 0.01 MB.
TABLE S3, XLSX file, 0.1 MB.
TABLE S4, XLSX file, 0.02 MB.

ACKNOWLEDGEMENTS

REFERENCES

1. Jannasch HW, Mottl MJ. 1985. Geomicrobiology of deep-sea hydrothermal vents. Science 229:717–725. https://doi.org/10.1126/science.229.4715.717.
2. Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. 2007. Microbial population structures in the deep marine biosphere. Science 318:97–100. https://doi.org/10.1126/science.1146689.
3. Meier DV, Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, Amann R, Meyerdierks A. 2017. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. ISME J 11:1545–1558. https://doi.org/10.1038/ismej.2017.37.
4. Fortunato CS, Larson B, Butterfield DA, Huber JA. 2018. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ Microbiol 20:769–784. https://doi.org/10.1111/1462-2920.14011.
5. McNichol J, Stryhanyuk H, Sylva SP, Thomas F, Musat N, Seewald JS, Sievert SM. 2018. Primary productivity below the seafloor at deep-sea hot springs. Proc Natl Acad Sci U S A 115:6756–6761. https://doi.org/10.1073/pnas.1804351115.
6. Trembath-Reichert E, Butterfield DA, Huber JA. 2019. Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. ISME J 13:2264–2279. https://doi.org/10.1038/s41396-019-0431-y.
7. Galambos D, Anderson RE, Reveillaud J, Huber JA. 2019. Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. Environ Microbiol 21:4395–4410. https://doi.org/10.1111/1462-2920.14806.
8. Suttle CA. 2005. Viruses in the sea. Nature 437:356–361. https://doi.org/10.1038/nature04160.
9. Suttle CA. 2007. Marine viruses–major players in the global ecosystem. Nat Rev Microbiol 5:801–812. https://doi.org/10.1038/nrmicro1750.
10. Chow C-ET, Suttle CA. 2015. Biogeography of viruses in the sea. Annu Rev Virol 2:41–66. https://doi.org/10.1146/annurev-virology-031413-085540.
11. Weinbauer MG, Rassoulzadegan F. 2004. Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11. https://doi.org/10.1046/j.1462-2920.2003.00539.x.
12. 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 Oceanogr Res Pap 52:1515–1527. https://doi.org/10.1016/j.dsr.2005.04.002.
13. Anderson RE, Brazelton WJ, Baross JA. 2011. Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. FEMS Microbiol Ecol 77:120–133. https://doi.org/10.1111/j.1574-6941.2011.01090.x.
14. Williamson SJ, Cary SC, Williamson KE, Helton RR, Bench SR, Winget D, Wommack KE. 2008. Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. ISME J 2:1112–1121. https://doi.org/10.1038/ismej.2008.73.
15. Labonté JM, Pachiadaki M, Fergusson E, McNichol J, Grosche A, Gulmann LK, Vetriani C, Sievert SM, Stepanauskas R. 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. https://doi.org/10.3389/fmicb.2019.01262.
16. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. Nat Microbiol 3:754–766. https://doi.org/10.1038/s41564-018-0166-y.
17. He T, Li H, Zhang X. 2017. Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. mBio 8:e00893-17. https://doi.org/10.1128/mBio.00893-17.
18. Anderson RE, Sogin ML, Baross JA. 2014. Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. PLoS One 9:e109696. https://doi.org/10.1371/journal.pone.0109696.
19. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. 2014. Sulfur oxidation genes in diverse deep-sea viruses. Science 344:757–760. https://doi.org/10.1126/science.1252229.
20. Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55:709–742. https://doi.org/10.1146/annurev.micro.55.1.709.
21. Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc Natl Acad Sci U S A 102:14332–14337. https://doi.org/10.1073/pnas.0504068102.
22. Fuchsman CA, Collins RE, Rocap G, Brazelton WJ. 2017. Effect of the environment on horizontal gene transfer between bacteria and archaea. PeerJ 5:e3865. https://doi.org/10.7717/peerj.3865.
23. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304. https://doi.org/10.1038/35012500.
24. Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732. https://doi.org/10.1038/nrmicro1235.
25. Brazelton WJ, Baross JA. 2009. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. ISME J 3:1420–1424. https://doi.org/10.1038/ismej.2009.79.

26. Stewart LC, Algar CK, Fortunato CS, Larson BI, Vallino JJ, Huber JA, Butterfield DA, Holden JF. 2019. Fluid geochemistry, local hydrology, and metabolic activity define methanogen community size and composition in deep-sea hydrothermal vents. ISME J 13:1711–1721. https://doi.org/10.1038/s41396-019-0382-3.

27. Opatkiewicz AD, Butterfield DA, Baross JA. 2009. Individual hydrothermal vents at Axial Seamount harbor distinct subseafloor microbial communities. FEMS Microbiol Ecol 70:413–424. https://doi.org/10.1111/j.1574-6941.2009.00747.x.

28. Anderson RE, Sogin ML, Baross JA. 2015. Biogeography and ecology of the rare and abundant microbial lineages in deep-sea hydrothermal vents. FEMS Microbiol Ecol 91:1–11. https://doi.org/10.1093/femsec/fiu016.

29. Han Y, Perner M. 2015. The globally widespread genus Sulfurimonas: versatile energy metabolisms and adaptations to redox clines. Front Microbiol 6:989. https://doi.org/10.3389/fmicb.2015.00989.

30. Dick GJ. 2019. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol 17:271–283. https://doi.org/10.1038/s41579-019-0160-2.

31. Von Damm K. 1990. Seafloor hydrothermal activity: black smoker chemistry and chimneys. Annu Rev Earth Planet Sci 18:173–204. https://doi.org/10.1146/annurev.ea.18.050190.001133.

32. Kelley DS. 2005. A serpentinite-hosted ecosystem: the Lost City hydrothermal field. Science 307:1428–1434. https://doi.org/10.1126/science.1102556.

33. Anderson RE, Reveillaud J, Reddington E, Delmont TO, Eren AM, McDermott JM, Seewald JS, Huber JA. 2017. Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. Nat Commun 8:1114. https://doi.org/10.1038/s41467-017-01228-6.

34. Butterfield DA, Roe KK, Lilley MD, Huber JA, Baross JA, Embley RW, Massoth GJ. 2004. Mixing, reaction and microbial activity in the sub-seafloor revealed by temporal and spatial variation in diffuse flow vents at axial volcano. The Subseafloor Biosphere at Mid-Ocean Ridges, vol 144. Geophysical Monograph Series. American Geophysical Union, Washington, DC.

35. McDermott JM, Seewald JS, German CR, Sylva SP. 2015. Pathways for abiotic organic synthesis at submarine hydrothermal fields. Proc Natl Acad Sci U S A 112:7668–7672. https://doi.org/10.1073/pnas.1506295112.

36. German CR, Bowen A, Coleman ML, Honig DL, Huber JA, Jakuba MV, Kinsey JC, Kurz MD, Leroy S, McDermott JM, de Lépinay BM, Nakamura K, Seewald JS, Smith JL, Sylva SP, Van Dover CL, Whitcomb LL, Yoerger DR. 2010. Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. Proc Natl Acad Sci U S A 107:14020–14025. https://doi.org/10.1073/pnas.1009205107.

37. McDermott JM. 2015. Geochemistry of deep-sea hydrothermal vent fluids from the Mid-Cayman Rise, Caribbean Sea. PhD thesis. Massachusetts Institute of Technology, Cambridge, MA.

38. McDermott JM, Sylva SP, Ono S, German CR, Seewald JS. 2018. Geochemistry of fluids from Earth's deepest ridge-crest hot-springs: Piccard hydrothermal field, Mid-Cayman Rise. Geochim Cosmochim Acta 228:95–118. https://doi.org/10.1016/j.gca.2018.01.021.

39. Moulana A, Anderson RE, Fortunato CS, Huber JA. 2020. Selection is a significant driver of gene gain and loss in the pangenome of the bacterial genus in geographically distinct deep-sea hydrothermal vents. mSystems 5:e00673-19. https://doi.org/10.1128/mSystems.00673-19.

40. Reveillaud J, Reddington E, McDermott J, Algar C, Meyer JL, Sylva S, Seewald J, German CR, Huber JA. 2016. Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. Environ Microbiol 18:1970–1987. https://doi.org/10.1111/1462-2920.13173.

41. 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. https://doi.org/10.1186/s40168-017-0283-5.

42. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science 348:1261498. https://doi.org/10.1126/science.1261498.

43. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537:689–693. https://doi.org/10.1038/nature19366.

44. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. Nature 536:425–430. https://doi.org/10.1038/nature19094.

45. Flores GE, Campbell JH, Kirshtein JD, Meneghin J, Podar M, Steinberg JI, Seewald JS, Tivey MK, Voytek MA, Yang ZK, Reysenbach A-L. 2011. Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. Environ Microbiol 13:2158–2171. https://doi.org/10.1111/j.1462-2920.2011.02463.x.

46. Peters DL, Lynch KH, Stothard P, Dennis JJ. 2015. The isolation and characterization of two Stenotrophomonas maltophilia bacteriophages capable of cross-taxonomic order infectivity. BMC Genomics 16:664. https://doi.org/10.1186/s12864-015-1848-y.

47. Akerman NH, Butterfield DA, Huber JA. 2013. Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. Front Microbiol 4:185. https://doi.org/10.3389/fmicb.2013.00185.

48. Meyer JL, Huber JA. 2014. Strain-level genomic variation in natural populations of Lebetimonas from an erupting deep-sea volcano. ISME J 8:867–880. https://doi.org/10.1038/ismej.2013.206.

49. Yoshida-Takashima Y, Takaki Y, Shimamura S, Nunoura T, Takai K. 2013. Genome sequence of a novel deep-sea vent epsilonproteobacterial phage provides new insight into the co-evolution of Epsilonproteobacteria and their phages. Extremophiles 17:405–419. https://doi.org/10.1007/s00792-013-0529-5.

50. Emerson JB, Andrade K, Thomas BC, Norman A, Allen EE, Heidelberg KB, Banfield JF, 2013. 2013. Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. Archaea 2013:370871. https://doi.org/10.1155/2013/370871.

51. Held NL, Whitaker RJ. 2009. Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. Environ Microbiol 11:457–466. https://doi.org/10.1111/j.1462-2920.2008.01784.x.

52. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712. https://doi.org/10.1126/science.1138140.

53. Breier JA, Gomez-Ibanez D, Reddington E, Huber JA, Emerson D. 2012. A precision multi-sampler for deep-sea hydrothermal microbial mat studies. Deep Sea Res Part I Oceanogr Res Pap 70:83–90. https://doi.org/10.1016/j.dsr.2012.10.006.

54. Breier JA, Rauch CG, McCartney K, Toner BM, Fakra SC, White SN, German CR. 2009. A suspended-particle rosette multi-sampler for discrete biogeochemical sampling in low-particle-density waters. Deep Sea Res Part I Oceanogr Res Pap 56:1579–1589. https://doi.org/10.1016/j.dsr.2009.04.005.

55. Crump BC, Hopkinson CS, Sogin ML, Hobbie JE. 2004. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. Appl Environ Microbiol 70:1494–1505. https://doi.org/10.1128/AEM.70.3.1494-1505.2004.

56. Zhou J, Bruns MA, Tiedje JM. 1996. DNA recovery from soils of diverse composition. Appl Environ Microbiol 62:316–322. https://doi.org/10.1128/aem.62.2.316-322.1996.

57. Fortunato CS, Huber JA. 2016. Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. ISME J 10:1925–1938. https://doi.org/10.1038/ismej.2015.258.

58. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. PLoS One 8:e66643. https://doi.org/10.1371/journal.pone.0066643.

59. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.

60. Skennerton CT, Imelfort M, Tyson GW. 2013. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. Nucleic Acids Res 41:e105. https://doi.org/10.1093/nar/gkt183.

61. van Dongen SM. 1969. Graph clustering by flow simulation. PhD dissertation. Utrecht University, Utrecht, the Netherlands.

62. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.

63. Darling AE, Jospin G, Lowe E, Matsen FA, IV, Bik HM, Eisen JA. 2014. Phylo-Sift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243. https://doi.org/10.7717/peerj.243.

64. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.

65. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196. https://doi.org/10.1093/nar/gkm864.

66. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.

67. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

68. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.

69. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.

70. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veget Sci 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

71. Waskom M, Botvinnik O, O'Kane D, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Cole JB, Warmenhoven J, de Ruiter J, Pye C, Hoyer S, Vanderplas J, Villalba S, Kunter G, Quintero E, Bachant P, Martin M, Meyer K, Miles A, Ram Y, Yarkoni T, Williams ML, Evans C, Fitzgerald C, Brian, Fonnesbeck C, Lee A, Qalieh A. 2017. mwaskom/seaborn: v0.8.1. Zenodo. https://doi.org/10.5281/zenodo.883859.

72. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li C, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol 3:870–880. https://doi.org/10.1038/s41564-018-0190-y.

73. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome datasets are compositional: and this is not optional. Front Microbiol 8:2224. https://doi.org/10.3389/fmicb.2017.02224.

74. Gloor GB, Reid G. 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can J Microbiol 62:692–703. https://doi.org/10.1139/cjm-2015-0821.

75. Palarea-Albaladejo J, Martín-Fernández JA. 2015. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. Chemometrics Intell Lab Syst 143:85–96. https://doi.org/10.1016/j.chemolab.2015.02.019.

76. Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE, Hoyt DW, Meulia T, Wolfe RA, Hanson AJ, Mouser PJ, Moore JD, Wunch K, Sullivan MB, Wrighton KC, Wilkins MJ. 2019. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. Nat Microbiol 4:352–361. https://doi.org/10.1038/s41564-018-0312-6.

77. Garrett RA, Vestergaard G, Shah SA. 2011. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol 19:549–556. https://doi.org/10.1016/j.tim.2011.08.002.

78. Manica A, Zebec Z, Teichmann D, Schleper C. 2011. In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol Microbiol 80:481–491. https://doi.org/10.1111/j.1365-2958.2011.07586.x.

79. RStudio Team. 2015. RStudio: integrated development for R. RStudio, Inc, Boston, MA.

80. Wickham H, Hester J, Francois R. 2017. Readr: Read rectangular text data. R package version 1.

81. Wickham H, Bryan J. 2017. readxl: Read Excel files. R package version 1.

82. Wickham H, Henry L. 2018. tidyr: easily tidy data with 'spread()' and 'gather()' functions. R package version 0.8.2.

83. Wickham H. 2017. stringr: simple, consistent wrappers for common string operations. R package version 1.

84. Wickham H, Francois R. 2015. dplyr: a grammar of data manipulation. R package version 0.4.3.

85. Briatte F. 2016. ggnetwork: geometries to plot networks with 'ggplot2'. R package version 0.5.1.

86. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. 2008. statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24:1548−7660. https://doi.org/10.18637/jss.v024.i01.

87. Kassambara A. 2017. ggpubr:"ggplot2" based publication ready plots. R package version 0.1.6.

88. Wickham H. 2011. ggplot2. WIREs Comp Stat 3:180–185. https://doi.org/10.1002/wics.147.

89. Luciani TJ, Decorde M, Lise V. 2009. RSvgDevice: an R SVG graphics device. R package version 0.6.4.