Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Method paper

Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis

Hugo G. Castelán-Sánchez[a], Itzel Lopéz-Rosas[b], Wendy A. García-Suastegui[c], Raúl Peralta[a], Alan D.W. Dobson[d,e], Ramón Alberto Batista-García[a], Sonia Dávila-Ramos[a,*]

[a] Centro de Investigación en Dinámica Celular, Instituto de Investigaciones en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Morelos. Av. Universidad 1001. Col. Chamilpa. Cuernavca, Morelos. C.P, Cuernavaca 62209, Mexico

[b] CONACyT Research fellow-Colegio de Postgraduados Campus Campeche, Carretera Haltunchén - Edzná Km 17.5. Colonia Sihochac. Champotón, Campeche 24450, Mexico

[c] Laboratorio de Toxicología Molecular, Departamento de Biología y Toxicología de la Reproducción, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Puebla C.P., 72570, Mexico

[d] School of Microbiology, University College Cork. Cork, Ireland

[e] Environmental Research Institute, University College, Cork, Ireland

ARTICLE INFO

Keywords:
Deep-sea virome
Natural selection
Auxiliary metabolic gene
Cytitidyltransferase
Ribonucleotide reductase

ABSTRACT

Ten publicly available metagenomic data sets from hydrothermal vents were analyzed to determine the taxonomic structure of the viral communities present, as well as their potential metabolic functions. The type of natural selection on two auxiliary metabolic genes was also analyzed. The structure of the virome in the hydrothermal vents was quite different in comparison with the viruses present in sediments, with specific populations being present in greater abundance in the plume samples when compared with the sediment samples. ssDNA genomes such as *Circoviridae* and *Microviridae* were predominantly present in the sediment samples, with *Caudovirales* which are dsDNA being present in the vent samples. Genes potentially encoding enzymes that participate in carbon, nitrogen and sulfur metabolic pathways were found in greater abundance, than those involved in the oxygen cycle, in the hydrothermal vents. Functional profiling of the viromes, resulted in the discovery of genes encoding proteins involved in bacteriophage capsids, DNA synthesis, nucleotide synthesis, DNA repair, as well as viral auxiliary metabolic genes such as cytitidyltransferase and ribonucleotide reductase. These auxiliary metabolic genes participate in the synthesis of phospholipids and nucleotides respectively and are likely to contribute to enhancing the fitness of their bacterial hosts within the hydrothermal vent communities. Finally, evolutionary analysis suggested that these auxiliary metabolic genes are highly conserved and evolve under purifying selection, and are thus maintained in their genome.

1. Introduction

Hydrothermal vents are cracks or fissures in the seafloor from which geothermally heated water emerges in a column form as the seawater meets the magma (Ledesma, 2011; Tarasov et al., 2005). Despite hydrothermal vents having temperatures of up to 400 °C and a highly reducing chemical nature (Kelley et al., 2005; Martin et al., 2008), they are a source of bacteria and archaea with a high level of biodiversity; which has been investigated using both culture dependent (Cary et al., 1997; Harmsen et al., 1997; Jeanthon, 2000) and independent approaches (Xie et al., 2010; Anderson et al., 2011a; Anantharaman et al., 2015; Zhang et al., 2016a; Pjevac et al., 2018; Cerqueira et al., 2018) A recent report shows an inverse relationship between the abundance and diversity levels in the microbial populations inhabiting hydrothermal vents, suggesting the presence of specific microbial groups which are very well established in these hyperthermophilic environments (Anderson et al., 2017). Prokaryotes have to date been the best studied microorganisms in vents (Huber et al., 2007; Dick et al., 2013; Sheik et al., 2015; Poli et al., 2017; Dávila-Ramos et al., 2014), with reports showing that bacteria and archaea communities residing in hydrothermal plumes are quite different from those present in sediments primarily due to the fact that the plume is much colder and much more strongly influenced by the background seawater (Dick et al., 2013; Ding et al., 2017; Christakis et al., 2018).

In the case of the virome in these polyextremophilic ecosystems, little is known about their taxonomic structure, their metabolism or

* Corresponding author at: Ave. Universidad 1001. Col. Chamilpa, Cuernavaca, 62209, Mexico.
E-mail address: sonia.davila@uaem.mx (S. Dávila-Ramos).

https://doi.org/10.1016/j.margen.2019.03.001
Received 6 December 2018; Received in revised form 25 January 2019; Accepted 1 March 2019

Table 1
Metagenome data considered in the structural and functional analysis.

Name of site	Location	Ocean	Deep (m)	ID No. SRA	Sample/Collection date	Sequencing platform	Insert size (bp)	Sequenced bases (Gb)	Reference
Lau Basin Tahi Moana	S 20 40. 927,843 W 176 11.001806	Pacific Ocean	800	SRR 1217461	Plume water/ June 2009	Illumina HiSeq 2500 (paired)	N/A	18.3	Anantharaman et al. (2014)
Mid Cayman 1	N/A	Atlantic Ocean	2238	SRR 2046236	Plume water/ January 2012	Illumina HiSeq 2000 (paired)	N/A	30.7	Li et al. (2015)
Mid Cayman 2	N/A	Atlantic Ocean	4869	SRR 2046238	Plume water/ January 2012	Illumina HiSeq 2000 (paired)	202	36.1	Li et al. (2015)
Axial Seamount	N 27.025 W 111.400	Pacific Ocean	1500	ERR 2021511	Plume water/ August 2015	Illumina NextSeq 500 (paired)	275	7.38	Fortunato et al. (2018)
Guaymas Mexico	N 27.025 W 111.400	Pacific Ocean	1993	SRR 3577362	Plume water/ July 2004	Illumina HiSeq 2000 (paired)	N/A	23.5	Anantharaman et al. (2014)
Menez Gwen	N/A	Atlantic Ocean	828	ERR 1078302	Fluid/ October 2009	Illumina MiSeq (paired)	N/A	5.3	Meier et al. (2016)
South Mid Atlantic	S 15.16 W 13.350	Atlantic Ocean	2500	SRR 4028170	Sediment/ August 2012	Illumina HiSeq 2000 (paired)	N/A	4.07	N/A
Southwest Indian Ridge 1	S 37.55 E 51.00	Indian Ocean	2400	SRR 3136143	Sediment/ March 2014	Illumina HiSeq 2000 (paired)	N/A	5.34	He et al. (2017)
Southwest Indian Ridge 2	S 27.85 E 63.94	Indian Ocean	2400	SRR 3133481	Sediment/ December 2013	Illumina HiSeq 2000 (paired)	N/A	7.30	He et al. (2017)
Southern Mariana Japan	N 12.93 E 143.62	Pacific Ocean	3024	DRR 093004	Sediment/ June 2010	Illumina MiSeq (paired)	300	9.08	Kato et al. (2018)

N/A: No apply.

their overall ecology (Sime-Ngando, 2014). Given the important of viruses as microbial predators that are known to influence global biogeochemical cycles and to impact microbial evolution (Rohwer et al., 2009), it is likely that they play an important role in the ecological relationships with these unique microbial communities inhabiting deep-sea hydrothermal vents. Viral mediated horizontal gene transfer is known to occur on a widespread basis in the oceans (McDaniel et al., 2010). Viruses are known to encode auxiliary metabolic genes (AMGs), which play a crucial role in promoting biochemical and metabolic processes (Beiko et al., 2005; Breitbart et al., 2007).

The viral abundance in active vents has been estimated to be 3.5×10^6 and 2.94×10^6 viruses per ml^{-1} from plume and sediment samples, respectively (Ortmann and Suttle, 2005; Manini et al., 2007). In spite of this, few viruses have to date been isolated from hydrothermal vents using classical techniques; those that have include; the bacteriophages *Bacillus virus* W1 (BVW1), *Geobacillus virus* E1 (GVE1), *Geobacillus virus* E2 (GVE2), *Nitratiruptor phage* (NRS-1) and *Marinitoga piezophila virus* (MPV-1), TPV1 (*Thermococcus prieurii virus* 1) (Gorlas et al., 2012; Romancer et al., 2006; Prangishvili, 2003; Lossouarn et al., 2015). With advances in next generation sequencing based approaches, it is clear that metagenomics will allow us gain a greater appreciation of the virome in these hydrothermal vents. Metagenomic studies have already shown that *Siphoviridae*, *Myoviridae* and *Podoviridae* are the predominant viral families present in these ecosystems (Breitbart et al., 2007; Millard et al., 2014; Anderson et al., 2017; Strazzulli et al., 2017), and that viruses which infect archaea are present in high abundance (Rice et al., 2001; Prangishvili, 2003; Geslin et al., 2003). Despite this, further efforts are needed to increase our knowledge relating to the virosphere that is present in hydrothermal vents; particularly with a view to determining the potential function that these viruses may play in these environments.

As previously mentioned viruses are known to play an essential role in biogeochemical cycles (Rohwer et al., 2009; Weitz and Wilhelm, 2012; Mizuno et al., 2016). Viral AMGs can complement metabolic pathways that are present in bacteria, and following acquisition can remain in the prokaryotic genomes by natural selection, and consequently enhance the fitness of bacterial strains that host these viral genes (Anderson et al., 2011b; Anderson et al., 2014; He et al., 2017). However there is little knowledge about the kind of natural selection that is important in the evolution of these auxiliary metabolic genes. In addition, lysogenic viruses can also have a significant impact on their bacterial hosts, by inducing cell lysis process within the host; and thereby modifying the microbial food web and energy transfer to higher trophic levels (Williamson et al., 2008; Rastelli et al., 2017).

In the last few years a good deal of metagenomic sequence data has been generated from the microbiota of hydrothermal vents (Zhang et al., 2016b). However, generating detailed analysis that allows us to exploit all the information that is available is currently a bottleneck for scientists. With this in mind, we focused on ten publically available metagenomic data sets deposited in the National Center Biotechnology Information (NCBI), which we investigated to analyze the taxonomic structure of the viromes in hydrothermal vents from which they data had been obtained. We found that ssDNA viruses predominate in sediments, whereas dsDNA are more abundant in plumes. In addition we analyze these metagenomes for potential metabolic functions and for the presence and integrity of metabolic pathways potentially involved biogeochemical cycles, and uncovered some specific and complete pathways involved in nitrogen, sulfur, carbon and iron biogeochemical cycles. Following the functional analysis of the viral sequences the most abundant were capsid sequences of bacteriophages, and few carbohydrates and AMGs. Finally the type of natural selection was analyzed on capsids and on two AMGs, which appear to generally evolve under purifying selection, but with a few sites in the AMGs appearing to be under episodic selection.

2. Materials and methods

2.1. Metagenomic data from hydrothermal vents

Ten publically available metagenomic data sets obtained from sequence-based metagenomic studies in hydrothermal vents were collected from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (Leinonen et al., 2010). Table 1 summarizes details of these metagenomic sequences. These datasets were chosen because they were derived from shotgun metagenomic projects in hydrothermal vents and were generated from the Illumina Next-Generation Sequencing platform, that typically produces homogeneous and high quality sequences.

2.2. Sequence quality analysis

Sequence quality control (QC) analysis was performed using the FastQC program (Bioinformatics, 2011). The quality trimming threshold was set to a 30 Q score (corresponds to an 0.001 error rate). Adapters were removed using Trimmomatic software (Bolger et al., 2014). Subsequently, duplicated sequences were eliminated using CD-HIT-DUP software with a maximum mismatch number of -e 0.03 (Huang et al., 2010) since during amplification processes some sequences are artificially produced (Gomez-Alvarez et al., 2009). After sequence quality control a sequence set was obtained and used in the subsequent analysis.

2.3. Viral taxonomic analysis of metagenomes

Sequences were assembled and contigs were obtained using MegaHit software, which uses an assembly algorithm based on Bruijn graphs, using paired-end mode, k min = 21, k max = 131, k step = 10 (Li et al., 2016). The assembled sequences were compared with a previously constructed database that contains approximately 6000 virus genomes available in NCBI (non-redundant nucleotide database). Briefly, to generate the local database, the virus genomes collected from NCBI were indexed using standalone BLASTn.

For the identification of the viral sequences from all the assembled sequences, a comparison with the local database was performed using the following parameters: number of aligments = 20, e-value = 0.0001 and word size = 11. The best twenty scoring BLAST hits were parsed and taxonomically assigned using MEGAN 5.10.6 software (Huson et al., 2007). For the virus taxonomic classification, the method of the lowest common ancestor (LCA) was used using the following parameters: minimum support = 2, minimum score = 70, top percent = 10; this reduced the risk of obtaining false positive or false negative taxonomic assignments (Huson et al., 2016). When taxonomic profiles were obtained, matrix abundances were generated and later processed in R software (version 3.2.3). Finally, plots were also done in R software with the libraries ggplot2 and RColorBrewer 175 (Team, 2013) (www.ColorBrewer.org).

Hierarchical clustering analysis was addressed to compare virome relative abundance of the data with other viromes deposited in Metavir2 (Roux et al., 2014). The hierarchical clusters were obtained using the heatmap2 package in R based in the hclust library, which evaluates the dissimilarity between the relative abundances of the virus families using distance matrix methods (Becker et al., 1988). To conduct this hierarchical clustering analysis, 21 viromes previously deposited in Metavir2 were collected (Table 2). Overall, the clustering analysis included the virus relative abundance obtained from the ten hydrothermal vent metagenomes (previously mentioned in section 2.1) and these 21 viromes.

2.4. Functional analysis of metagenomic sequences

The assembled contigs were uploaded to the MG-RAST server (Glass and Meyer, 2011) and the functional annotation was obtained in the classification of subsystem technology platform (SEED), which is a categorize system that includes five hierarchical levels of functional annotation. For the viral functional annotation on MG-RAST server, only the viral domain sequences in the RefSeq database were selected to avoid the inclusion of contigs from microbial sequences in the functional analysis. The viral functions were obtained according to the classification at level 3 of the subsystems database with an e-value threshold of $1\text{-}e^{-5}$. The AMGs were identified using the same viral functional analysis.

Multigenomic entropy based score (MEBS) software (Anda et al., 2017) (FDR 0.0001) was used to identify the completeness of metabolic pathway involved in the biogeochemical cycles within the metagenomic datasets, all contigs from each sample were used in this analysis.

2.5. Natural selection analysis

Using only viral sequences obtained from the MG-RAST server we evaluated the natural selection sites by different methods. To obtain the virus sequences from the metagenomics contigs, the RefSeq database was used.

Fixed-effects likelihood (FEL) and random effects likelihood (REL) were used to conduct the natural selection analysis. These algorithms use the principles of maximum likelihood to estimate the proportion of synonymous and non-synonymous rates of each nucleotide site (Pond and Frost, 2005). These methods detect natural selection in a coding gene, identifying higher non-synonymous substitution rates (dN when an amino acid changes) in relation to synonymous substitution rates (dS, silent mutations) that are considered neutral. This relation is represented as:

$$\omega = dS/dN \quad (1)$$

while mixed effect model of evolution (MEME) allows the distribution of ω to vary site by site (fixed effects) and also branch by branch in a site. In addition, the method identifies the two types of episodic and constant natural selection (Murrell et al., 2012). The presence of natural selection was also evaluated, which makes a global comparison of dS/dN rates with “evolutionary fingerprint” software in Datamonkey server. This software is based on certain sites in genes evolving rapidly or resisting the change of natural election. These sites are typically called an “evolutionary fingerprint” (Pond et al., 2009).

3. Results and discussion

3.1. Viral communities in sediments and deep-water from hydrothermal vents

In an attempt to gain additional knowledge on the viral communities present in hydrothermal vents, for which there is currently quite limited information available; this study focused on comparing the viral populations in ten virospheres from metagenomics datasets available for ten different hydrothermal vents located in different geographical zones. The structure of the viral populations in each of the ten locations is sampled divided in plumes and sediments (Fig. 1). The most representative viral communities in both sample types belong to the *Caudovirales* order, with 70–80% of the assembled contigs being classified into three main families namely the *Siphoviridae*, *Podoviridae* and *Myoviridae*. Bacteriophages (*Myoviridae* family) showed the higher relative abundance in samples ranging from 50 to 60% in the plumes and 30–70% in the sediments.

It is well established that bacteriophages are the most abundant viruses found in environmental samples from soil, freshwater and marine ecosystems, and are known to actively regulate the ecological dynamics of the bacteria populations within these environments (Dick et al., 2013; Sepulveda et al., 2016; Hayes et al., 2017; Tetz and Tetz, 2018). The results here appear to indicate that based on the levels of

Table 2
Summary of samples collection from Metavir 2.

Name of site	Location	Ocean	Deep (m)	ID No. Metavir 2	Sample/ Collection date	Sequencing Platform	Insert size (bp)	Sequences number	Reference
Atlantic	41°43′51.2394″ N − 10°40′56.568″ E	Atlantic Ocean	3530	Atl_Vir 1157	Sediment/ October 2008	454 Roche Pyrosequencing	N/A	26,826	Corinaldesi et al. (2017)
Atlantic	41°43′51.2394″ N − 10°40′56.568″ E	Atlantic Ocean	3530	Atlantic-Extra 2125	Sediments/ August 2005	454 Roche Pyrosequencing	N/A	107,090	Corinaldesi et al. (2017)
Arctic	79°8′0.5994″ N 2°50′32.2794″ E	Arctic Ocean	5571	Arct_Vir 1159	Sediment/ August 2005	454 Roche Pyrosequencing	N/A	63,869	Corinaldesi et al. (2017)
Black Sea	42°59′54.204″ N 31°30′58.644″ E	Black Sea	1970	Black Sea 1155	Sediment/ September 2006	454 Roche Pyrosequencing	N/A	78,436	Corinaldesi et al. (2017)
Mediterranean	39°31′04.1880″ N 6°10′32.4012″ E	Mediterranean	2850	Mediterranean 1161	Sediment/ April 2002	454 Roche Pyrosequencing	N/A	65,340	Corinaldesi et al. (2017)
Atlantic	39°30′24.18″ N − 9°50′0.604″ E	Atlantic Ocean	3530	NE_Atlantic_2 1156	Sediment/ August 2005	454 Roche Pyrosequencing	N/A	165,517	Corinaldesi et al. (2017)
Arctic	79°8′0.5994″ N 2°50′32.2794″ E	Arctic Ocean	5571	Arct_Ocean 1158	Sediment/ September 2006	454 Roche Pyrosequencing	N/A	79,646	Corinaldesi et al. (2017)
Izu-Ogasawara Trench	29°09′ N 142°49′ E	Pacific Ocean	9760	Izu-Ogasawara Trench 164	Sediment/ December 2007	454 Roche Pyrosequencing	N/A	46,458	Yoshida et al. (2013)
Mariana Trench	11°22′ N 42°42′ E	Pacific Ocean	10,332	Mariana Trench 165	Sediment/ May 2008	454 Roche Pyrosequencing	N/A	49,584	Yoshida et al. (2013)
Shimokita Peninsula	41°10′ N 142°12′ E	Pacific Ocean	1181	Shimokita Peninsula 166	Sediment/ January 2006	454 Roche Pyrosequencing	N/A	76,498	Yoshida et al. (2013)
Brazos Trinity Basin	27°18.0809′ N 94°,23.2537′ W	Gulf of Mexico	1470	Brazos-Trinity 8mbsf 3813	Sediment/ N/A	454 Roche Pyrosequencing	N/A	270,730	Biddle et al. (2011)
LineP transect, ocean station	48°58′8” N -130°,40′12” W	Pacific Ocean	1000	LJ12O 1394	Water/ June 2009	454 Roche Pyrosequencing	N/A	122,565	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°58′8” N -130°,40′12” W	Pacific Ocean	2000	LJ12D 1395	Water/ June 2010	454 Roche Pyrosequencing	N/A	49,914	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LA26O 1398	Water/ August 2009	454 Roche Pyrosequencing	N/A	70,596	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N, −144°,59′56” W	Pacific Ocean	2000	LA26D 1400	Water/ August 2009	454 Roche Pyrosequencing	N/A	68,516	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LF26O 1403	Water/ Febrary 2009	454 Roche Pyrosequencing	N/A	147,537	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LF26D 1404	Water/ Febrary 2009	454 Roche Pyrosequencing	N/A	125,896	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LJ26O 1407	Water/ June 2009	454 Roche Pyrosequencing	N/A	101,179	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	2000	LJ26D 1408	Water/ June 2009	454 Roche Pyrosequencing	N/A	55,332	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°38′58” N -126°39′52” W	Pacific Ocean	1000	LJ4O 1414	Water/ June 2009	454 Roche Pyrosequencing	N/A	97.126	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°38′58” N -126°39′52” W	Pacific Ocean	1300	LJ4D 1415	Water/ June 2009	454 Roche Pyrosequencing	N/A	98,478	Hurwitz and Sullivan (2013)
LineP transect, ocean station	33°17′13” N -129°25′42” W	Pacific Ocean	1000	M6O1K 1432	Water/ Octuber 2009	454 Roche Pyrosequencing	N/A	225,833	Hurwitz and Sullivan, 2013
LineP transect, ocean station	33°17′13” N -129°25′42” W	Pacific Ocean	4300	M7O4K 1433	Water/ Octuber 2009	454 Roche Pyrosequencing	N/A	144,588	Hurwitz and Sullivan, 2013
LineP transect, ocean station	50°6” N, −144°,59′56” W	Pacific Ocean	500	LF26A 1402	Water/ Febrary 2009	454 Roche Pyrosequencing	N/A	167,616	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N, −144°,59′56” W	Pacific Ocean	500	LA26A 1397	Water/ August 2009	454 Roche Pyrosequencing	N/A	42,118	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°58′8” N -130°,40′12” W	Pacific Ocean	500	LJ12A 1393	Water/ June 2010	454 Roche Pyrosequencing	N/A	58,108	Hurwitz and Sullivan (2013)

(*continued on next page*)

Table 2 (*continued*)

Name of site	Location	Ocean	Deep (m)	ID No. Metavir 2	Sample/ Collection date	Sequencing Platform	Insert size (bp)	Sequences number	Reference
LineP transect, ocean station	48°38′58” N -126°39′52” W	Pacific Ocean	10	LJ4S 1409	Water/ June 2009	454 Roche Pyrosequencing	N/A	92,415	Hurwitz and Sullivan (2013)

N/A: No apply.

bacteriophage present that they are also likely to play a significant ecological contribution in hydrothermal vent ecosystems. The second cluster of viral families with the highest representation in the virome profiles, were the *Mimiviridae, Phycodnaviridae* and *Poxviridae* (Fig. 1). This is not surprising perhaps given that *Mimiviridae* and *Phycodnaviridae* have previously been reported in aquatic environments, with some members of these families being discovered in hot spring environments such as in Yellowstone (Zhang et al., 2015).

Single strand DNA (ssDNA) viruses were exclusively detected in the sediment samples, with *Microviridae* being the family with highest abundance. This finding is consistent with reports of ssDNA viruses being found in freshwater sediments (Hewson et al., 2012; Roux et al., 2012a) and in deep sea samples (Yoshida et al., 2018). ssDNA viruses are considered as allochthonous viruses in marine sediments since it is believed that they are deposited in the benthic zones through sedimentation (Hewson et al., 2012). dsDNA viruses tend to be preferentially detected in environmental samples due to a methodological *bias* in the multiple displacement amplification technique employed, which results in a preferential amplification of dsDNA viruses (Anderson et al., 2014). In the two metagenome data sets from the Southwest Indian Ocean (SRR3136143 and S3133481) included in this work, this type of amplification was not used, and consequently ssDNA viruses were detected. In the Indian samples, *Gokushovirus* was one of the ssDNA virus subfamily which was found, that correlates with the 16S ribosomal gene analysis carried out by Anderson et al. (2014) where *Chlamydiales*, the natural host of *Gokushovirus* (Roux et al., 2012b; Labonté and Suttle, 2013) were shown to be present (Anderson et al., 2014).

The way in which the presence of viral genomes has been calculated has often involved staining the viral capsids with SYBR green, which allows a direct quantification of the amount of viral particles in any given sample. However recent studies have shown that quantification of viral particles with a ssDNA genome has been underestimated in different environments, particularly in marine sediments (Yoshida et al., 2018). It has been estimated that there are between 1×10^8 to 3×10^9 copies of viral genomes per cm^3 in sediments, an amount which is higher than for dsDNA viruses. Given that ssDNA viruses are likely to play an important role in regulating bacterial mortality levels and in ecological succession occurring in prokaryotic communities in deep-sea sediments, further efforts should be made to study the taxonomic structure of these viral populations without methodological biases.

There is currently a lack of knowledge about the structure and ecology of ssDNA viruses in these deep-sea environments. However, it could be inferred that the viral communities within the hydrothermal vents are likely to be stratified in the same way that has been reported for bacterial communities (Dick et al., 2013).

There is however little data currently available regarding virosphere stratification in deep sea hydrothermal vents. To address this, we performed a comparison in virome composition between samples belonging to deep-hydrothermal vents (plumes and sediments), with other samples from deep-sea water (water and sediments) (Fig. 2). The clustering analysis included 21 new datasets from the Metavir2 database, and ten metagenomics datasets from the analysis of deep hydrothermal vents, and aimed to establish whether stratification of the virome is maintained independently of their hydrothermal vent origin. The clustering clearly indicates different viral families in the deep sea. While ssDNA viruses (*Circoviridae Microviridae* and *Geminiviridae*) are predominantly present in the sediment samples, the dsDNA viruses (*Myoviridae, Siphoviridae* and *Podoviridae*) are present in high abundance in the deep water samples (Fig. 2).

Regardless of the clustering of sediments and water samples, there are differences in the viromes that allow the samples that come from hydrothermal vents to be distinguished from other deep sea samples. This suggests that the viral communities of hydrothermal vents are distinct and represent unique extremophilic systems from which novel viruses may be discovered. In addition, we observed an absence of

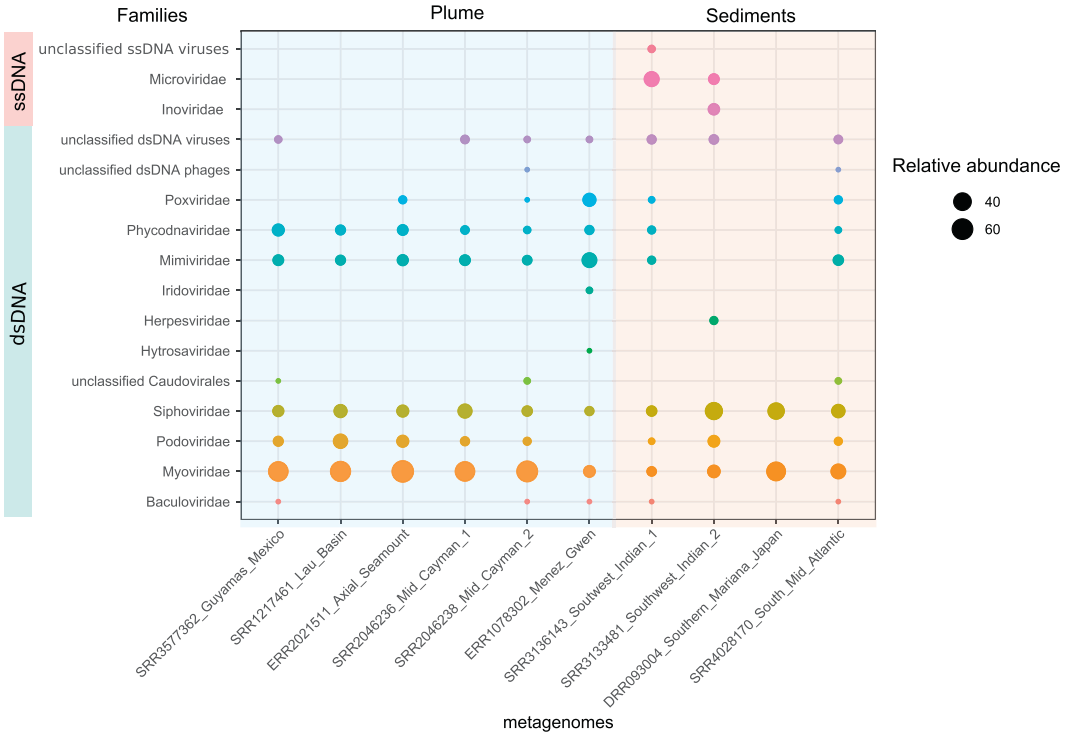

Fig. 1. Viral taxonomic composition in ten samples, which were obtained from SRA database, where correspond to six of plumes and four of sediments of deep hydrothermal vents. Relative abundance (number of contigs) of viral families dsDNA and ssDNA viruses are shown.

archaeal viruses (*Fuselloviridae*) in the metagenome data sets from the plume, sediment and deep-water samples; which have previously been reported in terrestrial high temperature environments such as the hot springs in the Yellowstone National Park (Munson-McGee et al., 2018). Therefore, it appears likely that hydrothermal vents may harbor a different viral structure that terrestrial hydrothermal ecosystems.

3.2. Biogeochemical cycles in hydrothermal vents

Coupled with the lack of information regarding the structure and composition of the virome in hydrothermal vents, the functional analysis of these communities has to date also received very little attention. However, it is clear that the functional analysis of these communities is likely to uncover a vast array of novel genes from viruses inhabiting hyperthermophilic environments such as hydrothermal vents, which may have important biotechnological applications, such as biocatalysts that are active at elevated temperatures (Frock and Kelly, 2012). In recent years, a number of genes has been identified and characterized from deep-sea viromes, including the (*PsbA*) gene encoding the D1 protein in photosystem II and the *NarG, NarH,* and *NarJ* viral nitrate reductase genes involved in the biogenesis of respiratory nitrate from hydrothermal vents (He et al., 2017; Garin-Fernandez et al., 2018). Thus, functional bioprospecting of these environments could provide an opportunity to discovery truly novel proteins.

The metagenomics data sets were firstly analyzed for the completeness of metabolic pathways which are involved in biogeochemical cycles involving carbon, nitrogen, iron and sulfur (Fig. 3), given that as previously mentioned viruses as known to play an important role in the natural recycling of these chemical elements (Weitz and Wilhelm, 2012). This analyses allowed identification of the main pathways used by the microbial community in hydrothermal vents in both the plumes and the sediments samples. This is likely to be directly related to the metabolic activities of microbial communities which are present, which mainly consist of populations involved in carbon fixation and, those involved in redox reactions of nitrogen, sulfur and iron (Eecke et al., 2012; Dick et al., 2013).

This analysis revealed 30 metabolic pathways mainly for nitrogen and sulfur with a completeness of between 80 and 100% in all the metagenomes analyzed (Fig. 3) indicating that the relevant biochemical pathways were present in both the plume water and sediment samples. Amongst these are pathways that are likely to play key ecological roles in the degradation of sulfured compounds in nature, such as: dimethylsulfoniopropionate (DMSP) oxidation, sulfoacetate oxidation, dimethylsulfone oxidation, cysteate oxidation, alkanesulfonate degradation, tetrathionate oxidation and carbon disulfide oxidation. Interestingly, others sulfur-related pathways such as those involved in sulfoquinovosyl diaglycerol (SQDG) biosynthesis and homotaurine degradation were also present in the metagenomes from both the sediments and the plumes. Genes involved in the pathway for sulfite oxidation was also present with a completeness of between 60 and 70% in all the samples with the exception of the sediment sample from Southwest India (SRR2133481).

Regarding the pathways involved in nitrogen metabolism a completeness of between 80 and 100% was observed in all metagenomes, with pathways involved in nitrate reduction (I-X), ammonia oxidation II, nitrate reduction, nitrate reductase (*nirBD*), and the superpathway ammonia being present. Genes involved in methanogenesis pathways were also present in all the metagenomics datasets including those for methanogenesis energy conversion, methanopterin (MTP) methanogenesis, dimethylsulfide (DMS) methanogenesis. However, other complete methanogenesis pathways were poorly represented in all datasets samples with only the 30% of the genes distinguishable for these pathways, while the completeness of the biochemical pathway involved in the conversion of CO_2 into methane was in the range of 80–90% of the pathway being present.

These results demonstrate that nitrogen, carbon and sulfur biogeochemical cycles participate in metabolic processes with a high ecological significance in microbial communities present in hydrothermal vents since the bacterial genomes contain a high percentage all the enzymes that participate as biocatalysts in these catabolic and anabolic

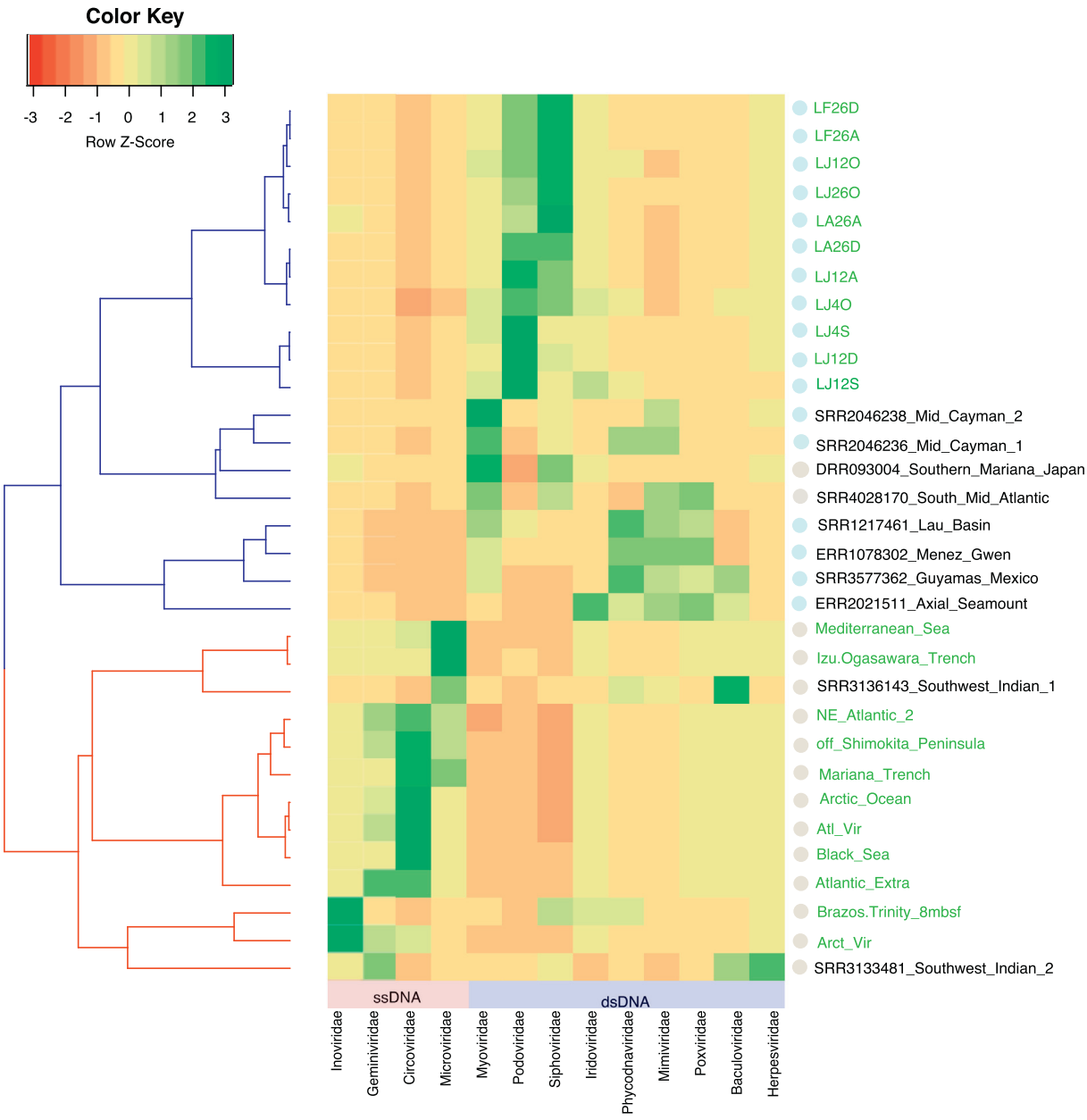

Fig. 2. Heat map of the viral communities, in deep-hydrothermal vents, and samples from deep sea. This clustering reveals two clades, one corresponding to sediments (red) and the other clade correspond to water (blue). Samples with IDs in green font are from Metavir2 and samples with IDs in black font are from the SRA database. Circles in light blue are samples from water, while circles in brown are samples from sediments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

processes (Anantharaman et al., 2015).

3.3. Functional analysis viral in plumes and sediments from hydrothermal vents

A functional analysis using MG-RAST was then performed to specifically examine the metabolic profiles in the viral communities. In general, the prophage (capsid) was the more abundant category observed (*e*-value of $1e^{-5}$), which is to be expected since these protein structures are abundant in viruses (Brum et al., 2016). This metabolic function for phase capsid synthesis was dominant in both, plume and sediment samples with 3750 and 230 sequences being present, respectively.

Genes encoding other functions including those involved in lytic and lysogenic viral cycles and, those involved in DNA repair and replication, such as the Rlt-like protein and genes involved in phosphate metabolism were also commonly found (*e*-value of $1e^{-5}$) (Fig. 4). There was a lesser diversity in the metabolic functions in the viral genomes recovered from the sediment samples.

Not only were the metabolic functions less diverse, but the number of contigs (genes) associated with a specific metabolic function was lower. While in the plume samples there were 1496 contigs associated with virion structure, only 214 were identified in the sediment samples. The same was true for DNA metabolism, with 883 and 347 coding sequences being identified in plume and sediment samples, respectively; with a *e*-value of $1e^{-5}$.

Finally, some biochemical functions were exclusively observed in viral communities from plumes, such as phosphorus uptake, folate biosynthesis, macromolecular synthesis, ribonucleotide reductase and Type II ATP dependent DNA topoisomerases, amongst others. All these

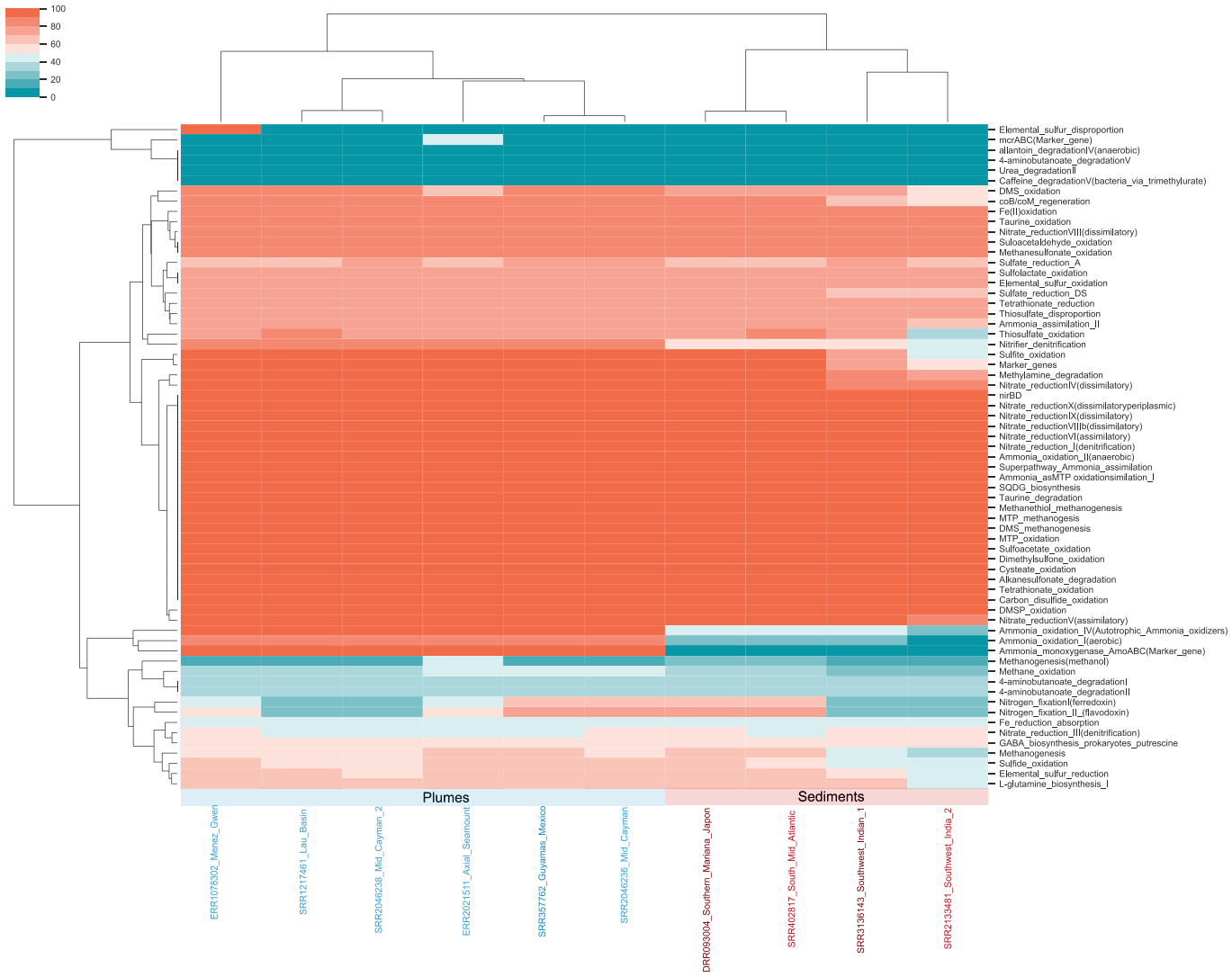

Fig. 3. Clustering hierarchical of pathways of the carbon and sulfur metabolism. Hierarchical heat map is shown, where the most pathway in biogeochemical cycles is marked in red colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are relevant for DNA replication in viral particles.

Genes involved in folate biosynthesis (thymidylate synthase thyX, (TS) was another predominant function observed in the plume datasets. This enzyme is necessary to catalyze the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), and participates in the folate cycle. It is also essential in the synthesis of methionine, together with methylation reactions (Graziani et al., 2004; Leduc et al., 2007). Thymidylate synthases have previously been reported in viruses such as *Phycodnaviridae* (Graziani et al., 2004), *Herpesviridae* and *Caudovirales* (Stern et al., 2010).

Furthermore while proteins involved in carbohydrate metabolism were not abundant; we did however identify a GDP-L-fucose synthase from *Prochlorococcus* phage P-SSM2, which is known to be involved in the biochemical synthesis of oligosaccharides (Han et al., 2012). These enzymes have been reported in *Caudovirales*, *Herpesviruses*, *Poxviruses*, *Baculoviruses* and *Phycodnaviruses* (Markine-Goriaynoff et al., 2004; Graves et al., 2001); but have not been reported from extreme environments as the virosphere from hydrothermal vents has to date not been extensively studied from a biotechnological standpoint.

This functional analysis also allowed the identification in high abundance of genes involved in auxiliary metabolic functions, with cytitidyltransferase and ribonucleotide reductase genes being the most abundant within the class Clustering-based subsystem. The former encode for nucleotidyl transferases which are typically involved in the transfer of phosphorus-containing groups, and have been reported in *Prochlorococcus* phage P-SSM2 (Sullivan et al., 2005; Sullivan et al., 2010; Aylward et al., 2017). Ribonucleotide reductases are involved in nucleotide biosynthesis and have previously been reported in many viral genomes (Sakowski et al., 2014). Moreover, *phoH* genes which are related to the acquisition of phosphate (Goldsmith et al., 2011) were observed but at lower abundances. These genes have also been previously widely reported in bacteriophages (Lindell et al., 2004).

3.4. Analysis of natural selection on auxiliary metabolic genes

In general, in the functional analysis of viruses the most abundant genes that were identified corresponded to those encoding for structural parts of the virion together with those involved in some metabolic auxiliary functions. There is particular interest in the latter sets of genes since it is known that they encode for AMGs, which are known to be involved in promoting biochemical processes and in doing so improving the fitness of their bacterial hosts; by potentially facilitating adaptation within these bacteria/archaea due to the adverse conditions present in the hydrothermal vents ecosystems (Anderson et al., 2011b; He et al., 2017). An example of the role that AMGs play has been reported in cyanobacteria, where cyanophage express their host's photosynthetic

Fig. 4. Viral functional annotation from plumes and sediments of hydrothermal vents. The most abundant correspond to capsids of phages.

genes during the lytic cycle of infection (Clokie et al., 2006), to potentially either support the host or to redirect the cyanobacterial host metabolism to support phage DNA biosynthesis (Thompson et al., 2011). Interestingly AMGs were found in high abundance during our functional analysis and given that it has been suggested that these genes may have evolved under purifying selection in hydrothermal vents (Anderson et al., 2014); a selection analysis was performed to determine which evolutionary processes dominate in the evolution of AMGs in extreme deep-sea environment. There have been very few reports of this in the literature and thus there is a deficit in our knowledge in relation to the evolution of viral genes in deep-sea ecological niches.

Concerning the AMGs, two genes with the higher relative abundance in the viral genomes were chosen, namely the cytitidyltransferase and ribonucleotide reductase genes. Fifty-one and sixty three genes encoding for these enzymes respectively were found and subsequently subjected to potential natural selection analysis. While no potential sites were identified using the REL software; however the use of MEME resulted in the detection of three and six potential sites respectively in the cytitidyltransferase genes. All these sites were predicted with

Table 3
Sites under positive selection in Cytitidiltranferase with FEL p = 0.05.

Codon	dS	dN	dN/dS	*p*-value
52	0	0.7687	Infinite	0.03576
114	0	1.2061	Infinite	0.03568
407	0	1.1755	Infinite	0.00374

Table 4
Sites under episodic selection in Cytitidiltransferese with MEME p = 0.05.

Codon	*a*	β	Pr [β =β+]	β+	Pr [β =β+]	p-value
125	0.0507	0	0.5965	2.03945	0.403492	0.014053
126	0.0907	0.0285	0.7905	7.03581	0.209416	0.0303779
407	0	0	1.0001e-09	0.380733	1	0.00920052
659	0.1596	0.1022	0.82767	139.081	0.17233	0.0326676
753	0.2383	0	0.636865	79.5711	0.363135	0.00641633
965	0	0	0.644192	40.2061	0.355808	0.00698377
1213	0	0	0.605042	1.19307	0.394958	0.0426682

statistical significance levels ($p < 0.05$). The three codons that were identified that may be under positive selection (pervasive selection) using FEL were located in positions 52, 114 and 407 (Table 3); while with MEME (episodic selection), six positions were identified namely; 125, 126, 407, 659, 753 and 1213 (Table 4). In the case of the ribonucleotide reductase genes, no evidence of positive selection was found using MEME, REL and FEL software.

Given the difficult in identifying complete genes in the viral genomes, as well as considering the limited number of sequences present, the replacement rates using evolutionary fingerprint was compared with the data of those sequences encoding viral capsid proteins and cytitidyltransferase. This analysis revealed that the substitution rates *dN/dS* evolved under a purifying selection (negative selection), as predicted by FEL, REL and MEME. However, in the sequences analyzed some changes in the non-synonymous replacement rates were observed, but these variations were subtle without exceeding the value of neutrality (Fig. 5).

In the case of AMGs, evolution under negative selection was generally also observed. This is similar to previous reports indicating that viral genes in hydrothermal system are subject to purifying selection (Anderson et al., 2011a). However, it has also been reported that the AMGs when they are transferred from the virus to their host, can evolve under positive selection (Anderson et al., 2014).

Hydrothermal vents are dynamic and fluent ecosystems, but only a small number of positions in the genes analyzed were identified as having evolved under episodic selection. This indicates that there are periods where alternating conservative selection acts and, periods of change which favor the accumulation of non-synonymous mutations thereby allowing certain adaptive advantages in those genes. This is the first analysis confirming that some genes evolve under episodic selection, and that the frequency of non-synonymous substitution indicates episodes of rapid evolution.

4. Conclusions

The structure of the virome in the hydrothermal vents allows us to distinguish specific populations and those that were present in greater abundance in the plume samples when compared with those of the sediments. The main difference in the structure appears to be due to the presence of ssDNA genomes such as *Circoviridae* and *Microviridae* in the sediment samples. In addition the viromes of the vents are very similar to other samples that have previously been analyzed from deep waters, where *Caudovirales* are ubiquitous. Genes that participate in metabolic pathways that contribute to the production of carbon, nitrogen and sulfur in the hydrothermal vents, were found in greater abundance, when compared with those involved in the oxygen cycle; indicating the types of viral populations that may be participating directly or indirectly in these cycles. On the other hand, in the functional profile of the viromes, we found that the most represented genes were those encoding for proteins involved in bacteriophage capsid synthesis, phage packaging machinery, DNA synthesis, nucleotide synthesis, DNA repair, as well as auxiliary metabolic functions. The AMGs (cytitidyltransferase and ribonucleotide reductase from viruses) which participate in the synthesis of phospholipids and are essential for the synthesis of

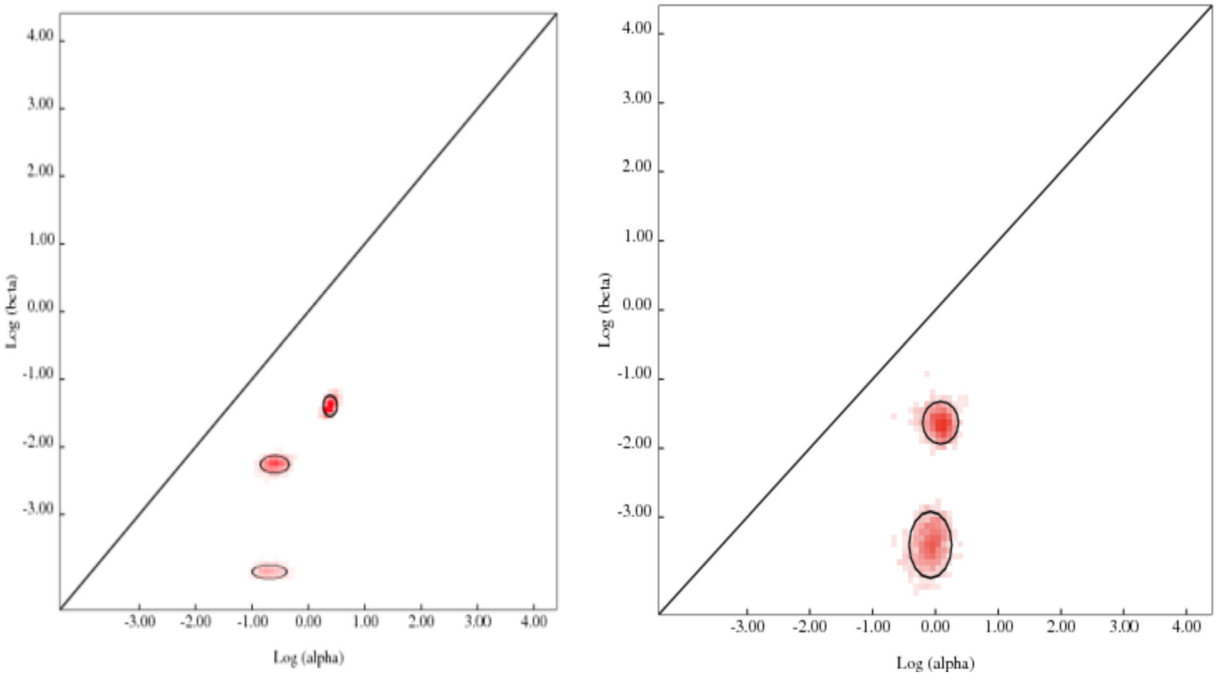

Fig. 5. Comparison of the evolutionary fingerprint *dS/dN* rates. The estimated distribution of the *dS/dN* values is shown. The diagonal indicates the neutrality state. Values located above the diagonal indicative positive selection, while values located below the diagonal indicate negative selection. A) Cytitidyltransferase genes B) Ribonucleotide reductase genes. All *dN/dS* rates are under purifying selection (negative selection).

nucleotides respectively, are likely to contribute to enhancing the fitness of their hosts within the hydrothermal vents as previously proposed (Anderson et al., 2011b; Anderson et al., 2014). The evolutionary analysis suggests that these AMGs are highly conserved and evolve under purifying selection, and are thus maintained in their genome.

Author contributions

HGCS and SDR designed and performed the analysis. RP, ILR, WAGS participated in the analysis. HGCS, SDR, ADWD and RABG analyzed and interpreted the data. All authors contributed in the preparation of the manuscript. All authors read and approved the final manuscript.

Competing interest/Conflict of interest statement

Authors declare that the research was conducted in the absence of any commercial or financial relationships. Thus, any conflict of interest exists.

Acknowledgements

Authors grateful to the Unidad de Secuenciación Masiva y Bioinformática (Instituto de Biotecnología-UNAM) and Centro de Investigación en Dinámica Celular for the access to its computer cluster. HGCS has a scholarship from CONACyT (Mexico). Authors especially grateful to the researchers that obtained the metagenomic sequences used in this manuscript, which are totally public and are deposited in NCBI web site.

References

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K., Toner, B.M., Dick, G.J., 2014. Sulfur oxidation genes in diverse deep-sea viruses. Science 344, 757–760. https://doi.org/10.1126/science.1252229.

Anantharaman, K., Breier, J.A., Dick, G.J., 2015. Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J. 10, 225–239. https://doi.org/10.1038/ismej.2015.81.

Anda, V.D., Zapata-Penasco, I., Hernandez, A.C., Fruns, L.E., Moreira, B.C., Souza, V., 2017. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: Unraveling the sulfur cycle. Gigascience 6, 1–17. https://doi.org/10.1101/191288.

Anderson, R.E., Brazelton, W.J., Baross, J.A., 2011a. Is the genetic landscape of the deep subsurface biosphere affected by viruses? Front. Microbiol. 2, 219. https://doi.org/10.3389/fmicb.2011.00219.

Anderson, R.E., Brazelton, W.J., Baross, J.A., 2011b. 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.

Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., ... Huber, J.A., 2017. Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. Nat. Commun. 8 (1), 1114.

Anderson, R.E., Sogin, M.L., Baross, J.A., 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.

Aylward, F.O., Boeuf, D., Mende, D.R., Wood-Charlson, E.M., Vislova, A., Eppley, J.M., Delong, E.F., 2017. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. Proc. Natl. Acad. Sci. 114, 11446–11451. https://doi.org/10.1073/pnas.1714821114.

Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The New S Language. Wadsworth Brooks/Cole (S version.).

Beiko, R.G., Harlow, T.J., Ragan, M.A., 2005. Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 14332–14337. https://doi.org/10.1073/pnas.0504068102.

Biddle, J.F., White, J.R., Teske, A.P., House, C.H., 2011. Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. ISME J. 5, 1038. https://doi.org/10.1038/ismej.2010.199.

Bioinformatics, B., 2011. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Institute. Biology, Cambridge, UK. https://doi.org/10.1371/journal.pbio.0030144.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Breitbart, M., Thompson, L., Suttle, C., Sullivan, M., 2007. Exploring the vast diversity of marine viruses. Oceanography 20, 135–139. https://doi.org/10.5670/oceanog.2007.58.

Brum, J.R., Ignacio-Espinoza, J.C., Kim, E.H., Trubl, G., Jones, R.M., Roux, S., Sullivan, M.B., 2016. Illuminating structural proteins in viral “dark matter” with metaproteomics. Proc. Natl. Acad. Sci. 113, 2436–2441. https://doi.org/10.1073/pnas.1525139113.

Cary, S.C., Cottrell, M.T., Stein, J.L., Camacho, F., Desbruyeres, D., 1997. Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *Alvinella pompejana*. Appl. Environ. Microbiol. 63, 1124–1130.

Cerqueira, T., Barroso, C., Froufe, H., Egas, C., Bettencourt, R., 2018. Metagenomic signatures of microbial communities in deep-sea hydrothermal sediments of Azores vent fields. Microb. Ecol. 76, 1–17. https://doi.org/10.1007/s00248-018-1144-x.

Christakis, C.A., Polymenakou, P.N., Mandalakis, M., Nomikou, P., Kristoffersen, J.B., Lampridou, D., ... Magoulas, A., 2018. Correction to: microbial community differentiation between active and inactive sulfide chimneys of the Kolumbo submarine volcano, Hellenic Volcanic Arc. Extremophiles 22 (5). https://doi.org/10.1007/s00792-018-1017-8. (825-825).

Clokie, M.R.J., Shan, J., Bailey, S., Jia, Y., Krisch, H.M., West, S., Mann, N.H., 2006. Transcription of a “photosynthetic” T4-type phage during infection of a marine cyanobacterium. Environ. Microbiol. 8, 827–835. https://doi.org/10.1111/j.1462-2920.2005.00969.x.

Corinaldesi, C., Tangherlini, M., Dell'Anno, A., 2017. From virus isolation to metagenome generation for investigating viral diversity in deep-sea sediments. Sci. Rep. 7, 8355. https://doi.org/10.1038/s41598-017-08783-4.

Dávila-Ramos, S., Estradas-Romero, A., Prol-Ledesma, R.M., Juárez-López, K., 2014. Bacterial populations (first record) at two shallow hydrothermal vents of the Mexican Pacific West coast. Geomicrobiol J. 32, 657–665. https://doi.org/10.1080/01490451.2014.980526.

Dick, G.J., Anantharaman, K., Baker, B.J., Li, M., Reed, D.C., Sheik, C.S., 2013. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. Front. Microbiol. 4, 124. https://doi.org/10.3389/fmicb.2013.00124.

Ding, J., Zhang, Y., Wang, H., Jian, H., Leng, H., Xiao, X., 2017. Microbial community structure of deep-sea hydrothermal vents on the ultraslow spreading southwest Indian ridge. Front. Microbiol. 8, 1012. https://doi.org/10.3389/fmicb.2017.01012.

Eecke, H.C., Butterfield, D.A., Huber, J.A., Lilley, M.D., Olson, E.J., Roe, K.K., Evans, L.J., Merkel, A.Y., Cantin, H.B., Holden, J.F., 2012. Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. Proc. Natl. Acad. Sci. 109, 13674–13679. https://doi.org/10.1073/pnas.1206632109.

Fortunato, C.S., Larson, B., Butterfield, D.A., Huber, J.A., 2018. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ. Microbiol. 20 (2), 769–784.

Frock, A.D., Kelly, R.M., 2012. Extreme thermophiles: moving beyond single-enzyme catalysis. Curr. Opin. Chem. Eng. 1, 1–10. https://doi.org/10.1016/j.coche.2012.07.003.

Garin-Fernandez, A., Pereira-Flores, E., Glöckner, F., Wichels, A., 2018. The North Sea goes viral: occurrence and distribution of North Sea bacteriophages. Mar. Genomics 41, 31–41. https://doi.org/10.1016/j.margen.2018.05.004.

Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G., Prieur, D., 2003. Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. Res. Microbiol. 154, 303–307. https://doi.org/10.1016/S0923-2508(03)00075-5.

Glass, E.M., Meyer, F., 2011. The Metagenomics RAST Server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. In: Handbook of Molecular Microbial Ecology I, pp. 325–331. https://doi.org/10.1002/9781118010518.ch37.

Goldsmith, D.B., Crosti, G., Dwivedi, B., Mcdaniel, L.D., Varsani, A., Suttle, C.A., Breitbart, M., 2011. Development of pho has a novel signature gene for assessing marine phage diversity. Appl. Environ. Microbiol. 77, 7730–7739. https://doi.org/10.1128/aem.05531-11.

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317. https://doi.org/10.1038/ismej.2009.72.

Gorlas, A., Koonin, E.V., Bienvenu, N., Prieur, D., Geslin, C., 2012. TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. Environ. Microbiol. 14, 503–516. https://doi.org/10.1111/j.1462-2920.2011.02662.x.

Graves, M.V., Bernadt, C.T., Cerny, R., Etten, J.L., 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. Virology 285, 332–345. https://doi.org/10.1006/viro.2001.0937.

Graziani, S., Xia, Y., Gurnon, J.R., Van Etten, J.L., Leduc, D., Skouloubris, S., Liebl, U., 2004. Functional analysis of FAD-dependent thymidylate synthase ThyX from *Paramecium bursaria Chlorella* virus-1. J. Biol. Chem. 279, 54340–54347. https://doi.org/10.1074/jbc.M409121200.

Han, N.S., Kim, T.J., Park, Y.C., Kim, J., Seo, J.H., 2012. Biotechnological production of human milk oligosaccharides. Biotechnol. Adv. 30, 1268–1278. https://doi.org/10.1016/j.biotechadv.2011.11.003.

Harmsen, H., Prieur, D., Jeanthon, C., 1997. Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment culture of thermophilic subpopulations. Appl. Environ. Microbiol. 63, 2876–2883.

Hayes, S., Mahony, J., Nauta, A., Sinderen, D.V., 2017. Metagenomic approaches to assess bacteriophages in various environmental niches. Viruses 9, 127. https://doi.org/10.3390/v9060127.

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.

Hewson, I., Barbosa, J.G., Brown, J.M., Donelan, R.P., Eaglesham, J.B., Eggleston, E.M.,

Labarre, B.A., 2012. Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes. Appl. Environ. Microbiol. 78, AE-01705. https://doi.org/10.1128/aem.01705-12.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Huber, J.A., Welch, D.B.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., Sogin, M.L., 2007. Microbial population structures in the deep marine biosphere. Science 318, 97–100. https://doi.org/10.1126/science.1146689.

Hurwitz, B.L., Sullivan, M.B., 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS One 8, e57355. https://doi.org/10.1371/journal.pone.0057355.

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. https://doi.org/10.1101/gr.5969107.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Tappu, R., 2016. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput. Biol. 12, e1004957. https://doi.org/10.1371/journal.pcbi.1004957.

Jeanthon, C., 2000. Molecular ecology of hydrothermal vent microbial communities. Antonie Van Leeuwenhoek 77, 117–133. https://doi.org/10.1023/A:1002463825025.

Kato, S., Shibuya, T., Takaki, Y., Hirai, M., Nunoura, T., Suzuki, K., 2018. Genome-enabled metabolic reconstruction of dominant chemosynthetic colonizers in deep-sea massive sulfide deposits. Environ. Microbiol. 20, 862–867. https://doi.org/10.1111/1462-2920.14032.

Kelley, D.S., Karson, J.A., Früh-Green, G.L., Yoerger, D.R., Shank, T.M., Butterfield, D.A., Jakuba, M., 2005. A serpentinite-hosted ecosystem: the Lost City hydrothermal field. Science 307, 1428–1434. https://doi.org/10.1126/science.1102556.

Labonté, J.M., Suttle, C.A., 2013. Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. Front. Microbiol. 4, 404. https://doi.org/10.3389/fmicb.2013.00404.

Ledesma, R.M., 2011. Nuevos ambientes submarinos: Ventilas hidrotermales. UNAM. México, pp. 54.

Leduc, D., Escartin, F., Nijhout, H.F., Reed, M.C., Liebl, U., Skouloubris, S., Myllykallio, H., 2007. Flavin-dependent thymidylate synthase ThyX Activity: implications for the folate cycle in bacteria. J. Bacteriol. 189, 8537–8545. https://doi.org/10.1128/jb.01380-07.

Leinonen, R., Sugawara, H., Shumway, M., 2010. The sequence read archive. Nucleic Acids Res. 39, D19–D21. https://doi.org/10.1093/nar/gkq1019.

Li, M., Baker, B.J., Anantharaman, K., Jain, S., Breier, J.A., Dick, G.J., 2015. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat. Commun. 6, 8933. https://doi.org/10.1038/ncomms9933.

Li, D., Luo, R., Liu, C., Leung, C., Ting, H., Sadakane, K., Lam, T., 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., Chisholm, S.W., 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc. Natl. Acad. Sci. 101, 11013–11018. https://doi.org/10.1073/pnas.0401526101.

Lossouarn, J., Dupont, S., Gorlas, A., Mercier, C., Bienvenu, N., Marguet, E., Geslin, C., 2015. An abyssal mobilome: viruses, plasmids and vesicles from deep-sea hydrothermal vents. Res. Microbiol. 166, 742–752. https://doi.org/10.1016/j.resmic.2015.04.001.

Manini, E., Luna, G.M., Corinaldesi, C., Zeppilli, D., Bortoluzzi, G., Caramanna, G., Raffa, F., Danovaro, R., 2007. Prokaryote diversity and virus abundance in shallow hydrothermal vents of the Mediterranean Sea (Panarea Island) and the Pacific Ocean (North Sulawesi-Indonesia). Microb. Ecol. 55, 626–639. https://doi.org/10.1007/s00248-007-9306-2.

Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., Vanderplasschen, A., 2004. Glycosyltransferases encoded by viruses. J. Gen. Virol. 85, 2741–2754. https://doi.org/10.1099/vir.0.80320-0.

Martin, W., Baross, J., Kelley, D., Russell, M.J., 2008. Hydrothermal vents and the origin of life. Nat. Rev. Microbiol. 6, 805–814. https://doi.org/10.1038/nrmicro1991.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Richie, K.B., Paul, J.H., 2010. High frequency of horizontal gene transfer in the oceans. Science 330, 50. https://doi.org/10.1126/science.1192243.

Meier, D.V., Bach, W., Girguis, P.R., Gruber-Vodicka, H.R., Reeves, E.P., Richter, M., Vidoudez, C., Amann, R., Meyerdierks, A., 2016. Heterotrophic Proteobacteria in the vicinity of diffuse hydrothermal venting. Environ. Microbiol. 18, 4348–4368. https://doi.org/10.1111/1462-2920.13304.

Millard, A.D., Hands-Portman, I., Zwirglmaier, K., 2014. Morphotypes of virus-like particles in two hydrothermal vent fields on the East Scotia Ridge, Antarctica. Bacteriophage 4, e20732. https://doi.org/10.4161/bact.28732.

Mizuno, C.M., Ghai, R., Saghaï, A., López-García, P., Rodriguez-Valera, F., 2016. Genomes of abundant and widespread viruses from the deep ocean. MBio. 7, e00805–e00816. https://doi.org/10.1128/mBio.00805-16.

Munson-McGee, J.H., Snyder, J.C., Young, M.J., 2018. Archaeal viruses from high-temperature environments. Genes 9, 128. https://doi.org/10.3390/genes9030128.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Pond, S.L., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, 1002762. https://doi.org/10.1371/journal.pgen.1002764.

Ortmann, A.C., Suttle, C.A., 2005. High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. Deep-Sea Res. I Oceanogr. Res. Pap. 52, 1515–1527. https://doi.org/10.1016/j.dsr.2005.04.002.

Pjevac, P., Meier, D.V., Markert, S., Hentschker, C., Schweder, T., Becher, D., Gruber-Vodicka, H.R., Richter, M., Bach, W., Amann, R., Meyerdierks, A., 2018. Metaproteogenomic profiling of microbial communities colonizing actively venting hydrothermal chimneys. Front. Microbiol. 9, 680. https://doi.org/10.3389/fmicb.2018.00680.

Poli, A., Finore, I., Romano, I., Gioiello, A., Lama, L., Nicolaus, B., 2017. Microbial diversity in extreme marine habitats and their biomolecules. Microorganisms 5, 25. https://doi.org/10.3390/microorganisms5020025.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533. https://doi.org/10.1093/bioinformatics/bti320.

Pond, S.L., Scheffler, K., Gravenor, M.B., Poon, A.F., Frost, S.D., 2009. Evolutionary fingerprinting of genes. Mol. Biol. Evol. 27, 520–536. https://doi.org/10.1093/molbev/msp260.

Prangishvili, D., 2003. Evolutionary insights from studies on viruses of hyperthermophilic archaea. Res. Microbiol. 154, 289–294. https://doi.org/10.1016/s0923-2508(03)00073-1.

Rastelli, E., Corinaldesi, C., Dell'Anno, A., Tangherlini, M., Martorelli, E., Ingrassia, M., Chiocci, F., Lo Martire, M., Danovaro, R., 2017. High potential for temperate viruses to drive carbon cycling in chemoautotrophy-dominated shallow-water hydrothermal vents. Environ. Microbiol. 19, 4432–4446. https://doi.org/10.1111/1462-2920.13890.

Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T., Young, M.J., 2001. Viruses from extreme thermal environments. Proc. Natl. Acad. Sci. 98, 13341–13345. https://doi.org/10.1073/pnas.231170198.

Rohwer, F., Prangishvili, D., Lindell, D., 2009. Roles of viruses in the environment. Environ. Microbiol. 11, 2771–2774. https://doi.org/10.1111/j.1462-2920.2009.02101.x.

Romancer, M.L., Gaillard, M., Geslin, C., Prieur, D., 2006. Viruses in extreme environments. Life Extreme Environ. 99–113. https://doi.org/10.1007/978-1-4020-6285-8_6.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., Colombet, J., Sime-Ngando, T., Debroas, D., 2012a. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One 7, e33641. https://doi.org/10.1371/journal.pone.0033641.

Roux, S., Krupovic, M., Poulet, A., Debroas, D., Enault, F., 2012b. Evolution and diversity of the microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One 7, e40418. https://doi.org/10.1371/journal.pone.0040418.

Roux, S., Tournayre, J., Mahul, A., Debroas, D., Enault, F., 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinform. 15, 76. https://doi.org/10.1186/1471-2105-15-76.

Sakowski, E.G., Munsell, E.V., Hyatt, M., Kress, W., Williamson, S.J., Nasko, D.J., Polson, S.W., Wommack, K.E., 2014. Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. Proc. Natl. Acad. Sci. 111, 15786–15791. https://doi.org/10.1073/pnas.1401322111.

Sepulveda, B.P., Redgwell, T., Rihtman, B., Pitt, F., Scanlan, D.J., Millard, A., 2016. Marine phage genomics: the tip of the iceberg. FEMS Microbiol. Lett. 363, fnw158. https://doi.org/10.1093/femsle/fnw158.

Sheik, C.S., Anantharaman, K., Breier, J.A., Sylvan, J.B., Edwards, K.J., Dick, G.J., 2015. Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back-arc basin. ISME J. 9, 1434–1445. https://doi.org/10.1038/ismej.2014.228.

Sime-Ngando, T., 2014. Environmental bacteriophages: viruses of microbes in aquatic ecosystems. Front. Microbiol. 5, 355. https://doi.org/10.3389/fmicb.2014.00355.

Stern, A., Mayrose, I., Penn, O., Shaul, S., Gophna, U., Pupko, T., 2010. An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. Syst. Biol. 59, 212–225. https://doi.org/10.1093/sysbio/syp104.

Strazzulli, A., Fusco, S., Cobucci-Ponzano, B., Moracci, M., Contursi, P., 2017. Metagenomics of microbial and viral life in terrestrial geothermal environments. Rev. Environ. Sci. Biotechnol. 16, 425–454. https://doi.org/10.1007/s11157-017-9435-0.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., Chisholm, S.W., 2005. Three *Prochlorococcus Cyanophage* genomes: signature features and ecological interpretations. PLoS Biol. 3, e144. https://doi.org/10.1371/journal.pbio.0030144.

Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., DeFrancesco, A.S., Kern, S.E., Thompson, L.R., Young, S., Yandava, C., 2010. Genomic analysis of oceanic cyanobacterias myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ. Microbiol. 12, 3035–3056. https://doi.org/10.1111/j.1462-2920.2010.02280.x.

Tarasov, V., Gebruk, A., Mironov, A., Moskalev, L., 2005. Deep-sea and shallow-water hydrothermal vent communities: two different phenomena? Chem. Geol. 224, 5–39. https://doi.org/10.1016/j.chemgeo.2005.07.021.

Team, R.C., 2013. R: A Language and Environment for Statistical Computing.

Tetz, G., Tetz, V., 2018. Bacteriophages as new human viral pathogens. Microorganisms 6, 54. https://doi.org/10.3390/microorganisms6020054.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc. Natl. Acad. Sci. U. S. A. 108, E757–E764. https://doi.org/10.1073/pnas.1102164108.

Weitz, J., Wilhelm, S., 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol. Rep. 4, 17. https://doi.org/10.3410/b4-17.

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

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S.M., Meng, J., Huang, G., Li, Y., Yan, Q., Wu, S., Wang, X., Chen, S., He, G., Xiao, X., Xu, A., 2010. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with

contrasting chemistries. ISME J. 5, 414–426. https://doi.org/10.1038/ismej.2010.144.

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., Takai, K., 2013. Metagenomic analysis of viral communities in (hado) pelagic sediments. PLoS One 8, e57271. https://doi.org/10.1371/journal.pone.0057271.

Yoshida, M., Mochizuki, T., Urayama, S., Yoshida-Takashima, Y., Nishi, S., Hirai, M., Nomaki, H., Takaki, Y., Nunoura, T., Takai, K., 2018. Quantitative viral community DNA analysis reveals the dominance of single-stranded dna viruses in Offshore Upper Bathyal sediment from Tohoku, Japan. Front. Microbiol. 9, 75. https://doi.org/10.3389/fmicb.2018.00075.

Zhang, W., Zhou, J., Liu, T., Yu, Y., Pan, Y., Yan, S., Wang, Y., 2015. Four novel algal virus genomes discovered from Yellowstone Lake metagenomes. Sci. Rep. 5, 15131. https://doi.org/10.1038/srep15131.

Zhang, L., Kang, M., Xu, J., Xu, J., Shuai, Y., Zhou, X., Ma, K., 2016a. Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India Ridge. Sci. Rep. 6, 25982. https://doi.org/10.1038/srep25982.

Zhang, W., Sun, J., Cao, H., Tian, R., Cai, L., Ding, W., Qian, P., 2016b. Post-translational modifications are enriched within protein functional groups important to bacterial adaptation within a deep-sea hydrothermal vent environment. Microbiome 4, 49. https://doi.org/10.1186/s40168-016-0194-x.