CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	A review on viral metagenomics in extreme environments
Authors	Dávila-Ramos, Sonia;Castelán-Sánchez, Hugo G.;Martínez-Ávila, Liliana;Sánchez-Carbente, María del Rayo;Peralta, Raúl;Hernández-Mendoza, Armando;Dobson, Alan D. W.;Gonzalez, Ramón A.;Pastor, Nina;Batista-García, Ramón Alberto
Publication date	2019-10-18
Original Citation	Dávila-Ramos, S., Castelán-Sánchez, H. G., Martínez-Ávila, L., Sánchez-Carbente, M. d. R., Peralta, R., Hernández-Mendoza, A., Dobson, A. D. W., Gonzalez, R. A., Pastor, N. and Batista-García, R. A. (2019) 'A Review on Viral Metagenomics in Extreme Environments', Frontiers in Microbiology, 10, 2403. (19pp.) DOI: 10.3389/fmicb.2019.02403
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.frontiersin.org/articles/10.3389/fmicb.2019.02403/full - 10.3389/fmicb.2019.02403
Rights	©2019 Dávila-Ramos, Castelán-Sánchez, Martínez-Ávila, Sánchez-Carbente, Peralta, Hernández-Mendoza, Dobson, Gonzalez, Pastor and Batista-García. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms. - http://creativecommons.org/licenses/by/4.0/
Download date	2026-10-29 01:26:54
Item downloaded from	https://hdl.handle.net/10468/8841

UCC

University College Cork, Ireland

Coláiste na hOllscoile Corcaigh

frontiers in Microbiology

REVIEW
published: 18 October 2019
doi: 10.3389/fmicb.2019.02403

A Review on Viral Metagenomics in Extreme Environments

*Sonia Dávila-Ramos[1], Hugo G. Castelán-Sánchez[1], Liliana Martínez-Ávila[1], María del Rayo Sánchez-Carbente[2], Raúl Peralta[1], Armando Hernández-Mendoza[1], Alan D. W. Dobson[3,4], Ramón A. Gonzalez[1], Nina Pastor[1] and Ramón Alberto Batista-García[1]**

[1] Centro de Investigación en Dinámica Celular, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, [2] Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, [3] School of Microbiology, University College Cork, Cork, Ireland, [4] Environmental Research Institute, University College Cork, Cork, Ireland

OPEN ACCESS

Edited by:
Manuel Martinez Garcia,
University of Alicante, Spain

Reviewed by:
Alon Philosof,
California Institute of Technology,
United States
Mario López-Pérez,
Universidad Miguel Hernández
de Elche, Spain
Tiong Gim Aw,
Tulane University, United States

***Correspondence:**
Ramón Alberto Batista-García
rabg@uaem.mx;
rbatista25@yahoo.com

Specialty section:
This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 14 May 2019
Accepted: 04 October 2019
Published: 18 October 2019

Citation:
Dávila-Ramos S,
Castelán-Sánchez HG,
Martínez-Ávila L,
Sánchez-Carbente MdR, Peralta R,
Hernández-Mendoza A,
Dobson ADW, Gonzalez RA, Pastor N
and Batista-García RA (2019) A
Review on Viral Metagenomics
in Extreme Environments.
Front. Microbiol. 10:2403.
doi: 10.3389/fmicb.2019.02403

Viruses are the most abundant biological entities in the biosphere, and have the ability to infect Bacteria, Archaea, and Eukaryotes. The virome is estimated to be at least ten times more abundant than the microbiome with 10^7 viruses per milliliter and 10^9 viral particles per gram in marine waters and sediments or soils, respectively. Viruses represent a largely unexplored genetic diversity, having an important role in the genomic plasticity of their hosts. Moreover, they also play a significant role in the dynamics of microbial populations. In recent years, metagenomic approaches have gained increasing popularity in the study of environmental viromes, offering the possibility of extending our knowledge related to both virus diversity and their functional characterization. Extreme environments represent an interesting source of both microbiota and their virome due to their particular physicochemical conditions, such as very high or very low temperatures and >1 atm hydrostatic pressures, among others. Despite the fact that some progress has been made in our understanding of the ecology of the microbiota in these habitats, few metagenomic studies have described the viromes present in extreme ecosystems. Thus, limited advances have been made in our understanding of the virus community structure in extremophilic ecosystems, as well as in their biotechnological potential. In this review, we critically analyze recent progress in metagenomic based approaches to explore the viromes in extreme environments and we discuss the potential for new discoveries, as well as methodological challenges and perspectives.

Keywords: metagenomic, virosphere, extreme environment, viral gene bioprospection, extremophile virome

INTRODUCTION

Viruses are the most abundant biological entities in the planet, from the world oceans to the most extreme environments found in the biosphere (Zhang et al., 2018; Graham et al., 2019). Historically, the study of viral communities has been carried out by co-culture of viruses and their cellular hosts (Tennant et al., 2018), and more recently by viral metagenomic-based approaches (Nooij et al., 2018; Graham et al., 2019).

The exploration of viral populations in extreme environments has uncovered considerable genetic complexity and diversity. The biological organisms that inhabit extreme environments are

termed extremophiles, and are found in all three domains of life (Merino et al., 2019). Like all other organisms, extremophiles serve as hosts for viral replication (Castelán-Sánchez et al., 2019). As viruses depend on a cellular host for replication, the interactions with their hosts affect microbial diversity, population interactions and dynamics, and even the genomes of these hosts (Le Romancer et al., 2006; Zhang et al., 2018; Castelán-Sánchez et al., 2019). In extreme environments their impact extends from influencing microbial evolution to playing an indirect but significant role in the earth's biogeochemical cycles (Weitz and Wilhelm, 2012; Munson-McGee et al., 2018). However, despite their relevance, little is currently known about their ubiquity and diversity in extreme ecosystems (Paez-Espino et al., 2016; Berliner et al., 2018).

Nowadays the study of viruses can be carried out using metagenomic-based strategies that do not depend on cell culture approaches (Rose et al., 2016; Nooij et al., 2018; Zhang et al., 2018; Graham et al., 2019). Metagenomics represent a unique opportunity to describe the composition of viral communities in extreme environments, as well as to analyze the viral genetic reservoirs to characterize novel proteins and bioactive compounds of potential biotechnological utility.

Metagenomic studies are providing new sequences that in many cases do not share homology with sequences deposited in the reference databases (Hayes et al., 2017; Cantalupo and Pipas, 2019; Kinsella et al., 2019). It is evident that metagenomics from extreme habitats could be a powerful method to drastically increase the number of virus reported to date. It is surprising that despite the recent rapid advances in high-throughput sequencing techniques, there are still quite a limited number of studies describing the viromes of extreme environments in the literature. Here, we critically analyze recent progress in metagenomic-based approaches to explore the viromes in extreme environments, as well as methodological challenges and perspectives.

PERSPECTIVES ON SAMPLING AND PROCESSING: METHODOLOGICAL CHALLENGES FOR VIRAL METAGENOMICS IN EXTREME ENVIRONMENTS

Viral metagenomic studies are dependent on the ability to obtain sufficient amounts of nucleic acids from complex mixtures (Roux et al., 2019), particularly in extreme environmental samples that are as diverse as hot springs (Schoenfeld et al., 2008; Zablocki et al., 2017b), deep seawater, marine-sediments and oceanic basement (Breitbart et al., 2004; Hurwitz et al., 2013; Nigro et al., 2017), Antarctic and desert soils (Zablocki et al., 2014, 2017a), among others, to facilitate either the construction of metagenomic libraries or to perform direct sequencing.

The number of viral particles estimated to be present in a liter of water or kilogram of soil is in the order of 10^9–10^{11}, while the world's oceans are estimated to contain up to 10^6 viral particles per ml (Hara et al., 1991; Mokili et al., 2012). In contrast, the viral abundance from Octopus hot spring water from Yellowstone National Park or oceanic basement samples are at the lower range of $\sim 10^4$ and $\sim 10^5$ viral particles per ml respectively, compared with non-extreme aquatic environments (Schoenfeld et al., 2008; Nigro et al., 2017). In spite of viral particle abundance, during purification only subnanograms of viral DNA or RNA are typically recovered (Van Etten et al., 2010). Considering that a phage contains $\sim 10^{-17}$ g of DNA per particle, obtaining the amount of 1–5 μg of DNA required for standard pyrosequencing would implicate that $\sim 3 \times 10^{11}$ viral particles should be recovered (Thurber et al., 2009), and even for third generation PacBio technology 10 μg of DNA is required (Faino et al., 2015). With the development of new sequencing technologies it is likely that lower amounts of nucleic acids will be needed. For example, only 50 ng of DNA was required to sequence bacteriophages and archaeal viruses from hypersaline environments (Motlagh et al., 2017; Liu et al., 2019) and surprisingly only 1 ng of DNA was needed to explore the metavirome from deep sea and the chaotropic salt lake Salar de Uyuni using KAPA Hyper Prep Kit and Nextera XT kits with Illumina platforms (Hirai et al., 2017; Ramos-Barbero et al., 2019).

Filtration and Concentration of Viral Particles

To overcome the problem of obtaining sufficient viral nucleic acid amounts in extreme habitats, viral-particles must be concentrated by ultracentrifugation, flocculation, or filtration while minimizing contamination from prokaryotic or eukaryotic nucleic acids (Mokili et al., 2012; Liu et al., 2019; Ramos-Barbero et al., 2019; Roux et al., 2019). The use of classical size-selective ultrafiltration methods is not widely used, as the filters can often become blocked by impurities during concentration of the samples. Instead Tangential Flow Filtration (TFF) and/or ultracentrifugation were preferentially used in samples from hot springs and chaotropic salt lake Salar de Uyuni (Diemer and Stedman, 2012; Zablocki et al., 2017b; Ramos-Barbero et al., 2019). An excellent review by Lawrence and Steward (2010) on centrifugation highlights the efficiency of the methodology to sediment even the smallest viruses, where centrifugal separations can be divided in differential pelleting and zonal separations. The former has been successfully used to remove cell debris from samples obtained from enrichment cultures of archaeal viruses from an acidic hot spring Umi Jigoku in Beppu (Japan) (Liu et al., 2019) before concentrating and purifying viral particles and the latter has been tested using different gradient materials such as glycerol, OptiPrep and sucrose to isolate virus from a boreal lake in Finland (Laanto et al., 2017).

Due to the limitation of the volume size of aquatic samples, John et al. (2011) introduced the use of $FeCl_3$ flocculation to concentrate viruses from seawater that results in the recovery of 92–95% of viruses, which has been favorably used in samples from glacier waters and deep-sea (Bellas et al., 2015; Poulos et al., 2018). This compares favorably with traditional centrifugation or TFF, which results in recovery levels of 18–26% and 62–93% respectively (Furtak et al., 2016), as evaluated by SYBR Gold

staining, meaning that the use of $FeCl_3$ increases the efficiency of viral particle recovery from extreme environment samples to around 30–60%.

A variation of the ultracentrifugation technique uses certain compounds to precipitate viruses like PolyEthylene Glycol (PEG) and ethanol, followed by purification through CsCl gradient ultracentrifugation (Fancello et al., 2013; Kleiner et al., 2015; Rastrojo and Alcamí, 2017; Chatterjee et al., 2019). These strategies have been implemented during viral recovery from Artic and Antarctic polar habitats, hot spring lakes and hypersaline lakes (Diemer and Stedman, 2012; Motlagh et al., 2017; Rastrojo and Alcamí, 2017; Mizuno et al., 2019).

Filters in the range of 0.1–0.22 μm have been used to enrich samples from South African hot springs and igneous crust of the seafloor (Nigro et al., 2017; Zablocki et al., 2017b). However, in recent years giant viruses –*giruses*- (particle size of ∼720 nm) have been discovered (Thurber et al., 2009; Van Etten et al., 2010) leading some groups to use 0.45 μm filters that are effective in recovering these larger viral-particles (Van Etten et al., 2010; Hurwitz et al., 2013; Sangwan et al., 2015). Up to now, limited knowledge about giant viruses in extreme niches has been produced. For example, a new large DNA virus named Medusavirus, was isolated from hot spring water in Japan using a filter of 1.2 μm (Yoshikawa et al., 2019). In addition, 64 members of the *Mimiviridae* family were recently identified in Antarctic marine water (Andrade et al., 2018). Thus, selective filtration strategies should be considered to recover extreme giant viruses.

Despite the use of a variety of different approaches to enrich viral particles from extreme environments, systematic studies comparing these different concentration methods (TFF, FeCl3, PEG, commercial concentrators) are still lacking and likely the methods employed for viral enrichment may need to be adapted considering the nature of the sample.

Nuclease Treatment, Concentration and Viral Nucleic Acid Purification

Viral samples are usually treated with DNAse I, to avoid contamination with cellular genomic DNA that would, following sequence-based analysis, result in a large number of spurious DNA sequences from sources other than the virome. This treatment was used to obtain viral metagenomic DNA from Boiling Springs Lake (United States) and Great Salt Lake (thermophile and hypersaline ecosystems, respectively) (United States) (Diemer and Stedman, 2012; Motlagh et al., 2017). However, there are examples, such as in deep-sea ocean viral metagenomes (Hurwitz et al., 2013), desert perennial ponds (Fancello et al., 2013), hot springs (Zablocki et al., 2017b) among others, where despite the use of DNase treatment prior to viral genome purification, it was not possible to eliminate completely the cellular genome.

After DNase treatment, concentration steps are recommended using CsCl, sucrose or Cs_2SO_4 gradients by ultracentrifugation (Thurber et al., 2009; Fancello et al., 2013; Bellas et al., 2015). Additionally, once the viral particles have been concentrated and purified, the capsids have to be broken to release the viral genomes. The classical method is the use of formamide (Breitbart et al., 2002; Thurber et al., 2009; Fancello et al., 2013) followed by phenol:chloroform:isoamylic alcohol extraction (Diemer and Stedman, 2012; Nigro et al., 2017) or alternatively, through thermal shock (Bellas et al., 2015; Roux et al., 2016; Motlagh et al., 2017). However, in samples from hypersaline ponds, thermal shock may not be fully efficient to denude enveloped viral DNA, which may be the reason why in some studies the majority of viral DNA has been recovered from non-enveloped tailed viruses (Roux et al., 2016). Some single-stranded DNA (ssDNA) extreme viruses (e.g., HaloRubrum Pleomorphic ssDNA Virus 1, Haloarcula Hispanica Pleomorphic Virus 3, Aeropyrum Coil-shaped Virus) infecting hyperhalophile or hyperthermophile archaeal hosts, present a lipid envelope and multiprotein complexes or two criss-crossed halves of a circular nucleoprotein (Pietilä et al., 2010; Mochizuki et al., 2012; Demina et al., 2016) that could confer resistance to capsid disassembly. Thus, the capsid composition of extremophile viruses is a relevant feature of unknown viruses, as well as to access their genetic material, and consequently limits the identification of unusual extreme morphotypes.

Retrotranscription or Amplification Steps

The identification of viral genomes to date has mainly focused on ssDNA or double-stranded DNA (dsDNA) viruses and only small RNA genomes of 5–10 kb have been assembled from extreme metaviromes. For example, RNA viruses infecting archaea were discovered from an acidic hot spring in Yellowstone (United States) (Bolduc et al., 2012; Wang et al., 2015), as well as from alkaline hot springs (Schoenfeld et al., 2008). In addition, RNA cyanophages have been recently reported from Porcelana hot spring in Chilean Patagonia (Guajardo-Leiva et al., 2018).

Andrews-Pfannkoch et al. (2010) have implemented the use of hydroxyapatite chromatography to efficiently fractionate dsDNA, ssDNA, dsRNA, and ssRNA genomes of known bacteriophages from samples of marine environments. This methodology has been employed to study ssDNA viruses from deep-sea sediments, alkaline siliceous hot springs and Artic shelf seafloor (Yoshida et al., 2013; Nguyen and Landfald, 2015; Schoenfeld et al., 2015), but to our knowledge it has not been applied to study RNA viruses from extreme ecosystems.

When working with RNA viruses, a retro-transcription step is required previous to library preparation, and if the efficiency of the nucleic acid recovery is low, amplification strategies are required. Among these, phi29 polymerase-based multiple displacement amplification and random PCR using modified versions of Sequence Independent Single-Primer Amplification (SISPA) have been useful in the virome amplification in samples from hot acidic lakes, hot springs and polar aquatic environments (Diemer and Stedman, 2012; Mead et al., 2017; Yau and Seth-Pasricha, 2019). When the viral genome material is RNA, a Random-Priming SISPA (RP-SISPA) method is frequently used (Miranda et al., 2016). This approach was successfully conducted in the isolation of RNA viruses from seawater (Steward et al., 2013) and Antarctic virioplankton (Miranda et al., 2016).

Other strategies for viral amplification are also used when extremophile metaviromes are studied. While the implementation of the Linker Amplified Shotgun Library

methodology (LASL) is suggested to amplify dsDNA, Multiple Displacement Amplification (MDA) is employed to enrich ssDNA preferentially (Fancello et al., 2013). LASL has been performed to analyze viral metegenomes from Yellowstone hot springs and Antarctic virioplankton (Schoenfeld and Mead, 2015; Miranda et al., 2016), while MDA has been used to describe the virome present in deep-sea samples from Antarctica (Gong et al., 2018).

Zablocki et al. (2016) have argued that although viral amplification is commonly used in metavirome studies, especially for samples collected from extreme habitats such as hyperarid desert soils, this step should be avoided because it prevents the determination of viral particle abundance and diversity, and may promote a biased amplification of certain virus groups.

Thus, it is clear that further comparative methodological studies using samples from extreme environments are required to evaluate if purification, concentration and amplification methods have any impact in the virome structure obtained from metagenome analysis.

DATABASE AND BIOINFORMATIC ANALYSIS: GENERAL REMARKS

Up to now viral sequence search is conducted essentially on the NCBI database GenBank or RefSeq, according to their viral sequence classification criteria. The RefSeq database excludes some categories of data such as those that incorporate too much information that cannot be processed readily, such as metagenomes or genomes that have significant mismatch or indel variation compared to other closely related genomes. In addition, not all sequences have a taxonomic classification in the International Committee on Taxonomy of Viruses (ICTV) (O'Leary et al., 2015). The number of viral sequences reported in GenBank, reached almost two million by December 2018, of which only 3,279 were registered as genomes in RefSeq and of these, only 1,800 have a classification at the species level in the ICTV (Kang and Kim, 2018). The classification of viruses in ICTV has been based on the characteristics that can be used to distinguish one virus from another, such as the genome composition, the capsid structure, the gene expression program during viral replication, host range and pathogenicity, among others. Comparisons of both pairwise sequence similarity and phylogenetic relationships have become the primary guidelines used to define virus taxa (Simmonds, 2015). However, without the incorporation of metagenomic data in both the RefSeq and the ICTV databases, the comparison of sequences and their allocation is limited (Simmonds et al., 2017). Alternative viral sequence similarity search strategies, such as VirSorter (Roux et al., 2015) and VirFinder (Ren et al., 2017) have been developed. The former is designed to search protein-coding genes and the latter works with k-mer composition, both attempting to identify viral sequences in prokaryotic genomes. Integration of such strategies should reduce the number of unidentified sequences and the comparisons of viromes should then help to formulate more robust theories about their biological roles within a given community, thereby increasing the possibility of gaining a fuller understanding of the viromes in any environment (Simmonds et al., 2017).

Virus databases have been developed, such as Virome with 73 projects from viromes or metagenomes currently containing data from 270 libraries (Wommack et al., 2012); EBI metagenomics, a virome dataset and pipelines for analysis of metagenomes (Hunter et al., 2014); Metavir2, which is a web server for analysis of environmental viromes (Roux et al., 2014); and more recently, IMG/VR v.2.0 (Paez-Espino et al., 2018) that includes >600 extreme environmental metaviromes; Gut Virome Database (GVD) with 648 viral or microbial metagenomes (Gregory et al., 2019); iVirus (based on vConTACT as the main classification tool) which contains a dataset from 1,866 samples and 73 ocean expedition projects (Bolduc et al., 2017) (**Figure 1**). Some of them include viral sequences obtained through strategies such as the construction of fosmid libraries (Mizuno et al., 2013), cellular fraction of metagenomes (López-Pérez et al., 2017) or single viral genomics (Martinez-Hernandez et al., 2017), which enrich the virome sequences further. However, none of the above databases is particularly dedicated to viromes from extreme environments. Despite the limitations described above a comparative analysis of the population structure of viromes in extreme environments was carried out here, using publicly accessible virus metagenomic libraries as an attempt to exemplify the results that may be obtained using available tools and information. The data deposited in MetaVir2 until 2016 were selected because its user-friendly interface, which allows access to raw data or contig metagenome samples that contain well-classified metadata. We could select the 17 studies in MetaVir2 database that contain 66 viral metagenomes collected from most representative extreme environments: deep-sea (24), oxygen minimum zones (OMZ) (4), arid habitats (9), saline niches (23), cold environments (3) and hyperthermophile regions (3). The bioinformatic pipeline used was common to all data, so the comparison between environments relied on the same criteria. MetaVir2 followed two strategies to search the contigs in each sample: BLAST search in the RefSeq Virus database with the best-hit selection, and search for k-mer composition using di, tri or tetra nucleotides comparisons (Willner et al., 2009).

VIROMES IN EXTREME ENVIRONMENTS

Comparison of Viromes Between Extreme Habitats

The relative abundances from these data were analyzed comparing the similarities between environments. Their metadata are summarized in **Supplementary Table S1**.

The structure of the viral population in all the metagenomes analyzed by Metavir2 was compared (**Figures 2**, **3**). The 10 most abundant families are represented in **Figure 2** and the rest in **Figure 3** to visualize the differences in abundance of each family. Some families of the order *Caudovirales* are ubiquitous and the most abundant were the *Siphoviridae*, *Myoviridae,* and *Podoviridae* as expected, since the viruses that belong to these families infect a wide range of bacterial hosts from more than 140 prokaryotic genera (Konstantinidis et al., 2009;

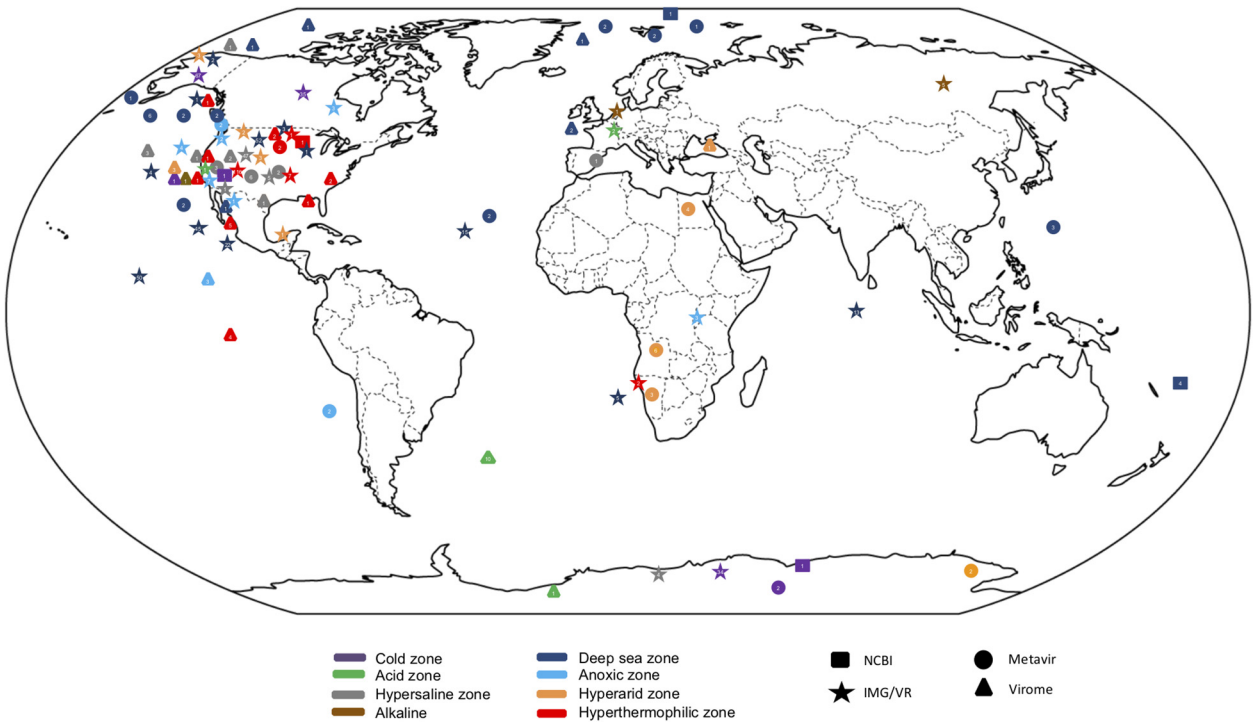

FIGURE 1 | Global distribution of metagenomics studies from extreme environments from public databases. Circles represent metagenomes deposited in Metavir database, stars in IMG/VR, triangles in Virome and squares in NCBI.

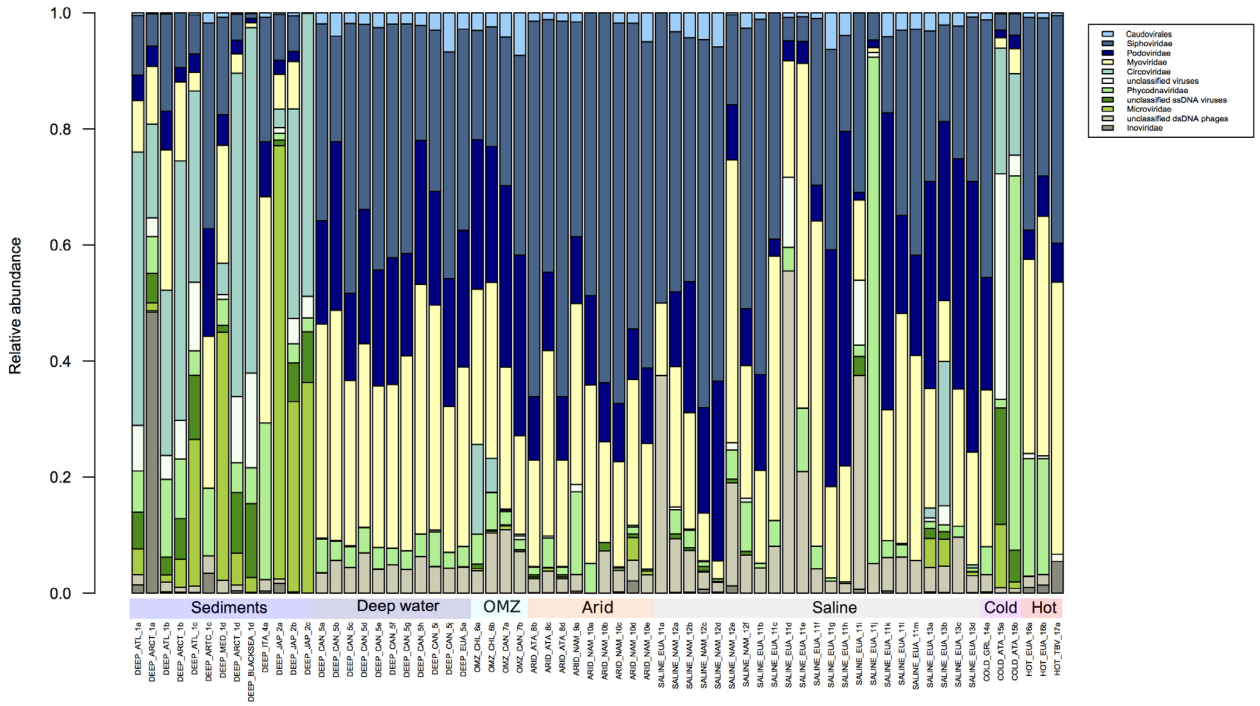

FIGURE 2 | Comparison of the 10 most abundant virus families according to the Metavir database. The taxonomic composition is expressed in relative abundance at the virus family level. The families *Siphoviridae*, *Podoviridae,* and *Myoviridae* are ubiquitous in extreme environments. The figure was constructed from an abundance matrix, using the number of sequences reported in the Metavir database, from which the relative abundance was obtained; using the R program.

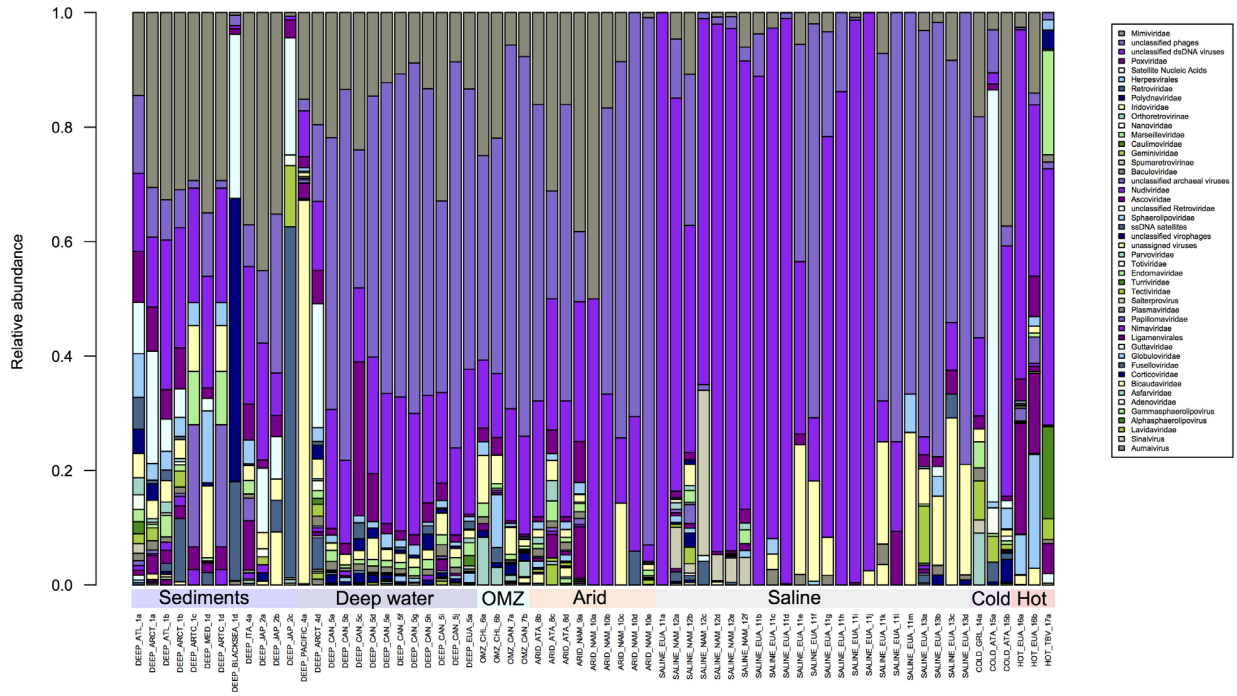

FIGURE 3 | Comparison of families with lower abundance in the Metavir database. Unclassified phages predominate in all environments, compared with other virus families. However, greater diversity is observed in the sediments, hyperthermophile and hypersaline environments compared to deep waters, Oxygen Minimum Zones (OMZ) or saline environments.

Hurwitz et al., 2013; Danovaro et al., 2016; Graham et al., 2019). It has been considered that the information from whole metagenomic analysis can give clues of potential model microorganisms to host virus replication, through the analysis of the Clusters of Regularly Interspaced Short Palindromic Repeats (CRISPR) loci from the cellular fraction of the metagenomes, that have been isolated from extreme environments (Gudbergsdóttir et al., 2016; Sharma et al., 2018; Liu et al., 2019; Martin-Cuadrado et al., 2019).

Next in relative abundance were *Circoviridae,* followed by *Phycodnaviridae, Microviridae,* and *Inoviridae* and high fractions of unidentified ssDNA and dsDNA viruses and phages. *Circoviridae* infect vertebrates and were particularly abundant in sediment samples (Dennis et al., 2018, 2019; Blanc-Mathieu et al., 2019). Members of *Phycodnaviridae* have been found at high levels in deep water samples (Mizuno et al., 2016; Gong et al., 2018; Blanc-Mathieu et al., 2019), which is curious given that they preferentially infect eukaryotic algae which require light to grow. It is possible therefore that these dsDNA viruses may infect other as yet unknown marine hosts in deep waters (Van Etten et al., 2010; Blanc-Mathieu et al., 2019). However, the predominance of the above families present obvious exceptions, such as in two samples from cold environments, a sample from a saline environment and most of the samples from deep marine sediments.

Within the metagenomes that correspond to deep sea environments (depths greater than 1,000 m), where the absence of light, oligotrophic conditions, low-oxygen concentrations, low temperatures and high hydrostatic pressure dominate (Le Romancer et al., 2006; Liang et al., 2019), two categories were considered based on the origin of the samples: sediments and deep water (**Figure 2**). It is clear that in metagenomes from sediments of the Atlantic, Arctic and Pacific Northwest, ssDNA viruses like *Circoviridae*, *Microviridae* and *Inoviridae* are more abundant than dsDNA viruses (**Figure 2**). This characteristic seems to be exclusive to samples from this environment. It should be noted that the two samples from cold environments show a similar composition to that of sediments, and all others include ssDNA viruses in low abundance (**Figure 2**). This is in agreement with a recent report by Yoshida and coworkers who reported that ssDNA viruses predominate in marine sediments and have an estimated abundance of 1×10^8 to 3×10^9 genome copies per cm^3 of sediment, clearly more abundant than dsDNA viruses which range from 3×10^6 to 5×10^6 genome copies per cm^3 (Yoshida et al., 2018).

In **Figure 3**, where the remaining viral families are shown, two general points can be highlighted: *Mimiviridae* are present in almost all environments, which is not surprising since some of their hosts are known polyextremophiles (Claverie et al., 2018; Yau and Seth-Pasricha, 2019). The second point is the abundance of unclassified sequences, which do not allow any conclusion to be made about the diversity observed by environment since these sequences could come from one or more than one family. A large part of the sequences obtained from different environments, except for sediments, have no similarity in the databases, an issue that should change with

the inclusion of additional metagenomic-derived sequences in databases (**Figure 3**). Overall analysis of the composition of viral families present in each extreme environment could at the very least allow a description of the families that are shared or that are exclusive to each environment.

Some environments are characterized by low-oxygen concentrations; these include those with high concentrations of greenhouse gases, which directly affect the biodiversity in those environments (Kiehl and Shields, 2005; Resplandy et al., 2018). There are three central oceanic regions which are considered to be Oxygen Minimum Zones (OMZ), namely the Eastern Tropical North Pacific (ETNP), the Eastern Tropical South Pacific (ETSP) and the Arabian Sea, within which the activity of anaerobic microorganisms is highly significant (Paulmier and Ruiz-Pino, 2009; Thamdrup, 2012). As expected, the viral population diversity closely reflects the microbial diversity in these environments (Cassman et al., 2012; Parvathi et al., 2018; Fuchsman et al., 2019), with the virome composition in OMZ being commonly composed of the *Myoviridae* and *Siphoviridae* families, followed by *Phycodnaviridae* (**Figure 2**).

OMZ were sampled at 200 m depths in Chile and Canada and virus composition was analyzed using the MDA (Genomiphi and GenomePlex) protocol. While the ssDNA *Circoviridae* family was predominantly observed in samples from Chile, this virus family was not observed in samples from the Canadian OMZ. In addition, in the samples from Canada (Chow et al., 2015) *Parvoviridae* (ssDNA) were highly abundant, but were totally absent in samples from Chile (**Figure 3**). In previous studies it was observed that the viral community along the vertical dissolved oxygen gradients was characterized by abundance taxa and diversity fluctuations. These differences could be related to changes in the viral replication strategy from lytic to lysogenic. It seems that oxygen reduction concurs with a decrease in viral abundance (Cassman et al., 2012; Parvathi et al., 2018). It should be noted that a large proportion of sequences obtained from these regions do not find similarity with other viruses in the databases, but those sequences could be from viruses that infect little known prokaryotic hosts, like ammonia-oxidizing archaea and anaerobic ammonia-oxidizing (anammox) bacteria which predominate in this environment (Parvathi et al., 2018).

Hyperarid environments exhibit conditions that are considered to be limiting for life, such as lack of water, high levels of UV radiation and extreme temperatures. However, both prokaryotic and eukaryotic organisms have adapted to live in these environments (Merino et al., 2019). Although low diversity might be expected in these environments, metagenomic studies performed with hypolithic communities have shown this not to be the case, with a high level of diversity being reported; particularly in bacterial communities from Antarctica (cold desert) and Namibia (desert), which are mainly *Actinobacteria*, *Proteobacteria,* and *Cyanobacteria* (Vikram et al., 2016). In the hypolithic viral communities from the Namibian desert and the Antarctic, metagenomic data has revealed the presence of *Caudovirales* which do not correlate with phages that infect *Cyanobacteria* species (Adriaenssens et al., 2015). The samples from the Antarctic hyperarid region displayed a greater diversity of unique viruses such as *Bicaudaviridae*, *Asfarviridae*, *Lavidaviridae*, *Tectiviridae,* and *Sphaerolipoviridae* when compared with the families found in the Namibian desert (**Figure 3**). Zablocki and coworkers have previously reported a higher viral diversity in the Arctic when compared with the Namibian desert, and it has been observed that Antarctic desert soils contain higher proportions of free extracellular virus-like particles compared to hot hyperarid desert soils, where a lysogenic lifestyle seems to prevail (Zablocki et al., 2016).

In **Figure 3** the variability in the composition of viral families in hypersaline habitats is evident. Such environments are widely distributed throughout the world and are present in salt lakes, salt flats and salt deposits. In these environments, the low water activity directly affects the composition of the microbial communities (Le Romancer et al., 2006; Ma et al., 2010; Merino et al., 2019). Viruses that have been identified in these ecosystems are haloviruses and a large number of these infect Archaea, Bacteria and Eukaryotes (Atanasova et al., 2018; Plominsky et al., 2018; Ramos-Barbero et al., 2019). About 64 archaeal viruses have been isolated from the two kingdoms, *Crenarchaeota* and *Euryarchaeota* (Porter et al., 2007). These samples are also those that have a greater abundance in unassigned or not classified viruses, which prevents determination of the real diversity of that group of archaea viruses, probably because they are the least studied and have low representation in the databases (Atanasova et al., 2018; Ramos-Barbero et al., 2019).

In addition, unclassified dsDNA viruses have also been observed (**Figure 3**), while haloviruses such as HGV-1, HTVAV-4 and HSTV-1 have also been identified at high levels. On the other hand, ssDNA viruses which mostly infect eukaryotes such as colpodellids, nematodes, arthropods, chlorophytes, among others, are present at low levels in hypersaline habitats (Feazel et al., 2008; Heidelberg et al., 2013).

The thermophile environments are characterized by high temperatures, where thermophilic microorganisms thrive at 65–80°C as their optimal growth temperature, and >80°C for hyperthermophiles (Merino et al., 2019). Viruses that infect bacteria and archaea are abundant in these hyperthermophilic habitats (Schoenfeld et al., 2008; Strazzulli et al., 2017; Liu et al., 2019). The virome of hyperthermophile environments is composed of viruses that infect all three domains of life, with members of the *Turriviridae, Fuselloviridae, Bicaudaviridae,* and *Globuloviridae* families that infect Archaea (Krupovic et al., 2018). Moreover, the *Nudiviridae, Phycodnaviridae,* and *Poxviridae* families that infect eukaryotes are also present (**Figure 3**).

Within the hyperthermophile metagenomes analyzed here, the presence of ssDNA or RNA viruses was not observed, but in other studies from these environments the presence of picornavirus-like, alphavirus-like, and flavivirus-like RNA viruses has been reported (Bolduc et al., 2012). It is possible that ssDNA and RNA viruses were not detected in the samples we analyzed due to differences in sample processing (**Figures 2**, **3**). Thus, as previously mentioned, if comparative viral metagenomic studies are to be undertaken to allow an accurate comparison between viromes from different ecosystems and to potentially identify novel viral clusters, then standardized methodologies will need to be developed and employed.

The polar regions of the Earth are dominated by the polar ice caps, with the microbial diversity present in these regions being much higher than might be expected. It is well established that viruses play an important role in controlling microbial mortality in these habitats (López-Bueno et al., 2009; Cárcer et al., 2015; Yau and Seth-Pasricha, 2019). While it has been reported that different lakes located in the Arctic and Antarctic share similar virome compositions, marked differences have been found.

Although at this taxonomic level it is possible to differentiate some of the particularities described above, in terms of the virus composition in each environment studied, very little information is revealed at the genus or species level that would allow a better understanding of the virus–host relationship and its influence in the environment.

Therefore, two environments, OMZ and deep-sediments, which at the family level have a very similar structure (**Figures 2**, **3**) were selected in an attempt to determine if it is possible to obtain biologically meaningful information on the differences or similarities in virus-host interactions at the genus level. The genus composition of two well-known families were analyzed: *Podoviridae* that infect bacteria and

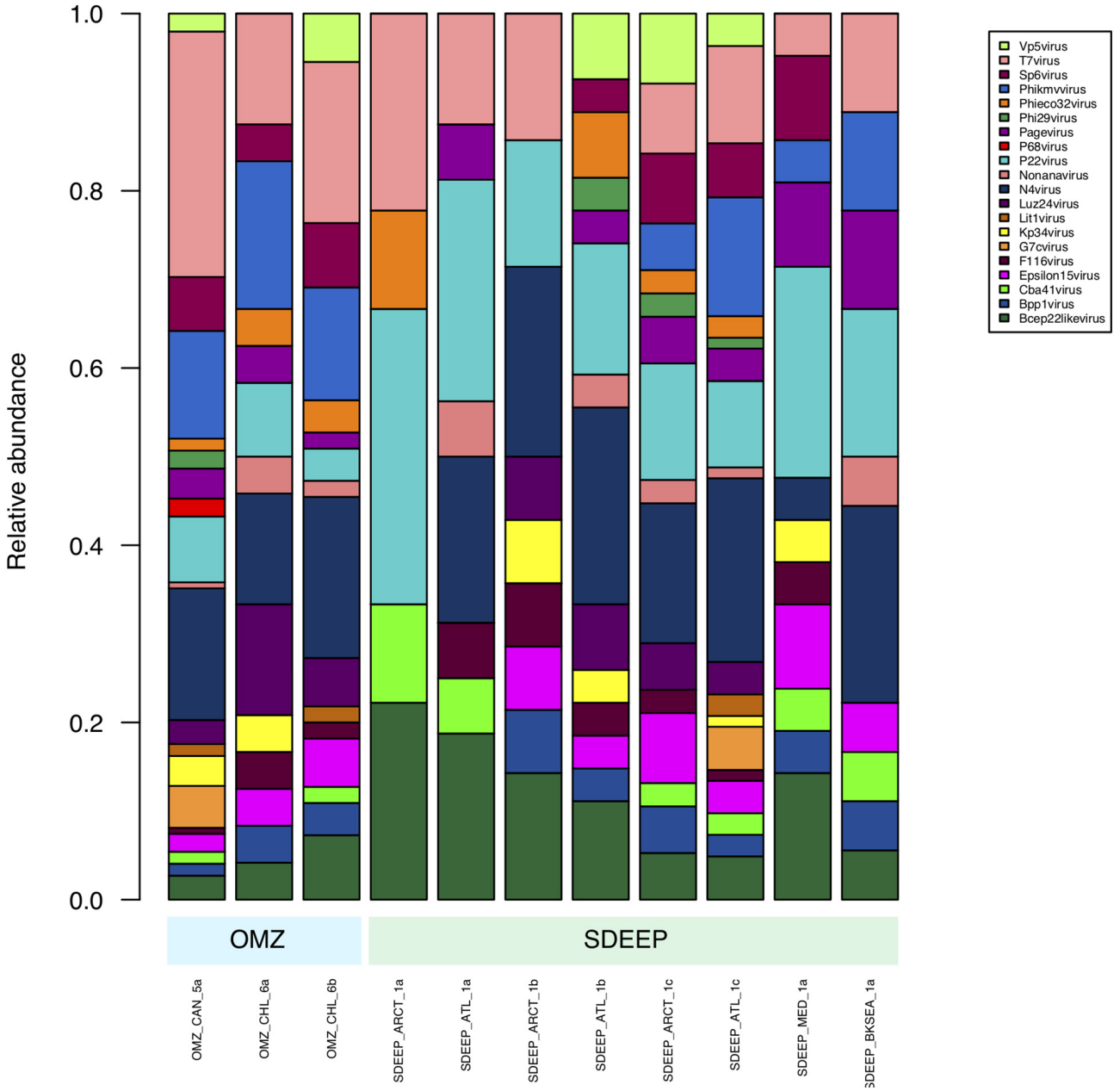

FIGURE 4 | Composition of *Podoviridae* family at genus level. Metagenomes from OMZ and deep-sediments were considered in the analysis.

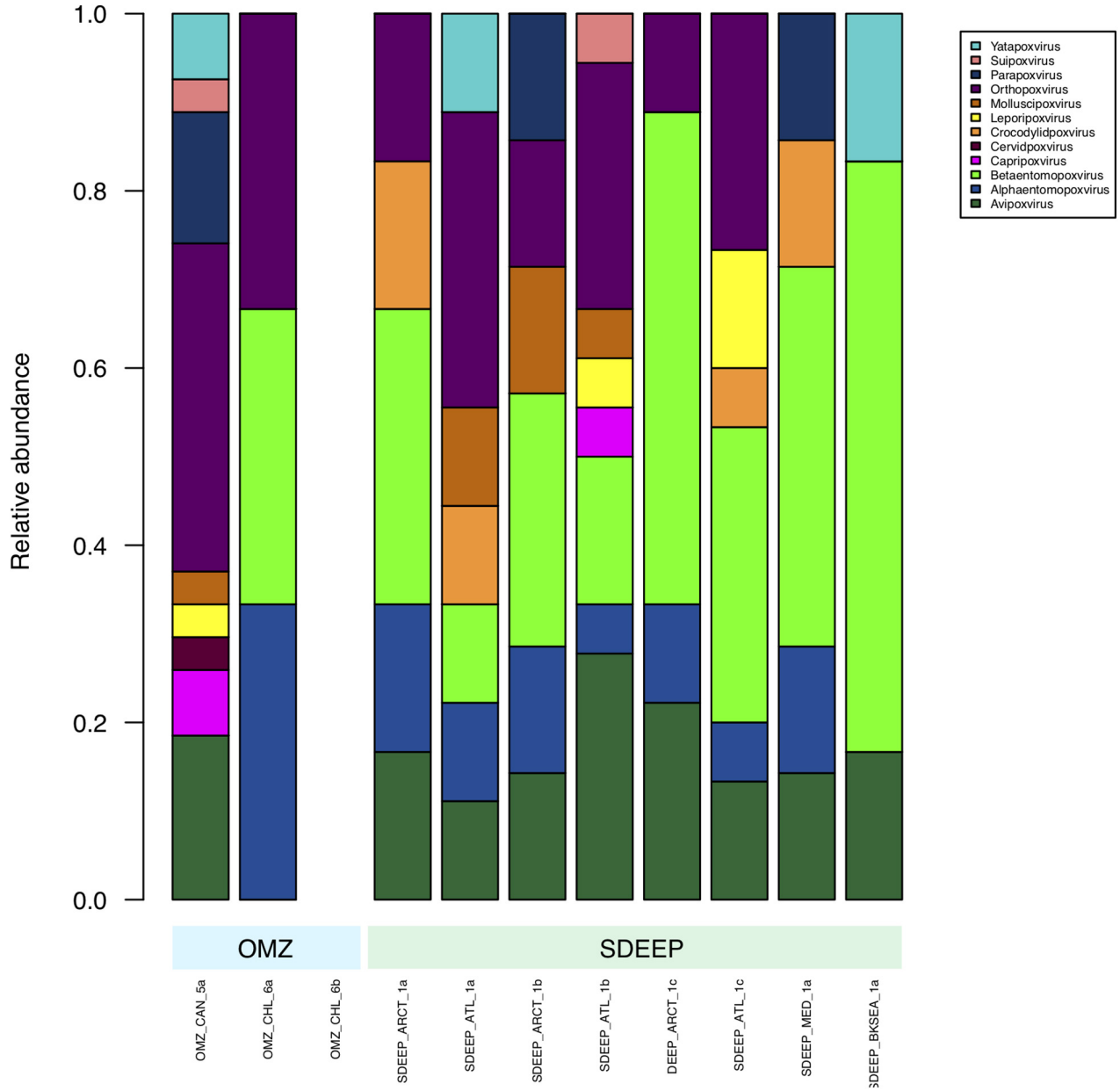

FIGURE 5 | Composition of *Poxviridae* family at genus level. Metagenomes from OMZ and deep-sediments were considered in the analysis.

are ubiquitous even in extreme environments (**Figure 4**), and *Poxviridae* (**Figure 5**) for which known hosts are terrestrial vertebrates and invertebrates and their presence in extreme environments, particularly in aquatic environments has not been reported. At this level only some viruses can be taxonomically identified and can be seen to vary in abundance, however, most of the genera were unidentified. The genera from *Podoviridae* identified by sequence were enterobacter phages, which infect bacteria that are known human pathogens, and would not be expected in these environments (**Figure 4**). In the case of *Poxviridae*, most genera found infect terrestrial vertebrates, suggesting that either the sequences of the *Poxviridae* family members obtained from aquatic niches are sufficiently similar to those from *Poxviridae* members that infect terrestrial hosts, or this is an artifact caused by the lack of sequences from *Poxviridae* found in the databases (**Figure 5**).

Clustering by Environmental Virome

Hierarchical clustering analysis was performed from abundant viral families previously published in the aforementioned metagenomic datasets (**Figure 6**). From this it was possible to conclude that some extreme environments have groups that indicate similarities in the viral communities present

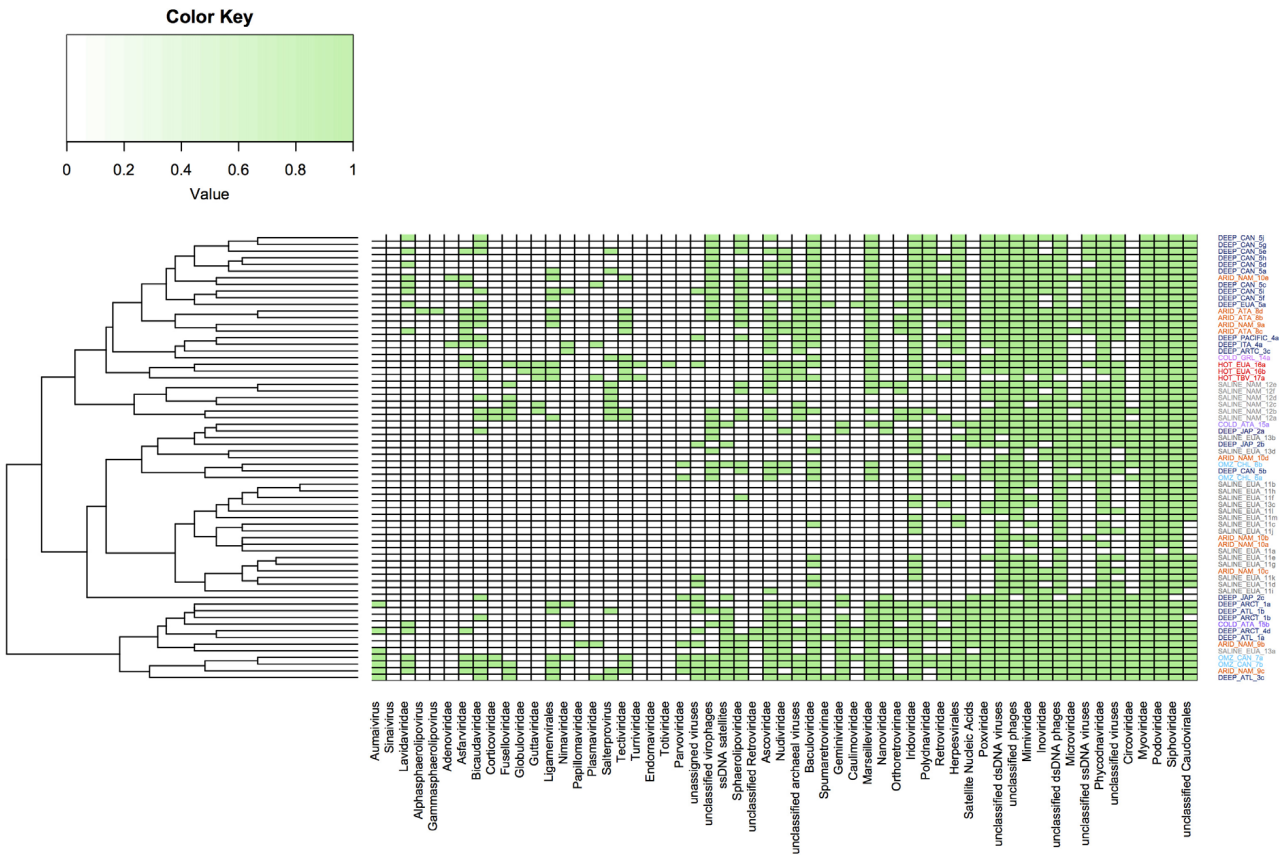

FIGURE 6 | Clustering dendrogram based on the abundance of virus families present in the metagenomes from extreme environments. Virus families are represented by presence or absence; the green color indicates the presence while the white color indicates absence. Families of the order of Caudovirales are present in all environments. The figure was constructed using a presence and absence matrix, using the heatmap library on R.

in these environments. This was particularly evident for some viromes obtained from hypersaline, deep-sea and hyperarid environments, while it was less evident in other extreme ecosystems which did not appear to show clustering, such as cold environments. However, this analysis again shows that some viral families are ubiquitous in all extreme environments, while the ssDNA viruses appear to predominate in sediments from deep-sea and cold environments.

In general, the virome structure from hypersaline samples reveals low levels of diversity, even in samples from different geographical areas (**Figure 6**). The high concentration of NaCl might limit viral diversity due to the shortage of prokaryotic hosts, since *Haloquadratum walsbyi*, *Salinibacter ruber,* and nanohaloarchaeas are the predominant organisms in these environments, with more than 90% of the contigs annotated to these taxa (Ventosa et al., 2015). Another factor that could determine the virome diversity that is observed in hypersaline environments is the dynamic switch between lytic and lysogenic replication cycles, since this represents a significant adaptation mechanism in environments with high salinity content (Roux et al., 2016).

This notwithstanding, from **Figure 6** it is clear that the virus family composition is quite similar in these environments, which could provide significant information related not only to viral evolution but also to physiological adaptation of microorganisms in response to high temperatures (Schoenfeld et al., 2008; Biddle et al., 2011).

Figure 7 shows the degree of overall similarity between the viral metagenomes in relation to the extreme environment from which the viromes were isolated. As previously described, some viral families belonging to the *Caudovirales* order are ubiquitous and display polyextremophilic adaptation. The hypersaline environments present a consistent clustering depending on the viral diversity, as well as the relative viral family abundance, which suggests that NaCl enriched environments provide strong constraints for the development of life that may restrict ecosystem diversity. Some viromes from hyperarid, deep-sea and saline environments are closely clustered (**Figure 7**) suggesting that the organisms and therefore the viral composition is partially shared, at least between these environments. Regarding deep-sea environments, **Figure 7** shows two clustered metagenome populations derived from the deep-sea, where those viromes

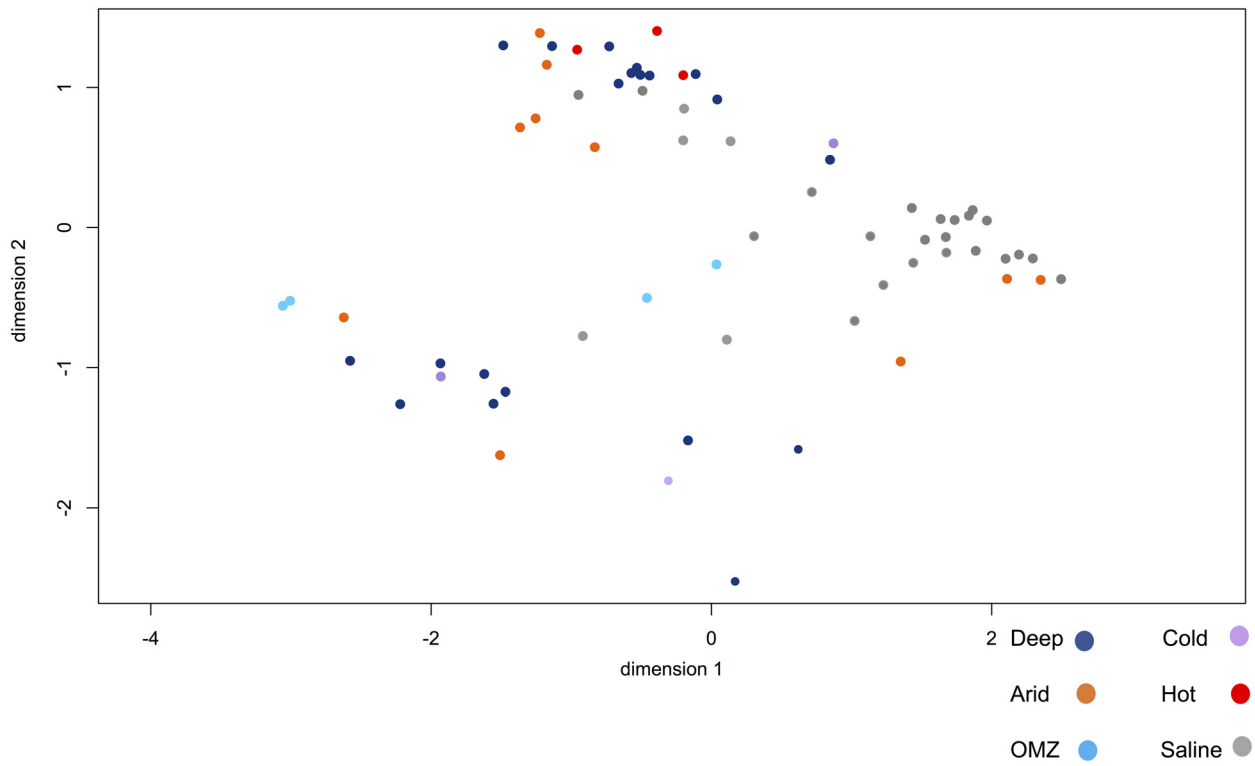

FIGURE 7 | Multidimensional scaling (NMDS) to visualize the degree of overall similarity between the metagenomes. Metagenomes from different habitats are shown: deep-sea (pink), OMZ (green), hyperarid (red), hypersaline (blue), psychrophile (orange), hyperthermophile (yellow). Hypersaline metagenomes have a strong clustering compared to other metagenomes. The figure was constructed from an abundance matrix, where the multidimensional scaling algorithm was applied to observe the general similarity of metagenomes, using the R software.

obtained from deep water are closely clustered, as well as those from sediments.

An interesting hypothesis to be investigated using metagenomics studies conducted in different geographical areas is the possibility of identifying specific viral clusters associated with a particular extreme environment. The large numbers of unclassified sequences in the databases is an important issue to consider with studies on viromes from extreme environments. The limitations of the bioinformatic pipelines to assign a taxonomic identity to a majority of the viral sequences, together with our limited understanding about viruses in extreme environments, has resulted in a lack of progress in our knowledge of extremophilic viromes. This has also negatively impacted our understanding in terms of evolution, gene horizontal transfer, ecology and virus-host interactions.

To meaningfully compare viromes from different environments it is necessary to at least partially answer the previous questions and provide new information about how viromes are potentially limited by extreme physicochemical characteristics, geographical area, or other artificial circumstances such as sampling methods, enrichment techniques and other technical biases. It should be possible to determine whether some viral populations could be closely related to a specific type of extreme ecosystem and consequently obtain more information about viral evolution (Simmonds, 2015).

FUNCTIONAL METAGENOMICS IN EXTREME ENVIRONMENTS: METHODOLOGICAL CHALLENGES, DISCOVERIES AND OPPORTUNITIES

Functional viral metagenomics focuses on exploring viral diversity to discover novel genes. Extreme environments harbor an enormous diversity of unknown viruses (Desnues et al., 2008; Dinsdale et al., 2008; Williamson et al., 2008; Rosario and Breitbart, 2011; Kristensen et al., 2012; Atanasova et al., 2016; Gudbergsdóttir et al., 2016; Nigro et al., 2017; Zablocki et al., 2017a,b; Sharma et al., 2018; Liu et al., 2019; Martin-Cuadrado et al., 2019; Mizuno et al., 2019; Roux et al., 2019) and, consequently, a potentially large number of unknown viral proteins. Functional viral metagenomics in these niches show a limited progress, with few reported recent advances (Schoenfeld et al., 2009; Schmitz et al., 2010; Moser et al., 2012; Heller et al., 2019) (**Figure 8**).

Viral-host systems (*in vitro* screening), sequence-based screening, activity-based screening (heterologous expression of

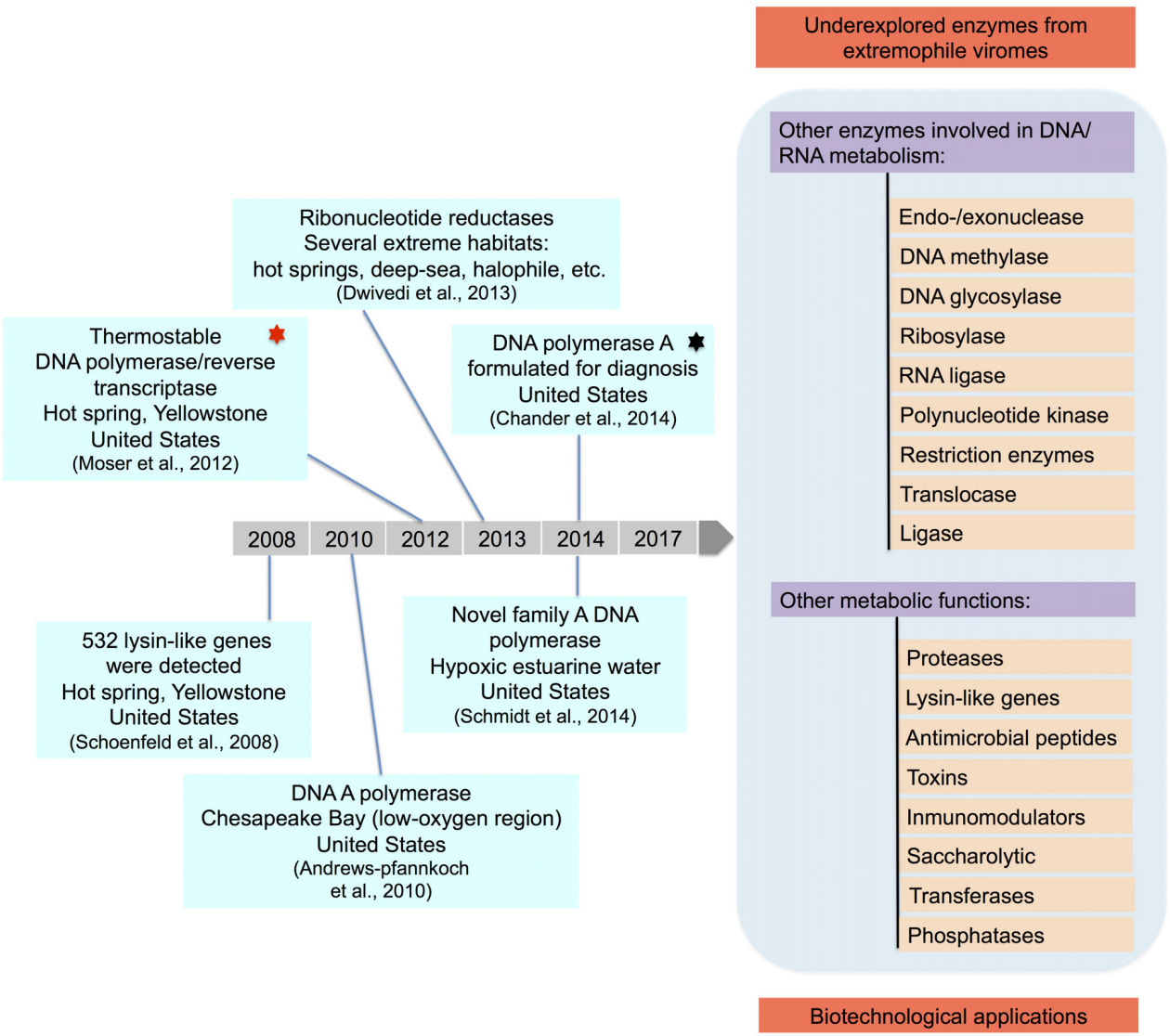

FIGURE 8 | Progress in function-based metagenomics. Functional bioprospections from extreme environments performed in the last years are summarized. Underexplored enzymes from extremophile viromes are listed.

viral proteins), and PCR- and hybridization-based screening, could be implemented for functional analysis from extreme viromes (Moser et al., 2012; Bzhalava and Dillner, 2013; Fancello et al., 2013; Heller et al., 2019). While sequence-based screenings have been responsible for the discovery of the majority of new viral enzymes (at least as annotated proteins) from extreme environments, PCR- and hybridization-driven methods have not been employed to date to functionally explore extremophile viromes.

In vitro screening from extremophile viromes is a challenge with respect to the co-cultivation of both hosts and viruses, and in particular in trying to mimic the conditions which are present in these habitats, thereby ensuring a better success rate regarding viral replication and viral protein expression. Thus, to help overcome this bottleneck, it will be important to develop new host systems (for prokaryotic and eukaryotic viruses) which grow under extreme pH, temperature, salinity, pressure and radiation (Schoenfeld et al., 2009).

Activity-based screening could be a very useful approach to identify novel enzymes. This method demands an efficient heterologous expression system for viral proteins. Thus, there are problems with the expression levels of many viral proteins in foreign host systems, particularly in genes isolated from extremophile viromes, which are dominated by rare genes, with issues such as codon usage together with promoter regulation/activation negatively impacting on enzyme production in different heterologous systems (Kristensen et al., 2012).

While the well-established *Escherichia coli* heterologous expression system is available, it is clear that additional systems

with a particular focus on extremophile bacteria and fungi will need to be developed to increase the chances of producing sufficient levels of viral extremoenzymes to allow their detection in function-based screens. These screens usually employ activity-based assays which involve colorimetric changes, typically following utilization of a substrate. However, these types of screenings are not particularly useful with viromes, since viral genes encoding enzymes involved in the metabolism of different substrates are quite rare. Due to this, there are no reports to date of the detection of viral enzymes from viromes through activity-based screening.

Despite the aforementioned disadvantages related to the heterologous expression of viral proteins, the activity-driven screenings allow functional gene annotation through an *in vitro* phenotypic-based test. This has an important advantage over sequencing-driven screening where a high number of genes are annotated as "*unknown function*," because the gene repertoires of the extremophile viromes are currently undersampled.

A number of viral enzymes with utility in scientific, diagnosis and therapeutic applications have been identified using sequencing-based screens from extremophile viromes (Schoenfeld et al., 2008, 2009; Moser et al., 2012; Dwivedi et al., 2013; Schmidt et al., 2014; Mead et al., 2017). Also, the genomes of extremophile viruses are likely to be a source of novel antimicrobial peptides that may have applications in the biopharmaceutical and molecular diagnostics areas (Rice et al., 2001; Le Romancer et al., 2006; Schoenfeld et al., 2009).

For example, a sequencing-driven metagenomic study from two mildly alkaline hot springs in Yellowstone, allowed identification of 532 lysin-like genes (Schoenfeld et al., 2008). In recent years, these lytic enzymes have gained increasing importance due to their potential use in biomedical science applications (Schmitz et al., 2010); however, no lysin-like genes from extreme environments have to date been experimentally characterized.

DNA polymerases (502 sequences) have been also detected in extreme metaviromes, particularly from hypoxic estuarine waters obtained in the Gulf of Maine, Dry Tortugas National Park and the Chesapeake Bay (Andrews-Pfannkoch et al., 2010; Schmidt et al., 2014). These shotgun metagenomic studies revealed a novel DNA polymerase A family in marine virioplankton, since some sequences were distantly grouped in a phylogeny comprising DNA polymerase A from virus and bacteria (Schmidt et al., 2014).

Ribonucleotide reductases (RNR) have also been found from viromes obtained from hypersaline, psychrophile and thermophile niches (Dwivedi et al., 2013). For example, a bioinformatics analysis demonstrated that viruses isolated from hot springs contained a high abundance of RNR. However, some habitats such as hydrothermal vents from the East Pacific Rise, a solar saltern pond and salterns from Alicante (Spain) were found to have fewer ($\leq$5) identifiable RNR viral homologs (Dwivedi et al., 2013).

Recent efforts to characterize new viral DNA polymerases from extreme environments have resulted in the identification of a thermostable polymerase in a viral metagenomic DNA library from a near-boiling thermal pool in a hot spring in Yellowstone (Moser et al., 2012; Heller et al., 2019). This was the first report describing the isolation of a polymerase from a viral metagenomic library. In this study 59 complete polymerase clones were identified as possessing thermostable DNA polymerase activity following a functional screen. One of these polymerases, namely PyroPhage 3173 Pol, also has 5′-3′ exonuclease activity, as well an innate reverse transcriptase activity. It was subsequently tested in high fidelity reverse transcription PCR (RT-PCR) reactions and compared with some commercially available enzyme systems (Moser et al., 2012; Heller et al., 2019). The PyroPhage 3173 Pol-based RT-PCR enzyme was found to have a higher specificity and sensitivity that the other enzymes. While the PyroPhage 3173 DNA polymerase shares amino acid identity (∼32%) with another bacterial polymerase, no significant similarity was found with other viral proteins (Moser et al., 2012). This highlights the potential diversity of enzymes that may be present in extremophile viromes. The enzyme has subsequently been characterized and shown to be effective in the molecular detection of certain viral and bacterial pathogens by loop-mediated isothermal amplification (Chander et al., 2014).

STRUCTURAL BIOLOGY OF VIROMES

Specific molecular-level adaptations to extreme environments can only be appreciated once the detailed molecular structures are known. In order to explore the available structural information of proteins belonging to extremophile viruses, we carried out a manual search in the Protein Data Bank[1] (Berman et al., 2000) of all the viral families and genera identified in the metagenomes analyzed in this review, and kept only those whose hosts were either marine viruses or frank extremophiles. The resulting proteins were then classified according to their annotated function, and are discussed below.

In general, all these structures are valuable from a biochemical and biotechnological perspective, as they contain the molecular representation of the required adaptation to the particular extreme environment favored by the virus host. For example, viruses that infect *Acidianus* or *Sulfolobus* archaea are subject to the combination of high temperature and acidic pH; their proteins tend to have many charged residues, in particular, acidic ones (see structure 3DF6, an orphan protein). They also tend to have compact folds with structured termini, short loops with prolines in specific positions to stabilize them (see structure 2BBD, a major capsid protein), and the absence of cavities. Despite being DNA-binding proteins, and therefore cytoplasmic, some of them also include disulfide bridges, intramolecular (see structure 2VQC) or intermolecular (in structure 2CO5), that can impart up to 14 degrees in thermal stability for the protein. The formation of these disulfide bonds requires the existence of a sulfhydryl oxidase, either encoded by the host or by the virus itself. When a mesophilic homolog exists, a direct comparison of the structural features of the proteins can guide protein engineering to improve stability and/or function.

[1] rcsb.org

Also, as some viruses have space limitations in their capsids, resulting in compact genomes, viral homologs in these cases tend to be the minimal possible version of the protein family, allowing for the identification of the critical residues that stabilize both structure and function. A nice example of this is the minimal catalytic integrase domain of *Sulfolobus* spindle-shaped virus 1 (structures 3VCF, 4DKS and 3UXU). On the other hand, for viruses that have less space limitations, viral proteins can have surprising combinations of domains, suggesting ways to engineer multidomain proteins. This is particularly notorious in *Mimivirus*, where identifiable catalytic domains can be linked to domains with no sequence or structure homology to any known protein, as in the sulfhydryl oxidase in structure 3TD7. Less dramatic examples are basic modules known to function as transcription factors, such as the ribbon-helix-helix domain, with an extra helix added as an embellishment that increases thermal stability, as in structure 4AAI from *Sulfolobus* virus Ragged Hills.

The analysis of the conservation of proteins amongst viruses of the same or different classes is instructive, and can help in establishing families and/or events of horizontal gene transfer. This conservation has been historically one of the criteria used to choose which proteins to study structurally from a particular virus. The wealth of information derived from identifying open reading frames in the data from sequencing endeavors can certainly be a source of novel activities, as described in the previous section. This functional annotation requires sequence homology to known proteins, something that does not happen frequently with extremophile viruses. As structure diverges more slowly than sequence, protein structural analysis allows for the inference of function when sequence homology is weak. In **Supplementary Table S2** we list viral protein structures that were obtained in this spirit, sometimes as part of Structural Genomics Initiatives (Oke et al., 2010). As can be seen from **Supplementary Table S2**, the goal of assigning function is not always achieved, as on occasion novel folds are found (see, for example, structures 4ART and 3DF6 discussed above), precluding the transfer of function. In other, happier cases, the structure instructs the experiments needed to functionally annotate the sequence (see structure 3O27, with functional DNA binding activity).

Most of the structures we found were obtained with a previous inkling of the function of the protein. For example, **Supplementary Table S3** lists structural proteins, such as capsids and tail spikes. The full capsid structures are interesting, for instance, as scaffolds for drug delivery, and as models to study capsid formation, propose infection mechanisms, and study the interactions with nucleic acids and membranes. In this regard, structure 5W7G proposes a model for the membrane envelope of *Acidianus* filamentous virus 1, composed of flexible tetraether lipids that are organized as horseshoes, including a mechanism for enrichment of the viral membrane with this particular lipid of low abundance in the host. Another important interaction is that of capsid proteins with DNA, and surprisingly, it appears that rod-shaped viruses (such as *Sulfolobus islandicus* rod-shaped virus 2 in structure 3J9X) organize their DNA in the A form, stabilized by alpha helices from the major capsid proteins. This is in stark contrast to icosahedral viruses, which pack their DNA in the B form. Turrets, tails and spikes are important to understand interactions with the hosts, as part of the ecological role that these viruses play.

Supplementary Table S4 lists proteins that bind either DNA or histones; the latter come from viruses that infect either fish or shrimp and are interesting because one of them is a DNA mimic (see structure 2ZUG). The architecture of these DNA-binding proteins is sometimes reminiscent of known bacterial classes (see structure 2CO5, a winged helix-turn-helix protein with an intramolecular disulfide bond, discussed above), or is a novel fold (see structure 2J85). Finally, **Supplementary Table S5** lists enzymes found in extremophile viruses. The range of activities is wide, going from DNA, protein and sugar metabolism, to reactive oxygen species management (see, for example, structure 4U4I, a superoxide dismutase that does not require chaperones to capture copper or oxidize its disulfide bridges). There is also interest in auxiliary metabolic genes that support more efficient phage replication, and are normally related to photosynthesis; in this class we find structure 5HI8, a phycobiliprotein lyase, and 3UWA, a peptide deformylase particularly selective for the D1 protein of photosystem II.

The relevance and utility of all these structures is multiple: as crystallographic-amenable homologs of difficult targets (see structure 3VK7, a DNA glycosylase) given their stability, as inspiration to improve mesophilic orthologs in their resistance to high temperature and low pH, as examples on how to trim these orthologs to minimal yet functional versions, in the identification of novel quaternary structures (see for example structure 5Y5O, a dUTPase with novel packing), and as examples on how to adapt new modules to them (see structure 3TD7, the sulfhydryl oxidase with a novel domain attached at the C-terminus). The field of structural biology of extremophile viruses is still young, and there is plenty of room for the exploration of orphan ORFs and for viruses subject to other extreme environments.

FINAL REMARKS

Metagenomics is a powerful approach to study the virome structure of extreme environments and its potential biotechnological applications in a number of fields. However, despite its potential few studies have been undertaken to characterize viral communities in these environments. Some methodological challenges need to be overcome to ensure that samples enriched in viral particles can be obtained, as well as increasing the yields of viral nucleic acids that can be isolated.

A comparative analysis of the population structure of viromes in extreme environments was carried out here, using the 17 publicly accessible virus metagenomic libraries deposited in MetaVir2. Viral communities from different extreme environments showed quite high levels of overall similarity, with viral families belonging to the *Caudovirales* order being ubiquitous and displaying seemingly polyextremophilic adaptation. The most abundant families of *Caudovirales* were *Siphoviridae*, *Myoviridae,* and *Podoviridae*, followed by *Circoviridae, Phycodnaviridae, Microviridae,* and *Inoviridae*. However, very high fractions of unidentified ssDNA and dsDNA viruses and phages were identified. Considering the large number

of unclassified viral sequences from extremophilic viromes, it is currently not possible to definitively identify novel virus families which are uniquely present in different extreme environments. Nor is it possible to correlate the presence of specific virus families or genera with any given environment.

Attempts to further explore specific virus–host relationships in the *Podoviridae* and *Poxviridae* present in OMZ and deep-sediments resulted in the identification of viruses whose known hosts are highly unlikely to reside in these environments. Although many more sequences were obtained when a similar comparison was made using newer, more up to date databases (RefSeq from NCBI, or IMG/VR), the viruses identified often correspond to those for which more sequences are available. Therefore, although MetaVir2 is no longer kept up to date, and richer, more recent databases exist, such as IMG/VR, a similar comparative analysis using such databases produced similar results (not shown), indicating that more accurate taxonomic assignments are required and ideally they should be in a common repository where all the viral metadata are collected. Also new tools should be developed to automatize sequence classification so that viral species assignment can be obtained.

Hierarchical clustering analysis was performed from abundant viral families previously mentioned and it was possible to conclude that some extreme environments, such as hypersaline, deep-sea and hyperarid niches, have groups that indicate similarities in the viral communities present in these environments, although as above, with the available data this comparison could only be evaluated at the family level. An important challenge from viral metagenomics is to establish specific viral clusters associated with particular extreme environments and describe their role in different extreme ecosystems. Although taxonomic allocation at the level of the genus or species in viruses is a challenge, new strategies for the classification of viruses are still in development from the use of genomic sequences without previous information or clustering of their coding sequences, that allow a more efficient classification process, that is scalable and user friendly.

In addition, the functional prospecting for viruses in extreme ecological niches has been almost exclusively limited to sequence-based screening to date. While some viral sequences have been annotated and assigned to specific functions, very few viral proteins discovered using metagenomics approaches have been subsequently cloned, heterologously expressed and biochemically characterized. Other functional-based methods such as activity-based screenings and PCR -or hybridization- based screenings are currently underexploited as approaches to identify viral proteins from extremophilic viromes, which may have utility in biotechnological applications. Considering that our current knowledge of viromes associated with extreme ecosystems is quite limited, we still cannot fully appreciate the great biotechnological potential that they may represent. Thus, further efforts should be made to screen extremophile viral metagenomes for novel proteins and biomolecules if we are to advance our understanding of their biological impact and to capitalize on the unique viral diversity that is present within these novel ecosystems.

AUTHOR CONTRIBUTIONS

SD-R, RG, AD, and RB-G designed and wrote the manuscript. MS-C prepared the Section "Perspectives on Sampling and Processing: Methodological Challenges for Viral Metagenomics in Extreme Environments." HC-S, SD-R, RP, and AH collected and processed the metagenomic data and conducted the analysis of viral communities. RB-G and LM-Á prepared Section "Functional Metagenomics in Extreme Environments: Methodological Challenges, Discoveries and Opportunities" related to functional bioprospection in extreme viromes, while NP prepared Section "Structural Biology of Viromes."

FUNDING

We acknowledge the graduate fellowships that HC-S and LM-Á received from CONACyT-Mexico. The authors appreciate the support received from the Unidad de Secuenciación Masiva y Bioinformática (Instituto de Biotecnología UNAM-Mexico) for the access to computer facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02403/full#supplementary-material

REFERENCES

Adriaenssens, E. M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., et al. (2015). Metagenomic analysis of the viral community in NAMIB DESERT hypoliths. *Environ. Microbiol.* 17, 480–495. doi: 10.1111/1462-2920.12528

Andrade, A. C., Arantes, T. S., Rodrigues, R. A., Machado, T. B., Dornas, F. P., Landell, M. F., et al. (2018). Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. *Virol. J.* 15:22. doi: 10.1186/s12985-018-0930-x

Andrews-Pfannkoch, C., Fadrosh, D. W., Thorpe, J., and Williamson, S. J. (2010). Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. *Appl. Environ. Microbiol.* 76, 5039–5045. doi: 10.1128/AEM.00204-10

Atanasova, N. S., Bamford, D. H., and Oksanen, H. M. (2016). Virus-host interplay in high salt environments. *Environ. Microbiol. Rep.* 8, 431–444. doi: 10.1111/1758-2229.12385

Atanasova, N. S., Demina, T. A., Shanthi, S. N. K. R., Oksanen, H. M., and Bamford, D. H. (2018). Extremely halophilic pleomorphic archaeal virus HRPV9 extends the diversity of pleolipoviruses with integrases. *Res. Microbiol.* 169, 500–504. doi: 10.1016/j.resmic.2018.04.004

Bellas, C. M., Anesio, A. M., Barker, G., and Pearce, D. A. (2015). Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. *Front. Microbiol.* 6:656. doi: 10.3389/fmicb.2015.00656

Berliner, A. J., Mochizuki, T., and Stedman, K. M. (2018). Astrovirology: viruses at large in the universe. *Astrobiology* 18, 207–223. doi: 10.1089/ast.2017.1649

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242.

Biddle, J. F., White, J. R., Teske, A. P., and House, C. H. (2011). Metagenomics of the subsurface brazos-trinity basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J.* 5, 1038–1047. doi: 10.1038/ismej.2010.199

Blanc-Mathieu, R., Kaneko, H., Endo, H., Chaffron, S., Hernández-Velázquez, R., Nguyen, C. H., et al. (2019). *Viruses of the Eukaryotic Plankton are Predicted to Increase Carbon Export Efficiency in the Global Sunlit Ocean. biorxiv.* [preprint]. doi: 10.1101/710228

Bolduc, B., Shaughnessy, D. P., Wolf, Y. I., Koonin, E. V., Roberto, F. F., and Young, M. (2012). Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J. Virol.* 86, 5562–5573. doi: 10.1128/JVI.07196-11

Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B. L., and Sullivan, M. B. (2017). iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J.* 11, 7–14. doi: 10.1038/ismej.2016.89

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., et al. (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc. R. Soc. Lond. B* 271, 565–574. doi: 10.1098/rspb.2003.2628

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *PNAS* 99, 14250–14255. doi: 10.1073 pnas.202488399

Bzhalava, D., and Dillner, J. (2013). Data mining in genomics proteomics. *Data Min. Genom. Proteom.* 4, 3–7. doi: 10.4172/2153-0602.1000134

Cantalupo, P. G., and Pipas, J. M. (2019). Detecting viral sequences in NGS data. *Curr. Opin. Virol.* 39, 41–48. doi: 10.1016/j.coviro.2019.07.010

Cárcer, D. A., De, López-bueno, A., Pearce, D. A., and Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Sci. Adv.* 2, 1–9. doi: 10.1126/sciadv.1400127

Cassman, N., Prieto-davó, A., Walsh, K., Silva, G. G. Z., Angly, F., Akhter, S., et al. (2012). Oxygen minimum zones harbour novel viral communities with low diversity. *Environ. Microbiol.* 14, 3043–3065. doi: 10.1111/j.1462-2920.2012.02891.x

Castelán-Sánchez, H. G., Lopéz-Rosas, I., García-Suastegui, W. A., Peralta, R., Dobson, A. D., Batista-García, R. A., et al. (2019). Extremophile deep-sea viral communities from hydrothermal vents: structural and functional analysis. *Mar. Genom.* 46, 16–28. doi: 10.1016/j.margen.2019.03.001

Chander, Y., Koelbl, J., Puckett, J., Moser, M. J., Klingele, A. J., Liles, M. R., et al. (2014). A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (AMP). *Front. Microbiol.* 5:395. doi: 10.3389/fmicb.2014.00395

Chatterjee, A., Sicheritz-Pontén, T., Yadav, R., and Kondabagil, K. (2019). Genomic and metagenomic signatures of giant viruses are ubiquitous in water samples from sewage, inland lake, waste water treatment plant, and municipal water supply in Mumbai. *India. Sci. Rep.* 9:3690. doi: 10.1038/s41598-019-40171-y

Chow, C. E. T., Winget, D. M., White, R. A. III, Hallam, S. J., and Suttle, C. A. (2015). Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Front. Microbiol.* 6:265. doi: 10.3389/fmicb.2015.00265

Claverie, J. M., Abergel, C., and Legendre, M. (2018). Giant viruses that create their own genes. *Med. Sci.* 34, 1087–1091. doi: 10.1051/medsci/2018300

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Rastelli, E., Cavicchioli, R., Krupovic, M., et al. (2016). Virus-mediated archaeal hecatomb in the deep seafloor. *Sci. Adv.* 2:e1600492. doi: 10.1126/sciadv.1600492

Demina, T. A., Atanasova, N. S., Pietilä, M. K., Oksanen, H. M., and Bamford, D. H. (2016). Vesicle-like virion of Haloarcula hispanica pleomorphic virus 3 preserves high infectivity in saturated salt. *Virology* 499, 40–51. doi: 10.1016/j.virol.2016.09.002

Dennis, T. P., de Souza, W. M., Marsile-Medun, S., Singer, J. B., Wilson, S. J., and Gifford, R. J. (2019). The evolution, distribution and diversity of endogenous circoviral elements in vertebrate genomes. *Virus Res.* 262, 15–23. doi: 10.1016/j.virusres.2018.03.014

Dennis, T. P., Flynn, P. J., De Souza, W. M., Singer, J. B., Moreau, C. S., Wilson, S. J., et al. (2018). Insights into circovirus host range from the genomic fossil record. *J. Virol.* 92, e145–e118. doi: 10.1128/JVI.00145-18

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452, 340–343. doi: 10.1038/nature06735

Diemer, G. S., and Stedman, K. M. (2012). A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Dir.* 7:13. doi: 10.1186/1745-6150-7-13

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., et al. (2008). Functional metagenomic profiling of nine biomes. *Nat. Lett.* 452, 629–633. doi: 10.1038/nature06810

Dwivedi, B., Xue, B., Lundin, D., Edwards, R. A., and Breitbart, M. (2013). A bioinformatic analysis of ribonucleotide reductase genes in phage genomes andmetagenomes. *BMC Evol. Biol.* 13:1–17. doi: 10.1186/1471-2148-13-33

Faino, L., Seidl, M. F., Datema, E., van den Berg, G. C., Janssen, A., Wittenberg, A. H., et al. (2015). *Single-Molecule Real-Time Sequencing Combined With Optical Mapping Yields Completely Finished Fungal Genome. MBio*. [preprint]. doi: 10.1128/mBio.00936-15

Fancello, L., Trape, S., Catherine, R., Mickael, B., Nikolay, P., Raoult, D., et al. (2013). Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J.* 7, 359–369. doi: 10.1038/ismej.2012.101

Feazel, L. M., Spear, J. R., Berger, A. B., Harris, J. K., Frank, D. N., Ley, R. E., et al. (2008). Eucaryotic diversity in a hypersaline microbial mat. *Appl. Environ. Microbiol.* 74, 329–332. doi: 10.1128/AEM.01448-07

Fuchsman, C. A., Palevsky, H. I., Widner, B., Duffy, M., Carlson, M. C., Neibauer, J. A., et al. (2019). Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. *ISME J.* 1, 1–13. doi: 10.1038/s41396-019-0452-6

Furtak, V., Roivainen, M., Mirochnichenko, O., Zagorodnyaya, T., Laassri, M., Zaidi, S., et al. (2016). Environmental surveillance of viruses by tangential flow filtration and metagenomic reconstruction. *Eurosurveillance* 21:30193. doi: 10.2807/1560-7917.ES.2016.21.15.30193

Gong, Z., Liang, Y., Wang, M., Jiang, Y., Yang, Q., Xia, J., et al. (2018). Viral diversity and its relationship with environmental factors at the surface and deep sea of Prydz Bay. Antarctica. *Front. Microbiol.* 9:2981. doi: 10.3389/fmicb.2018.02981

Graham, E. B., Paez-Espino, D., Brislawn, C., Neches, R. Y., Hofmockel, K. S., Wu, R., et al. (2019). *Untapped Viral Diversity in Global Soil Metagenomes. biorxiv*. [preprint]. doi: 10.1101/583997

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., et al. (2019). Marine DNA viral macro-and microdiversity from pole to pole. *Cell* 177, 1109–1123. doi: 10.1016/j.cell.2019.03.040

Guajardo-Leiva, S., Pedrós-Alió, C., Salgado, O., Pinto, F., and Díez, B. (2018). Active crossfire between *Cyanobacteria* and Cyanophages in phototrophic mat communities within hot springs. *Front. Microbiol.* 9:2039. doi: 10.3389/fmicb.2018.02039

Gudbergsdóttir, S. R., Menzel, P., Krogh, A., Young, M., and Peng, X. (2016). Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. *Environ. Microbiol.* 18, 863–874. doi: 10.1111/1462-2920.13079

Hara, S., Terauchi, K., and Koike, I. (1991). Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy. *Appl. Environ. Microbiol.* 57, 2731–2734.

Hayes, S., Mahony, J., Nauta, A., and Van Sinderen, D. (2017). Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* 9:E127. doi: 10.3390/v9060127

Heidelberg, K. B., Nelson, W. C., Holm, J. B., Eisenkolb, N., Andrade, K., Emerson, J., et al. (2013). Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell. Australia. *Front. Microbiol.* 4:115. doi: 10.3389/fmicb.2013.00115

Heller, R., Chung, S., Crissy, K., Dumas, K., Schuster, D., and Schoenfeld, T. (2019). Engineering of a thermostable viral polymerase using metagenome-derived diversity for highly sensitive and specific RT-PCR. *Nucleic Acids Res.* 1, 1–12. doi: 10.1093/nar/gkz104

Hirai, M., Nishi, S., Tsuda, M., Sunamura, M., Takaki, Y., and Nunoura, T. (2017). Library construction from subnanogram DNA for pelagic sea water and deep-sea sediments. *Microb. Environ.* 4, 336–343. doi: 10.1264/jsme2.ME17132

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., et al. (2014). EBI metagenomics—a new resource for the analysis and archiving

of metagenomic data. *Nucleic Acids Res.* 42, D600–D606. doi: 10.1093/nar/gkt961

Hurwitz, B. L., Deng, L., Poulos, B. T., and Sullivan, M. B. (2013). Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* 15, 1428–1440. doi: 10.1111/j.1462-2920.2012.02836.x

John, S. G., Mendez, C. B., Deng, L., Poulos, B., Kauffman, A. K. M., Kern, S., et al. (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* 3, 195–202. doi: 10.1111/j.1758-2229.2010.00208.x

Kang, S., and Kim, Y. C. (2018). Identification of Viral Taxon-Specific Genes (VTSG): application to Caliciviridae. *Genom. Inform.* 16:e23. doi: 10.5808/GI.2018.16.4.e23

Kiehl, J. T., and Shields, C. A. (2005). Climate simulation of the latest Permian: implications for mass extinction. *Geology* 33, 757–760. doi: 10.1130/G21654.1

Kinsella, C. M., Deijs, M., and van der Hoek, L. (2019). Enhanced bioinformatic profiling of VIDISCA libraries for virus detection and discovery. *Vir. Res.* 263, 21–26. doi: 10.1016/j.virusres.2018.12.010

Kleiner, M., Hooper, L. V., and Duerkop, B. A. (2015). Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 16:7. doi: 10.1186/s12864-014-1207-4

Konstantinidis, K. T., Braff, J., Karl, D. M., and DeLong, E. F. (2009). Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl. Environ. Microbiol.* 75, 5345–5355. doi: 10.1128/AEM.00473-09

Kristensen, D. M., Mushegian, A. R., Dolja, V. V., and Koonin, E. V. (2012). New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11–19. doi: 10.1016/j.tim.2009.11.003.New

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., and Koonin, E. V. (2018). Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Vir. Res.* 244, 181–193. doi: 10.1016/j.virusres.2017.11.025

Laanto, E., Hoikkala, V., Ravantti, J., and Sundberg, L. (2017). Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat. Com.* 8, 2–8. doi: 10.1038/s41467-017-00158-7

Lawrence, J. E., and Steward, G. F. (2010). "Purification of viruses by centrifugation," in *Manual of Aquatic Viral Ecology*, eds S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle, (Waco, TX: ASLO), 166–181. doi: 10.4319/mave.2010.978-0-9845591-0-7.166

Le Romancer, M., Gaillard, M., Geslin, C., and Prieur, D. (2006). Viruses in extreme environments. *Rev. Environ. Sci. Biotechnol.* 6, 17–31. doi: 10.1007/s11157-006-0011-2

Liang, Y., Wang, L., Wang, Z., Zhao, J., Yang, Q., Wang, M., et al. (2019). Metagenomic analysis of the diversity of DNA viruses in the surface and deep-sea of the South China Sea. *Front. Microbiol.* 10:1951. doi: 10.3389/fmicb.2019.01951

Liu, Y., Brandt, D., Ishino, S., Ishino, Y., Koonin, E. V., Kalinowski, J., et al. (2019). New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. *Environ. Microbiol.* 21, 2002–2014. doi: 10.1111/1462-2920.14479

López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., and Alcamí, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* 80, 858–861. doi: 10.1126/science.1179287

López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M., and Rodriguez-Valera, F. (2017). Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genetics* 13:e1007018. doi: 10.1371/journal.pgen.1007018

Ma, Y., Galinski, E. A., Grant, W. D., Oren, A., and Ventosa, A. (2010). Halophiles 2010: life in Saline Environments. *Appl. Environ. Microbiol.* 76, 6971–6981. doi: 10.1128/AEM.01868-10

Martinez-Hernandez, F., Fornas, O., Gomez, M. L., Bolduc, B., de La Cruz Peña, M. J., Martínez, J. M., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* 8:15892. doi: 10.1038/ncomms15892

Martin-Cuadrado, A. B., Senel, E., Martínez-García, M., Cifuentes, A., Santos, F., Almansa, C., et al. (2019). Prokaryotic and viral community of the sulfate-rich crust from Peñahueca ephemeral lake, an astrobiology analogue. *Environ. Microbiol.* 21, 3577–3600. doi: 10.1111/1462-2920.14680

Mead, D. A., Monsma, S., Mei, B., Gowda, K., Lodes, M., Schoenfeld, T. W., et al. (2017). "Functional metagenomics of a replicase from a novel hyperthermophilic aquificales virus," in *Functional Metagenomics: Tools and Applications*, eds T. C. Charles, and M. R. Liles, (Cham: Springer), 217–242. doi: 10.1007/978-3-319-61510-3_13

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780

Miranda, J. A., Culley, A. I., Schvarcz, C. R., and Steward, G. F. (2016). RNA viruses as major contributors to Antarctic virioplankton. *Environ. Microbiol.* 18, 3714–3727. doi: 10.1111/1462-2920.13291

Mizuno, C. M., Ghai, R., Saghai, A., Lopez-Garcia, P., and Rodriguez-Valera, F. (2016). *Genomes of Abundant and Widespread Viruses from the Deep Ocean. MBio*. [preprint]. doi: 10.1128/mBio.00805-16.

Mizuno, C. M., Prajapati, B., Lucas-Staat, S., Sime-Ngando, T., Forterre, P., Bamford, D. H., et al. (2019). Novel haloarchaeal viruses from lake retba infecting haloferax and halorubrum species. *Environ. Microbiol.* 21, 2129–2147. doi: 10.1111/1462-2920.14604

Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., and Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genetics* 9:e1003987. doi: 10.1371/journal.pgen.1003987

Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., Sako, Y., Forterre, P., and Prangishvili, D. (2012). Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc. Natl. Acad. Sci.* 109, 13386–13391. doi: 10.1073/pnas.1203668109

Mokili, J. L., Rohwer, F., and Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2, 63–77. doi: 10.1016/j.coviro.2011.12.004

Moser, M. J., Difrancesco, R. A., Gowda, K., Klingele, A. J., Sugar, D. R., Stocki, S., et al. (2012). Thermostable DNA Polymerase from a Viral Metagenome Is a Potent RT-PCR Enzyme. *PLoS One* 7:e38371. doi: 10.1371/journal.pone.0038371

Motlagh, A. M., Bhattacharjee, A. S., Coutinho, F. H., Dutilh, B. E., Casjens, S. R., and Goel, R. K. (2017). Insights of phage-host interaction in hypersaline ecosystem through metagenomics analyses. *Front. Microbiol.* 8:352. doi: 10.3389/fmicb.2017.00352

Munson-McGee, J., Snyder, J., and Young, M. (2018). Archaeal viruses from high-temperature environments. *Genes* 9:E128. doi: 10.3390/genes9030128

Nguyen, T. T., and Landfald, B. (2015). Polar front associated variation in prokaryotic community structure in Arctic shelf seafloor. *Front. Microbiol.* 6:17. doi: 10.3389/fmicb.2015.00017

Nigro, O. D., Jungbluth, S. P., Lin, H. T., Hsieh, C. C., Miranda, J. A., Schvarcz, C. R., et al. (2017). *Viruses in the Oceanic Basement MBio.* [preprint]. doi: 10.1128/mBio.02129-16

Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M. P. (2018). Overview of virus metagenomic classification methods and their biological applications. *Front. Microbiol.* 9:749. doi: 10.3389/fmicb.2018.00749

Oke, M., Carter, L. G., Johnson, K. A., Liu, H., McMahon, S. A., Yan, X., et al. (2010). The scottish structural proteomics facility: targets. methods and outputs. *J. Struct. Funct. Gen.* 11, 167–180. doi: 10.1007/s10969-010-9090

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., et al. (2016). Uncovering Earth's virome. *Nature* 536, 425–430. doi: 10.1038/nature19094

Paez-Espino, D., Roux, S., Chen, I., Palaniappan, K., Ratner, A., Chu, K., et al. (2018). IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.* 47, D678–D686. doi: 10.1093/nar/gky1127

Parvathi, A., Jasna, V., Aparna, S., Pradeep Ram, A., Aswathy, V., Balachandran, K., et al. (2018). High incidence of lysogeny in the oxygen minimum zones of the arabian sea (Southwest Coast of India). *Viruses* 10:588. doi: 10.3390/v10110588

Paulmier, A., and Ruiz-Pino, D. (2009). Progress in oceanography oxygen minimum zones (OMZs) in the modern ocean. *Prog. Oceanogr.* 80, 113–128. doi: 10.1016/j.pocean.2008.08.001

Pietilä, M. K., Laurinavičius, S., Sund, J., Roine, E., and Bamford, D. H. (2010). The single-stranded DNA genome of novel archaeal virus Halorubrum pleomorphic virus 1 is enclosed in the envelope decorated with glycoprotein spikes. *J. Virol.* 8, 788–798. doi: 10.1128/JVI.01347-09

Plominsky, A., Henríquez, C. A., Delherbe, N., Podell, S., Ramirez, S., Ugalde, J. A., et al. (2018). Distinctive archaeal composition of an artisanal crystallizer pond and functional insights into salt-saturated hypersaline environment adaptation. *Front. Microbiol.* 9:1800. doi: 10.3389/fmicb.2018.01800

Porter, K., Russ, B. E., and Dyall-smith, M. L. (2007). Virus–host interactions in salt lakes. *Curr. Opin. Microbiol.* 10, 418–424. doi: 10.1016/j.mib.2007.05.017

Poulos, B. T., John, S. G., and Sullivan, M. B. (2018). "Iron Chloride Flocculation of Bacteriophages from Seawater," in *Bacteriophages. Methods in Molecular Biology*, eds M. Clokie, A. Kropinski, and R. Lavigne, (New York, NY: Humana Press), 1681.

Ramos-Barbero, M. D., Martínez, J. M., Almansa, C., Rodríguez, N., Villamor, J., Gomariz, M., et al. (2019). Prokaryotic and viral community structure in the singular chaotropic salt lake salar de uyuni. *Environ. Microbiol.* 21, 2029–2042. doi: 10.1111/1462-2920.14549

Rastrojo, A., and Alcamí, A. (2017). Aquatic viral metagenomics: Lights and shadows. *Virus Res.* 239, 87–96. doi: 10.1016/j.virusres.2016.11.021

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69. doi: 10.1186/s40168-017-0283-5

Resplandy, L., Keeling, R. F., Eddebbar, Y., Brooks, M. K., Wang, R., Bopp, L., et al. (2018). Quantification of ocean heat uptake from changes in atmospheric O_2 and CO_2 composition. *Nature* 563:105. doi: 10.1038/s41586-018-0651-8

Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001). Viruses from extreme thermal environments. *PNAS* 98, 13341–13345. doi: 10.1073 pnas.231170198

Rosario, K., and Breitbart, M. (2011). Exploring the viral world through metagenomics karyna rosario and mya breitbart. *Curr. Opin. Virol.* 1, 289–297. doi: 10.1016/j.coviro.2011.06.004

Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L., and Prosperi, M. (2016). Challenges in the analysis of viral metagenomes. *Virus Evol.* 2:vew0022. doi: 10.1093/ve/vew022

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., et al. (2019). Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* 37:29. doi: 10.1038/nbt.4306

Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. doi: 10.7717/peerj.985

Roux, S., Enault, F., Ravet, V., Colombet, J., Bettarel, Y., Auguet, J., et al. (2016). Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environ. Microbiol.* 18, 889–903. doi: 10.1111/1462-2920.13084

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

Sangwan, N., Lambert, C., Sharma, A., Gupta, V., Khurana, P., Khurana, J. P., et al. (2015). Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator–prey genotypes. *Environ. Microbiol. Rep.* 7, 812–823. doi: 10.1111/1758-2229.12297

Schmidt, H. F., Sakowski, E. G., Williamson, S. J., Polson, S. W., and Wommack, K. E. (2014). Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. *ISME J.* 8, 103–114. doi: 10.1038/ismej.2013.124

Schmitz, J. E., Schuch, R., and Fischetti, V. A. (2010). Identifying active phage lysins through functional. *Appl. Energy* 76, 7181–7187. doi: 10.1128/AEM.00732-10

Schoenfeld, T., Liles, M., Wommack, K. E., Polson, S. W., Godiska, R., and Mead, D. (2009). Functional viral metagenomics and the next generation of molecular tools. *Cell Press* 18, 20–30. doi: 10.1016/j.tim.2009.10.001

Schoenfeld, T. W., and Mead, D. (2015). "Use of viral metagenomes from yellowstone hot springs to study phylogenetic relationships and evolution," in *Encyclopedia of Metagenomics*, ed. K. E. Nelson (New York, NY: Springer), 681–700.

Schoenfeld, T. W., Moser, M. J., and Mead, D. (2015). "Functional viral metagenomics and the development of new enzymes for DNA and RNA amplification and sequencing," in *Encyclopedia of Metagenomics*, ed. K. E. Nelson (Boston, MA: Springer), 198–218.

Schoenfeld, T., Patterson, M., Paul, M., Wommack, K. E., and Young, M. (2008). Assembly of viral metagenomes from Yellowstone Hot Springs. *Appl. Enrion. Microbiol.* 74, 4164–4174. doi: 10.1128/AEM.02598-07

Sharma, A., Schmidt, M., Kiesel, B., Mahato, N. K., Cralle, L., Singh, Y., et al. (2018). Bacterial and archaeal viruses of Himalayan Hot Springs At Manikaran Modulate Host Genomes. *Front. Microbiol.* 9:3095. doi: 10.3389/fmicb.2018.03095

Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J. Gen. Virol.* 96, 1193–1206. doi: 10.1099/vir.0.000016

Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15:161. doi: 10.1038/nrmicro.2016.177

Steward, G. F., Culley, A. I., Mueller, J. A., Wood-Charlson, E. M., Belcaid, M., and Poisson, G. (2013). Are we missing half of the viruses in the ocean? *ISME J.* 7, 672–679. doi: 10.1038/ismej.2012.121

Strazzulli, A., Fusco, S., Moracci, M., and Contursi, P. (2017). Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev. Environ. Sci. Bio/Techno.* 16, 425–454. doi: 10.1007/s11157-017-9435-0

Tennant, P., Fermin, G., and Foster, J. E. (eds) (2018). *Viruses: Molecular Biology, Host Interactions, and Applications to Biotechnology*. Cambridge, MA: Academic Press.

Thamdrup, B. (2012). New pathways and processes in the global nitrogen cycle. *Annu. Rev. Ecol. Evol. Syst.* 43, 407–428. doi: 10.1146/annurev-ecolsys-102710-145048

Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470–483. doi: 10.1038/nprot.2009.10

Van Etten, J. L., Lane, L. C., and Dunigan, D. D. (2010). DNA viruses: the really big ones (giruses). *Annu. Rev. Microbiol.* 64, 83–99. doi: 10.1146/annurev.micro.112408.134338

Ventosa, A., de la Haba, R. R., Sánchez-Porro, C., and Papke, R. T. (2015). Microbial diversity of hypersaline environments: a metagenomic approach. *Curr. Opin. Microb.* 25, 80–87. doi: 10.1016/j.mib.2015.05.002

Vikram, S., Guerrero, L. D., Makhalanyane, T. P., Le, P. T., Seely, M., and Cowan, D. A. (2016). Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environ. Microbiol.* 18, 1875–1888. doi: 10.1111/1462-2920.13088

Wang, H., Yu, Y., Liu, T., Pan, Y., Yan, S., and Wang, Y. (2015). Diversity of putative archaeal RNA viruses in metagenomic datasets of a yellowstone acidic hot spring. *Springer Plus* 18 4:189. doi: 10.1186/s40064-015-0973-z

Weitz, J. S., and Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* 4, 17. doi: 10.3410/B4-17

Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., et al. (2008). The sorcerer ii global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* 3:e1456. doi: 10.1371/journal.pone.0001456

Willner, D., Thurber, R. V., and Rohwer, F. (2009). Metagenomic signatures of 86 microbial and viral metagenomes. *Environ. Microbiol.* 11, 1752–1766. doi: 10.1111/j.1462-2920.2009.01901.x

Wommack, K. E., Bhavsar, J., Polson, S. W., Chen, J., Dumas, M., Srinivasiah, S., et al. (2012). VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand. Genomic Sci.* 6, 421–433. doi: 10.4056/sigs.2945050

Yau, S., and Seth-Pasricha, M. (2019). Viruses of polar aquatic environments. *Viruses* 11:189. doi: 10.3390/v11020189

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

Yoshida, M., Mochizuki, T., Urayama, S. I., Yoshida-Takashima, Y., Nishi, S., et al. (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. doi: 10.3389/fmicb.2018.00075

Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., et al. (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *J. Virol.* 93, e2130–e2118. doi: 10.1111/1462-2920.14549

Zablocki, O., Adriaenssens, E. M., and Cowan, D. (2016). Diversity and ecology of viruses in hyperarid desert soils. *Appl. Environ. Microbiol.* 82, 770–777. doi: 10.1128/AEM.02651-15

Zablocki, O., Adriaenssens, E. M., Frossard, A., and Seely, M. (2017a). Metaviromes of extracellular soil viruses along a Namib desert aridity gradient. *Genome Announc.* 5, 4–5. doi: 10.1128/genomeA.01470-16

Zablocki, O., van Zyl, L., Kirby, B., and Trindade, M. (2017b). Diversity of dsDNA viruses in a South African hot spring assessed by metagenomics and microscopy. *Viruses* 9:348. doi: 10.3390/v9110348

Zablocki, O., Zyl, L. Van, Adriaenssens, E. M., Rubagotti, E., Tuffin, M., and Cary, C. (2014). High-Level diversity of tailed phages, eukaryote-associated viruses,and virophage-like elements in the metaviromes of Antarctic soils. *Appl. Environ. Microbiol.* 80, 6888–6897. doi: 10.1128/AEM.01525-14

Zhang, Y. Z., Shi, M., and Holmes, E. C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172. doi: 10.1016/j.cell.2018.02.043

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dávila-Ramos, Castelán-Sánchez, Martínez-Ávila, Sánchez-Carbente, Peralta, Hernández-Mendoza, Dobson, Gonzalez, Pastor and Batista-García. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.