

## Article

# Viral Diversity in Benthic Abyssal Ecosystems: Ecological and Methodological Considerations

Umberto Rosani <sup>1,\*</sup>, Cinzia Corinaldesi <sup>2</sup>, Gabriella Luongo <sup>3</sup>, Marco Sollitto <sup>4,5</sup>, Simeone Dal Monego <sup>6</sup>, Danilo Licastro <sup>6</sup>, Lucia Bongiorno <sup>7</sup>, Paola Venier <sup>1,†</sup>, Alberto Pallavicini <sup>4,†</sup> and Antonio Dell'Anno <sup>3,\*,†</sup>

<sup>1</sup> Department of Biology, University of Padova, Via U. Bassi 58/b, 35121 Padova, Italy; paola.venier@unipd.it

<sup>2</sup> Department of Materials, Environmental Sciences and Urban Planning, Polytechnic University of Marche, Via Brecce Bianche, 60131 Ancona, Italy; c.corinaldesi@staff.univpm.it

<sup>3</sup> Department of Life and Environmental Sciences, Polytechnic University of Marche, Via Brecce Bianche, 60131 Ancona, Italy; g.luongo@pm.univpm.it

<sup>4</sup> Department of Life Sciences, University of Trieste, Via Licio Giorgeri 5, 34127 Trieste, Italy; marco.sollitto@famnit.upr.si (M.S.); pallavic@units.it (A.P.)

<sup>5</sup> Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, 6000 Koper, Slovenia

<sup>6</sup> Laboratorio di Genomica ed Epigenomica, AREA Scienze Park, Padriciano 99, 34149 Trieste, Italy; simeone.dalmonego@areasciencepark.it (S.D.M.); danilo.licastro@areasciencepark.it (D.L.)

<sup>7</sup> Consiglio Nazionale delle Ricerche, Istituto di Scienze Marine, Tesa 104–Arsenale, Castello 2737/F, 30122 Venezia, Italy; lucia.bongiorno@ve.ismar.cnr.it

\* Correspondence: umberto.rosani@unipd.it (U.R.); a.dellanno@staff.univpm.it (A.D.)

† These authors contributed equally to this work.

**Abstract:** Viruses are the most abundant ‘biological entities’ in the world’s oceans. However, technical and methodological constraints limit our understanding of their diversity, particularly in benthic abyssal ecosystems (>4000 m depth). To verify advantages and limitations of analyzing virome DNA subjected either to random amplification or unamplified, we applied shotgun sequencing-by-synthesis to two sample pairs obtained from benthic abyssal sites located in the North-eastern Atlantic Ocean at ca. 4700 m depth. One amplified DNA sample was also subjected to single-molecule long-read sequencing for comparative purposes. Overall, we identified 24,828 viral Operational Taxonomic Units (vOTUs), belonging to 22 viral families. Viral reads were more abundant in the amplified DNA samples (38.5–49.9%) compared to the unamplified ones (4.4–5.8%), with the latter showing a greater viral diversity and 11–16% of dsDNA viruses almost undetectable in the amplified samples. From a procedural point of view, the viromes obtained by direct sequencing (without amplification step) provided a broader overview of both ss and dsDNA viral diversity. Nevertheless, our results suggest that the contextual use of random amplification of the same sample and long-read technology can improve the assessment of viral assemblages by reducing off-target reads.

**Keywords:** deep-sea viromes; shotgun viromics; ssDNA viruses



**Citation:** Rosani, U.; Corinaldesi, C.; Luongo, G.; Sollitto, M.; Dal Monego, S.; Licastro, D.; Bongiorno, L.; Venier, P.; Pallavicini, A.; Dell'Anno, A. Viral Diversity in Benthic Abyssal Ecosystems: Ecological and Methodological Considerations. *Viruses* **2023**, *15*, 2282. <https://doi.org/10.3390/v15122282>

Academic Editor: Jozef Nissimov

Received: 10 October 2023

Revised: 13 November 2023

Accepted: 18 November 2023

Published: 21 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Increasing evidence based on high-throughput sequencing (HTS) techniques indicates that marine viruses are highly diverse [1–6], and that benthic deep-sea environment can be a hotspot of viral diversity [7–14]. However, technical and methodological constraints still limit our understanding of the diversity of viruses present in deep-sea sediments. In fact, unlike cellular organisms which universally have double-stranded (ds) DNA genomes, viral genomes consist of either RNA or DNA molecules in ds or single-stranded (ss) forms, circular or linear, and encoded in single-unsegmented or multiple-segmented molecules, and display a remarkable diversity in terms of size and gene composition [15,16], hampering an unbiased overview of the viral diversity. So far, two different approaches have been used to analyse the diversity of DNA viruses in deep-sea sediments: one based on

the extraction of the viral particles from the sediment and subsequent concentration of the viral-size fraction through filtration and/or ultracentrifugation, before DNA extraction and virome analysis [2,7–9] and the other based on the extraction from the sediment of total DNA to be used directly for high-throughput shotgun sequencing, without any separation step of viral particles [10,11,13,14].

Both these approaches have intrinsic limitations and can introduce biases in assessing the actual viral diversity. Indeed, viruses can be extracted from the sediment with a different efficiency [8,17,18] and the total amount of DNA recovered is often very low and requires a random amplification step before the library preparation and sequencing, which can result in an over-representation of ssDNA viruses [8]. At the same time, the extraction of total DNA from environmental samples can underestimate viral diversity [19], as sample processing and sequencing can limit the identification of a large fraction of viruses with a single-strand genome [10–13], which have been repeatedly reported as an important fraction of the overall viral diversity in deep-sea sediments [2,7–9]. At present, methodological advancements allow preparing libraries starting from very low amounts of DNA (e.g., from 1 pg to 1 ng of total DNA extracted from surface water and deep-sea sediments [20]), possibly circumventing the problem of the pre-amplification step. Moreover, besides short-read sequencing platforms, such as Illumina, long-read sequencing platforms can also be used to facilitate the reconstruction of viral genomes [21,22], despite being more error-prone [23].

Here, to evaluate advantages and limitations of analysing unamplified (native) and randomly amplified viral-enriched DNA samples, we applied short-read Illumina shotgun sequencing to two sample pairs obtained from the upmost sediment of two abyssal sites (ca. 4700 m depth) of the North-eastern Atlantic Ocean. Long-read Nanopore sequencing was applied to one amplified sample for comparison and the taxonomic composition of the DNA viromes obtained with the different methodologies was compared and discussed to expand our knowledge of viral diversity in the largest ecosystem of Earth.

## 2. Materials and Methods

### 2.1. Sediment Collection and Viral DNA Extraction

Undisturbed sediment samples were collected using a multi-corer between May and June 2018 at two sites (hereafter defined as Site 39, 48°58.939' N and 16°33.026' W, and Site 57, 48°58.964' N and 16°32.866' W) ca. 1.8 km from each other and at a depth of ca. 4700 m in the Abyssal Hill province, located 480 km southwest of Ireland in the North-eastern Atlantic Ocean. The Abyssal Hill province is characterized by an elevation of ~300 m from the abyssal plain (the Porcupine Abyssal Plain) and current speeds up to ~10 cm s<sup>-1</sup> [24]. Sediments of the Abyssal Hill province are generally characterised by a coarser grain size compared to abyssal sediments [24,25] and the C input to the seafloor is estimated to be ca. 1.1 mmol C m<sup>-2</sup> d<sup>-1</sup>, largely composed of labile organic matter [26]. The C inputs to the Abyssal Hill seafloor sustain high biomass and diversified megafauna assemblages (ca. 9 g m<sup>-2</sup>, largely represented by suspension feeders [27]), and a relatively high C consumption rate (0.479 mmol C m<sup>-2</sup> d<sup>-1</sup> vs. 0.340 mmol C m<sup>-2</sup> d<sup>-1</sup> of bacteria [26]). Sediment samples were collected on the flank of the abyssal hill at ca. 30 m depth above the abyssal plain with a median flank slope angle of 7.6° [25]. After recovery, the superficial sediment layer (0–1 cm) was snap-frozen and maintained at –80 °C until further use. Viruses were extracted from the sediment according to a previously published procedure [8]. Before virus extraction, a similar amount of the top 0–1 cm of sediment collected by three independent multi-corer deployments at each benthic site was pooled together to provide a more realistic view of the actual viral diversity present therein. Briefly, a total amount of 50 g of wet sediment was mixed with 50 mL of virus-free seawater (pre-filtered onto 0.02-µm-pore-size filters and autoclaved) into a sterile jar and aliquots of 2 mL of the sediment slurry were split into 50 sterile tubes to increase the viral recovery efficiency. Epifluorescence microscopy observations using Sybr Gold as stain revealed the lack of contamination by viral particles of seawater used to prepare the sediment slurries.

Each sample was then supplemented with tetrasodium pyrophosphate solution (5 mM final concentration), incubated for 15 min in the dark and sonicated 3 times (Branson Sonifier 2200, 60 W, Thermofisher, Waltham, MA, USA) for 1 min. After centrifugation ( $800\times g$  for 10 min), the supernatants were collected and the sediments were resuspended again with 4 mL of virus-free seawater, centrifuged ( $800\times g$  for 10 min) and the supernatant collected. This step was repeated three times. The supernatants were combined and incubated with DNase I (1 U/mL, Merck, St. Luis, MO, USA) in the dark at room temperature for 15 min to remove extracellular DNA. The supernatant was pre-filtered through 0.2  $\mu\text{m}$  pore size filters (Whatman Anopore, Merck, Darmstadt, Germany) to remove cellular organisms before concentrating viral particles onto 0.02  $\mu\text{m}$  pore size  $\text{Al}_2\text{O}_3$  filters (Anodisc; diameter 47 mm, Merck). Filters containing the collected viruses were subjected to DNA extraction by using the DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) following manufacturer instructions. The viral DNA obtained was split into two aliquots, one directly used for library preparation and sequencing, whereas the other was subjected to amplification using the highly processive, proof-reading proficient, Phi29 polymerase (GenomiPhi kit, Merck) and random hexamer primers (Thermofisher), before further processing. Samples amplified with the GenomiPhi kit were then processed with the DNeasy PowerClean Cleanup Kit (Qiagen) following manufacturer instructions. The DNA concentrations in all samples were quantified fluorometrically with SYBR Gold on a NanoDrop ND-3000 (Thermofisher). The size distribution of the DNA samples was determined by microfluidics-based electrophoresis using a DNA high-sensitivity (HS) kit in an Agilent 2100 instrument (Agilent, Santa Clara, CA, USA).

## 2.2. Library Preparation and Illumina Short-Read Sequencing

The DNA used to prepare libraries for short read sequencing was fragmented with a Bioruptor<sup>TM</sup> UCD-200 instrument (Diagenode, Denville, NJ, USA), using the tuning intensity and fragmentation times described below. Amplified DNA samples were fragmented for 15 min, with 45''/15'' on/off steps, at maximum intensity with samples maintained on ice. Unamplified DNA samples were fragmented with the same cycle for 2 min. DNA libraries were prepared using the Swift Accel-NGS 1S Plus DNA Library Kit (Swift Bio-Sciences, Washtenaw County, MI, USA), following manufacturer recommendations. A starting DNA amount of 10 ng or approximately 20–50 pg was used for the amplified and unamplified samples, respectively (Table 1). The number of PCR cycles was set to 5 for amplified samples and 16 for the unamplified ones. DNA libraries were quantified with a Qubit 2.0 instrument (Thermofisher) using a DNA HS kit and the size distributions were determined as reported above. Libraries were pooled in equimolar quantities and sequenced on an Illumina NovaSeq6000 instrument (Area Science Park, Trieste, Italy), with a 150 bp paired-end layout.

**Table 1.** Summary of samples, sequencing and assembly details. Sample ID and description, sequencing technology, the starting amount of DNA used to prepare the libraries, number of deduplicated reads, number of contigs longer than 1.5 kb and size of the longest contig (kb) are reported for each sample.

Sample ID	Description	Technology	Starting DNA Quantity	No. of Deduplicated Reads [ $\times 10^6$ ]	No. of Contigs Longer Than 1.5 kb	Longest Contig (kb)
39NA	Site 39, not amplified	Illumina	20–50 pg	157.3	15,199	72,980
39A	Site 39, amplified	Illumina	10 ng	488.0	79,174	151,451
57NA	Site 57, not amplified	Illumina	20–50 pg	211.0	35,891	135,810
57A	Site 57, amplified	Illumina	10 ng	471.0	62,878	135,810
		Nanopore	1 $\mu\text{g}$	0.331	1009	

### 2.3. Nanopore Library Preparation and Sequencing

The amplified DNA obtained from Site 57 has also been sequenced by means of long-read Oxford Nanopore Technology platform. It was not possible to apply the same technology to the non-amplified DNA since the minimal amount of required DNA is currently 1 µg. The amplified DNA used for long-read sequencing was not fragmented and the sequencing library was prepared with the standard ligation sequencing kit LSK109 (Oxford Nanopore Technologies, Oxford, UK), adjusting the DNA repair and end-prep incubation step to 20 min. Sequencing was conducted on a MinION equipped with a FLO-MIN106 (R 9.4.1) flowcell (Oxford Nanopore Technologies) for 48 h with the base calling mode set as active.

### 2.4. Sequencing Data Processing

Raw Illumina readouts were demultiplexed and reads were trimmed for quality and to remove sequencing adaptors using TrimGalore (<https://github.com/FelixKrueger/TrimGalore> (accessed on 1 May 2023)). The minimal quality was set to PHRED25 and the last 10 nt at the 5'-end of each read were removed to avoid quality drop-offs as well as the presence of low complexity tails, as suggested in the library kit manufacturer's instructions. Reads were deduplicated using the dedupe tool implemented in the bmap suite [28] and reads were assembled using MEGAHIT [29], applying the a set of kmers in the 21–119 range (21, k + 10, 119) and selecting a minimum contig length of 1500 nt. For long reads, the raw electric signals were base called using guppy v.6.2.1 in high confidence mode to generate fastq files, which were then merged per sample. Critical parameters, such as read length and average quality of the readouts, were evaluated using Nanoplot v1.40.2 [30]. Nanopore reads were screened for the presence of chimeric regions using yacrdr v0.6.2 [31] and only reads passing coverage and chimeric filters were further used as input for read correction based on racon, using the 'normal' mode with a window size of 500 nt. Filtered nanopore reads were then assembled using metaFlye [32].

### 2.5. Identification and Analysis of Viral Sequences

To identify putative viruses we applied the Metaphage v.2 beta pipeline (<https://github.com/MattiaPandolfoVR/MetaPhage/> (accessed on 1 May 2023)); [33], with several modifications. The databases utilised for the data analysis have been downloaded as indicated (all databases have been accessed on 1 May 2023): Inphared v.Jan2022 (<https://github.com/RyanCook94/inphared>), Phigaro v2.3.0 (<https://github.com/bobeobibo/phigaro>), KEGG (<ftp://ftp.genome.jp/pub/db/kofam/archives/2019-03-20/>), Pfam v.32 (<https://pfam.xfam.org> (<ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam32.0/>)), VOG release 94 (<http://vogdb.org/> (<http://fileshare.csb.univie.ac.at/vog/vog94/>)), Virsorter2 v2.2.3 (<https://osf.io/v46sc/download>), checkV v1.5 (<https://portal.nersc.gov/CheckV/>), CAT/BAT ([tbb.bio.uu.nl/bastiaan/CAT\\_prepare/CAT\\_prepare\\_20210107.tar.gz](http://tbb.bio.uu.nl/bastiaan/CAT_prepare/CAT_prepare_20210107.tar.gz)) and NCBI nr (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz>, accessed on 1 February 2023). PHROGs, CARD and VFDB databases were downloaded from [https://zenodo.org/record/7563578/files/pharokka\\_v1.2.0\\_database.tar.gz](https://zenodo.org/record/7563578/files/pharokka_v1.2.0_database.tar.gz). Five different tools were used to predict viral sequences, including virsorter2 v2.2.3 [34], deepvirfinder v.1.0 [35], phigaro [36], VIRBRANT v1.2.1 [37] and VirFinder v1.1 [38], setting a minimal length of 2 kb for the tools allowing this parameter. Cd-hit-est [39] was used to dereplicate the resulting putative viral sequences applying 0.95 of similarity over 0.85 of contig length, thus providing viral Operational Taxonomic Units (vOTUs) as output. The vOTUs were further filtered with checkV v0.8.1 [40] to trim possible host genes and handle duplicate segments of circular contigs, removing the vOTUs with no detected viral genes. To evaluate the abundances of the vOTUs in the different samples, the quality-trimmed reads were mapped back on the vOTUs using bowtie v1.3.1 [41]. For Nanopore-derived vOTUs, minimap2 [42] was used for read back-mapping with standard parameters (-x ava-ont). The counts of mapped reads were used to compute relative abundance values as Count Per Millions (CPMs), thus normalizing for the contig length and the total mapped reads in each virome dataset. A

rarefaction analysis using 6,895,930 reads per sample has been performed with the R package *vegan* v2.6. The taxonomic classification of the vOTUs was assessed by *vpf-tools* [43], which uses viral protein families for taxonomic determination and by *CAT* v5.0.4 [44], based on *prodigal* v2.6.3 [45] and *DIAMOND* v2.0.6 [46]. To evaluate the result of *vpf-tools*, we considered the taxonomic assignment with a confidence score (CS) of  $\geq 0.2$  and membership ratio (MR) of  $\geq 0.2$ , as previously reported [42]. Functional annotation of the vOTUs was performed with *pharokka* v1.2.1 [47] based on the proteins predicted with *prodigal*, using PHROGs (38,880 protein orthologous groups), CARD (5170 Antimicrobial Resistance Detection Models) and VFDB (32,439 virulence factor-related genes), as reference databases applying an E-value to 0.01 to detect conserved domains. Predicted proteins were also blasted against the NCBI nr database (downloaded February 2023) using *DIAMOND*. The alpha diversity was computed across samples using a dedicated R script, as described in the Metaphage pipeline. Putative auxiliary metabolic genes were identified with *VIBRANT*. Data comparison and plotting were performed with R studio 23.03.0 based on R-4.2.1 using the following packages: *tidyverse* v2.0.0, *magrittr* v2.0.3, *phyloseq* v1.42.0, *plotly* v4.10.2 and *matrixStats* v1.0.0.

### 3. Results

#### 3.1. Comparison between Unamplified and Randomly Amplified Benthic Abyssal Viromes

Shotgun short-read sequencing of four DNA libraries generated from pg to ng of virus-enriched unamplified (NA) or amplified (A) DNA samples yielded 2.678 billion reads, which were reduced to 1.327 billion reads by removing exact replicates of the same read (i.e., deduplication, Table 1). Non-amplified samples were characterized by a higher number of identical reads, resulting in a lower read number after deduplication (57.3 M for 39NA, 211.0 M for 57NA), compared to amplified samples (488.0 M for 39A and 471.0 M for 57A). Read assemblies resulted in 15,199, 79,174, 35,891 and 62,878 contigs longer than 1.5 kb for samples 39NA, 39A, 57NA, and 57A, respectively (Table 1). These contigs were separately used as input for five different bioinformatic tools allowing for virus identification based on machine learning or protein similarity approaches. Overall, a total of 88,086 unique contigs were identified as putative viruses, called vOTUs after dereplication. In more detail, 4 out of 5 tools identified more than 40,000 vOTUs, while *phigaro* identified a few hits only (Figure S1). Although 11,463 vOTUs were identified by all 4 tools, the majority of vOTUs were predicted by single tools or by different combinations of them. *CheckV* was used to further screen the putative vOTUs for the presence of provirus contamination, duplicated segments of circular sequences and to remove contigs with no viral gene detected, resulting in 24,828 high-confidence vOTUs, which defined the virome of the 2 investigated benthic abyssal sites. In all the analysed samples the sequencing depth was adequate to reach saturation (Figure S2).

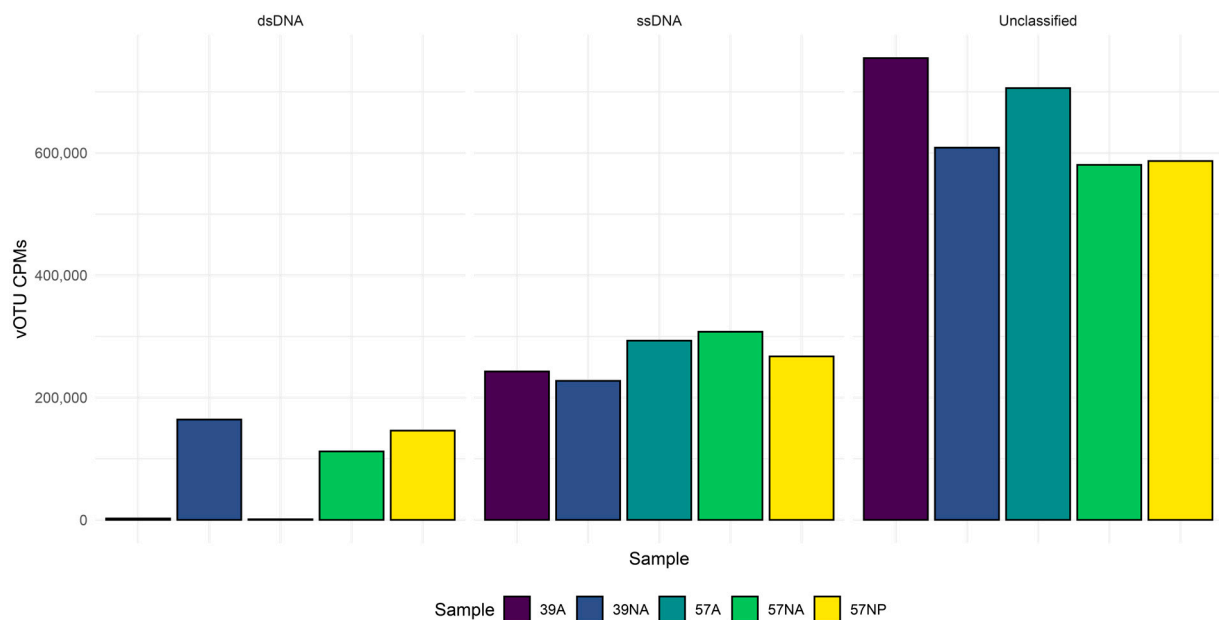
Back-mapping of the reads on the 24,828 vOTUs was used to determine their relative abundances in the different samples, to evaluate the viral diversity among samples as well as the taxonomic composition of the viromes in terms of Baltimore classification, family, and genus. Notably, the number of the reads mapped on the vOTUs revealed that the amplified DNA samples were considerably enriched in viral sequences (38.5 and 49.9% of total reads for sample 39A and 57A, respectively), compared to the non-amplified samples (4.4 and 5.8%, for sample 39NA and 57NA, respectively, Table 2). Applying a cut off 3 CPMs to remove poorly represented vOTUs, the total number of vOTUs per sample was slightly higher in the non-amplified samples (11,031, 12,354, 10,346, and 10,393 counted vOTUs for 39NA, 57NA, 57A, and 39A samples, respectively; Table 2). Accordingly, viral richness values estimated by Chao1 as well as values of viral diversity by Shannon and Fisher indexes were higher in non-amplified than in amplified samples (Table 2).

**Table 2.** Overview of the benthic abyssal viromes. The number of the high-confidence vOTUs identified, together with the percentages of mapped reads, viral richness estimated by Chao1 and alpha diversity values, calculated according to Shannon and Fisher indexes, are reported per sample. NA: non-amplified; A: amplified.

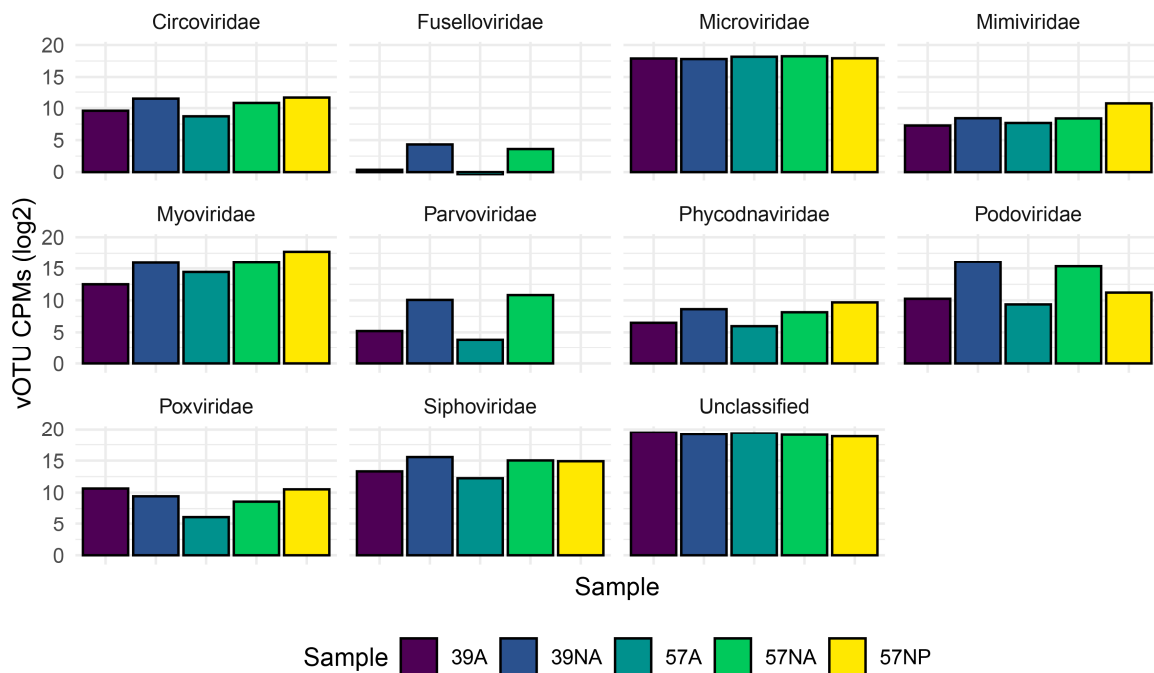
Sample ID	No. of vOTUs	Total Reads Mapped Back on vOTUs	Richness Chao1	Alpha Diversity Shannon	Alpha Diversity Fisher
39NA	11,031	4.4%	10,626	7.11	1661.12
39A	10,393	38.5%	9662	6.96	1487.58
57NA	12,354	5.8%	11,760	7.30	1875.05
57A	10,346	49.9%	9659	6.86	1486.90

### 3.2. Taxonomic Composition of the Benthic Abyssal Viromes

According to the taxonomic classification of vOTUs carried out with vpf-tools, we identified ssDNA and dsDNA viruses belonging to 22 viral families, including *Microviridae* (12,022), *Myoviridae* (352), *Siphoviridae* (324), *Podoviridae* (174), *Circoviridae* (28), *Parvoviridae* (3), *Poxviridae* (11), *Marseilleviridae* (3), *Phycodnaviridae* (18), and *Papillomaviridae* (3) as the 10 most represented families (Table S1). Considering the abundances per sample, vOTUs with no Baltimore classification were the dominant group, followed by ssDNA viruses (Figure 1). dsDNA viruses were mainly detectable in the non-amplified samples (16 and 11% of the total abundance for sample 39NA and 57NA, respectively), whereas their relative abundance dropped below 2.2 and 0.97% for sample 39A and 57A, respectively. At the family level, the most abundant vOTUs in all samples could not be taxonomically classified (54–73%), followed in decreasing abundance by those identified as *Microviridae*, *Podoviridae*, *Myoviridae*, and *Siphoviridae* (Figure 2).

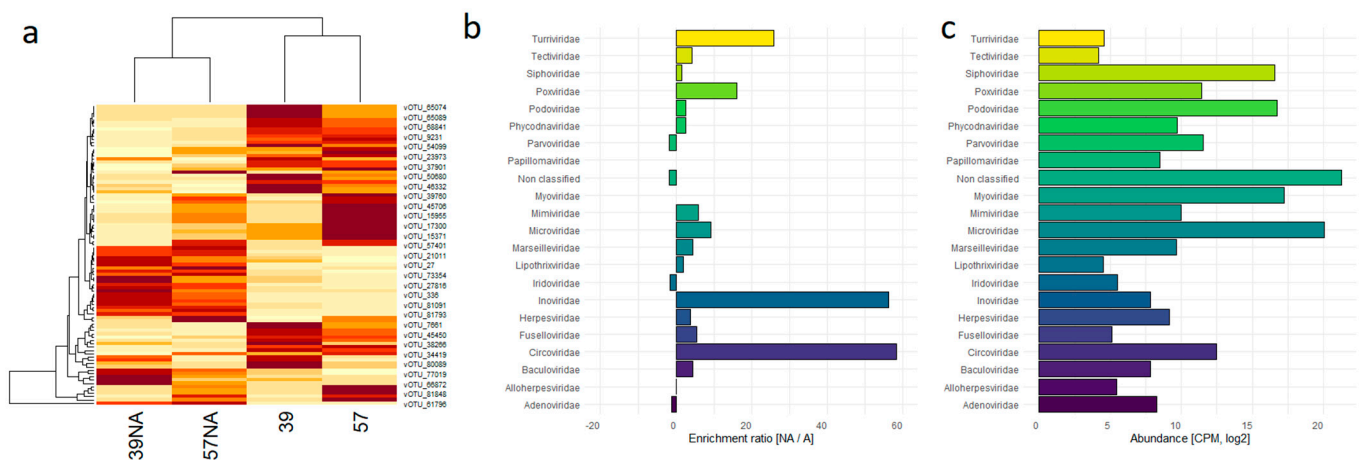


**Figure 1.** Taxonomic classification of the viromes in the two investigated benthic abyssal sites. The abundance of the vOTUs corresponding to dsDNA, ssDNA, and uncharacterized Baltimore groups are reported for the five reconstructed viromes. NA: non-amplified; A: amplified; NP: Nanopore.



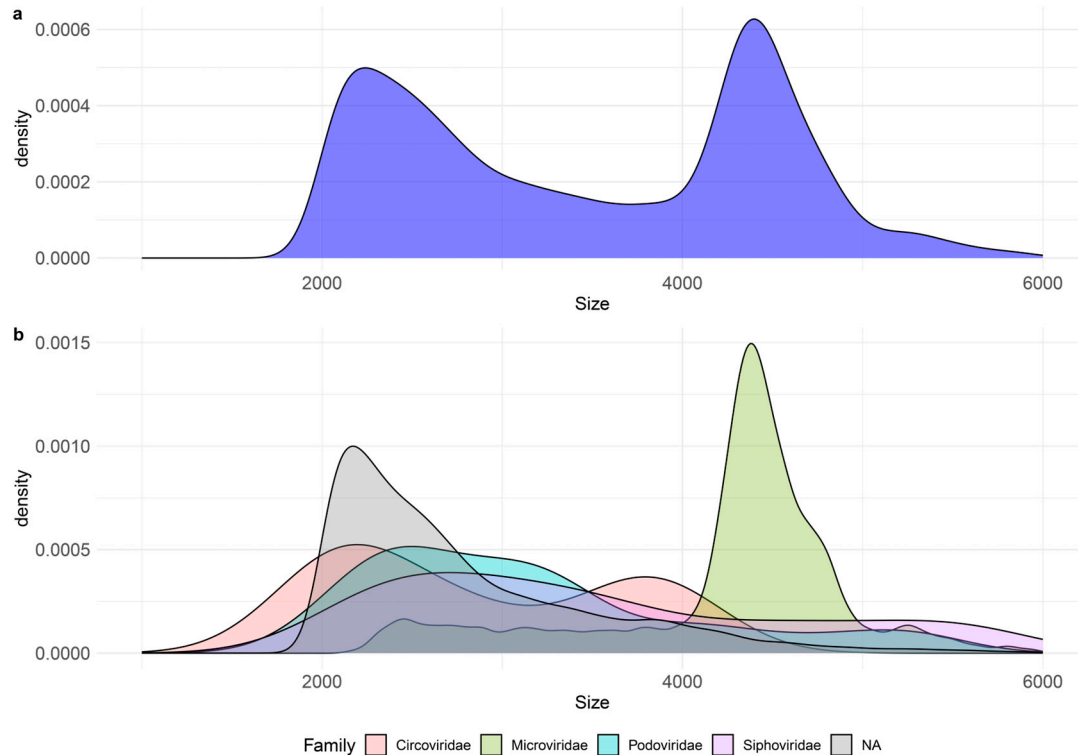
**Figure 2.** Taxonomic classification of the viromes in the two investigated benthic abyssal sites at the family level. The abundance of the vOTUs corresponding to the most represented families is reported in a log<sub>2</sub> scale for the five reconstructed viromes. NA: non-amplified; A: amplified; NP: Nanopore.

The heat map constructed on the most abundant vOTUs (>1000 CPMs) clustered the samples by DNA preparation method, nevertheless revealing hits with opposite trends among samples (Figure 3a). Notably, none of the relevant viral families (defined as the ones with average CPMs > 1000) resulted in being enriched by the amplification step, with the most represented family (*Microviridae*) maintaining the same relative abundance between the non-amplified and amplified samples. Conversely, the unclassified vOTUs increased their abundance in the amplified samples by 24% (Figure 3b,c).



**Figure 3.** Taxonomic distribution of vOTUs found in the five samples. (a) Heat map based on the 91 vOTUs with CPMs > 1000 and depicted by their relative abundances in the 4 samples. Dark-red boxes referred to the most abundant vOTUs; (b) enrichment ratio, computed by viral family, between non-amplified (NA) and amplified (A) samples; (c) total abundances of the identified viral families are reported as CPM in a log scale. The color codes of the families are maintained between panel (b,c).

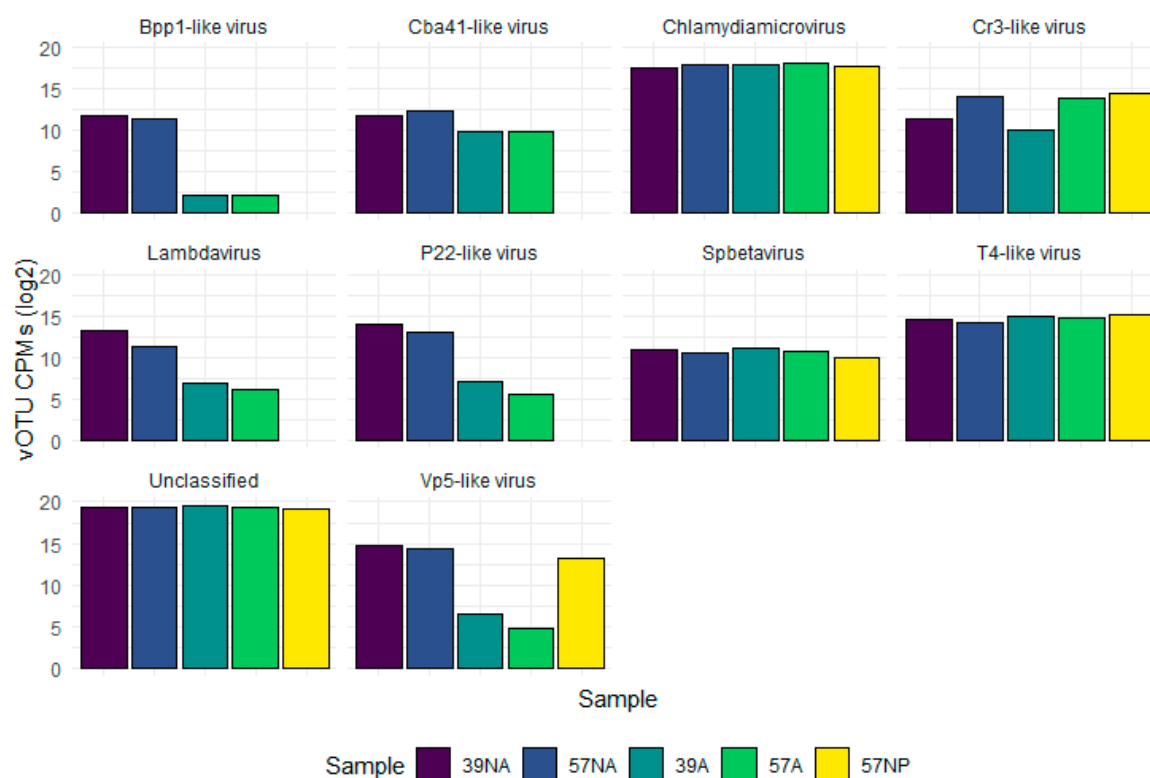
The size distribution of the whole virome was characterized by two main peaks, corresponding to 2.2 and 4.4 kb (Figure 4a), which can be associated to uncharacterized vOTUs and *Microviridae*, respectively (Figure 4b).



**Figure 4.** Size distribution of the benthic abyssal viromes reconstructed with short reads sequencing. (a) Total size distribution and (b) size distribution of the vOTUs belonging to the five most represented families.

Considering the genus-level taxonomic classification, 65–73% of the vOTUs remained unclassified, followed by *Clamydiamicrovirus* (18–27%) and T4-like phages (2–3%), with similar relative abundances in amplified and non-amplified samples (Figure 5 and Table S2). Differently, Vp5-like, P22-like, Lambdavirus, and Bpp1-like vOTUs showed higher relative abundances in the non-amplified samples.

A total of 91 vOTUs, referring mostly to non-classified vOTUs ( $N = 71$ ) and *Microviridae* ( $N = 14$ ), displayed an average abundance > 1000 CPM in the 4 short-read datasets. Notably, 14 of these vOTUs were classified at genus level as putative *Clamydiamicrovirus*. Following a different taxonomic classification approach based on CAT/BAT, we noticed some annotation inconsistencies (Table S2) as we identified 6 vOTUs similar to McMurdo Ice Shelf pond-associated circular DNA virus 8 and 16 vOTUs similar to *Circovirus* spp. Although not listed among the most abundant hits, we identified vOTUs likely referring to the family *Fuselloviridae*, as for instance vOTU\_578 (length 27.8 kb, 66 predicted ORFs) which was erroneously classified under the “dsDNA/*Myoviridae*” category by the classification tool, although two ORFs showed similarity (<40% of aminoacidic similarity) to a *Nitrosopumilus* spindle-shaped virus. Finally, the functional annotation of the 151,800 proteins encoded by the 24,828 high-confidence vOTUs revealed a majority of “uncharacterized proteins” (96%), followed by tail ( $N = 2756$ ), head and packaging ( $N = 2237$ ) and DNA, RNA and nucleotide metabolism ( $N = 2572$ , Figure S3a). A total of 135 proteins were identified as possible auxiliary metabolic genes (AMGs) by VIBRANT (Figure S3b).



**Figure 5.** Taxonomic classification at the genus level of the viromes in the two investigated abyssal sites. The abundance of the vOTUs corresponding to the 10 most represented genera is reported for the 5 reconstructed viromes in a log<sub>2</sub> scale. NA: non-amplified; A: amplified; NP: Nanopore.

### 3.3. Output of the Long Read Sequencing of the Amplified Virome

To further investigate the effect of the amplification process and evaluate the possibility of adopting single molecule sequencing to quickly profile a virome sample, we subjected the amplified DNA of site 57 to Nanopore sequencing. The resulting 301,522 reads (N<sub>50</sub> = 9.6 kb, median length = 6.2 kb) were reduced to 36,152 error-corrected, non-chimeric reads and assembled into 1009 contigs (N<sub>50</sub> = 9.9 kb). Applying the same pipeline as for short reads, a total of 555 vOTUs were identified, with the uncharacterized hits being the most abundant (46%), followed by vOTUs belonging to *Microviridae* (25%) and *Myoviridae* (21%) families. The size distribution of these vOTUs showed two-peaks, with the first peak (4.6 kb) reasonably associated to *Microviridae*, whereas the second peak was associated with both *Microviridae* and *Myoviridae* (Figure S4). In terms of taxonomic composition, the comparison of the long-read virome with the paired datasets obtained from short reads revealed a preserved ratio of ss/dsDNA viruses with the non-amplified dataset (Figure 1) and a similar distribution of families (Figure 2).

## 4. Discussion

Information on viral diversity in benthic abyssal ecosystems (>4000 m depth) and the contribution of ds- and ssDNA viruses is still extremely limited. In the present study, we analysed the viromes of two benthic abyssal sites of the North-eastern Atlantic Ocean using different sample preparation and sequencing approaches and evaluated their performance to investigate viral diversity, including the contribution of ssDNA and dsDNA viruses.

Starting from 10 ng of amplified DNA or from the unamplified DNA in the picogram-range, we successfully prepared sequencing libraries using the Accel-NGS 1S Plus kit. This kit has been previously tested with DNA quantities diluted up to picograms as well as from 100 ng of virome DNA [48], allowing the capture of both ss- and dsDNA molecules [49,50]. The higher number of PCR cycles required for the picogram DNA quantities resulted in a

higher number of identical reads in the unamplified samples, compared to the amplified ones, which need to be removed to avoid possibly assembly and mapping biases [51].

By integrating different bioinformatic tools, we retrieved a total of 24,828 vOTUs, with the amplified samples considerably enriched in viral sequences (ca. 40–50% of total sequences in amplified samples vs. <6% in the non-amplified samples).

Overall, the taxonomic classification of vOTUs based on viral-specific profiles allowed the identification of 22 viral families, with *Microviridae*, *Podoviridae*, *Myoviridae*, and *Siphoviridae* being the most represented ones, as reported in other deep-sea sediment viromes [2,9,11,13]. The bimodal size distribution of our viromes likely depends upon the great abundance of *Microviridae* that we recovered, plus taxonomically uncharacterized small viruses contributing to defining the smaller size peak. Irrespective of the procedure used to prepare the viromes, most of vOTUs had no matches in the database, reinforcing previous findings that benthic deep-sea ecosystems represent a hotspot of novel viruses [2,8,52]. Despite the lower number of viral reads in the non-amplified samples, viral diversity in terms of represented vOTUs and detected viral families was higher than in the amplified samples. The higher percentage of dsDNA-vOTUs in the unamplified samples (11–16%) compared to the amplified samples (ca. 1–2%) represents a further consistent difference. It is known that DNA amplification step by Phi29 polymerase can determine a preferential amplification of circular ssDNA to dsDNA [53], and thus, the relative abundance of ssDNA viral sequences can be overrepresented [2,8]. On the other side, studies indicate that dsDNA viruses can be an important component of the benthic deep-sea viral diversity. For example, a metagenomic study performed on deep-sea sediments of the Southwest Indian Ocean without prior separation of viral particles highlighted almost exclusively the presence of dsDNA viruses mostly affiliated with the families *Myoviridae*, *Podoviridae* and *Siphoviridae* [11]. Other findings obtained from the analyses of viromes generated with different procedures from other benthic deep-sea ecosystems revealed the presence of a relevant fraction of ssDNA viruses and dsDNA viruses [2,8,9]. This discrepancy could be due to the natural variability of the investigated deep-sea ecosystems, but also to the applied methodologies, which could profoundly influence the assessment of viral diversity. In our study, the virome was mainly composed of ssDNA viruses irrespective of the amplification step. Since the adopted methodology for library preparation allows for the recovery of ssDNA molecules, this might have contributed to a “more genuine” assessment of the different components. Intriguingly, we also obtained a similar ratio of ds/ssDNA viruses by applying long-read sequencing to one amplified DNA sample. This is possibly linked to limitations in the assembly process of amplified datasets, which failed to correctly resolve complex contigs, suggesting that long-read sequencing can be used to profile the virome composition before massive sequencing. Advancing the technology towards low-input DNA samples will likely make long-read sequencing pivotal for metagenomic/viromic analyses [21].

In our study, the genus-level taxonomic annotation by vpf-tools reveals that the majority of vOTUs in all samples were unclassified, mirroring the result obtained at the family level. Notably, the genus *Chlamydiamicrovirus* belonging to the *Microviridae* family was the most abundant both in amplified and non-amplified samples. Its presence is not surprising since its natural host (i.e., *Chlamydiae*) has already been reported in different environments, including deep-sea sediments from the Arctic Mid-Ocean Ridge [54]. Other identified viral genera showed differential abundances, typically with higher values in the non-amplified samples, as reported at the family-level classification. Although genus-level classification has been obtained using previously utilised confidence scores [43], these findings should be viewed with caution since the limited length of the viral contigs may affect the correct taxonomic classification [55].

Among viruses known to infect prokaryotic hosts, we found an almost complete genome of a putative archaeal virus belonging to the family *Fuselloviridae*, exclusively present in unamplified samples. Virus-induced lysis of archaea has been reported to be a key process in deep-sea sediments, as it contributes to the decrease of up to one-third of

the total prokaryotic biomass [56]. Since only a few archaeal viruses have been studied in detail so far [57], shedding light on the viruses that contribute to archaea dynamics and related ecosystem processes is critical to advancing science.

In the two investigated abyssal sites, a number of different contigs were identified as circoviruses. Although not abundant, their presence, which has been already reported in different benthic deep-sea habitats [58,59], poses questions regarding their ecological relevance as viruses infecting metazoans [60]. Among other possible eukaryotic viruses, we identified 18 vOTUs belonging to the *Phycodnaviridae* family, which includes NCLDV infecting photosynthetic eukaryotes (proposed order, “Megavirales”). These viruses have been previously reported to occur in high abundance in deep-sea ecosystems as result of their transport associated with sinking particles from the surface waters [61]. A few vOTUs, likely representing fragments of viruses belonging to *Iridoviridae*, *Herpesviridae*, and *Marseilleviridae* families were also identified.

Mirroring the presence of abundant unclassified vOTUs, also the large majority of proteins encoded by the vOTUs were uncharacterized. Only a minimal part of these proteins putatively harboured auxiliary metabolic functions, mainly related to carbohydrate and aminoacid metabolism. Nowadays, a “functional” role of viromes has been proposed [62,63]. Although the most represented AMGs that we found recalled the ones reported in a deep sea virome of the Mariana trench [64], the limited extension of both our dataset and of the retrieved AMGs prevents us from drawing general conclusions regarding their functional role.

## 5. Conclusions

Overall, our findings indicate that the analysis of viromes obtained from abyssal sediments by using direct sequencing (without amplification step) can provide a broad overview of viral richness and the composition of the viral assemblages in terms of ssDNA- and dsDNA-vOTUs. At the same time, the contextual use of a random amplification step of the same samples and Nanopore long-read technology can improve the assessment of viral assemblages and, thus, it represents a suitable approach for expanding our knowledge on viral diversity in the largest ecosystem on Earth and towards a better understanding of their interactions with potential hosts, both prokaryotes and eukaryotes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15122282/s1>, Table S1. List of the identified vOTUs. vOTU ID, contig size (bp) abundances in the four datasets as CPM and taxonomic classification (Baltimore, Family and Genus) are reported. Table S2. Most represented vOTUs with their taxonomic annotation obtained using vpf-tools and CAT/BAT. Figure S1. Number of contigs identified as viral by the five bioinformatic tools in the four datasets and their overlapping. Figure S2. Rarefaction curves of the four viromes reconstructed with Illumina short reads. A total of 6,895,930 reads per sample were sampled. Figure S3. (a) Relative abundances of the putative functions identified in the ORFs of the reconstructed viromes. (b) Counts of the different categories of AMGs, as classified by VIBRANT. Figure S4. Size distribution of the benthic virome of Site 57 reconstructed with long reads sequencing. (a) Total size distribution and (b) size distribution of the vOTUs belonging to the two most represented families.

**Author Contributions:** U.R., C.C. and A.D. conceived the study. G.L. prepared DNA samples. U.R. prepared the sequencing libraries. P.V. assured computational resources, equipment, and materials for library preparation and Nanopore sequencing, and supported the development of this study. U.R., M.S., A.P., S.D.M. and D.L. supervised sequencing activity. U.R., M.S. and A.P. carried out bioinformatics and data elaboration. All authors contributed to data interpretation. U.R. and A.D. drafted the first version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was financially supported by the Italian PRIN 2017 project VIRIDE (2017EKFA98).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data generated for this study are available in the NCBI SRA database under the project accession ID PRJNA909474.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Brum, J.R.; Ignacio-Espinoza, J.C.; Roux, S.; Doucier, G.; Acinas, S.G.; Alberti, A.; Chaffron, S.; Cruaud, C.; de Vargas, C.; Gasol, J.M.; et al. Ocean Plankton. Patterns and Ecological Drivers of Ocean Viral Communities. *Science* **2015**, *348*, 1261498. [[CrossRef](#)]
2. Yoshida, M.; Mochizuki, T.; Urayama, S.-I.; Yoshida-Takashima, Y.; Nishi, S.; Hirai, M.; Nomaki, H.; Takaki, Y.; Nunoura, T.; Takai, K. Quantitative Viral Community DNA Analysis Reveals the Dominance of Single-Stranded DNA Viruses in Offshore Upper Bathyal Sediment from Tohoku, Japan. *Front. Microbiol.* **2018**, *9*, 75. [[CrossRef](#)]
3. Gregory, A.C.; Zayed, A.A.; Conceição-Neto, N.; Temperton, B.; Bolduc, B.; Alberti, A.; Ardyna, M.; Arkhipova, K.; Carmichael, M.; Cruaud, C.; et al. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **2019**, *177*, 1109–1123.e14. [[CrossRef](#)]
4. Sunagawa, S.; Acinas, S.G.; Bork, P.; Bowler, C.; Eveillard, D.; Gorsky, G.; Guidi, L.; Iudicone, D.; Karsenti, E.; Lombard, F.; et al. Tara Oceans: Towards Global Ocean Ecosystems Biology. *Nat. Rev. Microbiol.* **2020**, *18*, 428–445. [[CrossRef](#)]
5. Guo, J.; Bolduc, B.; Zayed, A.A.; Varsani, A.; Dominguez-Huerta, G.; Delmont, T.O.; Pratama, A.A.; Gazitúa, M.C.; Vik, D.; Sullivan, M.B.; et al. VirSorter2: A Multi-Classifer, Expert-Guided Approach to Detect Diverse DNA and RNA Viruses. *Microbiome* **2021**, *9*, 37. [[CrossRef](#)]
6. Zhang, Q.-Y.; Ke, F.; Gui, L.; Zhao, Z. Recent Insights into Aquatic Viruses: Emerging and Reemerging Pathogens, Molecular Features, Biological Effects, and Novel Investigative Approaches. *Water Biol. Secur.* **2022**, *1*, 100062. [[CrossRef](#)]
7. Nguyen, T.T.; Robertsen, E.M.; Landfald, B. Viral Assemblage Variation in an Arctic Shelf Seafloor. *Aquat. Microb. Ecol.* **2017**, *78*, 135–145. [[CrossRef](#)]
8. Corinaldesi, C.; Tangherlini, M.; Dell'Anno, A. From Virus Isolation to Metagenome Generation for Investigating Viral Diversity in Deep-Sea Sediments. *Sci. Rep.* **2017**, *7*, 8355. [[CrossRef](#)]
9. Yoshida, M.; Takaki, Y.; Eitoku, M.; Nunoura, T.; Takai, K. Metagenomic Analysis of Viral Communities in (Hado)Pelagic Sediments. *PLoS ONE* **2013**, *8*, e57271. [[CrossRef](#)]
10. Li, Z.; Pan, D.; Wei, G.; Pi, W.; Zhang, C.; Wang, J.-H.; Peng, Y.; Zhang, L.; Wang, Y.; Hubert, C.R.J.; et al. Deep Sea Sediments Associated with Cold Seeps Are a Subsurface Reservoir of Viral Diversity. *ISME J.* **2021**, *15*, 2366–2378. [[CrossRef](#)]
11. Zheng, X.; Liu, W.; Dai, X.; Zhu, Y.; Wang, J.; Zhu, Y.; Zheng, H.; Huang, Y.; Dong, Z.; Du, W.; et al. Extraordinary Diversity of Viruses in Deep-Sea Sediments as Revealed by Metagenomics without Prior Virion Separation. *Environ. Microbiol.* **2021**, *23*, 728–743. [[CrossRef](#)]
12. Zhao, J.; Jing, H.; Wang, Z.; Wang, L.; Jian, H.; Zhang, R.; Xiao, X.; Chen, F.; Jiao, N.; Zhang, Y. Novel Viral Communities Potentially Assisting in Carbon, Nitrogen, and Sulfur Metabolism in the Upper Slope Sediments of Mariana Trench. *mSystems* **2022**, *7*, e0135821. [[CrossRef](#)]
13. Bäckström, D.; Yutin, N.; Jørgensen, S.L.; Dharamshi, J.; Homa, F.; Zaremba-Niedwiedzka, K.; Spang, A.; Wolf, Y.I.; Koonin, E.V.; Ettema, T.J.G. Virus Genomes from Deep Sea Sediments Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism. *mBio* **2019**, *10*, e02497-18. [[CrossRef](#)]
14. Chen, P.; Zhou, H.; Huang, Y.; Xie, Z.; Zhang, M.; Wei, Y.; Li, J.; Ma, Y.; Luo, M.; Ding, W.; et al. Revealing the Full Biosphere Structure and Versatile Metabolic Functions in the Deepest Ocean Sediment of the Challenger Deep. *Genome Biol.* **2021**, *22*, 207. [[CrossRef](#)]
15. Krupovic, M.; Dolja, V.V.; Koonin, E.V. Origin of Viruses: Primordial Replicators Recruiting Capsids from Hosts. *Nat. Rev. Microbiol.* **2019**, *17*, 449–458. [[CrossRef](#)]
16. Koonin, E.V.; Dolja, V.V.; Krupovic, M.; Varsani, A.; Wolf, Y.I.; Yutin, N.; Zerbini, F.M.; Kuhn, J.H. Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev.* **2020**, *84*, e00061-19. [[CrossRef](#)]
17. Pan, D.; Morono, Y.; Inagaki, F.; Takai, K. An Improved Method for Extracting Viruses From Sediment: Detection of Far More Viruses in the Subseafloor Than Previously Reported. *Front. Microbiol.* **2019**, *10*, 878. [[CrossRef](#)]
18. Danovaro, R.; Middelboe, M. Separation of free virus particles from sediments in aquatic systems. In *Manual of Aquatic Viral Ecology*; Wilhelm, S.W., Weinbauer, M.G., Suttle, C.A., Eds.; ASLO: Washington, DC, USA, 2010; pp. 74–81.
19. Malki, K.; Rosario, K.; Sawaya, N.A.; Székely, A.J.; Tisza, M.J.; Breitbart, M. Prokaryotic and Viral Community Composition of Freshwater Springs in Florida, USA. *mBio* **2020**, *11*, e00436-20. [[CrossRef](#)]
20. Hirai, M.; Nishi, S.; Tsuda, M.; Sunamura, M.; Takaki, Y.; Nunoura, T. Library Construction from Subnanogram DNA for Pelagic Sea Water and Deep-Sea Sediments. *Microbes Environ.* **2017**, *32*, 336–343. [[CrossRef](#)]
21. Zablocki, O.; Michelsen, M.; Burris, M.; Solonenko, N.; Warwick-Dugdale, J.; Ghosh, R.; Pett-Ridge, J.; Sullivan, M.B.; Temperton, B. VirION2: A Short- and Long-Read Sequencing and Informatics Workflow to Study the Genomic Diversity of Viruses in Nature. *PeerJ* **2021**, *9*, e11088. [[CrossRef](#)]
22. Willenbücher, K.; Wibberg, D.; Huang, L.; Conrady, M.; Ramm, P.; Gätcke, J.; Busche, T.; Brandt, C.; Szewzyk, U.; Schlüter, A.; et al. Phage Genome Diversity in a Biogas-Producing Microbiome Analyzed by Illumina and Nanopore GridION Sequencing. *Microorganisms* **2022**, *10*, 368. [[CrossRef](#)]

23. Beaulaurier, J.; Luo, E.; Eppley, J.M.; Uyl, P.D.; Dai, X.; Burger, A.; Turner, D.J.; Pendelton, M.; Juul, S.; Harrington, E.; et al. Assembly-Free Single-Molecule Sequencing Recovers Complete Virus Genomes from Natural Microbial Communities. *Genome Res.* **2020**, *30*, 437–446. [[CrossRef](#)]
24. Turnewitsch, R.; Lahajnar, N.; Haeckel, M.; Christiansen, B. An Abyssal Hill Fractionates Organic and Inorganic Matter in Deep-Sea Surface Sediments. *Geophys. Res. Lett.* **2015**, *42*, 7663–7672. [[CrossRef](#)]
25. Durden, J.M.; Bett, B.J.; Jones, D.O.B.; Huvenne, V.A.I.; Ruhl, H.A. Abyssal Hills—Hidden Source of Increased Habitat Heterogeneity, Benthic Megafaunal Biomass and Diversity in the Deep Sea. *Prog. Oceanogr.* **2015**, *137*, 209–218. [[CrossRef](#)]
26. Durden, J.M.; Ruhl, H.A.; Pebody, C.; Blackbird, S.J.; van Oevelen, D. Differences in the Carbon Flows in the Benthic Food Webs of Abyssal Hill and Plain Habitats. *Limnol. Oceanogr.* **2017**, *62*, 1771–1782. [[CrossRef](#)]
27. Durden, J.M.; Bett, B.J.; Ruhl, H.A. Subtle Variation in Abyssal Terrain Induces Significant Change in Benthic Megafaunal Abundance, Diversity, and Community Structure. *Prog. Oceanogr.* **2020**, *186*, 102395. [[CrossRef](#)]
28. Bushnell, B.; Rood, J.; Singer, E. BBMerge—Accurate Paired Shotgun Read Merging via Overlap. *PLoS ONE* **2017**, *12*, e0185056. [[CrossRef](#)]
29. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* **2015**, *31*, 1674–1676. [[CrossRef](#)]
30. De Coster, W.; D’Hert, S.; Schultz, D.T.; Cruets, M.; Van Broeckhoven, C. NanoPack: Visualizing and Processing Long-Read Sequencing Data. *Bioinformatics* **2018**, *34*, 2666–2669. [[CrossRef](#)]
31. Marijon, P.; Chikhi, R.; Varré, J.-S. Yacrd and Fpa: Upstream Tools for Long-Read Genome Assembly. *Bioinformatics* **2020**, *36*, 3894–3896. [[CrossRef](#)]
32. Kolmogorov, M.; Bickhart, D.M.; Behsaz, B.; Gurevich, A.; Rayko, M.; Shin, S.B.; Kuhn, K.; Yuan, J.; Polevikov, E.; Smith, T.P.L.; et al. metaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs. *Nat. Methods* **2020**, *17*, 1103–1110. [[CrossRef](#)]
33. Pandolfo, M.; Telatin, A.; Lazzari, G.; Adriaenssens, E.M.; Vitulo, N. MetaPhage: An Automated Pipeline for Analyzing, Annotating, and Classifying Bacteriophages in Metagenomics Sequencing Data. *mSystems* **2022**, *7*, e0074122. [[CrossRef](#)]
34. Roux, S.; Enault, F.; Hurwitz, B.L.; Sullivan, M.B. VirSorter: Mining Viral Signal from Microbial Genomic Data. *PeerJ* **2015**, *3*, e985. [[CrossRef](#)]
35. Ren, J.; Song, K.; Deng, C.; Ahlgren, N.A.; Fuhrman, J.A.; Li, Y.; Xie, X.; Poplin, R.; Sun, F. Identifying Viruses from Metagenomic Data Using Deep Learning. *Quant. Biol. Beijing China* **2020**, *8*, 64–77. [[CrossRef](#)]
36. Starikova, E.V.; Tikhonova, P.O.; Prianichnikov, N.A.; Rands, C.M.; Zdobnov, E.M.; Ilna, E.N.; Govorun, V.M. Phigaro: High-Throughput Prophage Sequence Annotation. *Bioinformatics* **2020**, *36*, 3882–3884. [[CrossRef](#)]
37. Kieft, K.; Zhou, Z.; Anantharaman, K. VIBRANT: Automated Recovery, Annotation and Curation of Microbial Viruses, and Evaluation of Viral Community Function from Genomic Sequences. *Microbiome* **2020**, *8*, 90. [[CrossRef](#)]
38. Ren, J.; Ahlgren, N.A.; Lu, Y.Y.; Fuhrman, J.A.; Sun, F. VirFinder: A Novel k-Mer Based Tool for Identifying Viral Sequences from Assembled Metagenomic Data. *Microbiome* **2017**, *5*, 69. [[CrossRef](#)]
39. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
40. Nayfach, S.; Camargo, A.P.; Schulz, F.; Eloë-Fadrosh, E.; Roux, S.; Kyrpides, N.C. CheckV Assesses the Quality and Completeness of Metagenome-Assembled Viral Genomes. *Nat. Biotechnol.* **2021**, *39*, 578–585. [[CrossRef](#)]
41. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* **2009**, *10*, R25. [[CrossRef](#)]
42. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
43. Pons, J.C.; Paez-Espino, D.; Riera, G.; Ivanova, N.; Kyrpides, N.C.; Llabrés, M. VPF-Class: Taxonomic Assignment and Host Prediction of Uncultivated Viruses Based on Viral Protein Families. *Bioinformatics* **2021**, *37*, 1805–1813. [[CrossRef](#)]
44. von Meijenfheldt, F.A.B.; Arkhipova, K.; Cambuy, D.D.; Coutinho, F.H.; Dutilh, B.E. Robust Taxonomic Classification of Uncharted Microbial Sequences and Bins with CAT and BAT. *Genome Biol.* **2019**, *20*, 217. [[CrossRef](#)]
45. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinform.* **2010**, *11*, 119. [[CrossRef](#)]
46. Buchfink, B.; Xie, C.; Huson, D.H. Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)]
47. Bouras, G.; Nepal, R.; Houtak, G.; Psaltis, A.J.; Wormald, P.-J.; Vreugde, S. Pharokka: A Fast Scalable Bacteriophage Annotation Tool. *Bioinformatics* **2022**, *39*, btac776. [[CrossRef](#)]
48. Kurihara, L.; Banks, L.; Chupreta, S.; Couture, C.; Kelchner, V.; Laliberte, J.; Sandhu, S.; Spurbeck, R.; Makarov, V. Abstract 3564: A New Method for Preparation of Low-Input, PCR-Free next Generation Sequencing Libraries. *Cancer Res.* **2014**, *74*, 3564. [[CrossRef](#)]
49. Roux, S.; Solonenko, N.E.; Dang, V.T.; Poulos, B.T.; Schwenck, S.M.; Goldsmith, D.B.; Coleman, M.L.; Breitbart, M.; Sullivan, M.B. Towards Quantitative Viromics for Both Double-Stranded and Single-Stranded DNA Viruses. *PeerJ* **2016**, *4*, e2777. [[CrossRef](#)]
50. Aigrain, L.; Gu, Y.; Quail, M.A. Quantitation of next Generation Sequencing Library Preparation Protocol Efficiencies Using Droplet Digital PCR Assays—A Systematic Comparison of DNA Library Preparation Kits for Illumina Sequencing. *BMC Genom.* **2016**, *17*, 458. [[CrossRef](#)]

51. Zhou, W.; Chen, T.; Zhao, H.; Eterovic, A.K.; Meric-Bernstam, F.; Mills, G.B.; Chen, K. Bias from Removing Read Duplication in Ultra-Deep Sequencing Experiments. *Bioinformatics* **2014**, *30*, 1073–1080. [[CrossRef](#)]
52. Cai, L.; Weinbauer, M.G.; Xie, L.; Zhang, R. The Smallest in the Deepest: The Enigmatic Role of Viruses in the Deep Biosphere. *Natl. Sci. Rev.* **2023**, *10*, nwad009. [[CrossRef](#)]
53. Edwards, R.A.; Rohwer, F. Viral Metagenomics. *Nat. Rev. Microbiol.* **2005**, *3*, 504–510. [[CrossRef](#)]
54. Dharamshi, J.E.; Tamarit, D.; Eme, L.; Stairs, C.W.; Martijn, J.; Homa, F.; Jørgensen, S.L.; Spang, A.; Ettema, T.J.G. Marine Sediments Illuminate Chlamydiae Diversity and Evolution. *Curr. Biol.* **2020**, *30*, 1032–1048.e7. [[CrossRef](#)]
55. Pratama, A.A.; Bolduc, B.; Zayed, A.A.; Zhong, Z.-P.; Guo, J.; Vik, D.R.; Gazitúa, M.C.; Wainaina, J.M.; Roux, S.; Sullivan, M.B. Expanding Standards in Viromics: In Silico Evaluation of dsDNA Viral Genome Identification, Classification, and Auxiliary Metabolic Gene Curation. *PeerJ* **2021**, *9*, e11447. [[CrossRef](#)]
56. Danovaro, R.; Dell’Anno, A.; Corinaldesi, C.; Rastelli, E.; Cavicchioli, R.; Krupovic, M.; Noble, R.T.; Nunoura, T.; Prangishvili, D. Virus-Mediated Archaeal Hecatomb in the Deep Seafloor. *Sci. Adv.* **2016**, *2*, e1600492. [[CrossRef](#)]
57. Snyder, J.C.; Bolduc, B.; Young, M.J. 40 Years of Archaeal Virology: Expanding Viral Diversity. *Virology* **2015**, *479–480*, 369–378. [[CrossRef](#)]
58. Castelán-Sánchez, H.G.; López-Rosas, I.; García-Suastegui, W.A.; Peralta, R.; Dobson, A.D.W.; Batista-García, R.A.; Dávila-Ramos, S. Extremophile Deep-Sea Viral Communities from Hydrothermal Vents: Structural and Functional Analysis. *Mar. Genom.* **2019**, *46*, 16–28. [[CrossRef](#)]
59. Dell’Anno, A.; Corinaldesi, C.; Danovaro, R. Virus Decomposition Provides an Important Contribution to Benthic Deep-Sea Ecosystem Functioning. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2014–E2019. [[CrossRef](#)]
60. Krupovic, M.; Varsani, A.; Kazlauskas, D.; Breitbart, M.; Delwart, E.; Rosario, K.; Yutin, N.; Wolf, Y.I.; Harrach, B.; Zerbini, F.M.; et al. Cressnaviricota: A Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *J. Virol.* **2020**, *94*, e00582-20. [[CrossRef](#)]
61. Gong, Z.; Liang, Y.; Wang, M.; Jiang, Y.; Yang, Q.; Xia, J.; Zhou, X.; You, S.; Gao, C.; Wang, J.; et al. Viral Diversity and Its Relationship With Environmental Factors at the Surface and Deep Sea of Prydz Bay, Antarctica. *Front. Microbiol.* **2018**, *9*, 2981. [[CrossRef](#)]
62. Gazitúa, M.C.; Vik, D.R.; Roux, S.; Gregory, A.C.; Bolduc, B.; Widner, B.; Mulholland, M.R.; Hallam, S.J.; Ulloa, O.; Sullivan, M.B. Potential Virus-Mediated Nitrogen Cycling in Oxygen-Depleted Oceanic Waters. *ISME J.* **2021**, *15*, 981–998. [[CrossRef](#)]
63. Chevallereau, A.; Pons, B.J.; van Houte, S.; Westra, E.R. Interactions between Bacterial and Phage Communities in Natural Environments. *Nat. Rev. Microbiol.* **2022**, *20*, 49–62. [[CrossRef](#)]
64. Gao, C.; Liang, Y.; Jiang, Y.; Paez-Espino, D.; Han, M.; Gu, C.; Wang, M.; Yang, Y.; Liu, F.; Yang, Q.; et al. Virioplankton Assemblages from Challenger Deep, the Deepest Place in the Oceans. *iScience* **2022**, *25*, 104680. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.