life

MDPI

Article

An Interplay between Viruses and Bacteria Associated with the White Sea Sponges Revealed by Metagenomics

Anastasiia Rusanova [1], Victor Fedorchuk [2], Stepan Toshchakov [3], Svetlana Dubiley [1,4] and Dmitry Sutormin [1,4,*]

[1] Institute of Gene Biology, Russian Academy of Sciences, 119334 Moscow, Russia; rusanova@genebiology.ru (A.R.); S.Dubiley@skoltech.ru (S.D.)
[2] The Faculty of Geology, Lomonosov Moscow State University, 119234 Moscow, Russia; V.Fedorchuk@skoltech.ru
[3] Kurchatov Center for Genome Research, National Research Center "Kurchatov Institute", 123182 Moscow, Russia; Toschakov_SV@nrcki.ru
[4] Skolkovo Institute of Science and Technology, 121205 Moscow, Russia
* Correspondence: Dmitry.Sutormin@skoltech.ru

Abstract: Sponges are remarkable holobionts harboring extremely diverse microbial and viral communities. However, the interactions between the components within holobionts and between a holobiont and environment are largely unknown, especially for polar organisms. To investigate possible interactions within and between sponge-associated communities, we probed the microbiomes and viromes of cold-water sympatric sponges *Isodictya palmata* (n = 2), *Halichondria panicea* (n = 3), and *Halichondria sitiens* (n = 3) by 16S and shotgun metagenomics. We showed that the bacterial and viral communities associated with these White Sea sponges are species-specific and different from the surrounding water. Extensive mining of bacterial antiphage defense systems in the metagenomes revealed a variety of defense mechanisms. The abundance of defense systems was comparable in the metagenomes of the sponges and the surrounding water, thus distinguishing the White Sea sponges from those inhabiting the tropical seas. We developed a network-based approach for the combined analysis of CRISPR-spacers and protospacers. Using this approach, we showed that the virus–host interactions within the sponge-associated community are typically more abundant (three out of four interactions studied) than the inter-community interactions. Additionally, we detected the occurrence of viral exchanges between the communities. Our work provides the first insight into the metagenomics of the three cold-water sponge species from the White Sea and paves the way for a comprehensive analysis of the interactions between microbial communities and associated viruses.

Keywords: sponge-associated communities; 16S metagenomics; shot-gun metagenomics; metavirome; phage-host interaction; bacterial antiphage defense system; CRISPR-Cas

check for updates

Citation: Rusanova, A.; Fedorchuk, V.; Toshchakov, S.; Dubiley, S.; Sutormin, D. An Interplay between Viruses and Bacteria Associated with the White Sea Sponges Revealed by Metagenomics. *Life* **2022**, *12*, 25. https://doi.org/10.3390/life12010025

Academic Editor: Francois Lefort

Received: 12 November 2021
Accepted: 21 December 2021
Published: 24 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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

The majority of multicellular organisms, such as plants and animals, are considered a combination of the host and the associated communities of microbes and viruses rather than individual units. Together, they make up "holobionts" [1] or "metaorganisms" [2]. Members of the microbial community contribute to the nutrition, protection, immunity, and development of the host they inhabit [3].

One of the most ancient and remarkable examples of holobionts are sponges (Porifera) that can host diverse communities of bacteria, archaea, microalgae, unicellular fungi, as well as viruses and virus-like particles. Associated organisms can occupy up to 40% of the sponge volume [4], inhabit their mesohyl matrix [3], or localize intracellularly [5]. The composition of sponge-associated microbial communities differs from surrounding seawater communities, is host lineage-specific, and weakly depends on its geographical location [6–8]. The most frequently detected bacterial phyla in sponges are Proteobacteria

(mainly Alpha-, Delta-, and Gamma-), Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Nitrospirae, Thaumarchaeota [9–11], along with *Candidatus* phylum Poribacteria, which was first discovered in sponges [12].

The majority of sponges are marine organisms that live an attached lifestyle and typically feed on microscopic particles by filtering them from surrounding waters [13]. Since a sponge can pump thousands of liters a day, it can absorb a significant number of viruses and thus cause a potential threat for associated microbial communities [14,15]. This attracts attention to the metagenomics of sponges, especially for the organisms from tropical and temperate regions. At the same time, sponges from polar regions have remained understudied [10]. An expanding corpus of data indicates that genomes of sponge-associated bacteria are enriched with genes of diverse defense systems providing resistance against mobile genetic elements. Adaptive immune systems comprised of clustered regularly interspaced repeats (CRISPR) and CRISPR-associated proteins (Cas proteins), restriction-modification systems (RM), toxin-antitoxin systems (TA), DNA phosphorothioation system (DND), phage exclusion systems (Pgl), DISARM, and Zorya were found to be enriched in comparison to the metagenomes of the surrounding seawater [16–21].

A diversity of associated viruses of several sponges from temperate and tropical locations was investigated [15,22–26]. In all studies, the most commonly identified viruses were tailed dsDNA bacteriophages belonging to Podoviridae, Siphoviridae, and Myoviridae families of Caudovirales. Identification of auxiliary metabolic genes (AMGs) in sponge-associated viruses indicates that phages may complement the bacterial metabolic pathways [27]. Recently, a group of Ankyphages, encoding auxiliary ankyrin repeats (ANKs), was discovered. It was demonstrated in vitro, that ankyrin proteins from these phages reduce the phagocytosis of bacteria by macrophages, implicating that phages may mediate the symbiosis between bacteria and sponges [15]. Despite the several mechanisms discovered to date, the roles of viruses as a part of a sponge holobiont are not completely understood.

In this study, we aimed to characterize the interplay between microbial communities and viruses associated with three species of the White Sea sponges, namely *Isodictya palmata*, *Halichondria panicea*, and *Halichondria sitiens* (Demospongiae: Heteroscleromorpha.), and the surrounding seawater. Using the 16S metagenomics analysis, we show that studied species of the polar sponges are inhabited by highly diverse yet distinct microbial communities. Using shot-gun sequencing and assembly of metagenomes of the bacterial and viral fractions, we analyzed both the diversity of viral contigs found in the metagenomes and the arsenal of bacterial defense systems. Our study shows that, in contrast to previously studied tropical and temperate sponges, the abundance of defense systems appeared to be comparable for metagenomes of the White Sea sponges and surrounding water. Using the network-based approach for analysis of CRISPR-spacers found in CRISPR arrays and protospacers from viral genomes, we revealed possible phage bursts that originated in *I. palmata* and were further spread to *H. sitiens* and *H. panicea*. Our work provides the first insight into the metagenomics of the three cold-water sponges from the White Sea and paves the way for a comprehensive analysis of the interactions between microbial communities and associated viruses.

2. Materials and Methods

2.1. Sample Collection

Two (for *Isodictya palmata*) or three (for *Halichondria sitiens* and *Halichondria panicea*) marine sponge individuals were collected by SCUBA divers at 5–7 m water depth at the N. Pertsov White Sea Biological Station (WSBS, 66.5527° N, 33.1033° E) in the Kandalaksha Bay of the White Sea (Russia) from 10–20 August 2018. Samples were held separately in 5 L of sterile marine water (filtered through the 0.22 µm Sartorius filter) for 2 h at 5 °C. Identification of sponge species was performed by zoologist Dr. Boris Osadchenko and further confirmed by 18S rRNA gene region amplification and sequencing. 3 L of surrounding seawater at the sampling site were collected in a sterile container and were immediately processed in the WSBS laboratory.

2.2. Bacterial and Viral Fractions Isolation

For bacterial fraction isolation from sponges, a 1 cm^3 fragment of sponge tissue was diced and fragmented by razor and forceps in 25 mL of sterile marine water. The obtained suspension was centrifuged at 200× *g* for 5 min. The supernatant was collected and centrifuged further at 3500× *g* for 10 min. The obtained pellet was defined as bacterial fraction and used for DNA isolation. Depletion of sponge cells and spicules in the bacterial fraction was confirmed by microscopy. The supernatant collected after the second round of centrifugation was further filtered using the 0.2 μm filter (Sartorius, Göttingen, Germany). Collected flow-throughs were pooled for replicates of the same sponge species and concentrated with Pellicon XL50 tangential filtering cassette with 100 kDa cutoff membrane (Millipore, MA, USA), which is frequently used for concentration of viral particles from environmental samples [28]. The resultant concentrate (15–20 mL) was defined as a viral fraction.

To isolate bacterial fraction from marine water, 3 L of water were pre-filtered through the 5 μm filter (Sartorius). The flow-through was further filtered using the 0.2 μm filter (Sterivex, MA, USA). The fraction bound to the 0.2 μm membrane was defined as a bacterial fraction and was used for DNA isolation. The second flow-through was concentrated with the Pellicon XL50 cassette with 100 kDa cutoff (Millipore). The resultant concentrate (15–20 mL) was defined as viral fraction.

For collection of viral particles, PEG 8000 was added to 10% (*w*/*v*) and NaCl to 1 M to the viral fractions [29]. Samples were incubated at 4 °C for 1 h and centrifuged at 3500× *g* for 30 min. The supernatant was discarded, and the pellets were resuspended in 400 μL of STM buffer (100 mM NaCl, 10 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5). Samples were vortexed with 400 μL of chloroform and then centrifuged at 4 °C for 5 min at 200 g. The resultant aqueous phase contained concentrated viral particles and was used for DNA isolation.

2.3. DNA Extraction

DNA was extracted from bacterial fractions using the Diatom DNA Prep kit (Galart Diagnosticum, Moscow, Russia, catalog number 100 D1024) according to the manufacturer's manual. For the bacterial fraction of marine water, settled in the 0.2 μm Sterivex filter, the filtering unit was opened, and the membrane was fragmented using the sterile razor. DNA was purified from the fragmented filter with the Diatom DNA prep kit. Then, DNA was extracted twice with phenol–chloroform and chloroform and precipitated in ethanol for the additional purification.

To purify DNA from the viral fractions, the samples (400 μL) were incubated with 1 μL of RNase A (Thermo Fisher Scientific, Waltham, MA, USA, catalog number R1253) for 15 min at 37 °C. Then, 50 μL of a lysis buffer was added (10% SDS, 20 mM EDTA, 200 mM Tris-HCl pH 7.5) and the samples were incubated with Proteinase K (Thermo Fisher Scientific, catalog number EO0491), final concentration 100 μg/mL, for 1 h at 55 °C. DNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma–Aldrich, St. Louis, MO, USA, catalog number P2069), and traces of phenol were removed by triple extraction with equal volumes of chloroform:isoamyl alcohol (24:1). Finally, DNA was precipitated with ethanol. DNA precipitates were dissolved in 20 μL of TE buffer and stored at −20 °C. DNA concentration was assessed with Qubit 3 (Invitrogen) with Qubit dsDNA HS Assay Kit chemistry (Invitrogen, Waltham, MA, USA, catalog number Q32851) and its integrity was checked by agarose electrophoresis.

2.4. High-Throughput Sequencing

For bacterial fractions from sponges and marine water, a V3–V4 region of 16S rRNA gene was amplified using standard degenerate primers fused with sequencing adapters (see Illumina guide for 16S Metagenomic Sequencing Library Preparation, Part number 15,044,223 Rev. B). The amplicon libraries were prepared and sequenced at Kurchatov Institute Core Sequencing Center using the 250 + 250 bp paired-end protocol with Illumina MiSeq.

DNA of bacterial or viral fraction replicates was pooled and submitted for shotgun sequencing. Libraries were prepared at Skoltech Genomics Core Facility and were sequenced with 150 + 150 bp paired-end protocol using Illumina HiSeq.

2.5. 16S rRNA Data Analysis

Raw reads were trimmed and filtered using Trimmomatic version 0.39 (SE -phred 33 HEADCROP 17 ILLUMINACLIP:2:30:10 MINLEN:150) [30]. Survived forward reads were processed with DADA2 pipeline v. 3.6.2 [31] (including additional trimming, denoising, and errors correction) giving amplicon sequence variants (ASVs). The ASVs were clustered using MMseqs2 v. 10-6d92c [32] (coverage > 0.95, identity > 0.98) and representative sequences of clusters were further treated as operative taxonomic units (OTUs). OTUs were returned to DADA2 and taxonomy was assigned to OTUs using the SILVA SSU taxonomic training data formatted for DADA2 v.138 [33]. Finally, sequences classified as eukaryotes were removed. PCoA (Principal coordinates analysis), alpha-diversity, and taxonomic analyses were performed with the phyloseq package v. 1.30.0 [34].

2.6. Shot-Gun Metagenomes Assembly and Annotation

Raw reads were trimmed and filtered using Trimmomatic v. 0.39 (PE -phred 33 LEADING:3 TRAILING:3 ILLUMINACLIP:2:30:10 MINLEN:36). Raw and processed reads were inspected with FastQC v. 0.11.9 [35] to control for the read length and quality and presence of adapter sequences. Survived read pairs and forward unpaired reads were assembled with SPAdes v. 3.15.3 (with metaspades option on and k-mer length 55, 99, 121, 127) for de novo assembly [36]. The resultant assemblies were analyzed with QUAST [37]. Contigs longer than 5 kb were selected for further analyses. For taxonomy identification, contigs were aligned against NCBI's nr database using the DIAMOND blastx v. 0.9.24 [38]. The results were transferred to MEGAN v. 6.19.7 [39] and analyzed by the LCA algorithm. For accurate assembly of viral sequences, the filtered reads were assembled with metaviralSPAdes tool [40]. In all selected bacterial and viral contigs ORFs were predicted by MetaGeneMark v. 3.25 [41]. ORFs were annotated using DIAMOND blastp against NCBI's nr database (date of access Apr 2020). Average coverage depth of contigs was obtained by mapping of reads used for de novo assembly on selected contigs with Bowtie 2 v. 2.4.4 [42], and computed from the bam files with SAMtools depth v. 1.10 both programs with default parameters [43].

2.7. Identification of Viral Sequences and Taxonomic Assignment

To detect viral sequences in metagenomic data, contigs assembled with metaviralSPAdes were processed using ViralVerify v. 1.1 [40], VirSorter2 v. 2.2.3 [44], and CheckV v. 0.8.1 [45]. For ViralVerify detected sequences were analyzed with ViralComplete v. 1.1 [40]. All programs were run with default parameters. Sequences, recognized as viruses by at least two out of three pipelines were considered viral. The final set of sequences was clustered using Cd-hit v. 4.8.1 (coverage > 0.8, identity > 0.95) [46] to assemble a non-redundant set. Taxonomy of viral contigs was assigned by gene-sharing network approach using vConTACT2 v. 0.9.22 with default database Prokaryotic Viral RefSeq 207 and with '-s 'MetaGeneMark' –rel-mode 'DIAMOND' parameters [47]. The network was visualized using Cytoscape v. 3.8.0 [48]. Phylogenetic trees were constructed for viral contigs with the ViPTree v. 2.0 web server [49].

2.8. Prediction and Quantification of Phage Defense Systems

Antiphage defense systems were detected in contigs using the PADS arsenal database [50]. Briefly, groups of orthologous sequences from PADS arsenal database were clustered using MMseqs2 v. 13-45111 with coverage > 0.7, identity > 0.85. Representative sequences were aligned using MAFFT v. 7.487 with mafft-linsi (–maxiterate 100) [51] and obtained alignments were polished with trimAl v. 1.2 (gap threshold 50%) [52]. HMM profiles were constructed from the polished alignments with HHMER v. 3.3.2 with hmmbuild algorithm with default parameters [53]. ORFs from selected contigs were scanned with hmm profiles using hmm-

search and hits with *E*-value < 10×10^{-10} were selected. If an ORF was detected by several HMM profiles, a profile with the lowest E-value was considered. Putative defense systems were detected as at least two ORFs separated by less than 5 ORFs, which were recognized by two different HMM profiles belonging to one defense system in the PADS arsenal database. Additionally, defense systems were detected by PADLOC v. 1.0.2 and DefenseFinder v. 0.0.11 web versions [54,55].

Raw counts of defense system genes and putative defense systems were obtained by simple counting of the genes and systems in the metagenomes.

For weighted and normalized quantification of putative defense systems or genes, assembly length, total number of mapped reads, and coverage depth of contigs were considered, and the raw counts of defense system genes and putative defense systems were transformed according to the formula:

$\frac{covdepth \times 10^{12}}{L \times N}$, where L is the total length of a metagenome assembly, N is the total number of reads mapped to the assembly, covdepth is an average coverage depth of a contig, 10^{12}—is a scaling coefficient universal for all metagenomes used to obtain numerical value in a range of raw counts for defense genes and putative defense systems.

2.9. CRISPR Spacer Detection and Analysis

CRISPR-Cas spacer arrays were detected with CRISPRCasTyper web version v. 1.2.3 [56]. Retrieved sequences were clustered with Cd-hit v. 4.8.1 (coverage > 0.95, identity > 0.95) to obtain a non-redundant dataset and to identify spacers shared between metagenomes. To identify spacers shared with CRISPRCasdb [57], they were clustered with sequences from the database using Cd-hit (coverage > 0.95, identity > 0.95). To detect possible protospacers, spacers were searched against the RefSeq viral sequences (date of access Jun 2020) and viral assemblies derived from the White Sea metagenomes using BLASTN [58] with search parameters -word_size 8 -dust no -qcov_hsp_perc 95.

3. Results

3.1. Samples Description and Processing

Visually healthy individuals of marine sponges *Halichondria sitiens* (three samples), *Halichondria panicea* (three samples), and *Isodictya palmata* (two samples), as well as one sample of surrounding seawater, were collected in August 2018 at the single site near N. Pertsov White Sea Biological Station (WSBS, Figure 1A,B). Bacterial fractions were extracted from the sponges and the seawater by step-wise centrifugation and filtration, respectively. Viral fractions were concentrated by tangential filtration from the flow-through after the bacterial fraction isolation. The composition of bacterial communities was investigated by amplification and sequencing (250 + 250 bp, Illumina MiSeq) of the V3-V4 region of 16S rRNA genes resulting in an average sequencing depth of 165 thousand reads per sample (Figure 2D). Sequencing data were processed by a combination of DADA2 pipeline with the subsequent analysis using the phyloseq R package (Figure 1C). Due to the inefficient merging of read pairs by DADA2, analysis was performed for forward reads only. To remove microvariations of ASVs and to work at a genus level, the ASVs were clustered to OTUs with a 98% identity threshold. The rarefaction curve analysis demonstrated that the communities were sequenced with a saturating sequencing depth (Figure S1A).

Total shot-gun metagenomes of the bacterial and viral fractions were sequenced using the Illumina HiSeq; on average, 120 mln high-quality reads were generated per metagenome (Table 1). Analysis of the cumulative length of metagenomes demonstrated the signs of length saturation, indicating the appropriate sequencing depth for the samples (Figure S1B). The overall data analysis pipeline, including assembly and annotation of metagenomes, identification of viral sequences and antiphage systems, and analysis of 16S data is shown in Figure 1C.

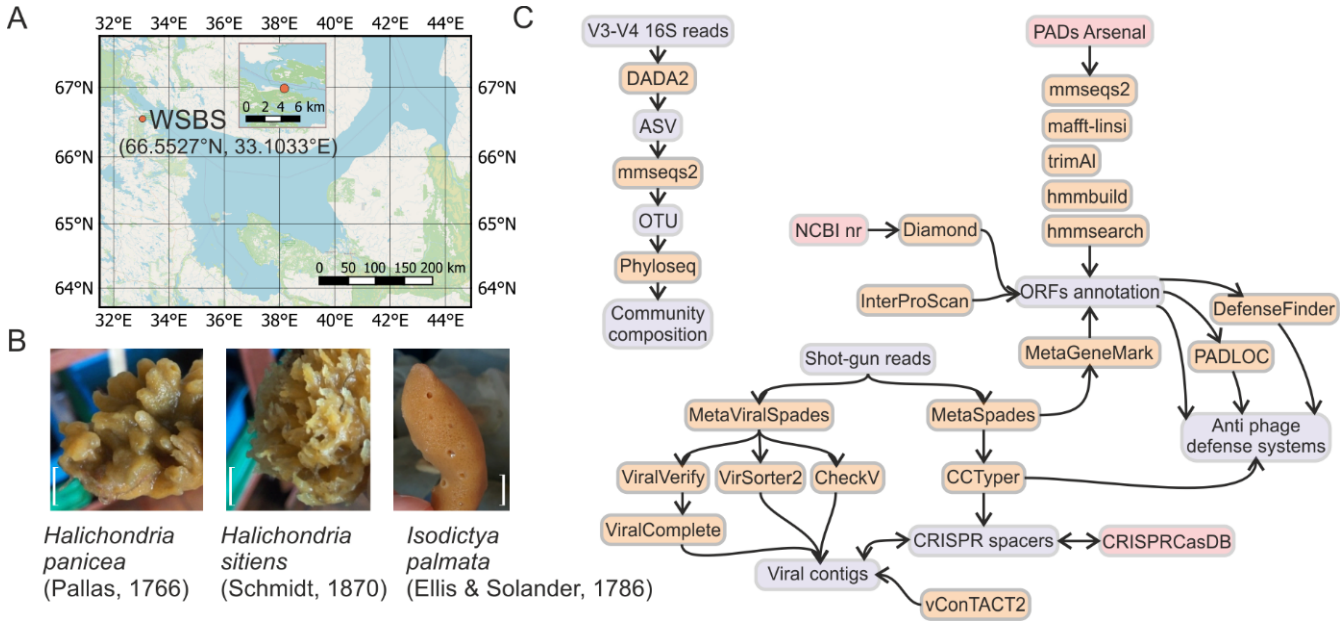
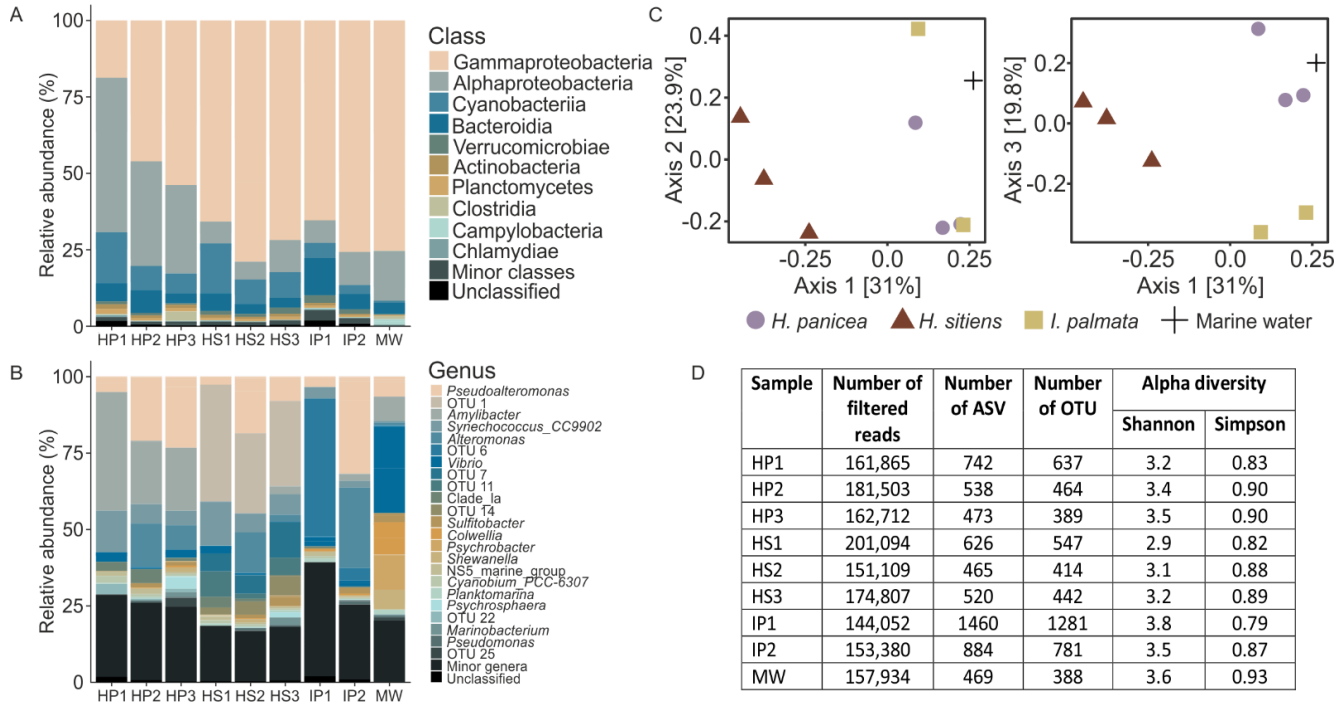


Figure 1. Samples collection and data analysis pipeline. (**A**) Samples of marine water and marine sponges were collected near the N. Pertsov White Sea Biological Station (WSBS), the Kandalaksha Bay of the White Sea, Russia. Scale bars 1 cm. (**B**) Representative fragments of the collected sponges—*H. sitiens*, *H. panicea*, *I. palmata*. (**C**) Program pipelines used for data analysis.



D

Sample	Number of filtered reads	Number of ASV	Number of OTU	Alpha diversity	
				Shannon	Simpson
HP1	161,865	742	637	3.2	0.83
HP2	181,503	538	464	3.4	0.90
HP3	162,712	473	389	3.5	0.90
HS1	201,094	626	547	2.9	0.82
HS2	151,109	465	414	3.1	0.88
HS3	174,807	520	442	3.2	0.89
IP1	144,052	1460	1281	3.8	0.79
IP2	153,380	884	781	3.5	0.87
MW	157,934	469	388	3.6	0.93

Figure 2. Diversity of White Sea sponge-associated and surrounding seawater microbial communities. (**A**) Taxonomic distribution of operational taxonomic units OTUs at the level of classes. *H. panicea* (three replicates, HP), *H. sitiens* (three replicates, HS), *I. palmata* (two replicates, IP), marine water (MW). OTUs that were not assigned at the phylum level were grouped as Unclassified. (**B**) Taxonomic distribution of OTUs at the level of genera. OTUs that were not assigned at a genus level are indicated by numbers. The unclassified group as in A. (**C**) PCoA of Bray-Curtis dissimilarity between the studied microbial communities. (**D**) 16S sequencing statistics and alpha diversity metrics for the samples probed.

Table 1. Shot-gun metagenomes sequencing and assembly statistics. Bct and Vrs are for metagenomes of bacterial and viral fractions, respectively.

Sample	Number of Reads after Filtering	Total Number of Contigs	Total Assembly Length, Mb	N50	L50	Number of Contigs > 5 kb	Total Assembly Length (Contigs > 5 kb), Mb
H.panicea (Bct)	156,744,899	1,048,333	704.2	1678	62,534	10,147	102.6
H. sitiens (Bct)	151,978,145	660,860	515.4	2309	36,966	11,252	105.4
I. palmata (Bct)	151,705,427	984,734	626.6	1646	49,855	8239	87.0
Marine water (Bct)	139,611,995	1,409,157	783.8	1271	74,205	7222	82.2
H.panicea (Vrs)	105,797,172	773,423	477.2	1253	54,605	3723	50.6
H. sitiens (Vrs)	81,791,648	416,792	250.4	1313	21,446	2579	33.8
I. palmata (Vrs)	74,156,030	214,337	111.1	1227	8941	965	11.1
Marine water (Vrs)	107,464,178	1,078,397	691.0	1437	56,832	7208	100.6

3.2. 16S Metagenomics Revealed Distinct Complex Bacterial Communities Associated with the White Sea Sponges and Marine Water

Metagenomic analysis of the V3–V4 fragment of 16S rRNA gene revealed the difference in the compositions of communities associated with marine sponges and marine water. The most abundant phyla in sponges were Proteobacteria (80%), Cyanobacteria (9%), and Bacteroidetes (6%) (Figure 2A). The most abundant classes in *H. panicea* were Gamma- and Alphaproteobacteria (40% and 38%, respectively), followed by Cyanobacteria (10%). *H. sitiens* and *I. palmata* were largely dominated by Gammaproteobacteria (72% and 71%, respectively), Alphaproteobacteria (8% and 9%), and Cyanobacteria (11% and 4%). On a genus level, *Pseudoalteromonas* OTUs were present in all samples, while sponges were dominated by sponge-specific OTUs not abundant in marine water (Figure 2B). For *H. panicea* the characteristic OTU was OTU 2 classified as *Amylibacter* (27%), for *H. sitiens* it was OTU 1 (31%, order UBA10353), and for *I. palmata* it was OTU 6 (25%, unclassified Gammaproteobacteria). The *Amylibacter* OTU had a 100% identity with *Candidatus Halichondribacter symbioticus*, a recently described symbiont of *H. panicea* from Icelandic waters [59]. Correspondently, the results of the PCoA analysis using Bray-Curtis dissimilarity demonstrated that sponges are inhabited by specific communities differ from surrounding marine water (Figure 2C). On average, 497 OTUs were detected in *H. sitiens*, 468 in *H. panicea*, 1031 in *I. palmata*, and 388 in marine water (Figure 2D). Alpha-diversity Shannon and Simpson indexes indicated that all bacterial communities of sponges and sea water had a complex structure, but they were dominated by specific groups of microorganisms (Figure 2D).

3.3. Enrichment and Diversity of Viromes from the White Sea

To obtain viral particles from the sponge and marine water samples, an enrichment strategy was developed, which is based on the fractionation of samples by consecutive centrifugation steps and/or filtering (see Materials and Methods). Shot-gun metagenomes of the resultant viral fractions were sequenced and assembled (Table 1), and viral sequences were identified by several algorithms including ViralVerify, VirSorter2, and CheckV. To reduce the number of false positives, sequences detected by at least two pipelines were considered as putative viral sequences and used in further analysis (Figure S2A). Applying this approach, in total 453 viral sequences were identified in metagenomes. An average number of retrieved viral sequences was considerably higher for the viral fractions than for the bacterial fractions (96 vs. 27, *t*-test *p*-value 0.027), indicating an efficient enrichment procedure (Figure 3A). Interestingly, the number of viral sequences shared between viral and bacterial metagenomes was low (six sequences, Figure 3B, Venn diagram) allowing us to suppose that the enrichment method could be selective for some viral particles. Viromes associated with particular sponge species and marine water were distinct, as indicated by the low number of viral sequences shared among samples (Figure 3B, heatmap). This

observation is not surprising given the distinctive profiles of bacterial communities revealed in these samples using 16S metagenomics (Figure 2B,C).

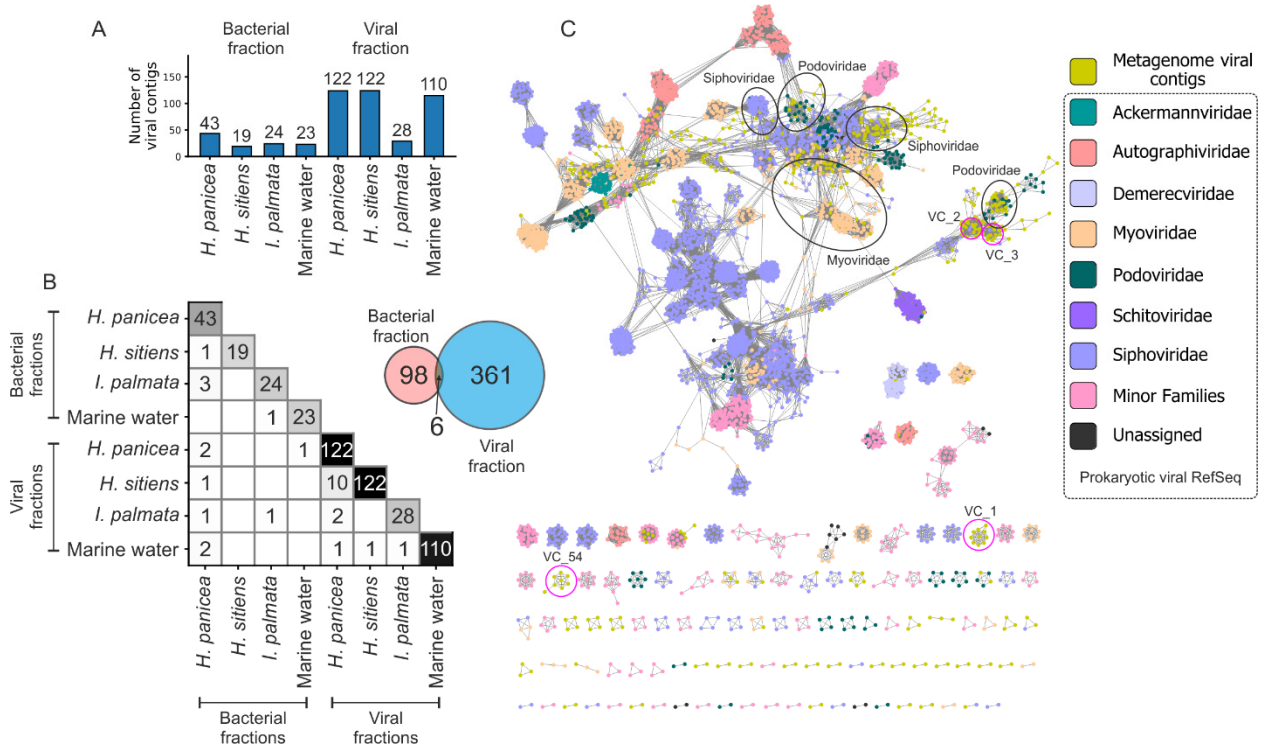


Figure 3. Analysis of viromes associated with marine sponges and sea water. (**A**) The number of viral contigs identified in different samples. Viral contigs were clustered using CD-hit (identity > 95%, coverage > 80%). (**B**) Viral contigs shared between the samples. The number of viral contigs shared between bacterial and viral fractions. (**C**) Sequence similarity network (SSN) constructed by vConTACT2 with Viral RefSeq-prokaryotes database clustered with viral contigs detected in the White Sea metagenomes. Nodes and edges represent the viral genomes and a significant similarity shared between them, respectively. Black ovals indicate viral contigs derived from the metagenomes in clusters with phages from the families of Caudovirales. Pink circles mark the most abundant viral clusters composed exclusively of viral contigs (VC) from the White Sea metagenomes. The network was visualized in Cytoscape.

Phylogeny of retrieved viruses was investigated using vConTACT2 (Figure 3C). From 453 identified viral contigs, only 10% (46) were clustered to the sequences from the reference database. These belong to the order Caudovirales and families Myoviridae (21 sequences), Siphoviridae (19 sequences), and Podoviridae (six sequences). 90% of viral sequences had no similarity to known viral sequences; among them, 261 sequences were clustered (81 clusters) and others were assigned as outliers or singletons (Figure S2B).

3.4. Antiphage Systems Detected in Metagenomes

Putative defense genes and defense systems were detected in assembled and annotated metagenomes using the custom HMM profiles. The profiles were built from the alignments constructed using the sequences from the PADS Arsenal database [50]. By simple counting, the most abundant putative defense genes belonged to Septu, BREX, Zorya, and DISARM prokaryotic antiphage defense systems [50] (Figure 4A). The most abundant putative defense systems included RM, Zorya, and CRISPR-Cas (Figure 4B). Normalization of metagenomes and adjustment of gene copy-number by the average coverage depth of contigs did not change the abundance of putative defense genes and systems dramatically, except that Abi replaced CRISPR-Cas in top-three defense systems (Figure 4C,D). In accordance with the custom search results, the most prevalent systems identified in

metagenomes by DefenseFinder were RM, followed by Abi and CRISPR-Cas (Figure S3A). Analysis with PADLOC also demonstrated the high prevalence of CRISPR-Cas systems in the metagenomes (Figure S3B).

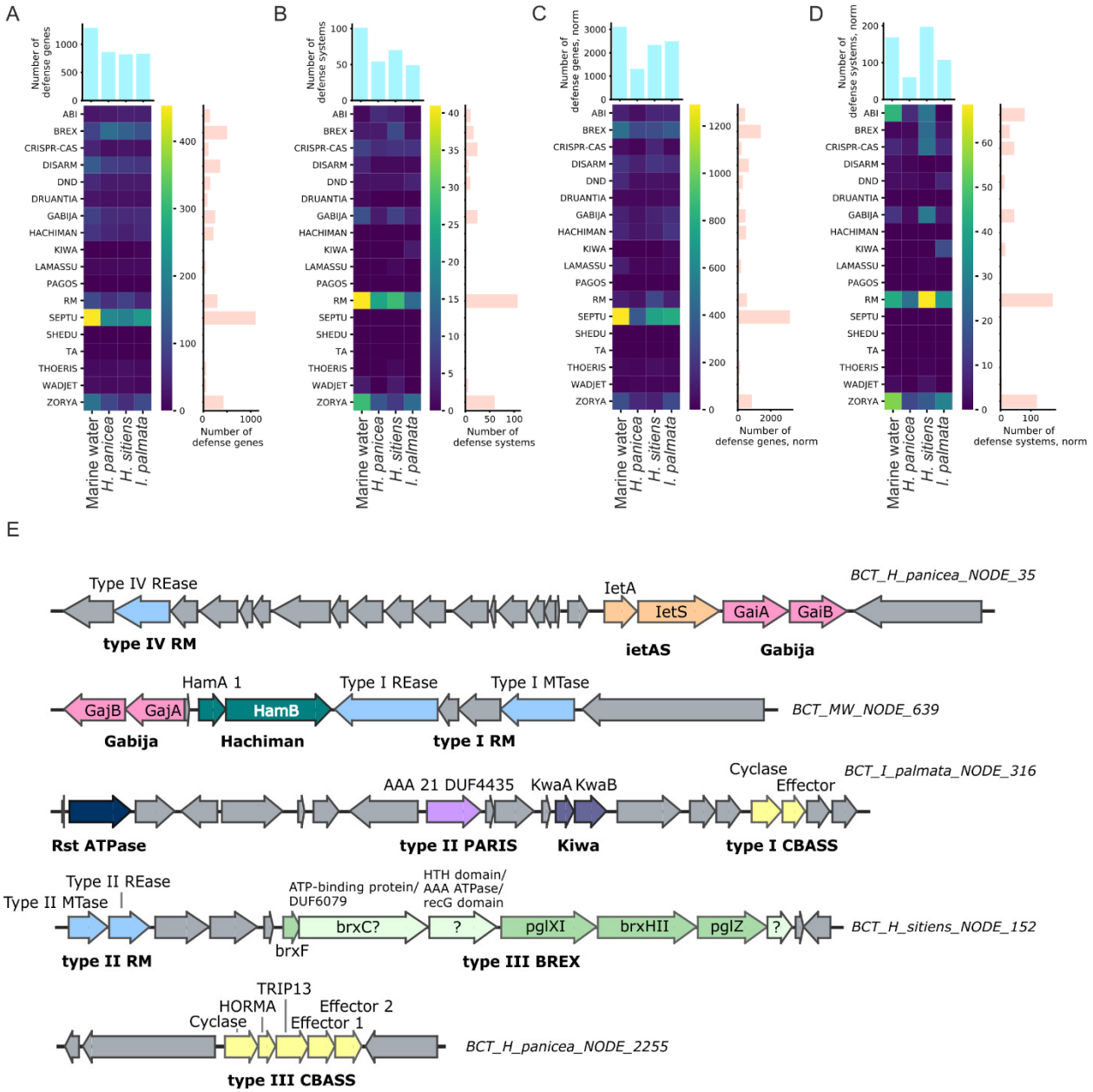


Figure 4. Heatmaps representing the abundance of putative defense system genes (**A**) and putative defense systems (**B**) found in metagenomes. Heatmaps representing the normalized abundance of putative defense system genes (**C**) and putative defense systems (**D**) found in metagenomes. The color legend indicating the numbers of genes (**A**,**C**) or systems (**B**,**D**) for the heatmaps is provided with colorbars on the side of heatmaps. (**E**) Examples of genomic islands enriched in defense systems genes and unusual defense systems identified in the metagenomes of White Sea sponges and marine water. ORFs are indicated as arrows; the length of arrows is proportional to the length of ORFs; defense systems of different types are marked with distinct colors; non-defense genes are in gray.

Overall, among the studied metagenomes, marine water had slightly increased raw counts of defense systems and is particularly rich in RM, Zorya, Gabija, and CRISPR-Cas systems (Figure 4B and Figure S3A). PADLOC also identified a substantial number of the

CBASS systems in the marine water metagenome (Figure S3B). After the normalization procedure, however, the metagenome from *H. sitiens* was found to have the highest number of defense systems (Figure 4D), thus indicating the presence of abundant bacteria enriched with the RM, CRISPR-Cas, Gabija, Gao, and QatABCD defense systems.

In the metagenomes from *H. panicea* and the marine water, two unusual type-III CBASS systems were found. These systems contain two effector genes with a limited similarity shared, a feature previously not described for the known CBASS systems [60] (Figure 4E). Also, for several contigs, most of which were assigned to Gammaproteobacteria by MEGAN, islands including co-clustered defense systems were identified. In one of the defense islands from the *H. sitiens* metagenome, a novel variation of the type-III BREX system was found. It contains a non-canonical *brxC* gene with limited homology to the classical C-genes and the insertion of an additional gene of unknown function between the *brxC* and *brxX* genes (Figure 4E).

3.5. Analysis of CRISPR Spacers Show the Flow of Viruses between Marine Communities

CRISPR immune systems are known to keep the record of invasions of bacteria by phages and other mobile elements in the associated CRISPR arrays [61]. Non-array sequences matching the spacers from arrays, known as protospacers, evidence the facts of invasion and the subsequent immune adaptation of the host.

We used CRISPRCasTyper to detect CRISPR arrays in metagenomes and derived spacers. It is noteworthy that the metagenomes of viral fractions contained microbial DNA. Totally, 2358 unique spacers were identified in 197 CRISPR arrays, of which 171 were found in the bacterial fractions, and 26 in the viral fractions. None of them matched spacers from the CRISPRCasDB database (contains 291,402 unique spacers) (Figure 5C). The spacer sets from different White Sea metagenomes poorly overlap, thus highlighting the distinct compositions of bacterial communities (Figure 5A). Interestingly, metagenomes of viral fractions had spacers not detected in bacterial fractions, which may originate from arrays of small-size bacteria passing through the 0.22 μm filter (Figure 5B).

To detect potential protospacers, we searched spacer sequences in the RefSeq Viruses database and the set of putative viral contigs. Only one reliable match was found with a genome of *Pseudomonas* phage phi2 from the RefSeq Viruses, and 121 matches were found with putative viral contigs discussed above. Next, we analyzed the sources of spacers and protospacers and constructed a bipartite community-community network of virus-host interactions (Figure 5D). As can be seen from the network, the number of productive interactions (captured by the spacer-acquisition events; black numbers on the edges of the network) were typically higher within communities than between different communities. As an exception, the spacers from the bacterial community of *H. sitiens* more frequently matched with viral contigs from *H. panicea* than with any other community, including the *H. sitiens* itself, potentially indicating the flow of viruses between the communities. Generally, *H. panicea* contained more viruses matched by spacers from other samples than any other community, possibly reflecting the ongoing spread of viruses. In contrast, *I. palmata* had more spacers matching viruses from other samples than any other community, likely highlighting the recent numerous events of a successful defense. Supporting this speculation, *I. palmata* metagenomes showed the smallest number of putative viral contigs (Figure 3A). Additionally, the average coverage depth of a ubiquitous 144–145 kb-long viral contig, closely related to *Pseudoalteromonas* phages from the Myoviridae group (Figure S4A), was considerably lower in the *I. palmata* metagenome (9 times less than in *H. panicea* and 36 times less than in *H. sitiens*). The drop in the contig abundance can be explained by efficient CRISPR adaptation and interference, as 10 different spacers, matching this contig (the CRISPR system I-F type, Figure S4B), were found in *I. palmata* metagenome (Figure S4C). Of note, nine out of 10 protospacers had an adjacent CC PAM classical for the I-F systems [62]. Taken together, our data indicate that a directional flow of viruses may exist between closely localized bacterial communities.

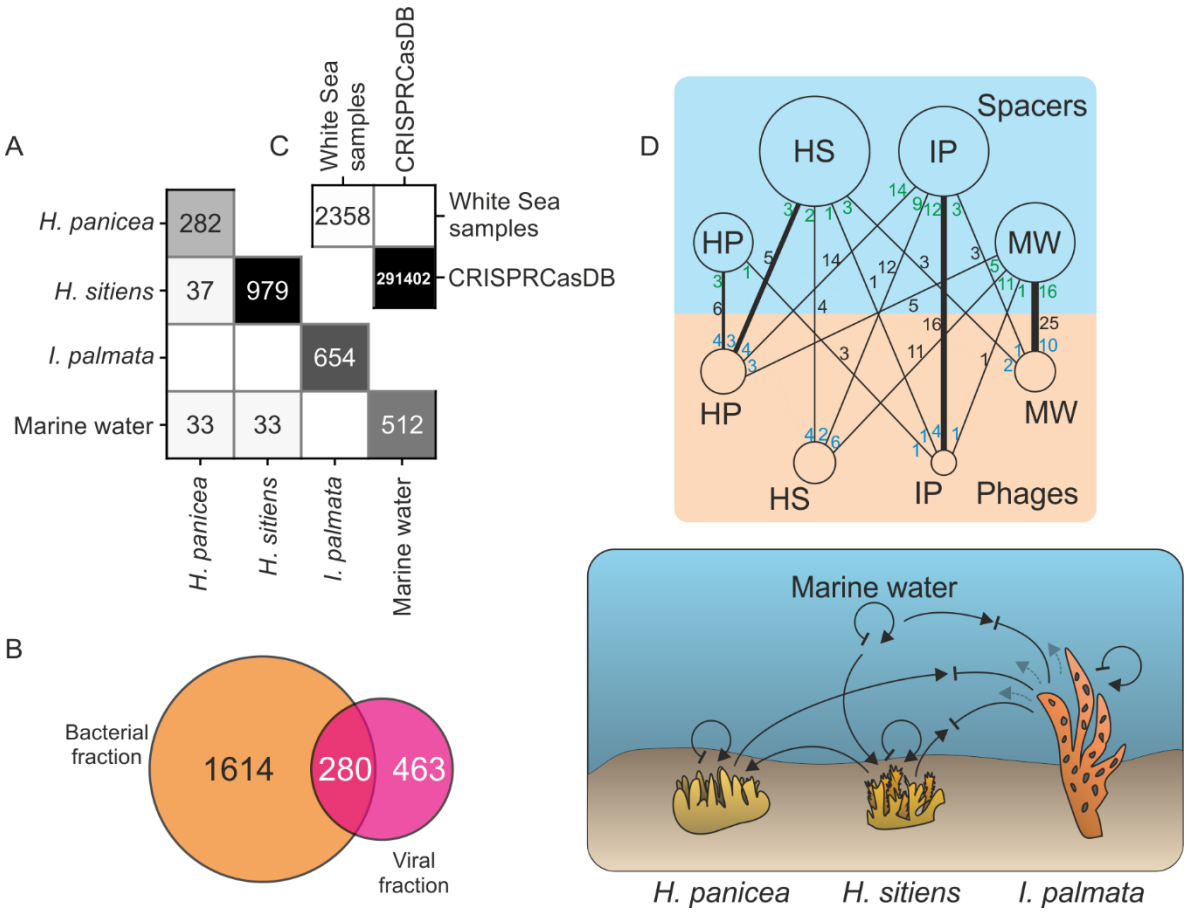


Figure 5. (**A**) Numbers of CRISPR spacers shared between studied bacterial communities. (**B**) The number of CRISPR spacers shared between bacterial and viral fractions. (**C**) The number of CRISPR spacers identified in the White Sea metagenomes shared with CRISPRCasDB database. (**D**) Top, a bipartite community–community network of virus–host interactions. The upper and the lower nodes represent the source of the spacers and the viral contigs, respectively. The area of nodes is proportional to the number of spacers or viral contigs identified in metagenomes. Black, green, and blue numbers on the edges of the network account for the total number of matches between spacers and protospacers, number of matching spacers, and number of viral contigs with protospacers, respectively. For each of the upper nodes, an edge with the largest number of spacers is shown in bold. Arrows represent a phage burst or a phage flow between communities; inhibition arrows indicate the defense against phages. Dark black and pale arrows correspond to more recent and older events, respectively.

4. Discussion

To investigate possible interactions between bacteria and viruses within and between sympatric marine sponges, bacterial and viral fractions from three species of White Sea sponges and surrounding marine water were isolated and their metagenomic DNA was sequenced.

4.1. Sponge Species Specificity of Associated Bacterial and Viral Communities

Microbiomes of *H. panicea* from different geographical locations were investigated earlier by strain culturing [63], 16S metagenomics [7,64], and whole metagenome sequencing [65]. In accordance with the previously published 16S metagenomics data, Alphaproteobacteria, in particular, a few sponge-specific OTUs from this class (e.g., *Amylibacter*), were prevalent in *H. panicea*-associated bacterial communities from the White Sea. The *Amylibacter* OTU sequence was identical to the 16S of the symbiotic bacterium *Candidatus Halichondribacter symbioticus*, recently described in *H. panicea* from Icelandic waters [59],

indicating the wide geographical association between the sponge and the symbiotic bacterium. It is worth noting that Gammaproteobacteria were unusually prevalent in the White Sea *H. panicea* samples (especially the genera *Pseudoalteromonas*, *Alteromonas*, and *Vibrio*). We speculate that this reflects the community profile shift due to the increased water temperature during the summer season of 2018 [66].

In the current study, the compositions of microbial communities of *H. sitiens* and *I. palmata* were investigated using metagenomic techniques for the first time. The communities were largely dominated by Gammaproteobacteria, Alphaproteobacteria, and Cyanobacteria, a pattern, observed for many sponge species [7,10]. At the genus level, communities were dominated by specific yet poorly taxonomically assigned OTUs (OTU 1 for *H. sitiens* and OTU 6 for *I. palmata*), which arguably correspond to novel putative sponge symbionts. Similar composition of associated bacterial communities and, particularly, the abundance of indicated OTUs were also observed in sponges sampled in 2016 (unpublished data), additionally supporting the existence of sponge-specific communities. The predominance of a single bacterium has been shown for several different sponges previously [59,67,68].

Only a minor fraction of viral contigs identified in the White Sea metagenomes were classified to known groups of viruses. These groups were mostly represented by Myoviridae, Podoviridae, and Siphoviridae, the phage families typically found in viromes of different sponges from various habitats [15,26,27,69]. The majority of viral contigs did not fall into clusters with known reference sequences, representing the viral "dark matter" [70].

The compositions of microbial metagenomes, viromes and spaceromes (the set of all CRISPR spacers) were compared between different sponge species. We revealed a prominent host species specificity in all cases, though, our analyses could be biased due to small sample size (two or three individuals). This observation highlights the distinct nature and independence of sponge-associated communities, even spatially closely localized [6,71].

4.2. Analysis of Defense Systems Repertoires of White Sea Sponges

Sponges filter thousands of liters of water a day, accumulating virus particles that can affect the composition of the associated bacterial communities [72]. Thus, sponge-associated microorganisms are forced to provide reliable protection against phages and plasmids in conditions of high viral load and high cell density [16,17,19,21,25,73]. To investigate the defense potential of the communities, we performed mining of the White Sea sponge and marine water metagenomes and revealed a variety of defense systems. Unexpectedly, the diversity and abundance of the systems in the sponge and marine water metagenomes were comparable. This observation argues with the published data that postulate the higher prevalence of various defense systems in the metagenomes of sponge-associated communities over the surrounding water (19 of 20 sponge metagenomes studied). Noteworthy, this rule was also correct for the two species of cold-water Antarctic sponges [20]. However, in the sponge *Scopalina ruetzleri*, collected in the equatorial Atlantic, CRISPR-Cas and RM systems were less abundant, indicating possible deviations from the general scheme [19]. One can speculate that the discrepancy can be explained by the skew of the earlier studies toward the detection and counting of CRISPR-Cas, RM, and TA systems. We hypothesize that the defense potential of the bacterial communities of cold-water sponges could be compensated by diverse defense systems belonging to other and/or yet unknown types. A unified analysis, including the detection of an extended set of defense systems, is needed for the correct comparison of the metagenomes.

Cyclic-oligonucleotide-based anti-phage signaling systems (CBASS) are a group of bacterial antiphage systems. Upon phage infection, the CBASS generates signaling cyclic oligonucleotides. In turn, the oligonucleotides activate an effector, which promotes cell death resulting in abortive infection. The effectors can contain various cell-killing domains: patatin-like phospholipase, endonuclease, peptidase, etc. [58]. We identified unusual type-III CBASS variants in two contigs from marine water and *H. panicea* sponge: clusters contained two effector genes, both encoding putative DNA endonucleases. Interestingly,

the effectors were distant homologs, which probably implies a different substrate specificity or, alternatively, the effectors can be activated by different signaling oligonucleotides.

A bacteriophage exclusion defense system (BREX) blocks phage replication and discriminates between host and phage DNA by methylation patterns [74,75]. A novel variation of the type-III BREX cluster with an unknown gene inserted between *brxC* (putative ATPase) and *brxX* (putative methyltransferase) was identified. Using HHpred we predicted that this protein contains a helix-turn-helix domain on the N-terminus and its C-terminal part is homologous to ATP-dependent helicases. Also, it contains a conserved triad typical for the Mg^{2+}-binding site, which is a common characteristic for some ribonucleases [76,77]. Recently, insertion of a type-IV RM nuclease named *brxU* was observed in a plasmid-borne type-I BREX system between the *brxC* and *brxX* genes [78]. We speculate that there could be a hotspot for the insertions of accessory proteins, such as nucleases, inside the BREX cluster. The hotspot may be mediated by the frequent rearrangements occurring at the *brxX* gene that might be due to the high toxicity of its product [74].

4.3. Tracking the Viral Flow between Communities by a Spacer-Protospacer Network

Bacteria acquire new spacers when a CRISPR-Cas adaptation complex captures fragments of a phage genome and incorporates them into CRISPR arrays during the process known as adaptation. When the phage infects the cells harboring the matching spacer, the viral genome is recognized and degraded by a CRISPR interference complex guided by the spacer transcript [79,80]. The information stored in CRISPR-Cas arrays can be used to reveal the history of the phage-host interactions [81]. Based on this, we investigated the exchange of viruses between sponge-associated communities. Using a network approach, we identified several examples of inter-community viral exchanges. In particular, in the *I. palmata* metagenome, a contig containing the type-IF CRISPR-Cas system and an associated array was found to carry spacers matching the abundant viral contigs identified in all investigated sponges. Although these viral contigs were highly prevalent in the case of *H. panicea* and *H. sitiens*, no matching spacers were found in the microbial metagenomes of these organisms. In contrast, the viral contigs were considerably less abundant in *I. palmata*, implying the effective interference by the CRISPR-Cas system. We speculate that this phage could have been first propagated in *I. palmata*, where it was suppressed by the activity of CRISPR-Cas. At the same time, it was spread and invaded the communities of other nearby sponges not yet adapted to it. Interestingly, the I-F array contained two groups of spacers targeting the phage genome—at the beginning and the end of the array. The first group contained more spacers perfectly matching the genome. This observation and the fact that new spacers are usually incorporated at the beginning of an array [82] suggest that the community of *I. palmata* faced the phage twice—earlier and probably, during the current phage outburst. Our network-based approach can be scaled up and applied for the investigation of viral exchanges within large sets of bacterial communities. This method in particular can be utilized for investigation of virus-mediated horizontal gene transfer between bacterial communities, which could play a role in adaptation for extreme cold environments [83].

5. Conclusions

This study presents the first insight into the bacterial and viral communities associated with three cold-water marine sponge species from the White Sea. We demonstrated that despite the close spatial proximity of the holobionts on the seafloor, the sponge-associated communities are species-specific and differ substantially by a composition of bacterial OTUs, diversity of associated viruses, and acquired CRISPR-spacers. At the same time, the frequent exchange of viruses between the communities exists, as indicated by the acquisition of CRISPR-spacers against viruses from another community. The abundances and diversity of defense systems in the metagenomes of sponges and the metagenome of water are comparable, which could be a specific feature of polar ecosystems. Finally, in the metagenomes, we identified unusual variations of BREX and CBASS bacterial defense

systems. This study enables us to begin to fill the gaps in the knowledge of complex interactions between bacterial and viral communities associated with polar marine organisms.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/life12010025/s1, Figure S1: (A) Rarefaction curves analysis of V3-V4 16S metagenomes sequences in this study. (B) Analysis of the cumulative length of metagenomes for contigs > 5 kb, Figure S2: (A) The Venn diagrams represent the concordance between the predictions of viral sequences by ViralVerify and ViralComplete pipeline, VirSorter2, and CheckV. (B) Phylogenetic tree of representative viral contigs to four of the most abundant viral clusters composed exclusively of viral contigs from the White Sea metagenomes, Figure S3: (A) Heatmaps representing the numbers of defense systems genes, numbers of putative defense systems, normalized numbers of defense systems genes, and normalized numbers of putative defense systems found in metagenomes using DefenseFinder. (B) Heatmaps representing the numbers of defense systems genes, numbers of putative defense systems, normalized numbers of defense systems genes, and normalized numbers of putative defense systems found in metagenomes using PADLOC, Figure S4: (A) Phylogenetic tree of viral sequences constructed with ViPTree, (B) CRISPR-Cas systems containing spacers for which matching protospacers were found in the putative viral contigs, (C) Annotation of a viral contig targeted by 10 spacers from I-F CRISPR-Cas system identified in the *I. palmata* metagenome.

Author Contributions: A.R. processed samples, analyzed data, and drafted the manuscript. V.F. collected and processed samples. S.T. prepared the 16S rRNA sequencing libraries and sequenced the samples. S.D. acquired funding and finalized the manuscript. D.S. conceptualized and designed the research, analyzed data, prepared figures, and drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by a grant from the Ministry of Science and Higher Education of Russian Federation (agreement No. 075-10-2021-114 from 11 October 2021). D.S. was supported by the Russian Foundation for Basic Research grant (project No. 20-34-90069). Sequencing at Skoltech Genomics Core Facility was supported by the Skoltech Life Sciences Program grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Sequencing data were deposited in SRA (SRR17015069-SRR17015085). Metagenome assemblies were deposited in GeneBank under the BioProject PRJNA781598. Custom code used for data analysis is available from https://github.com/sutormin94/WSBS_Sponges_metagenomics (accessed on 30 November 2021).

Acknowledgments: We thank the personnel of the White Sea Biological Station, particularly, Alexander Tzetlin and Tatyana Neretina for the opportunity to collect and process samples at WSBS, WSBS SCUBA diving team for samples collection, and Boris Osadchenko for identification of sponge species.

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

References

1. Bordenstein, S.R.; Theis, K.R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol.* **2015**, *13*, e1002226. [CrossRef]
2. Bosch, T.; McFall-Ngai, M. Metaorganisms as the New Frontier. *Zoology* **2011**, *114*, 185–190. [CrossRef] [PubMed]
3. Pita, L.; Rix, L.; Slaby, B.M.; Franke, A.; Hentschel, U. The Sponge Holobiont in a Changing Ocean: From Microbes to Ecosystems. *Microbiome* **2018**, *6*, 1–18. [CrossRef] [PubMed]
4. Freeman, C.J.; Easson, C.G.; Matterson, K.O.; Thacker, R.W.; Baker, D.M.; Paul, V.J. Microbial Symbionts and Ecological Divergence of Caribbean Sponges: A New Perspective on an Ancient Association. *ISME J.* **2020**, *14*, 1571–1583. [CrossRef]
5. Tianero, M.D.; Balaich, J.N.; Donia, M.S. Localized Production of Defence Chemicals by Intracellular Symbionts of Haliclona Sponges. *Nat. Microbiol.* **2019**, *4*, 1149–1159. [CrossRef]
6. Cleary, D.F.R.; de Voogd, N.J.; Polónia, A.R.M.; Freitas, R.; Gomes, N.C.M. Composition and Predictive Functional Analysis of Bacterial Communities in Seawater, Sediment and Sponges in the Spermonde Archipelago, Indonesia. *Microb. Ecol.* **2015**, *70*, 889–903. [CrossRef] [PubMed]
7. Thomas, T.; Moitinho-Silva, L.; Lurgi, M.; Björk, J.R.; Easson, C.; Astudillo-García, C.; Olson, J.B.; Erwin, P.M.; López-Legentil, S.; Luter, H.; et al. Diversity, Structure and Convergent Evolution of the Global Sponge Microbiome. *Nat. Commun.* **2016**, *7*, 11870. [CrossRef]

8. Souza, D.T.; Genuário, D.B.; Silva, F.S.P.; Pansa, C.C.; Kavamura, V.N.; Moraes, F.C.; Taketani, R.G.; Melo, I.S. Analysis of Bacterial Composition in Marine Sponges Reveals the Influence of Host Phylogeny and Environment. *FEMS Microbiol. Ecol.* **2017**, *93*, 204. [CrossRef]
9. de Voogd, N.J.; Cleary, D.F.R.; Polónia, A.R.M.; Gomes, N.C.M. Bacterial Community Composition and Predicted Functional Ecology of Sponges, Sediment and Seawater from the Thousand Islands Reef Complex, West Java, Indonesia. *FEMS Microbiol. Ecol.* **2015**, *91*, 19. [CrossRef]
10. Moitinho-Silva, L.; Nielsen, S.; Amir, A.; Gonzalez, A.; Ackermann, G.L.; Cerrano, C.; Astudillo-Garcia, C.; Easson, C.; Sipkema, D.; Liu, F.; et al. The Sponge Microbiome Project. *Gigascience* **2017**, *6*, gix077. [CrossRef]
11. Astudillo-García, C.; Bell, J.J.; Montoya, J.M.; Moitinho-Silva, L.; Thomas, T.; Webster, N.S.; Taylor, M.W. Assessing the Strength and Sensitivity of the Core Microbiota Approach on a Highly Diverse Sponge Reef. *Environ. Microbiol.* **2020**, *22*, 3985–3999. [CrossRef] [PubMed]
12. Podell, S.; Blanton, J.M.; Neu, A.; Agarwal, V.; Biggs, J.S.; Moore, B.S.; Allen, E.E. Pangenomic Comparison of Globally Distributed Poribacteria Associated with Sponge Hosts and Marine Particles. *ISME J.* **2018**, *13*, 468–481. [CrossRef] [PubMed]
13. Van Soest, R.W.M.; Boury-Esnault, N.; Vacelet, J.; Dohrmann, M.; Erpenbeck, D.; De Voogd, N.J.; Santodomingo, N.; Vanhoorne, B.; Kelly, M.; Hooper, J.N.A. Global Diversity of Sponges (Porifera). *PLoS ONE* **2012**, *7*, e35105. [CrossRef]
14. Weisz, J.B.; Lindquist, N.; Martens, C.S. Do Associated Microbial Abundances Impact Marine Demosponge Pumping Rates and Tissue Densities? *Oecologia* **2007**, *155*, 367–376. [CrossRef] [PubMed]
15. Jahn, M.T.; Arkhipova, K.; Markert, S.M.; Stigloher, C.; Lachnit, T.; Pita, L.; Kupczok, A.; Ribes, M.; Stengel, S.T.; Rosenstiel, P.; et al. A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. *Cell Host Microbe* **2019**, *26*, 542–550.e5. [CrossRef]
16. Horn, H.; Slaby, B.M.; Jahn, M.T.; Bayer, K.; Moitinho-Silva, L.; Förster, F.; Abdelmohsen, U.R.; Hentschel, U. An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes. *Front. Microbiol.* **2016**, *7*, 1751. [CrossRef]
17. Karimi, E.; Slaby, B.M.; Soares, A.R.; Blom, J.; Hentschel, U.; Costa, R. Metagenomic Binning Reveals Versatile Nutrient Cycling and Distinct Adaptive Features in Alphaproteobacterial Symbionts of Marine Sponges. *FEMS Microbiol. Ecol.* **2018**, *94*, fiy074. [CrossRef]
18. Fan, L.; Reynolds, D.; Liu, M.; Stark, M.; Kjelleberg, S.; Webster, N.S.; Thomas, T. Functional Equivalence and Evolutionary Convergence in Complex Communities of Microbial Sponge Symbionts. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1878–E1887. [CrossRef]
19. Rua, C.P.J.; Gregoracci, G.B.; Santos, E.O.; Soares, A.C.; Francini-Filho, R.B.; Thompson, F. Potential Metabolic Strategies of Widely Distributed Holobionts in the Oceanic Archipelago of St Peter and St Paul (Brazil). *FEMS Microbiol. Ecol.* **2015**, *91*, 43. [CrossRef]
20. Moreno-Pino, M.; Cristi, A.; Gillooly, J.F.; Trefault, N. Characterizing the Microbiomes of Antarctic Sponges: A Functional Metagenomic Approach. *Sci. Rep.* **2020**, *10*, 645. [CrossRef]
21. Karimi, E.; Ramos, M.; Gonçalves, J.M.S.; Xavier, J.R.; Reis, M.P.; Costa, R. Comparative Metagenomics Reveals the Distinctive Adaptive Features of the Spongia Officinalis Endosymbiotic Consortium. *Front. Microbiol.* **2017**, *8*, 2499. [CrossRef]
22. Pascelli, C.; Laffy, P.W.; Kupresanin, M.; Ravasi, T.; Webster, N.S. Morphological Characterization of Virus-like Particles in Coral Reef Sponges. *PeerJ* **2018**, *6*, e5625. [CrossRef]
23. Laffy, P.W.; Wood-Charlson, E.M.; Turaev, D.; Jutz, S.; Pascelli, C.; Botté, E.S.; Bell, S.C.; Peirce, T.E.; Weynberg, K.D.; van Oppen, M.J.H.; et al. Reef Invertebrate Viromics: Diversity, Host Specificity and Functional Capacity. *Environ. Microbiol.* **2018**, *20*, 2125–2141. [CrossRef] [PubMed]
24. Nguyen, M.; Wemheuer, B.; Laffy, P.W.; Webster, N.S.; Thomas, T. Taxonomic, Functional and Expression Analysis of Viral Communities Associated with Marine Sponges. *PeerJ* **2021**, *9*, e10715. [CrossRef]
25. Zhou, K.; Qian, P.-Y.; Zhang, T.; Xu, Y.; Zhang, R. Unique Phage–Bacterium Interplay in Sponge Holobionts from the Southern Okinawa Trough Hydrothermal Vent. *Environ. Microbiol. Rep.* **2021**, *13*, 675–683. [CrossRef]
26. Butina, T.V.; Khanaev, I.V.; Kravtsova, L.S.; Maikova, O.O.; Bukin, Y.S. Metavirome Datasets from Two Endemic Baikal Sponges *Baikalospongia bacillifera*. *Data Br.* **2020**, *29*, 105260. [CrossRef]
27. Zhou, K.; Zhang, R.; Sun, J.; Zhang, W.; Tian, R.-M.; Chen, C.; Kawagucci, S.; Xu, Y. Potential Interactions between Clade SUP05 Sulfur-Oxidizing Bacteria and Phages in Hydrothermal Vent Sponges. *Appl. Environ. Microbiol.* **2019**, *85*, e00992-19. [CrossRef]
28. Sun, G.; Xiao, J.; Wang, H.; Gong, C.; Pan, Y.; Yan, S.; Wang, Y. Efficient purification and concentration of viruses from a large body of high turbidity seawater. *MethodsX* **2014**, *1*, e197–e206. [CrossRef] [PubMed]
29. Sambrook, J.; Fritsch, E.; Maniatis, T. Bacteriophage λ growth, purification, and DNA extraction. In *Molecular Cloning*, 2nd ed.; Nolan, C., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 1989; Volume 1.
30. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
31. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat. Methods* **2016**, *13*, 581–583. [CrossRef]
32. Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef] [PubMed]
33. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef] [PubMed]

34. McMurdie, P.J.; Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [CrossRef]
35. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2015. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 30 November 2021).
36. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
37. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
38. Buchfink, B.; Xie, C.; Huson, D.H. Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* **2014**, *12*, 59–60. [CrossRef] [PubMed]
39. Huson, D.H.; Beier, S.; Flade, I.; Górska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.-J.; Tappu, R. MEGAN Community Edition—Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **2016**, *12*, e1004957. [CrossRef]
40. Antipov, D.; Raiko, M.; Lapidus, A.; Pevzner, P.A. MetaviralSPAdes: Assembly of Viruses from Metagenomic Data. *Bioinformatics* **2020**, *36*, 4126–4129. [CrossRef]
41. Zhu, W.; Lomsadze, A.; Borodovsky, M. *Ab Initio* Gene Identification in Metagenomic Sequences. *Nucleic Acids Res.* **2010**, *38*, e132. [CrossRef]
42. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
43. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
44. 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-Classifier, Expert-Guided Approach to Detect Diverse DNA and RNA Viruses. *Microbiome* **2021**, *9*, 1–13. [CrossRef] [PubMed]
45. Nayfach, S.; Camargo, A.P.; Schulz, F.; Eloe-Fadrosh, E.; Roux, S.; Kyrpides, N.C. CheckV Assesses the Quality and Completeness of Metagenome-Assembled Viral Genomes. *Nat. Biotechnol.* **2020**, *39*, 578–585. [CrossRef]
46. Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
47. Bin Jang, H.; Bolduc, B.; Zablocki, O.; Kuhn, J.H.; Roux, S.; Adriaenssens, E.M.; Brister, J.R.; Kropinski, A.M.; Krupovic, M.; Lavigne, R.; et al. Taxonomic Assignment of Uncultivated Prokaryotic Virus Genomes Is Enabled by Gene-Sharing Networks. *Nat. Biotechnol.* **2019**, *37*, 632–639. [CrossRef]
48. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.; Wang, J.; Ramage, D.; Schwikowski, B. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
49. Nishimura, Y.; Yoshida, T.; Kuronishi, M.; Uehara, H.; Ogata, H.; Goto, S. ViPTree: The Viral Proteomic Tree Server. *Bioinformatics* **2017**, *33*, 2379–2380. [CrossRef]
50. Zhang, Y.; Zhang, Z.; Zhang, H.; Zhao, Y.; Zhang, Z.; Xiao, J. PADS Arsenal: A Database of Prokaryotic Defense Systems Related Genes. *Nucleic Acids Res.* **2020**, *48*, D590–D598. [CrossRef]
51. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
52. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef] [PubMed]
53. Johnson, L.; Eddy, S.; Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **2010**, *11*, 431. [CrossRef]
54. Payne, L.J.; Todeschini, T.C.; Wu, Y.; Perry, B.J.; Ronson, C.W.; Fineran, P.C.; Nobrega, F.L.; Jackson, S.A. Identification and Classification of Antiviral Defence Systems in Bacteria and Archaea with PADLOC Reveals New System Types. *Nucleic Acids Res.* **2021**, *49*, 10868–10878. [CrossRef] [PubMed]
55. Tesson, F.; Hervé, A.; Touchon, M.; d'Humières, C.; Cury, J.; Bernheim, A. Systematic and Quantitative View of the Antiviral Arsenal of Prokaryotes. *bioRxiv* **2021**, 2021.09.02.458658. [CrossRef]
56. Russel, J.; Pinilla-Redondo, R.; Mayo-Muñoz, D.; Shah, S.A.; Sørensen, S.J. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *CRISPR J.* **2020**, *3*, 462–469. [CrossRef] [PubMed]
57. Pourcel, C.; Touchon, M.; Villeriot, N.; Vernadet, J.-P.; Couvin, D.; Toffano-Nioche, C.; Vergnaud, G. CRISPRCasdb a Successor of CRISPRdb Containing CRISPR Arrays and Cas Genes from Complete Genome Sequences, and Tools to Download and Query Lists of Repeats and Spacers. *Nucleic Acids Res.* **2020**, *48*, D535–D544. [CrossRef]
58. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef] [PubMed]
59. Knobloch, S.; Jóhannsson, R.; Marteinsson, V. Bacterial Diversity in the Marine Sponge Halichondria Panicea from Icelandic Waters and Host-Specificity of Its Dominant Symbiont "Candidatus Halichondribacter Symbioticus". *FEMS Microbiol. Ecol.* **2018**, *95*, fiy220. [CrossRef]

60. Millman, A.; Melamed, S.; Amitai, G.; Sorek, R. Diversity and Classification of Cyclic-Oligonucleotide-Based Anti-Phage Signalling Systems. *Nat. Microbiol.* **2020**, *5*, 1608–1615. [CrossRef]
61. Garrett, S.C. Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array. *Front. Microbiol.* **2021**, *12*, 739. [CrossRef]
62. Vorontsova, D.; Datsenko, K.A.; Medvedeva, S.; Bondy-Denomy, J.; Savitskaya, E.E.; Pougach, K.; Logacheva, M.; Wiedenheft, B.; Davidson, A.R.; Severinov, K.; et al. Foreign DNA Acquisition by the I-F CRISPR–Cas System Requires All Components of the Interference Machinery. *Nucleic Acids Res.* **2015**, *43*, 10848–10860. [CrossRef] [PubMed]
63. Wichels, A.; Würtz, S.; Döpke, H.; Schütt, C.; Gerdts, G. Bacterial Diversity in the Breadcrumb Sponge Halichondria Panicea (Pallas). *FEMS Microbiol. Ecol.* **2006**, *56*, 102–118. [CrossRef]
64. Naim, M.A.; Morillo, J.A.; Sørensen, S.J.; Waleed, A.A.S.; Smidt, H.; Sipkema, D. Host-Specific Microbial Communities in Three Sympatric North Sea Sponges. *FEMS Microbiol. Ecol.* **2014**, *90*, 390–403. [CrossRef]
65. Strehlow, B.; Schuster, A.; Francis, W.; Canfield, D. Metagenomic Data for *Halichondria panicea* from Illumina and Nanopore Sequencing and Preliminary Genome Assemblies for the Sponge and Two Microbial Symbionts. *bioRxiv* **2021**, 2021.10.18.464794. [CrossRef]
66. Ereskovsky, A.; Ozerov, D.A.; Pantyulin, A.N.; Tzetlin, A.B. Mass Mortality Event of White Sea Sponges as the Result of High Temperature in Summer 2018. *Polar Biol.* **2019**, *42*, 2313–2318. [CrossRef]
67. Kennedy, J.; Codling, C.E.; Jones, B.V.; Dobson, A.D.W.; Marchesi, J.R. Diversity of Microbes Associated with the Marine Sponge, *Haliclona simulans*, Isolated from Irish Waters and Identification of Polyketide Synthase Genes from the Sponge Metagenome. *Environ. Microbiol.* **2008**, *10*, 1888–1902. [CrossRef]
68. Croué, J.; West, N.J.; Escande, M.L.; Intertaglia, L.; Lebaron, P.; Suzuki, M.T. A Single Betaproteobacterium Dominates the Microbial Community of the Crambescidine-Containing Sponge *Crambe Crambe*. *Sci. Reports* **2013**, *3*, 2583. [CrossRef]
69. Pascelli, C.; Laffy, P.W.; Botté, E.; Kupresanin, M.; Rattei, T.; Lurgi, M.; Ravasi, T.; Webster, N.S. Viral Ecogenomics across the Porifera. *Microbiome* **2020**, *8*. [CrossRef]
70. Roux, S.; Hallam, S.J.; Woyke, T.; Sullivan, M.B. Viral Dark Matter and Virus-Host Interactions Resolved from Publicly Available Microbial Genomes. *Elife* **2015**, *4*, e08490. [CrossRef] [PubMed]
71. Steinert, G.; Busch, K.; Bayer, K.; Kodami, S.; Arbizu, P.M.; Kelly, M.; Mills, S.; Erpenbeck, D.; Dohrmann, M.; Wörheide, G.; et al. Compositional and Quantitative Insights into Bacterial and Archaeal Communities of South Pacific Deep-Sea Sponges (Demospongiae and Hexactinellida). *Front. Microbiol.* **2020**, *11*, 716. [CrossRef] [PubMed]
72. Welsh, J.E.; Steenhuis, P.; De Moraes, K.R.; Van Der Meer, J.; Thieltges, D.W.; Brussaard, C.P.D. Marine Virus Predation by Non-Host Organisms. *Sci. Rep.* **2020**, *10*, 5221. [CrossRef]
73. Thomas, T.; Rusch, D.; DeMaere, M.Z.; Yung, P.Y.; Lewis, M.; Halpern, A.; Heidelberg, K.B.; Egan, S.; Steinberg, P.D.; Kjelleberg, S. Functional Genomic Signatures of Sponge Bacteria Reveal Unique and Shared Features of Symbiosis. *ISME J.* **2010**, *4*, 1557–1567. [CrossRef] [PubMed]
74. Goldfarb, T.; Sberro, H.; Weinstock, E.; Cohen, O.; Doron, S.; Charpak-Amikam, Y.; Afik, S.; Ofir, G.; Sorek, R. BREX Is a Novel Phage Resistance System Widespread in Microbial Genomes. *EMBO J.* **2015**, *34*, 169–183. [CrossRef]
75. Gordeeva, J.; Morozova, N.; Sierro, N.; Isaev, A.; Sinkunas, T.; Tsvetkova, K.; Matlashov, M.; Truncaite, L.; Morgan, R.D.; Ivanov, N.V.; et al. BREX System of Escherichia Coli Distinguishes Self from Non-Self by Methylation of a Specific DNA Site. *Nucleic Acids Res.* **2019**, *47*, 253–265. [CrossRef] [PubMed]
76. Yang, J.; Deng, X.; Li, Y.; Ma, X.; Feng, J.; Yu, B.; Chen, Y.; Luo, Y.; Wang, X.; Chen, M.; et al. Structure of Schlafen13 Reveals a New Class of TRNA/RRNA- Targeting RNase Engaged in Translational Control. *Nat. Commun.* **2018**, *9*, 1165. [CrossRef] [PubMed]
77. Garvie, C.W.; Wu, X.; Papanastasiou, M.; Lee, S.; Fuller, J.; Schnitzler, G.R.; Horner, S.W.; Baker, A.; Zhang, T.; Mullahoo, J.P.; et al. Structure of PDE3A-SLFN12 Complex Reveals Requirements for Activation of SLFN12 RNase. *Nat. Commun.* **2021**, *12*, 4375. [CrossRef] [PubMed]
78. Picton, D.M.; Luyten, Y.A.; Morgan, R.D.; Nelson, A.; Smith, D.L.; Dryden, D.T.F.; Hinton, J.C.D.; Blower, T.R. The Phage Defence Island of a Multidrug Resistant Plasmid Uses Both BREX and Type IV Restriction for Complementary Protection from Viruses. *Nucleic Acids Res.* **2021**, *49*, 11257–11273. [CrossRef]
79. Koonin, E.V.; Makarova, K.S.; Wolf, Y.I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **2017**, *71*, 233. [CrossRef] [PubMed]
80. Isaev, A.B.; Musharova, O.S.; Severinov, K.V. Microbial Arsenal of Antiviral Defenses—Part, I. *Biochemistry* **2021**, *86*, 319–337. [CrossRef]
81. Dion, M.B.; Plante, P.L.; Zufferey, E.; Shah, S.A.; Corbeil, J.; Moineau, S. Streamlining CRISPR Spacer-Based Bacterial Host Predictions to Decipher the Viral Dark Matter. *Nucleic Acids Res.* **2021**, *49*, 3127–3138. [CrossRef]
82. Yosef, I.; Goren, M.G.; Qimron, U. Proteins and DNA Elements Essential for the CRISPR Adaptation Process in Escherichia Coli. *Nucleic Acids Res.* **2012**, *40*, 5569–5576. [CrossRef]
83. Pucciarelli, S.; Devaraj, R.R.; Mancini, A.; Ballarini, P.; Castelli, M.; Schrallhammer, M.; Petroni, G.; Miceli, C. Microbial Consortium Associated with the Antarctic Marine Ciliate *Euplotes focardii*: An Investigation from Genomic Sequences. *Microb. Ecol.* **2015**, *70*, 484–497. [CrossRef] [PubMed]