

dsRNA-seq Reveals Novel RNA Virus and Virus-Like Putative Complete Genome Sequences from *Hymeniacidon* sp. Sponge

SYUN-ICHI URAYAMA^{1,2,3*}, YOSHIHIRO TAKAKI⁴, DAISUKE HAGIWARA^{2,3}, and TAKURO NUNOURA¹

¹Research Center for Bioscience and Nanoscience (CeBN), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan; ²Laboratory of Fungal Interaction and Molecular Biology (donated by IFO), Department of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan; ³Microbiology Research Center for Sustainability (MiCS), University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan; and ⁴Super-cutting-edge Grand and Advanced Research (SUGAR) Program, JAMSTEC, 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

(Received October 17, 2019—Accepted January 17, 2020—Published online February 28, 2020)

Invertebrates are a source of previously unknown RNA viruses that fill gaps in the viral phylogenetic tree. Although limited information is currently available on RNA viral diversity in the marine sponge, a primordial multicellular animal that belongs to the phylum *Porifera*, the marine sponge is one of the well-studied holobiont systems. In the present study, we elucidated the putative complete genome sequences of five novel RNA viruses from *Hymeniacidon* sponge using a combination of double-stranded RNA sequencing, called fragmented and primer ligated dsRNA sequencing, and a conventional transcriptome method targeting single-stranded RNA. We identified highly diverged RNA-dependent RNA polymerase sequences, including a potential novel RNA viral lineage, in the sponge and three viruses presumed to infect sponge cells.

Key words: RNA virus, metagenome, dsRNA, sponge

The marine sponge is a primordial multicellular animal that belongs to the phylum *Porifera*. Sponges are recognized as hosts of diverse symbiotic lives (Taylor *et al.*, 2007), including bacteria, archaea, fungi, protists, and viruses (Webster and Thomas, 2016), and the relationships between the host sponge and symbionts vary from mutualism to parasitism (Webster and Taylor, 2012; Thomas *et al.*, 2016). To understand the ecosystems associated with sponges, sponges have been described as “holobionts,” which are defined as assemblages of multiple species that form ecological units (Webster and Taylor, 2012). Sponge-bacteria/-archaea interactions have been extensively examined (Schmitt *et al.*, 2012; Thomas *et al.*, 2016; Pita *et al.*, 2018); however, the impact of viruses on sponge holobionts remains unclear. Among the limited number of virus studies on sponge holobionts, the presence of diverse virus-like particles in sponge cells and symbiont compartments has been reported (Vacelet and Gallissian, 1978; Pascelli *et al.*, 2018). Furthermore, metagenomic analyses suggested the dominance of bacteriophage lineages in DNA viromes associated with reef sponge holobionts (Butina *et al.*, 2015; Laffy *et al.*, 2016; Laffy *et al.*, 2018). To the best of our knowledge, one metatranscriptome study on a sponge holobiont characterized the presence of the partial genome sequences of RNA viruses related to the families *Narnaviridae*, *Dicistroviridae*, *Partitiviridae*, *Picobirnaviridae*, *Picornaviridae*, *Tombusviridae*, *Nodaviridae*, and *Hepeviridae* as well as unclassified viruses (Waldron, *et al.* [2018] Metagenomic sequencing suggests arthropod antiviral RNA interference is highly

derived, and identifies novel viral small RNAs in a mollusc and a brown alga. *bioRxiv* doi: <https://doi.org/10.1101/166488>).

To access RNA viromes, conventional total RNA-seq (single-stranded RNA [ssRNA]-seq) is the most commonly used method to detect RNA viral sequences in diverse environments and organisms, such as oceanic waters (Steward *et al.*, 2013; Culley *et al.*, 2014), lake waters (Djikeng *et al.*, 2009), seafloor sediments (Engelhardt *et al.*, 2015), invertebrates (Shi *et al.*, 2016), and vertebrates (Shi *et al.*, 2018). In the latter two studies, deep RNA sequencing was performed on total RNAs from invertebrates and vertebrates and thousands of RNA viruses that appeared to infect them were identified (Shi *et al.*, 2016; 2018). However, this method does not have the capacity to obtain complete genome sequences from RNA viruses because of the principle of cDNA synthesis. The rapid amplification of the cDNA end (RACE) or alternative methods have been conducted to obtain complete RNA viral genomes/segments, including terminal sequences, (Roossinck *et al.*, 2010; Shi *et al.*, 2016). In contrast, double-stranded RNA (dsRNA) sequencing has the advantage of the superior enrichment of RNA viral genomes over the conventional RNA sequencing method (Roossinck *et al.*, 2010). Long cellular dsRNA consists of dsRNA virus genomes and replicative intermediates of ssRNA viruses (Morris and Dodds, 1979). Fragmented and primer-ligated dsRNA sequencing (FLDS), a method to obtain the full-length sequence of dsRNA including terminal sequences, was recently developed (Koyama *et al.*, 2016; Urayama *et al.*, 2016; 2018b; Fukasawa *et al.*, 2019; Higashiura *et al.*, 2019). The feature of this method is to retrieve the sequences from terminal regions that are needed to reconstruct segmented RNA virus genomes based on the

* Corresponding author. E-mail: urayama.shunichi.gn@u.tsukuba.ac.jp;
Tel: +81-29-853-6636; Fax: +81-29-853-4605.

conserved sequences of terminal regions in each genome segment (Urayama *et al.*, 2018a). In the present study, we applied the FLDS method in combination with ssRNA-seq to reveal sponge RNA viromes.

Materials and Methods

Sample collection

Sponges on tidal rocks in Tokyo Bay (35.3405°N, 139.6396°E) were sampled in April 2014 and 2015. After washing with distilled water, samples were stored at -80°C until further analyses.

Extraction and purification of dsRNA and ssRNA

Sponge samples were disrupted in liquid nitrogen in a mortar. Total nucleic acids were manually extracted from approximately 1 g of sponge powder. dsRNA was purified using a combination of the cellulose column method (Urayama *et al.*, 2015) and an enzymatic treatment, as described previously (Urayama *et al.*, 2018a). Some of the powder was processed to obtain total RNA using the TRIzol Plus RNA Purification Kit (Invitrogen) according to the manufacturer's protocol. Total RNA was treated with DNase I (Invitrogen) and further purified using the RNA Clean and Concentrator-5 Kit (Zymo Research).

dsRNA sequencing library construction

Nuclease-treated dsRNA was fragmented by ultrasound at 4°C in Snap-Cap microTUBEs (Covaris) using Covaris S220 (settings: run time 35 s, peak power 140.0 W, duty factor 2.0%, and 200 cycles burst⁻¹) (Covaris) (Urayama *et al.*, 2018a). After the purification of fragmented dsRNAs using the ZymoClean Gel RNA Recovery Kit (Zymo Research), double-stranded cDNA was synthesized as described previously (Urayama *et al.*, 2018a). Briefly, a ssDNA adapter was ligated to the 3'-ends of fragmented dsRNA and cDNA was synthesized using the SMARTer RACE 5'/3' Kit (Takara Bio) with an oligonucleotide primer that had a complementary sequence to the adapter sequence. After amplification, cDNA was quantified using the Qubit dsDNA HS Kit (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies).

ssRNA sequencing library construction

Double-stranded cDNA was synthesized from 2 μg of total RNA with the Prime Script Double Strand cDNA Synthesis Kit (Takara Bio) using random primers (9-mers). The resultant cDNA was quantified using the Qubit dsDNA HS Kit (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies).

Illumina sequencing

Double-stranded cDNAs were fragmented using Covaris S220 (settings: run time 55 s, peak power 175.0 W, duty factor 5.0%, and 200 cycles burst⁻¹), and fragmented cDNAs were used as templates for KAPA Hyper Prep Kit Illumina platforms (Kapa Biosystems). The resultant Illumina libraries were evaluated using the KAPA library quantification kit (Kapa Biosystems) and applied to the Illumina MiSeq platform (Illumina) according to the manufacturer's protocol (600-cycle kit to perform 300-bp paired-end sequencing).

Data processing

Raw sequencing reads for dsRNA-seq were processed as described previously (Urayama *et al.*, 2018a). rRNA reads in dsRNA-seq libraries were identified by SortMeRNA (Kopylova *et al.*, 2012) and excluded from further virome analyses. In RNA virome analyses, reads were subjected to *de novo* assembly using CLC Genomics Workbench ver. 11.0 (CLC Bio) with the following parameters: a minimum contig length of 500, word value and bubble size set to auto, and assemblies that were manually examined and extended using the Tablet viewer (Milne *et al.*, 2010). Using the mapping

tool, each contig was screened under the following conditions: at least $3\times$ average coverage and 500 bp in length. If >10 reads defined the same terminal position of a certain contig, the position was recognized as a terminal end. Contigs for which both ends were identified as termini were defined as potential genome segments. RNA viral genes in potential genome segments and contigs were identified based on sequence similarities to known RNA viral proteins in the NCBI non-redundant (nr) database using BLASTX (Camacho *et al.*, 2009) with an $e\text{-value}\leq 1\times 10^{-5}$. Sequences that matched a known RNA-dependent RNA polymerase (RdRp) gene by BLASTX with an $e\text{-value}\leq 1\times 10^{-5}$ were collected from RNA virus contigs and segments. The presence of the RdRp domain (PF00680, PF00978, and PF02123) was confirmed using PfamScan (Chojnacki *et al.*, 2017) with an $e\text{-value}\leq 1\times 10^{-5}$ when a sequence had significant similarity ($e\text{-value}\leq 1\times 10^{-5}$) to an RNA viral polyprotein or hypothetical protein. Nucleotide sequences encoding the RdRp gene were clustered at 90% identity using VSEARCH (Rognes *et al.*, 2016). The centroid sequence of each cluster was defined as a representative sequence. We also identified the signature amino acid residues conserved in the RdRp domain (Venkataraman *et al.*, 2018) in the ORFs of potential genome segments that had no ORF with significant similarities ($e\text{-value}\leq 1\times 10^{-5}$) to sequences in the nr database. Quality-based variant detection was performed using CLC Genomics Workbench ver. 11.0 (CLC Bio).

Raw sequencing reads from ssRNA-seq were processed as described previously (Urayama *et al.*, 2016). Small subunit (SSU) rRNA sequences were reconstructed from ssRNA-seq reads with EMERGE (Miller *et al.*, 2011).

Phylogenetic analyses

The deduced amino acid sequences of the putative RdRp genes obtained in the present study and their relatives in the NCBI nr database were aligned using MUSCLE (Edgar, 2004) in MEGA5 (Tamura *et al.*, 2011). To exclude ambiguous positions, the alignment was trimmed by trimAl (option: $-\text{gt } 1$) (Capella-Gutiérrez *et al.*, 2009). Phylogenetic analyses were conducted using RAxML (Stamatakis, 2014). Bootstrap tests were conducted with 1,000 samplings. The model of amino acid substitution for each phylogeny was selected by Aminosan (Tanabe, 2011), as judged by Akaike's information criterion (Sugiura, 1978), *i.e.*, rtREV+G for *Partitiviridae*, LG+F for *Reoviridae*, LG+F+G for *Picobirnaviridae*, and rtREV+G for *Dicistroviridae*. Tree-Graph2 (Stöver and Müller, 2010) and FigTree (Rambaut, 2014) were used to illustrate the resulting phylogenies.

Accession numbers

The sequences obtained in the present study are available in the GenBank database repository (accession nos. DDBJ: BKZT01000001–BKZT01000434, BKZU01000001–BKZU01000431, and LC504241–LC504250) and Short Read Archive database (accession nos. DDBJ: DRA008491).

Results

Overview of NGS reads

We obtained 1.8–4.1 M high-quality reads in four RNA-seq libraries (Fig. S1). Regarding dsRNA-seq libraries, reads associated with RNA viruses (see below) and the remaining rRNA reads occupied approximately 20 and 30% of dsRNA reads, respectively. The ratio of rRNA reads in dsRNA-seq libraries varied among the samples reported previously (Urayama *et al.*, 2016; 2018a; 2018b; Decker *et al.*, 2019). Some rRNAs potentially contaminated the dsRNA fraction due to their secondary structures. Thus, the relative amount of dsRNA to rRNA in samples may have affected the rRNA read ratio in dsRNA sequencing. Furthermore, the

formation of dsRNAs of sense/antisense transcripts from overlapping regions may influence the results of dsRNA sequencing (Decker *et al.*, 2019). In ssRNA-seq libraries, RNA viruses and rRNA reads occupied approximately 0.5 and 80% of ssRNA reads, respectively.

Diversity of cellular rRNA in sponges

To identify potentially active biomes in sponges obtained in 2014 and 2015, SSU rRNA sequences were reconstructed from total ssRNA-seq reads, and compositions were esti-

mated based on mapped read numbers on SSU rRNA sequences (Miller *et al.*, 2011). *Hymeniacion sp.* SSU rRNA reads occupied >90% of all SSU rRNA reads in total ssRNA libraries from sponges obtained in both 2014 and 2015 (Fig. 1A). The next abundant eukaryotic SSU rRNA sequence was identified as *Mytilus edulis* (0.8%) mussel, and the abundance of other eukaryotic SSU rRNA sequences was lower than 0.2% of all SSU rRNA reads. No archaeal or bacterial SSU rRNA sequence was found from reads that dominated >2% of all SSU rRNA reads.

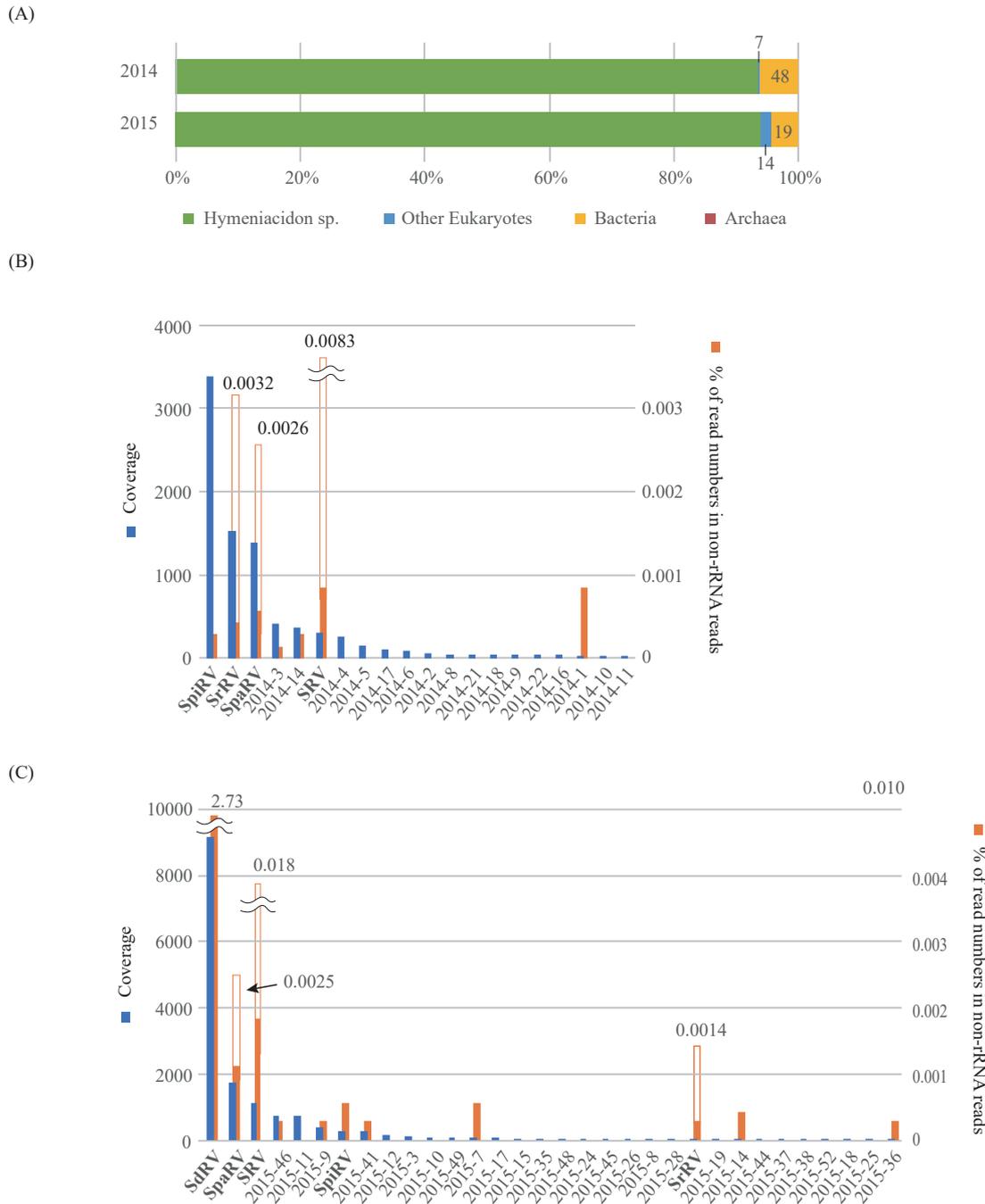


Fig. 1. Summary of dsRNA-seq and ssRNA-seq analyses. (A) Relative abundance of sequence reads mapped on SSU rRNA sequences constructed by EMIRGE. Numbers indicate OTU. (B, C) The average coverage (left axis) and relative abundance (right axis) of RdRp-encoding RNA viral genome segments (Table S1) estimated by the mapping of dsRNA-seq and ssRNA-seq data, respectively, from sponge samples collected in 2014 (B) and 2015 (C). Open bars indicate reads mapped on segments encoding CDS with no significant similarity (e-value ≤ 1 × 10⁻⁵) to known viral genes. Items on the horizontal axis represent the IDs of RdRp-encoding full-length sequences (Table S1) or virus names.

Sponge-associated non-retro RNA virome

The FLDS analysis (dsRNA-seq) succeeded in enriching viral reads and revealed RNA viromes associated with sponges collected in 2014 and 2015, while RNA virus reads were rare in the ssRNA-seq analysis, except for reads related to *Dicistroviridae* in 2015 samples (Table S1). We obtained 253 and 233 sequences encoding RdRp genes from the 2014 and 2015 samples, respectively (Table S1). These sequences included viral sequences related to 9 dsRNA virus families and 3 positive-sense ssRNA virus families based on a RdRp sequence similarity analysis, while 11 and 33 families were previously established in dsRNA and positive-sense ssRNA viruses, respectively (King *et al.*, 2012). We also found 78 sequences encoding RdRp genes belonging to unclassified RNA viral lineages based on the taxonomic lineage of BLASTX top hit sequences (Table S1), and 7.7% of RdRp reads in dsRNA libraries were mapped to these segments and contigs. Viral sequences related to families recognized as animal viruses, such as *Birnaviridae* and *Dicistroviridae*, were also identified from sponge RNA viromes. After the clustering of RdRp domain sequences, 167 representative sequences (>1.5 kb, <90% identity) were identified and defined as operational taxonomic units (OTUs). Based on the taxonomic lineage of BLASTX top hit sequences, the richness of OTUs was analyzed and many viral OTUs were suggested to be closely related with (or belonging to) *Totiviridae*, *Reoviridae*, and *Partitiviridae* (Fig. 2). The majority of OTUs in unclassified groups showed low amino acid sequence identity to the best hit sequence (Fig. S2). All of the top hit sequences in “unclassified RNA viruses” were RdRp identified from *Bryopsis cinicola* (Chlorophyta) (Koga *et al.*, 2003). In “unclassified dsRNA viruses” and “unclassified viruses”, several OTUs showed close sequence similarities to RdRp sequences identified from the Diatom colony (Urayama *et al.*, 2016). Dominance in the coverage of these dsRNA viruses was also confirmed based on the read abundance in dsRNA libraries (Table S1).

Major putative complete genomes

We obtained 118 and 180 full-length potential viral genome segments from 2014 and 2015 samples, respectively. Among RdRp-encoding potential viral genome seg-

ments (Fig. 1B and C), we focused on the five dominant viral genome segments that represented high average coverage (>1,000×) in each dsRNA-seq library. The five segments occupied more than 50% of the total average coverage of RdRp-encoding segments and contigs (253 and 233 OTU in 2014 and 2015, respectively). Following the identification of major RdRp sequences in dsRNA-seq libraries, we reconstructed RNA genome sets based on the similarity of terminal sequences among segments (Fig. 3 and S3). These five genome sets were considered to be “putative complete” because they had very high coverage (>1000×), and, thus, there was a low chance of missing other fragments of their genomes. Reconstructed viral genomes were also supported by the following parameters, in addition to the similarity of terminal sequences. The difference in GC% among sequences within a group sharing similar terminal sequences was <5%. Each ORF did not show amino acid sequence similarity (>50% similarity) to any ORFs encoded by other potential genome segments in a single group. We named these RNA viral genomes harboring ORFs that presented significant similarity ($e\text{-value} \leq 1 \times 10^{-5}$) with known viral sequences after the names of the most closely related virus family (Table S2 and see below); Sponge-associated dicistro-like RNA virus (SdRV, accession no. BKZU01000024), Sponge-associated partiti-like RNA virus (SpaRV, accession no. BKZT01000015), Sponge-associated reo-like RNA virus (SrRV, accession no. BKZT01000005), Sponge-associated picobirna-like RNA virus (SpiRV, accession no. BKZT01000016), and Sponge-associated RNA virus (SRV, accession no. LC504241–50).

No ORF of SRV showed significant similarity ($e\text{-value} \leq 1 \times 10^{-5}$) with sequences in the nr database. However, an ORF of SRV RNA2 showed insignificant similarity ($e\text{-value} = 0.01$) to the RT_like super family (cl02808) in the CDD search. Thus, we manually identified the signature amino acid residues conserved in the RdRp domain (Venkataraman *et al.*, 2018) including the “GDD” sequence in the ORF (Fig. S4). Accordingly, we concluded that the genome segment set corresponded to RNA viral genomes that belonged to a previously unknown viral family.

To estimate the potential host of these RNA viruses identified in the dsRNA-seq analysis, we evaluated the abundance of these viral sequences in ssRNA-seq according to the previously proposed definition that viral reads sharing

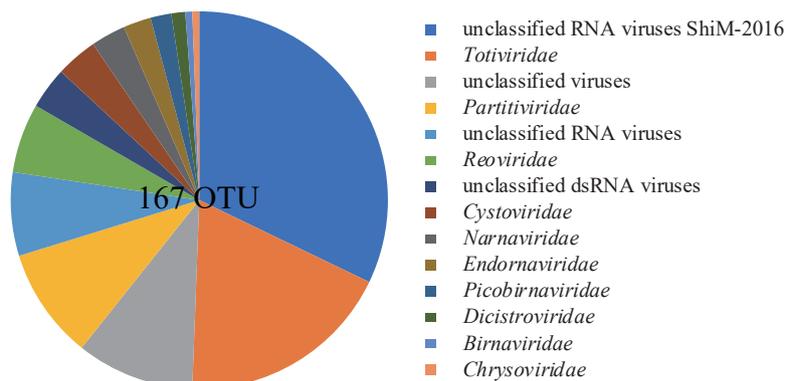
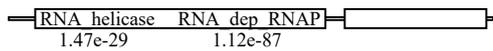


Fig. 2. Overview of the RNA virome identified from the sponge. Taxonomic composition of RdRp-based OTUs identified in the present study.

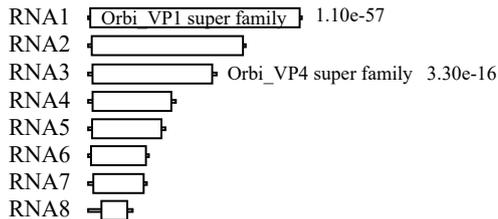
Group 1: Sponge-associated dicistrovirus-like RNA virus (SdRV)



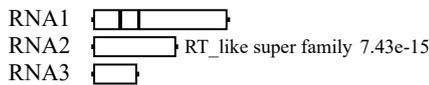
Group 2: Sponge-associated partitivirus-like RNA virus (SpaRV)



Group 3: Sponge-associated reovirus-like RNA virus (SrRV)



Group 4: Sponge-associated picobirnavirus-like RNA virus (SpiRV)



Group 5: Sponge-associated RNA virus (SRV)

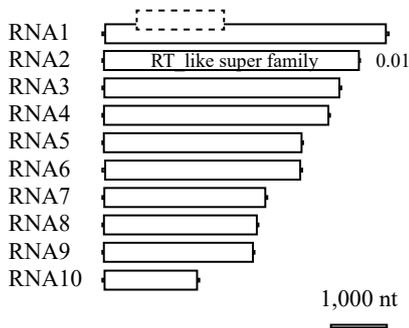


Fig. 3. Organization of reconstructed RNA viral or virus-like genomes. Values indicate the e-value with the viral top hit domain by a CD search.

>0.1% in non-rRNA reads in conventional RNA sequencing may be infected with the sampled organism (Shi *et al.*, 2016). Among the five major viral genomes, SdRV showed the highest read abundance, accounting for 2.7% of non-ribosomal reads in the ssRNA-seq library (Fig. 1C). The full-length sequence of SpaRV was obtained in 2014 and 2015 samples and only one base substitution was observed between sequences obtained from the two samplings. The read abundance in non-ribosomal reads was approximately 0.0025% in both ssRNA-seq libraries. SRV was also detected in 2014 and 2015 samples and occupied more than 0.008% in both ssRNA-seq libraries. Between the 2014 and 2015 SRV genomes, there were only 54 positions of a single-nucleotide variant and 90% (49/54) of them were identified as transitions. Therefore, SpaRV and SRV appeared to have infected the same or closely related organisms (host) within sponges obtained in both 2014 and 2015. In contrast, in both sponges, eukaryotes other than *Hymeniacidon* sponge did not dominate more than 0.0025% of the SSU rRNA composition, with the exception of *M. edulis* in 2015 samples, which shared 0.8% of SSU rRNA

reads. Thus, we concluded that *Hymeniacidon sp.* appeared to be a host of SdRV, SpaRV, and SRV. The read abundances of SpiRV and SrRV putative genomes were one to two orders of magnitude smaller than that of SdRV in ssRNA-seq data, and were only detected from either of the samples. Therefore, we were unable to define their host based on this definition. It is important to note that dsRNA viruses (SpaRV, SpiRV, and SrRV) showed lower read abundance than the ssRNA virus (SdRV) in ssRNA-seq data.

SdRV

Among reconstructed viral genomes, only SdRV presented significant similarity with the subject sequence Caledonia beadlet anemone dicistro-like virus 2 isolate I38 (accession no. MF189973, publication status: unpublished) (e-value=0.0, 71% identity) in the BLASTN analysis; its potential host Caledonia belongs to the phylum *Cnidaria*, not *Porifera*. The SdRV genome consists of one segment with two ORFs. The BLASTP search also revealed that the most similar sequence to SdRV RdRp was putative ORF1 of Caledonia beadlet anemone dicistro-like virus 2 isolate I38 (accession no. ASM93984) (e-value=0.0, 74% identity). A phylogenetic analysis of the RdRp sequence indicated that SdRV belonged to the positive-stranded ssRNA virus family *Dicistroviridae* (Fig. 4). The arrangement of ORFs on the SdRV genome was similar to those on known viruses in *Dicistroviridae* (Fig. S5). These results suggest that SdRV is a new ssRNA virus that belongs to *Dicistroviridae*.

SpaRV

In the BLASTP analysis for RdRp in the SpaRV genome, the most similar sequence in the nr database was Hubei partiti-like virus 53 (related to viruses in *Partitiviridae*) (e-value= 3×10^{-43} , 32% identity) found in the ssRNA-seq of invertebrates (Shi *et al.*, 2016). The genome of SpaRV consists of two segments with one ORF each and the genome structure is conserved in known partitiviruses. Although the RdRp sequence showed significant similarity (e-value $\leq 1 \times 10^{-5}$) with sequences in *Partitiviridae*, SpaRV was not classified into the established genus of *Partitiviridae* by the phylogenetic analysis (Fig. 5) (Nibert *et al.*, 2014). Therefore, SpaRV represents a novel genus in the family *Partitiviridae*. Viruses in *Partitiviridae* have been isolated from culturable organisms, such as plants, fungi, and protozoa (Nibert *et al.*, 2014). In addition, their relatives have been detected in both vertebrates and invertebrates in transcriptomic analyses (Liu *et al.*, 2012; Shi *et al.*, 2016); however, no partitivirus has been isolated from animals (Nibert *et al.*, 2014).

SrRV, SpiRV, and SRV

The sequence that was the most similar to RdRp in SrRV in the BLASTP analysis was Chuzan virus (genus *Orbivirus*) (e-value= 2×10^{-14} , 25% identity); however, a phylogenetic analysis of RdRp for *Reoviridae* revealed that SrRV was not positioned into the genus *Orbivirus* (Fig. 6). In addition, in the BLASTX analysis, an ORF of segment 3 showed significant similarity with VP3 (Orbivirus VP4 core protein) of Wad Medani virus (genus *Orbivirus*) (e-value= 2×10^{-14} , 25% identity), while other segments showed

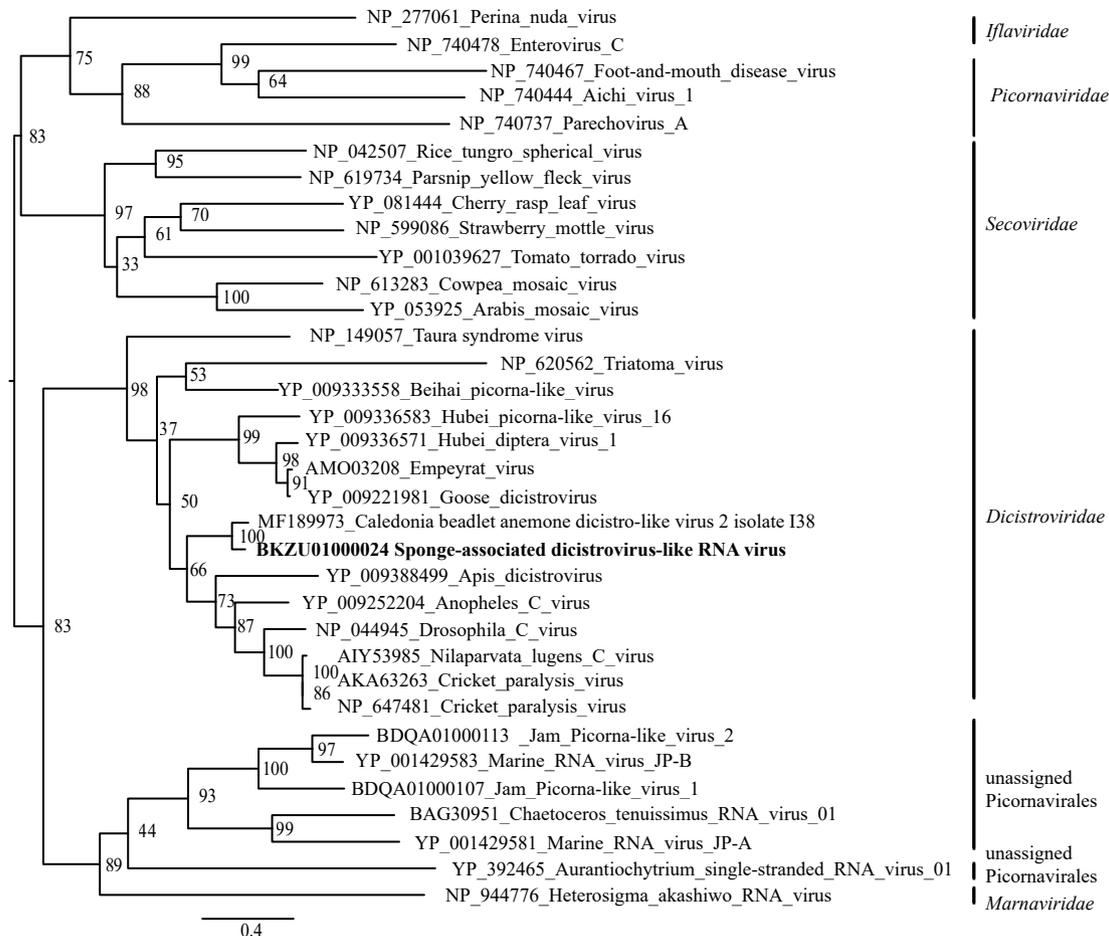


Fig. 4. Maximum likelihood tree of aligned RdRp amino acid sequences from SdRV and representative members of families *Dicistroviridae*, *Secoviridae*, *Picornaviridae*, and *Iflaviridae* using 305 residues. Numbers indicate bootstrap values (%) based on 1,000 RAXML calculations. The best-fitting amino acid substitution model was [rtREV+F+G].

no significant similarity ($e\text{-value} \leq 1 \times 10^{-5}$) with known viral sequences.

The most significantly similar sequence in the BLASTP analysis for RdRp in SpiRV was Hubei picobirna-like virus 2. A phylogenetic analysis of RdRp suggested that SpiRV belongs to an undefined and tentatively named α clade of picobirnavirus-like viruses derived from marine microorganisms, diatoms, and invertebrates (Urayama *et al.*, 2018a) (Fig. 7). In *Picobirnaviridae*, one sequence that originates from a sponge named Barns Ness breadcrumb sponge picobirna-like virus 1 (accession no. MF190026.1) was found in the nr database, but belongs to the β clade.

SRV was composed of 10 RNA segments with 35,412 nt that showed no significant similarity ($e\text{-value} \leq 1 \times 10^{-5}$) with sequences in the nr database in BLASTN and BLASTX analyses. The RdRp sequence was too distant from all known RdRp sequences to identify its phylogenetic position. However, as discussed above, the genome structure of SRV resembled that of *Reoviridae* and the ratio of SRV reads in non-rRNA ssRNA-seq was similar to those of dsRNA viruses (Fig. 1B and C). When considering the maximum segment number of known ssRNA viruses (viruses in *Orthomyxoviridae* have ~ 8 segments), SRV may be a reo-like novel dsRNA virus. To the best of our knowledge, the genome size of SRV is the largest among known RNA viruses.

Other viral segments and contigs

Although we focused on numerically abundant viruses with putative complete genome sequences in the dsRNA-seq libraries, there was a RdRp-encoding contig with similar average coverage ($1,839\times$) to these five genomes. The contig (ID: Sponge_RNAV429) was identified in 2014 samples, and showed significant similarity ($e\text{-value} = 4 \times 10^{-14}$, 24% identity) to known, but unclassified RdRp identified from the chloroplasts of *B. cinicola*. Other RdRp-encoding segments and contigs were lower than $1,000\times$ average coverage.

In the present study, 788 partial viral genome sequences were identified (Table S1). To clarify the phylogenetic distribution of these viruses, maximum likelihood trees for each established RNA virus family were calculated based on RdRp sequences encoded in the contigs of partial or putative complete genome segments (Fig. S6). Consistent with the similarity values obtained by the BLASTX analysis (Table S1), most viruses were phylogenetically distinct from known RNA viral genera, and RNA viruses related to the families *Partitiviridae*, *Reoviridae*, and *Totiviridae* were very diverse.

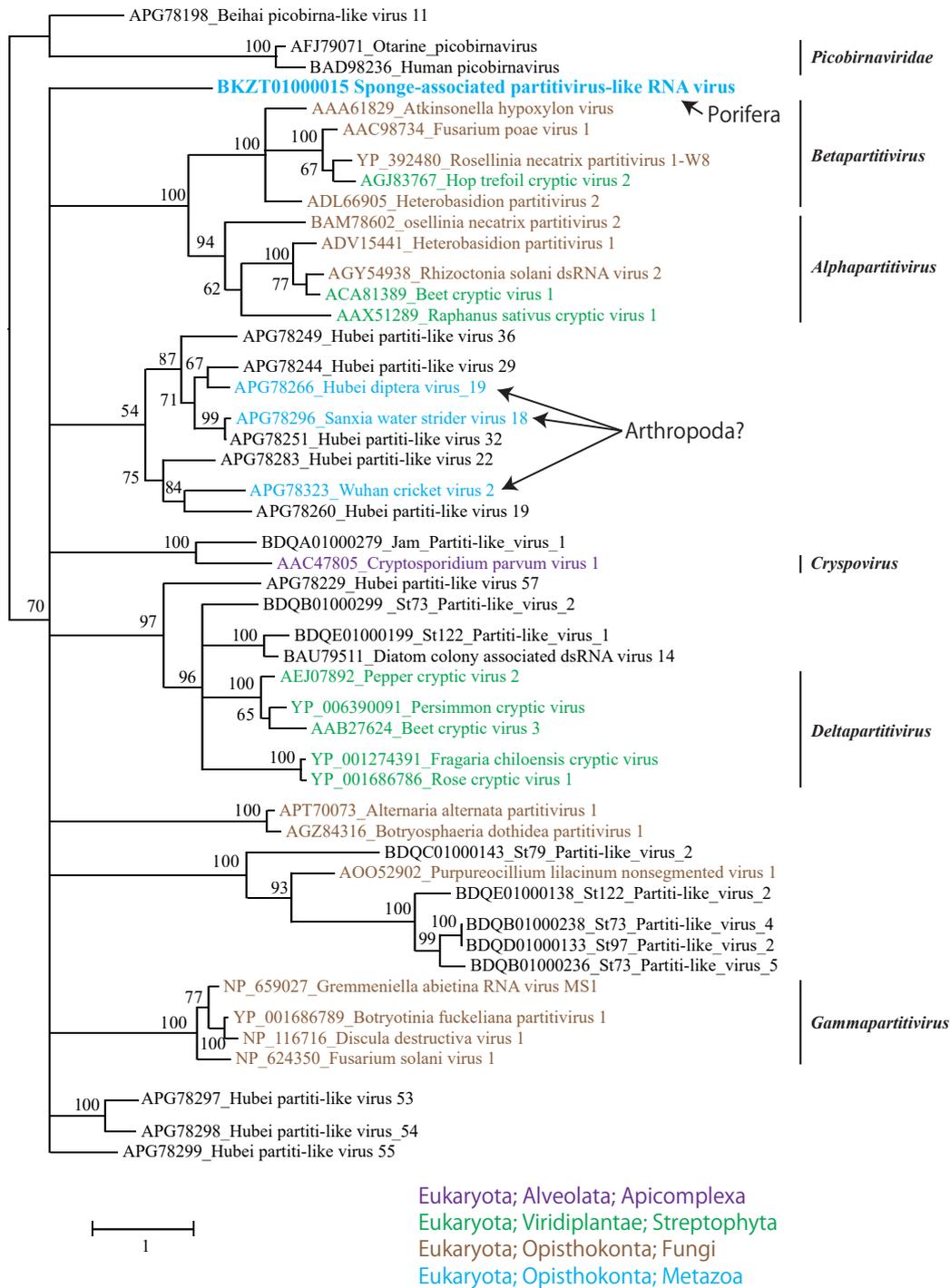


Fig. 5. Maximum likelihood tree of aligned RdRp amino acid sequences from SpaRV and representative members of families *Picobirnaviridae* and *Partitiviridae* using 302 residues. Numbers indicate bootstrap values (%) based on 1,000 RAXML calculations. The best-fitting amino acid substitution model was [rtREV+F+G]. Colors indicate the classification of the host organism.

Discussion

Recent RNA virome analyses suggest the extent of gene module shuffling among diverse viral genomes (Dolja and Koonin, 2017). Thus, high-quality information on RNA viral genomes is essential for understanding the diversity and evolution of RNA viruses. However, the conventional metagenomic ssRNA-seq approach has a limitation in

retrieving entire genomic structures because of the absence of the terminal sequences of each genome segment (Ozsolak and Milos, 2011). Most contigs encoding viral genes, except for RdRp genes, must be overlooked, except when the genome sequence presents significant similarity to previously reported viral genomes. For example, among the five possible RNA viral genomes from the two sponges examined in the present study, it was not possible to identify >50% of segments as RNA viral genome segments based on

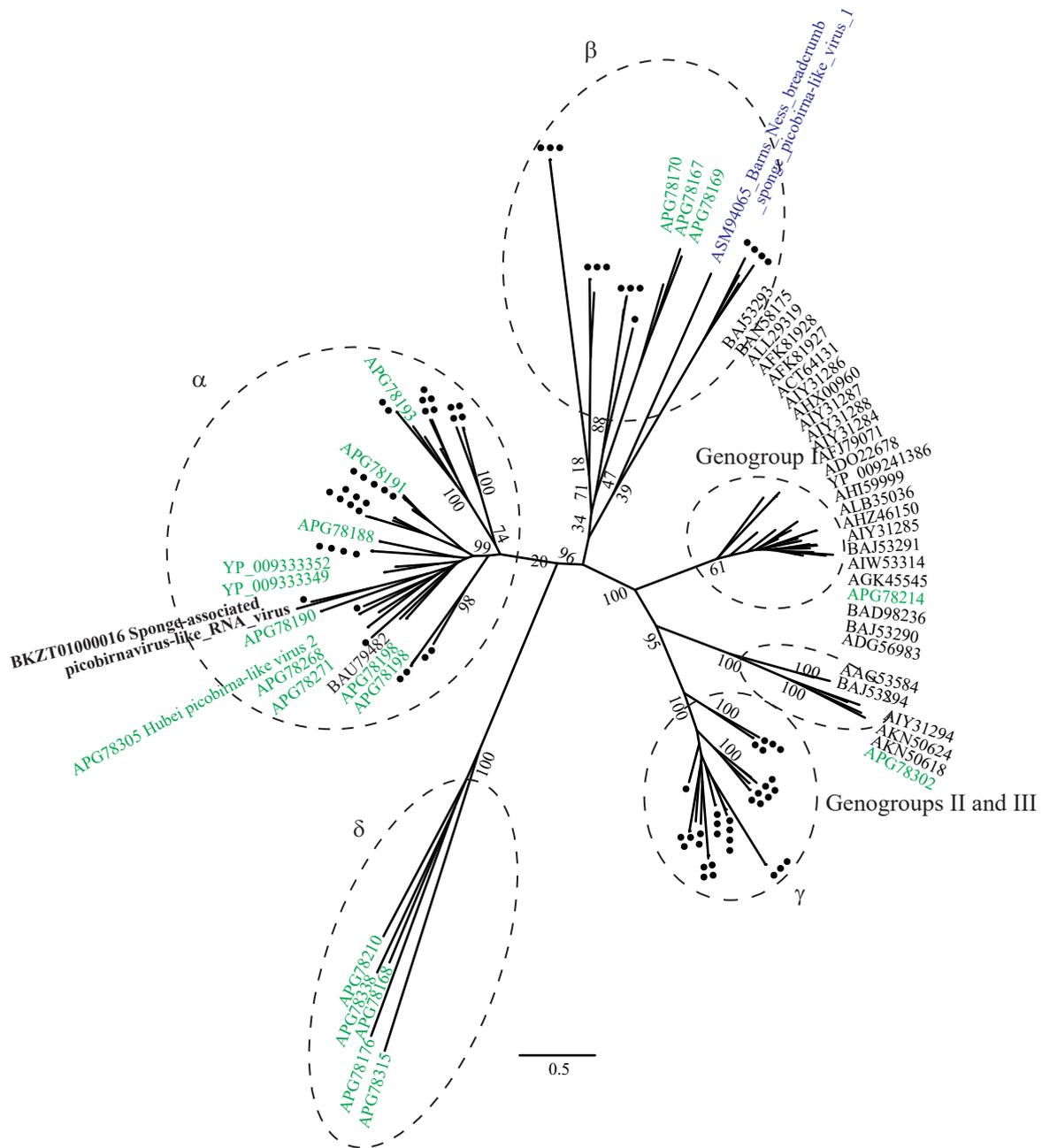


Fig. 7. Maximum likelihood tree of aligned RdRp amino acid sequences from SpiRV and representative members of the family *Picobirnaviridae* using 310 residues. Numbers indicate bootstrap values (%) based on 1,000 RAXML calculations. The best-fitting amino acid substitution model was [LG+F+G]. Black dots and green color represent OTUs derived from Urayama *et al.* (2018b) and Shi *et al.* (2016), respectively. Purple color indicates unpublished deposited sequences.

signed) with polyA tails were retrieved by FLDS (Urayama *et al.*, 2018a). The variance of the 3' terminal modification with a polyA tail among ssRNA viral genomes retrieved by FLDS implies variations in replication processes among ssRNA viruses.

In the present study, we assessed sponge RNA viromes using dsRNA- and ssRNA-seq and identified 167 RdRp genes, which represent RNA viral OTUs. The present results revealed that RNA viruses in sponge holobionts are highly diverse, similar to DNA viruses and prokaryotic populations (Taylor *et al.*, 2004; Laffy *et al.*, 2018). The present study suggests the importance of RNA viruses for the homeostasis and evolution of sponge holobiont systems,

while RNA viromes have been overlooked in holobiont systems. A high-quality list of sponge-associated RNA viruses will be a fundamental knowledge base for understanding the activity and ecological distribution of (sponge-associated) RNA viruses.

Acknowledgements

This research was supported in part by Grants-in-Aid for Scientific Research (Nos. 26892031 and 18H05368) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan and Grants-in-Aid for Scientific Research on Innovative Areas from MEXT of Japan (Nos. 16H06429, 16K21723, and 16H06437).

References

- Butina, T., Potapov, S., Belykh, O., and Belikov, S.I. (2015) Genetic diversity of cyanophages of the Myoviridae family as a constituent of the associated community of the Baikal sponge *Lubomirskia baicalensis*. *Genetika (Moscow)* **51**: 384–388.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- Chojnacki, S., Cowley, A., Lee, J., Foix, A., and Lopez, R. (2017) Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res* **45**: W550–W553.
- Culley, A.I., Mueller, J.A., Belcaid, M., Wood-Charlson, E.M., Poisson, G., and Steward, G.F. (2014) The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *mBio* **5**: e01210–01214.
- Decker, C.J., Steiner, H.R., Hoon-Hanks, L.L., Morrison, J.H., Haist, K.C., Stabell, A.C., *et al.* (2019) dsRNA-Seq: Identification of viral infection by purifying and sequencing dsRNA. *Viruses* **11**: pii: E943.
- Djikeng, A., Kuzmickas, R., Anderson, N.G., and Spiro, D.J. (2009) Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One* **4**: e7264.
- Dolja, V.V., and Koonin, E.V. (2017) Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res* **244**: 36–52.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Engelhardt, T., Orsi, W.D., and Jørgensen, B.B. (2015) Viral activities and life cycles in deep seafloor sediments. *Environ Microbiol Rep* **7**: 868–873.
- Fukasawa, F., Hirai, M., Takaki, Y., Shimane, Y., Thomas, C.E., Urayama, S., *et al.* (2019) A new polycipivirus identified in *Colobopsis shohki*. *Arch Virol* doi: 10.1007/s00705-019-04510-8.
- Higashiura, T., Katoh, Y., Urayama, S., Hayashi, O., Aihara, M., Fukuhara, T., *et al.* (2019) Magnaporthe oryzae chrysovirus 1 strain D confers growth inhibition to the host fungus and exhibits multifunctional viral structural proteins. *Virology* **535**: 241–254.
- King, A.M.Q., Adams, M.J., Carstens, E.B., and Lefkowitz, E.J. (2012) *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier.
- Koga, R., Horiuchi, H., and Fukuhara, T. (2003) Double-stranded RNA replicons associated with chloroplasts of a green alga, *Bryopsis cinicola*. *Plant Mol Biol* **51**: 991–999.
- Kopylova, E., Noé, L., and Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.
- Koyama, S., Sakai, C., Thomas, C.E., Nunoura, T., and Urayama, S. (2016) A new member of the family Totiviridae associated with arboreal ants (*Camponotus nipponicus*). *Arch Virol* **161**: 2043–2045.
- Laffy, P.W., Wood-Charlson, E.M., Turaev, D., Weyenberg, K.D., Botté, E.S., van Oppen, M.J., *et al.* (2016) HoloVir: a workflow for investigating the diversity and function of viruses in invertebrate holobionts. *Front Microbiol* **7**: 822.
- Laffy, P.W., Wood-Charlson, E.M., Turaev, D., Jutz, S., Pascelli, C., Botté, E.S., *et al.* (2018) Reef invertebrate viromics: diversity, host specificity and functional capacity. *Environ Microbiol* **20**: 2125–2141.
- Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S.A., Li, G., *et al.* (2012) Discovery of novel dsRNA viral sequences by in silico cloning and implications for viral diversity, host range and evolution. *PLoS One* **7**: e42147.
- Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., and Banfield, J.F. (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **12**: R44.
- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., and Marshall, D. (2010) Tablet—next generation sequence assembly visualization. *Bioinformatics* **26**: 401–402.
- Morris, T.J., and Dodds, J.A. (1979) Isolation and analysis of double-stranded-rna from virus-infected plant and fungal tissue. *Phytopathology* **69**: 854–858.
- Nibert, M.L., Ghabrial, S.A., Maiss, E., Lesker, T., Vainio, E.J., Jiang, D., and Suzuki, N. (2014) Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. *Virus Res* **188**: 128–141.
- Ozsolak, F., and Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* **12**: 87–98.
- Pascelli, C., Laffy, P.W., Kupresanin, M., Ravasi, T., and Webster, N.S. (2018) Morphological characterization of virus-like particles in coral reef sponges. *PeerJ* **6**: e5625.
- Pita, L., Rix, L., Slaby, B.M., Franke, A., and Hentschel, U. (2018) The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome* **6**: 46.
- Rambaut, A. (2014) FigTree 1.4. 2 software. Institute of Evolutionary Biology, Univ. Edinburgh. URL <http://tree.bio.ed.ac.uk/software/figtree/>.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.
- Roosinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., *et al.* (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol Ecol* **19** Suppl 1: 81–88.
- Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., *et al.* (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.
- Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., *et al.* (2016) Redefining the invertebrate RNA virosphere. *Nature* **540**: 539–543.
- Shi, M., Lin, X.D., Chen, X., Tian, J.H., Chen, L.J., Li, K., *et al.* (2018) The evolutionary history of vertebrate RNA viruses. *Nature* **556**: 197–202.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Steward, G.F., Culley, A.I., Mueller, J.A., Wood-Charlson, E.M., Belcaid, M., and Poisson, G. (2013) Are we missing half of the viruses in the ocean? *ISME J* **7**: 672–679.
- Stöver, B.C., and Müller, K.F. (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**: 7.
- Sugiura, N. (1978) Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. *Commun Stat Theory Methods* **7**: 13–26.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.
- Tanabe, A.S. (2011) Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol Ecol Resour* **11**: 914–921.
- Taylor, M.W., Schupp, P.J., Dahllöf, I., Kjelleberg, S., and Steinberg, P.D. (2004) Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environ Microbiol* **6**: 121–130.
- Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71**: 295–347.
- Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C., *et al.* (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* **7**: 11870.
- Urayama, S., Yoshida-Takashima, Y., Yoshida, M., Tomaru, Y., Moriyama, H., Takai, K., and Nunoura, T. (2015) A new Fractionation and recovery method of viral genomes based on nucleic acid composition and structure using tandem column chromatography. *Microbes Environ* **30**: 199–203.
- Urayama, S., Takaki, Y., and Nunoura, T. (2016) FLDS: A comprehensive dsRNA sequencing method for intracellular RNA virus surveillance. *Microbes Environ* **31**: 33–40.
- Urayama, S., Takaki, Y., Nishi, S., Yoshida-Takashima, Y., Deguchi, S., Takai, K., and Nunoura, T. (2018a) Unveiling the RNA virosphere associated with marine microorganisms. *Mol Ecol Resour* **18**: 1444–1455.

- Urayama, S., Takaki, Y., Nunoura, T., and Miyamoto, N. (2018b) Complete genome sequence of a novel RNA virus identified from a deep-sea animal, *osedax japonicus*. *Microbes Environ* **33**: 446–449.
- Vacelet, J., and Gallissian, M.-F. (1978) Virus-like particles in cells of the sponge *Verongia cavernicola* (Demospongiae, Dictyoceratida) and accompanying tissues changes. *J Invertebr Pathol* **31**: 246–254.
- Venkataraman, S., Prasad, B., and Selvarajan, R. (2018) RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses* **10**.
- Waldron, F.M., Stone, G.N., and Obbard, D.J. (2018) Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLoS Genet* **14**: e1007533.
- Webster, N.S., and Taylor, M.W. (2012) Marine sponges and their microbial symbionts: love and other relationships. *Environ Microbiol* **14**: 335–346.
- Webster, N.S., and Thomas, T. (2016) The sponge hologenome. *mBio* **7**: e00135-00116.
- York, A., and Fodor, E. (2013) Biogenesis, assembly, and export of viral messenger ribonucleoproteins in the influenza A virus infected cell. *RNA Biol* **10**: 1274–1282.