




## Unveiling the RNA virosphere associated with marine microorganisms

著者 (英)	Shunichi URAYAMA, Yoshihiro Takaki, Shinro Nishi, Yukari Yoshida-Takashima, Shigeru Deguchi, Ken Takai, Takuro Nunoura
journal or publication title	Molecular ecology resources
volume	18
number	6
page range	1444-1455
year	2018-09
権利	(C) 2018 The Authors. Molecular Ecology Resources Published by John Wiley & Sons Ltd This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
URL	<a href="http://hdl.handle.net/2241/00154145">http://hdl.handle.net/2241/00154145</a>

doi: 10.1111/1755-0998.12936



# Unveiling the RNA virosphere associated with marine microorganisms

Syun-ichi Urayama<sup>1,2</sup>  | Yoshihiro Takaki<sup>1,3,4</sup> | Shinro Nishi<sup>1,4</sup> |  
Yukari Yoshida-Takashima<sup>3</sup> | Shigeru Deguchi<sup>1</sup> | Ken Takai<sup>3</sup> | Takuro Nunoura<sup>1,4</sup>

<sup>1</sup>Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Kanagawa, Japan

<sup>2</sup>Laboratory of Fungal Interaction and Molecular Biology (donated by IFO), Department of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

<sup>3</sup>Department of Subsurface Geobiological Analysis and Research, JAMSTEC, Yokosuka, Kanagawa, Japan

<sup>4</sup>Ecosystem Observation and Evaluation Methodology Research Unit, Project Team for Development of New-generation Research Protocol for Submarine Resources, JAMSTEC, Japan

## Correspondence

Syun-ichi Urayama, Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Kanagawa, Japan. Laboratory of Fungal Interaction and Molecular Biology (donated by IFO), Department of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan.  
Email: urayama.shunichi.gn@u.tsukuba.ac.jp

## Funding information

Ministry of Education, Culture, Sports, Science and Technology, Grant/Award Number: 16H06429, 16H06437, 16K21723; Japan Society for the Promotion of Science, Grant/Award Number: JP26892031

## Abstract

The study of extracellular DNA viral particles in the ocean is currently one of the most advanced fields of research in viral metagenomic analysis. However, even though the intracellular viruses of marine microorganisms might be the major source of extracellular virus particles in the ocean, the diversity of these intracellular viruses is not well understood. Here, our newly developed method, referred to herein as fragmented and primer ligated dsRNA sequencing (FLDS) version 2, identified considerable genetic diversity of marine RNA viruses in cell fractions obtained from surface seawater. The RNA virus community appears to cover genome sequences related to more than half of the established positive-sense ssRNA and dsRNA virus families, in addition to a number of unidentified viral lineages, and such diversity had not been previously observed in floating viral particles. In this study, more dsRNA viral contigs were detected in host cells than in extracellular viral particles. This illustrates the magnitude of the previously unknown marine RNA virus population in cell fractions, which has only been partially assessed by cellular metatranscriptomics and not by contemporary viral metagenomic studies. These results reveal the importance of studying cell fractions to illuminate the full spectrum of viral diversity on Earth.

## KEYWORDS

dsRNA, marine microorganisms, metagenomics, RNA virus

## 1 | INTRODUCTION

Viruses are universal genetic elements associated with the three domains of life (Koonin, 2010) and impact the phenotype expression of hosts (Marquez, Redman, Rodriguez, & Roossinck, 2007; Suttle, 2005). Until the 1990s, pathogenic viruses were the major targets of study in virology (Roossinck, 2011), since the presence of a virus was

identified only when it caused recognizable symptoms in the host. However, pioneering studies in the 1990s showed that virus-like particles are present at approximately  $10^7$  particles per millilitre of seawater (Fuhrman, 1999). This finding suggested the significance of the extracellular virus community and opened a new era in the field of environmental virology. Advances in high-throughput sequencing technologies

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2018 The Authors. *Molecular Ecology Resources* Published by John Wiley & Sons Ltd

in the last decade have enabled us to identify the genetic diversity of environmental DNA virus particles (Breitbart et al., 2002; Edwards & Rohwer, 2005; Koch, 2016; Suttle, 2016) and have revealed that DNA viruses are the most abundant entities across all habitats and a major reservoir of genetic diversity (Paez-Espino et al., 2016).

Despite the development of RNA sequencing technologies, RNA viral diversity is not well understood. When evaluating the genetic diversity of viral particles from environments or host organisms (Culley, 2017; Culley et al., 2014; Culley, Lang, & Suttle, 2006; Djikeng, Kuzmickas, Anderson, & Spiro, 2009; Steward et al., 2013), RNA viruses that lack capsids and/or extracellular particles might be overlooked (King, Adams, Carstens, & Lefkowitz, 2012). In addition, although high-throughput cellular RNA sequencing has the ability to detect all RNA viruses, minor RNA populations including viral RNAs may be hidden by the much higher abundance of cellular RNAs (Shi et al., 2016; Zeigler Allen, McCrow, & Ininbergs, 2017). To resolve these issues, cellular double-stranded RNA (dsRNA) sequencing techniques have been developed to explore the diversity of nonretro RNA viruses in host cells (Decker & Parker, 2014; Roossinck et al., 2010), since long intracellular dsRNA populations consist of dsRNA viral genomes and replicative intermediates of nonretro single-stranded RNA (ssRNA) viruses (Morris & Dodds, 1979). However, there are still difficulties in retrieving the entire viral genome, especially the terminal regions, due to the performance limit of reverse transcriptase and/or principle of cDNA synthesis. Moreover, both classic RNA sequencing and dsRNA sequencing technologies are only effective for detecting sequences related to known viruses, and viral genome sequences without significant similarity to known viruses are undetectable.

To resolve these issues, we have developed a novel dsRNA-sequencing method named fragmented and primer ligated dsRNA sequencing version 1 (FLDS version 1). FLDS version 1 enables us to apply dsRNA sequencing to environmental dsRNA and to retrieve entire genome sequences including terminal regions (Koyama, Sakai, Thomas, Nunoura, & Urayama, 2016; Urayama, Takaki, & Nunoura, 2016). However, FLDS version 1 has several technical issues and cannot be applied to an RNA virosphere with high diversity and low amounts of dsRNA. Here, we further developed FLDS version 2 and applied this new technique to cellular RNA viromes associated with marine surface planktonic microbial communities from five pelagic and coastal sampling stations in the North Pacific. Moreover, identified dsRNA viral communities were compared with RNA viromes of surface seawater viral particles from the same sampling stations. This study illuminates the previously unknown marine RNA virus population in cell fractions and reveals the importance of studying cell fractions in order to unveil the full spectrum of RNA viral diversity on Earth.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection and purification of viral particles

The sample processing workflow is shown in Supporting information Figure S1. Surface seawater was collected from five stations in the

North Pacific during the JAMSTEC MR14-04 cruise; station (St.) 73, St. 79, St. 97, and St. 122 are located in subarctic pelagic regions, and St. Jam is a coastal site in front of JAMSTEC on Honshu, Japan (Supporting information Table S1). At each station, approximately 10 L of surface water was collected by a bucket, and each 3–4 L of seawater was filtered with 0.2- $\mu$ m-pore-size filters (cellulose acetate membrane, 47 mm diameter; Advantec, Tokyo, Japan). The filters were stored at  $-80^{\circ}\text{C}$  until nucleic acid extraction. Viral particles (VPs) in each flow through were concentrated by the iron chloride precipitation method (John et al., 2011), and iron precipitates were collected on 0.8- $\mu$ m-pore-size filters (polycarbonate membrane, 47 mm diameter; Advantec) and stored at  $4^{\circ}\text{C}$ . The precipitates were suspended in magnesium-EDTA-ascorbate buffer (0.1 M  $\text{Mg}_2\text{EDTA}$  and 0.2 M ascorbic acid, pH 6.0) at  $4^{\circ}\text{C}$ , and suspended VPs were ultracentrifuged at 165,000 g for 2 hr (Optima L-90K Ultracentrifuge and Type 45 Ti rotor; Beckman Coulter, Brea, CA, USA). The resultant precipitate was suspended in SM buffer (10 mM  $\text{MgSO}_4$  and 50 mM Tris-HCl, pH 7.5) containing 3% NaCl (w/v) at  $4^{\circ}\text{C}$ . VPs were further purified with CsCl density equilibrium centrifugation at 274,000 g for 48 hr (Optima L-90 K Ultracentrifuge and Type SW41 Ti rotor; Beckman Coulter). After ultracentrifugation, each 0.5-ml fraction was collected from top to bottom of the tube, and the density of each fraction was measured using an analytical balance. Fractions containing VPs with densities ranging 1.30–1.48 g/cm<sup>3</sup> (Steward et al., 2013) were pooled in a tube and diluted with SM buffer containing 3% NaCl (w/v). The solution was subjected to ultracentrifugation (165,000 g for 2 hr, Optima L-90K Ultracentrifuge and Type 70 Ti rotor; Beckman Coulter), and the pellets were used for further analyses.

### 2.2 | Nucleic acid extraction and RNA purification

Cells collected on a portion of the 0.2- $\mu$ m-pore-size filters in approximately 2 L of seawater were pulverized by a motor in liquid nitrogen and suspended in extraction buffer [20 mM Tris-HCl, pH 6.8, 200 mM NaCl, 2 mM EDTA, 1% SDS and 0.1% (v/v)  $\beta$ -mercaptoethanol]. Viral pellets obtained by ultracentrifugation were also suspended in extraction buffer. After conventional phenol/chloroform extraction of total nucleic acids, dsRNA and ssRNA were fractionated using the cellulose column chromatography method (Urayama et al., 2015). The dsRNA fraction was again purified through a microspin column containing cellulose powder. To remove remaining DNA and ssRNA, the eluted dsRNA was further treated with amplification grade DNase I (Invitrogen, Carlsbad, CA, USA) and S1 nuclease (Invitrogen) in nuclease buffer (62 mM  $\text{CH}_3\text{COONa}$ , 10 mM  $\text{MgCl}_2$ , 2 mM  $\text{ZnSO}_4$  and 209 mM NaCl) at  $37^{\circ}\text{C}$  for 2 hr.

The ssRNA fraction was also further purified by using the TRIzol Plus RNA Purification Kit (Invitrogen) according to the manufacturer's protocol. The ssRNA fraction was treated with DNase I (Invitrogen). Ribosomal RNA in ssRNA samples was depleted using the Ribo-Zero rRNA Removal Kit for bacteria (Illumina, San Diego, CA, USA) according to the manufacturer's protocol.

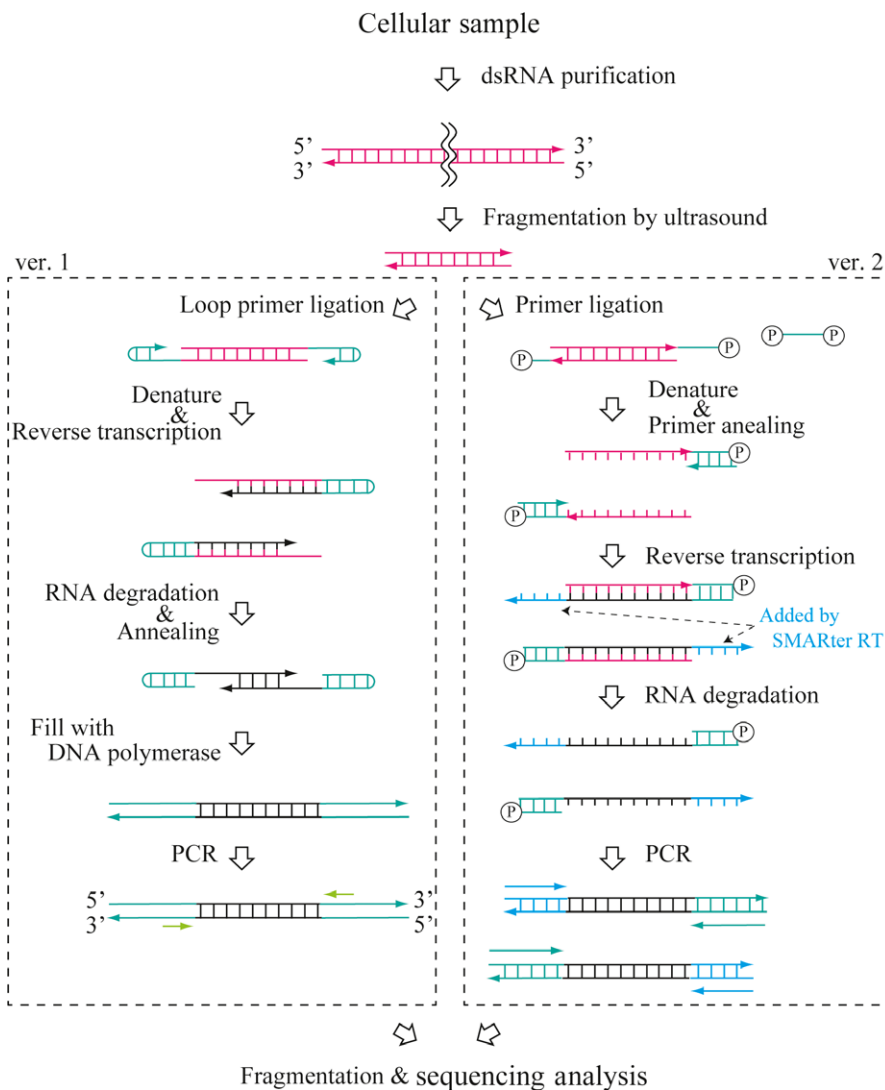
To determine the diversity of potential host microorganisms, total ssRNA was also extracted from a portion of the pulverized cell

sample described above using TRIzol Reagent (Life Technologies, Carlsbad, CA, USA) and the TRIzol Plus RNA Purification Kit (Invitrogen) according to the manufacturer's protocol. The total ssRNA fraction was treated with DNase I (Invitrogen).

### 2.3 | cDNA synthesis and amplification

ssRNAs were reverse-transcribed using the SMARTer Universal Low Input RNA Kit (Takara Bio, Kusatsu, Japan) according to the manufacturer's protocol. cDNA derived from fragmented dsRNA was synthesized as described previously with a few modifications (Okada, Kiyota, Moriyama, Fukuhara, & Natsuaki, 2015; Urayama et al., 2016, 2015) (Figure 1). Briefly, DNase/S1 nuclease-treated dsRNAs were fragmented by ultrasound using a Covaris S220 ultrasonicator (Woburn, MA, USA). The fragmentation conditions were as follows: run time 35 s, peak power 140.0 W, duty factor 2.0% and 200 cycles/burst. dsRNAs were purified using a Zymo Clean Gel RNA Recovery Kit (Zymo Research, Orange, CA, USA). U2 primer (5'-p-GAC GTA AGA ACG TCG CAC CA-p-3') designed in this study was

ligated to fragmented dsRNA in 50 mM HEPES/NaOH, pH 8.0, 18 mM MgCl<sub>2</sub>, 0.01% BSA, 1 mM ATP, 3 mM DTT, 10% DMSO, 20% polyethylene glycol 6,000, and 30 U T4 RNA ligase (Takara Bio) in a final volume of 30 µl at 37°C for 16 hr. The products were purified using a MinElute Gel Extraction Kit (Qiagen, Valencia, CA, USA). U2-comp primer (5'-OH-TGG TGC GAC GTT CTT ACG TC-OH-3'), which is the complementary sequence of U2 primer, was added to the elute. After denaturation at 95°C for 3 min and subsequent quenching in ice-water slurry, cDNA was synthesized using the SMARTer RACE 5'/3' Kit (Takara). After excess and hybrid RNA were digested with RNase H (Takara), cDNA was amplified with U2-comp and UPM (provided by the SMARTer RACE 5'/3' Kit) primers under the following conditions: 96°C for 2 min and 30–35 cycles of 98°C for 10 s, 60°C for 15 s, and 68°C for 2 min. Small cDNA and primer dimers were removed using the 1.25 × SPRIselect Reagent Kit (Beckman Coulter) according to the Left Side Size Selection procedure in the manufacturer's protocol. For comparative analysis, FLDS version 1 was performed as described by Urayama *et al.* (Urayama et al., 2016).



**FIGURE 1** Schematic workflow of FLDS version 1 (left side) and version 2 (right side). Details of the FLDS version 2 method are described in the Materials and Methods section [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## 2.4 | cDNA sequencing

cDNAs were fragmented by an ultrasonicator (Covaris S220). The fragmentation conditions were as follows: run time 55 s, peak power 175.0 W, duty factor 5.0% and 200 cycles/burst. The Illumina library was constructed with KAPA Hyper Prep Kit Illumina platforms (Kapa Biosystems, Woburn, MA, USA). The quantity of the library was evaluated using the Agilent 2100 bioanalyser (Agilent Technologies, Palo Alto, CA, USA) with a High-Sensitivity DNA chip and the KAPA library quantification kit (Kapa Biosystems). Approximately 300 bp of the paired-end sequences of each fragment was determined by the Illumina MiSeq platform (Illumina).

## 2.5 | Data assembly and processing

Adaptor and low-quality sequences were trimmed with TRIMMOMATIC version 0.32 (Bolger, Lohse, & Usadel, 2014). Primer sequences used for cDNA synthesis and amplification were trimmed with CUTADAPT version 1.10 (Martin, 2011). PhiX sequences derived from control libraries and experimentally contaminated sequences were also removed using a mapping tool, BOWTIE2 version 2.2.9 (Langmead & Salzberg, 2012). Reads shorter than 50 bases were removed with TRIMMOMATIC version 0.32. Low complexity reads were detected and removed with PRINSEQ version 0.20.4 (Schmieder & Edwards, 2011) using the DUST approach.

For RNA virome analyses, rRNA reads identified by SortMeRNA (Kopylova, Noé, & Touzet, 2012) were excluded from the resultant reads. To generate contig sequences for each station, the remaining reads of four types of metatranscriptomes were assembled together and subjected to de novo assembly using the CLC GENOMICS WORKBENCH version 9.0 (CLC Bio, Aarhus, Denmark) with the following parameters: a minimum contig length of 500, word value set to 20 and bubble size set to 500. Contigs with at least 3× average coverage were used for further analyses to reduce ambiguity in assembly. To obtain longer contigs, the ends of contigs described above were extended using PRICE version 1.2 (Ruby, Bellare, & Derisi, 2013) with paired reads. Processed reads were mapped on these extended contigs with the CLC GENOMICS WORKBENCH version 9.0 using the following parameters: mismatch cost of 2, insertion/deletion cost of 3, length fraction of 0.8 and similarity fraction of 0.8. To confirm the terminal regions of each contig, mapping data were manually examined (Urayama et al., 2016) using the TABLET version 1.14.10.20 (Milne et al., 2010). Contigs for which both ends were determined to be termini were identified as full-length genome segments. RNA viral genes in reference sequences were identified based on BLASTX (Camacho et al., 2009) analysis of protein sequences against the GenBank nonredundant (nr) database with an  $e$ -value  $\leq 1 \times 10^{-5}$ . An RNA viral contig was defined as one that encodes a coding sequence (CDS) with significant similarity to a known RNA viral protein sequence. Unmapped reads on reference sequences were compared to the NCBI protein database using DIAMOND BLASTX (Buchfink, Xie, & Huson, 2015) with an  $e$ -value  $\leq 1 \times 10^{-5}$ , and their origins were classified with MEGAN version 6.6.1 (Huson et al., 2016). In this analysis, reads and contigs

only related to environmental viral sequences, except for complete genome sequences, were classified into the category “other.” CDSs of (putative) RNA viral genome segments and RNA viral contigs were predicted with GENEMARKS (Besemer, Lomsadze, & Borodovsky, 2001).

## 2.6 | Overview of RNA viromes

Sequences that matched a known RNA-dependent RNA polymerase (RdRp) gene by BLAST with an  $e$ -value  $\leq 1 \times 10^{-5}$  were collected from RNA virus contigs and RNA virus segments. When a sequence had significant similarity to an RNA viral polyprotein, the presence of the RdRp domain was confirmed using Pfam (Finn et al., 2016) with an  $e$ -value  $\leq 1 \times 10^{-5}$ . Sequences encoding the RdRp gene were clustered at 90% identity using UCLUST (Edgar, 2010). The cluster's centroid sequences were selected as representative sequences and manually confirmed to cover more than 90% of other sequences in the cluster. Sequence coverage of representatives in each sample was estimated by mapping the reads to representative sequences using the short-read mapper BBmap (Bushnell, 2016) with a minimum identity of 90%. Subsequently, SAMTOOLS (Li et al., 2009) and BEDTOOLS (Quinlan & Hall, 2010) were used to calculate average sequence-wise coverage. To represent RdRp-based diversity, dsRNA viral operational taxonomic units (OTUs) were defined as representative contig sequences with more than 3× average coverage and with significant similarity to an RdRp sequence found in a known dsRNA virus. The CYTOSCAPE 2.8 program (Smoot, Ono, Ruscheinski, Wang, & Ideker, 2011) was used to compare the composition of RNA viromes among stations.

## 2.7 | Phylogenetic analysis

A subset of RdRp sequences including at least four of the seven conserved regions was used for phylogenetic analyses. Multiple alignments based on the deduced amino acid sequences of putative RdRp genes in the contigs were constructed by MUSCLE (Edgar, 2004) in MEGA5 (Tamura et al., 2011). The alignment was further trimmed by trimAl (option: -gt 1) (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) to exclude ambiguous positions. The best-fitting model of amino acid substitutions was tested in AMINOSAN (Tanabe, 2011) and judged by the corrected Akaike information criterion (Sugiura, 1978). Phylogenetic analyses were conducted using RAXML version 8.2.7 (Stamatakis, 2014) with a selected model and visualized using FIGTREE version 1.4.2 (Rambaut, 2014). Bayesian analyses with the covarion parameter were performed with one run and four chains for 1,000,000 generations using MRBAYES 3.2.3 (Ronquist & Huelsenbeck, 2003).

# 3 | RESULTS

## 3.1 | Development of FLDS version 2

We developed FLDS version 2 for conditions with lower amounts of template dsRNA and for RNA viral communities with high diversity.

In the FLDS version 1 protocol, inefficient annealing is expected for RNA viromes with high diversity, and thus, the method may be inadequate for library construction from a low amount of dsRNA (Figure 1). In addition, chimeric sequences can form in the annealing step among similar sequences using FLDS version 1. Accordingly, a reverse transcription template switching reaction was applied in FLDS version 2 (Scotto-Lavino, Du, & Frohman, 2006) instead of the annealing process of FLDS version 1 (Urayama et al., 2016) to achieve efficient cDNA synthesis and amplification and to reduce chimeric sequence formation. Both 3'- and 5'-ends of phosphorylated primers were also used in FLDS version 2 instead of loop primers to prevent concatemer formation in the primer ligation reaction. In the test analysis of FLDS version 2 using a diatom colony sample (Urayama et al., 2016), approximately 40 times more cDNA was synthesized than when using FLDS version 1 under the same conditions, and the entire range of RNA viral genome segments including both terminal regions was obtained (Supporting information Figure S2 and Supporting information Table S2). This result indicates the suitability of version 2 for lower amounts of dsRNA.

### 3.2 | Sample collection and diversity of cellular rRNA

To determine the coverage and diversity of the cellular RNA virome among the entire surface seawater RNA virome, North Pacific surface seawater was collected from four subarctic pelagic stations (St. 73, St. 79, St. 97, and St. 122) and one coastal station (St. Jam) (Supporting information Figure S3 and Supporting information Table S1), and cellular and cell-free VP fractions were obtained.

To reveal the composition of metabolically active potential host organisms of the RNA viromes, shotgun RNA-seq for total cellular RNA was conducted (Leininger et al., 2006). As a result, 25.4%–31.6% of trimmed reads were identified as parts of small subunit ribosomal RNA (SSU rRNA). Eukaryotic SSU rRNA reads predominated in all metatranscriptomes, and Alveolata, Stramenopiles, Holozoa and Prymnesiophyceae were detected as dominant rRNA groups (Supporting information Figure S4). In addition, Nucleotmycea and Rhizaria also dominated at St. Jam. The abundance of bacterial and archaeal SSU rRNA reads ranged from 17.0% to 32.3% of all SSU rRNA gene reads. In the bacterial population, Cyanobacteria and Proteobacteria were observed as the dominant rRNA groups.

### 3.3 | Abundance of RNA viral reads in metatranscriptomes

The compositions of each metatranscriptome are summarized in Supporting information Table S3. In the case of cellular dsRNA metatranscriptomes, reads associated with RNA viral reads (i.e., reads with significant similarity [ $e\text{-value} \leq 1 \times 10^{-5}$ ] to known RNA viral protein sequences), RNA viral contigs, and (putative) RNA viral genome segments occupied 11.3%–36.6% of the total reads. The definitions of (putative) RNA viral genome segments are summarized below. Read abundance of rRNA and other cellular functions ranged

from 23.1% to 33.3% of the total reads. The abundance of RNA virus reads in cellular ssRNA (rRNA-depleted) metatranscriptomes was only approximately 0.1% of the total reads in cellular ssRNA (rRNA-depleted) libraries.

The abundance of reads with cellular functions in VP ssRNA metatranscriptomes, most of which were identified as rRNA, was 3.4% in the St. Jam sample and 97.2%–98.8% in the four subarctic pelagic station samples. Similar trends were also observed in VP dsRNA metatranscriptomes. The abundance of reads with cellular functions in VP dsRNA metatranscriptomes was 3.8% in the St. Jam sample and 63.0%–82.3% in the four subarctic pelagic station samples. The ratio of RNA viral sequences was markedly increased in VP dsRNA metatranscriptomes compared with VP ssRNA metatranscriptomes (Supporting information Table S3), although in principle, ssRNA virus cannot be detected in VP dsRNA metatranscriptomes. The major SSU rRNA groups in VP ssRNA/dsRNA metatranscriptomes of the four subarctic pelagic station samples are shown in Supporting information Table S4.

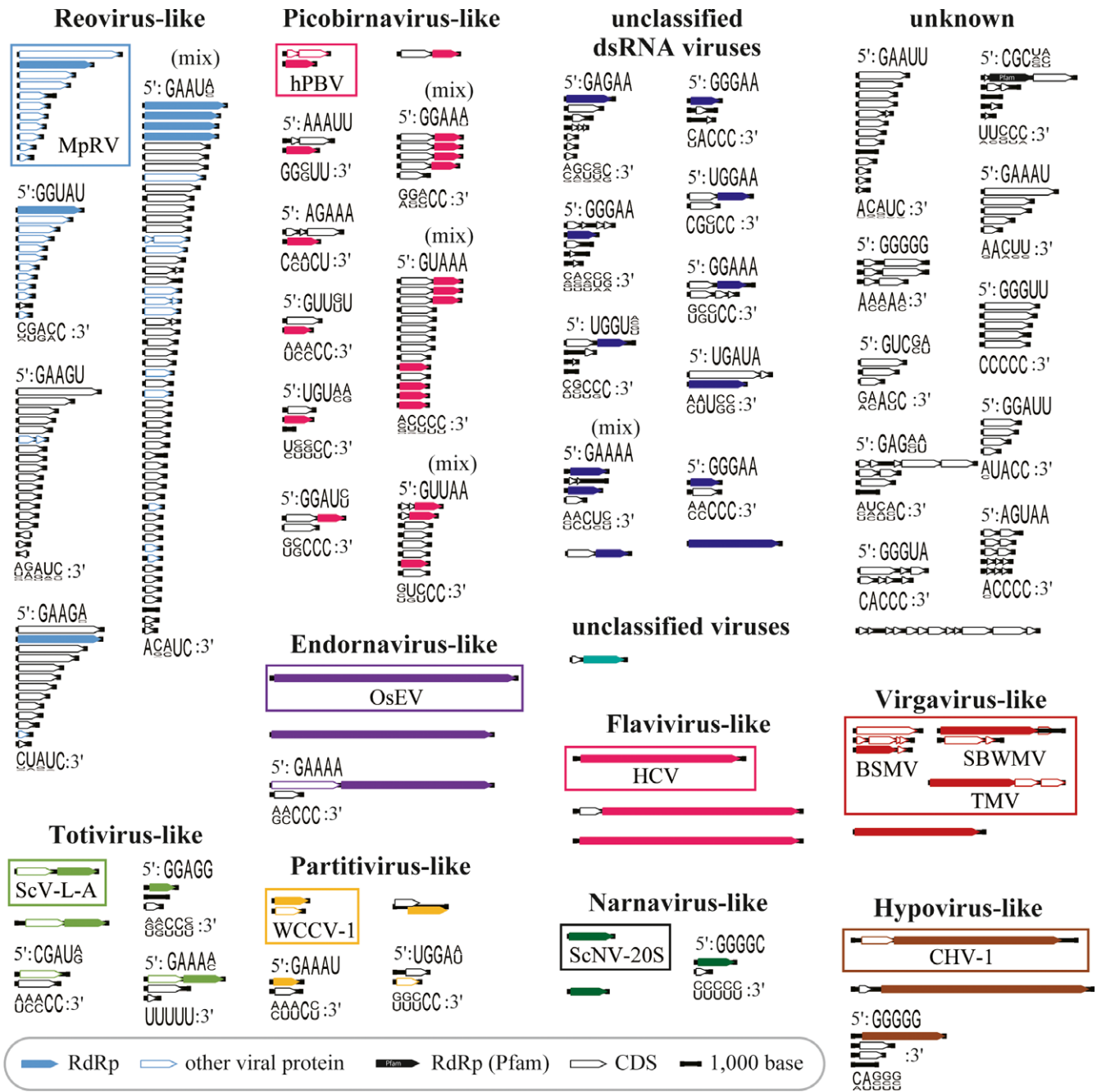
### 3.4 | RNA viromes in north pacific sites and coverage of each metatranscriptome

To elucidate RNA virus diversity, four metatranscriptomes obtained from each station were assembled. A summary of assembly statistics is shown in Supporting information Table S5. We identified a total of 1,270 RNA viral contigs based on BLASTX analysis (Supporting information Tables S6 and S7). To reveal the coverage of each metatranscriptome, mapping analysis with these contigs set as the reference sequences was performed for each station, and 1,266 of 1,270 contigs harboured more than three reads in dsRNA metatranscriptomes (Supporting information Tables S6–S8).

These contigs harboured CDSs with significant similarities ( $e\text{-value} \leq 1 \times 10^{-5}$ ) to genes found in the genomes of known RNA viral families. We identified genome sequences related to 10 dsRNA virus families of the 11 previously established families, 12 positive-sense ssRNA virus families of the 33 previously established families, and 1 negative-sense ssRNA virus family of the nine known families (Supporting information Tables S6 and S7).

Based on RdRp domain sequences, which are encoded as single proteins or polyproteins in their genomes (King et al., 2012), we identified sequences related to nine dsRNA and 10 ssRNA virus families. Phylogenetic analyses based on the RdRp domain amino acid sequence suggested that RdRp domain sequences found in this study were derived from viruses belonging to a known family or novel families in a known order (Supporting information Figures S5–S25). Note that only the maximum likelihood trees are presented in this manuscript because the maximum likelihood trees and Bayesian trees were very similar for all phylogenetic analyses.

As a result of read assembly, we also reconstructed putative complete RNA virus genomes or genome segments (Figure 2) and identified candidates for novel viral genes and viruses that could not be identified by BLASTX analysis. Based on read assembly and mapping



**FIGURE 2** Reconstructed putative genome structures consisted of the representative full-length contigs. Sequence logos represent sequence similarities for the 5' or 3' terminal region of the predicted genome segments. The coloured CDSs indicate RdRp genes that present significant similarities ( $e\text{-value} \leq 1 \times 10^{-5}$ ) with those found in known virus families or groups. Genome segment groups consisting only of possible genome segments and lacking genes presenting significant similarity with known viral genes are classified as the unknown group. Representative genome structures of the known RNA viruses are also shown in boxes. MpRV, *Micromonas pusilla reovirus*; hPBV, *human picobirnavirus*; OsEV, *Oryza sativa endornavirus*; WCCV-1, *White clover cryptic virus 1*; ScV-L-A, *Saccharomyces cerevisiae virus L-A*; HCV, *hepatitis C virus*; ScNV-20S, *Saccharomyces cerevisiae narnavirus 20S*; BSMV, *Barley stripe mosaic virus*; SBWMV, *soil-borne wheat mosaic virus*; TMV, *tobacco mosaic virus*; CHV-1, *Cryphonectria hypovirus 1* [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

analysis (Urayama et al., 2016), we identified full-length RNA virus genome segments and putative RNA virus genome segments that did not harbour CDSs with significant similarity to known genes. It is known that terminal sequences of genome segments are shared among segments in a single virus genome for viral RNA replication and/or encapsidation in some RNA viral lineages (Hutchinson,

Kirchbach, Gog, & Digard, 2010). Therefore, in this study, we defined putative RNA virus genome segments as those that satisfied the following conditions: (a) recognized as full-length based on the read mapping, (b) did not harbour CDSs with significant similarity to known genes, and (c) terminal sequences were shared with RNA virus genome segment(s) identified in this study. In fact, high

similarity was found among some of the (putative) RNA virus genome segments. In the case of viral genomes related to known viral families, we could reconstruct almost the entire genome including genome segments encoding CDSs presenting no significant similarity with the deposited genes in the NCBI database. In addition, putative RNA virus genomes of completely novel virus groups that lacked CDSs presenting significant similarity with known viral genes were identified based on similarity among the terminal sequences (Figure 2). Indeed, we found conserved domains for RdRp with the Pfam algorithm in a few CDSs of these unknown segments (Figure 2).

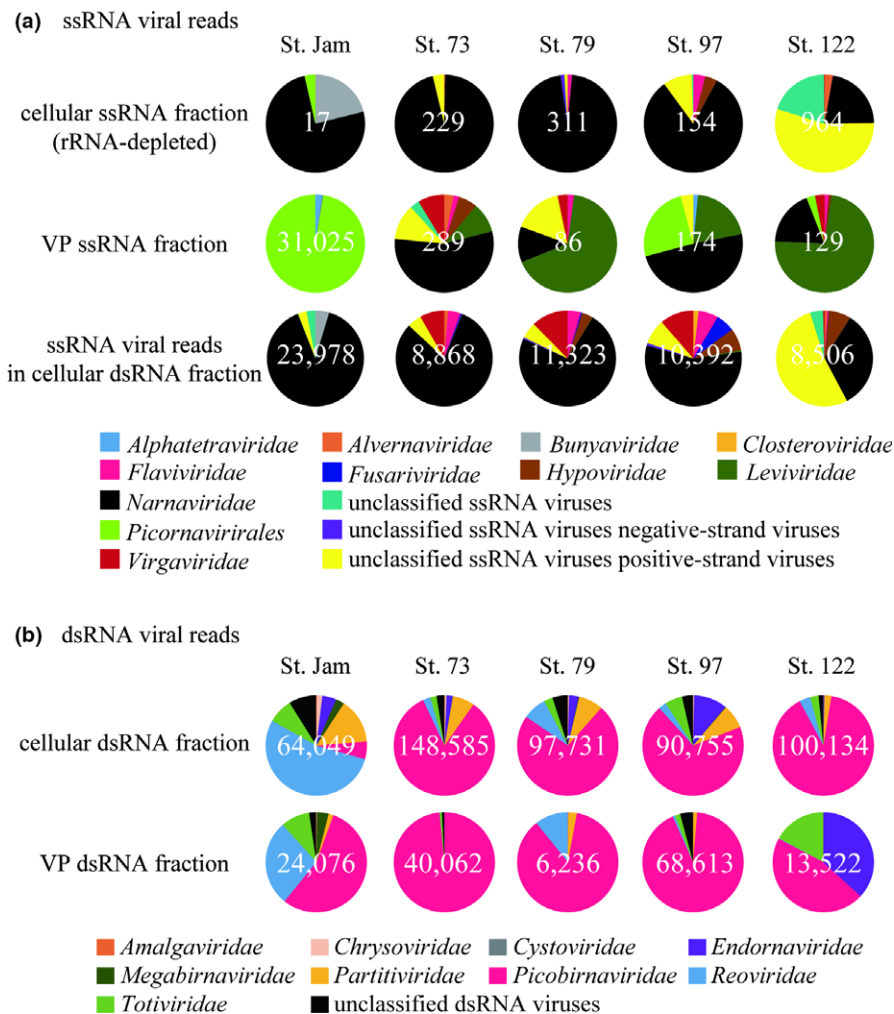
### 3.5 | Comparison of RNA viromes based on RdRp genes

The composition of RNA viromes in each metatranscriptome was compared based on the diversity of the RdRp domain sequences (Supporting information Table S6). Note that the abundance of reads related to ssRNA viruses was extremely low in ssRNA metatranscriptomes (Figure 3), and thus, ssRNA viruses were excluded from this analysis. At all stations, we detected 3.7–14.9 times more dsRNA viral operational taxonomic units (OTUs) in cellular dsRNA metatranscriptomes compared with VP dsRNA metatranscriptomes (Figure 4a).

Moreover, at pelagic station viromes, almost no viral sequences specific for VP dsRNA metatranscriptomes were detected (Figure 4b). In addition, most OTUs (105 of 130) detected in VP metatranscriptomes were related to viruses in the families Reoviridae and Picobirnaviridae, which are the only known eukaryotic dsRNA viruses with an extracellular phase (King et al., 2012).

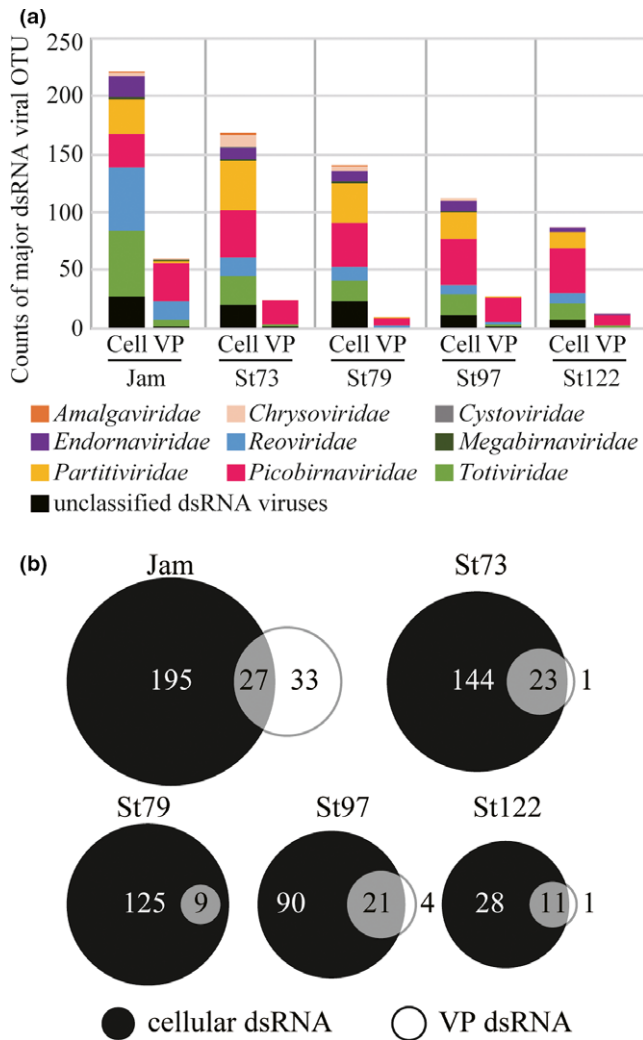
To clarify the relationship between the number of reads used for assembly and the number of obtained dsRNA virus OTUs, we also performed rarefaction analyses of the total length of assembled contigs and the count of mapped RdRp contigs. The results showed that more contigs encoding RdRp were identified in cellular dsRNA samples than in VP dsRNA samples for all the sampling stations (Supporting information Figure S26). In addition, the analyses suggest that most of the dsRNA viral lineages associated with the VP fraction were identified from St. Jam, while the sequence reads were insufficient to retrieve all the viral lineages associated with the VP fractions at the subarctic pelagic stations.

Based on the average coverage of OTUs, Reoviridae-like and Picobirnaviridae-like reads were also abundant in cellular dsRNA metatranscriptomes (Figure 3). Comparing RNA viromes among stations, some OTUs were detected from multiple stations (Figure 5). Among the subarctic pelagic stations, a total of 150 cell fraction-



**FIGURE 3** Taxonomic classification of reads mapped on contigs with a CDS encoding RdRp from ssRNA (a) and dsRNA virus (b) in the metatranscriptomic libraries constructed in this study based on BLASTX analysis. Numbers shown in white indicate the total read numbers. In this study, Megabirnaviridae and Hypoviridae were tentatively classified as dsRNA and ssRNA viruses, respectively [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



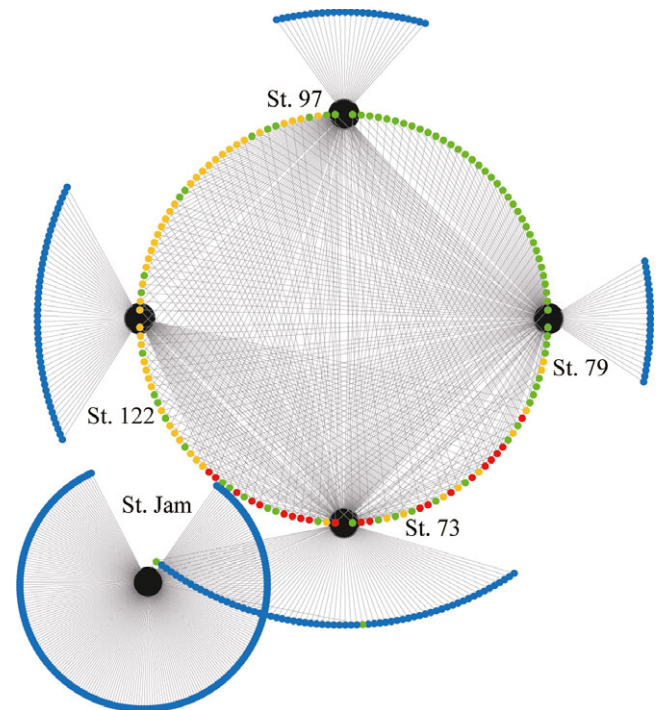


**FIGURE 4** OTU-based community structures of the marine dsRNA viromes in this study. (a) Counts of dsRNA viral RdRp OTUs in cellular and VP dsRNA metatranscriptomes. (b) Venn diagrams of the detected dsRNA viral OTUs between cellular dsRNA and VP dsRNA metatranscriptomes from each sampling station. Numbers indicate the number of dsRNA viral OTUs. The identification of each contig encoding RdRp is summarized in Supporting information Table S6 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

specific OTUs were detected, and 9.3% (14 sequences) of them were detected in the cell fraction from at least one other station (Supporting information Table S9). A total of 49 OTUs were detected in both cell and VP fractions from the same pelagic station, and 40.8% (20 sequences) of them were also found in the cell fraction from at least one other pelagic station (Supporting information Table S9).

### 3.6 | Diversity and genome structure of Picobirnavirus-like RNA viruses

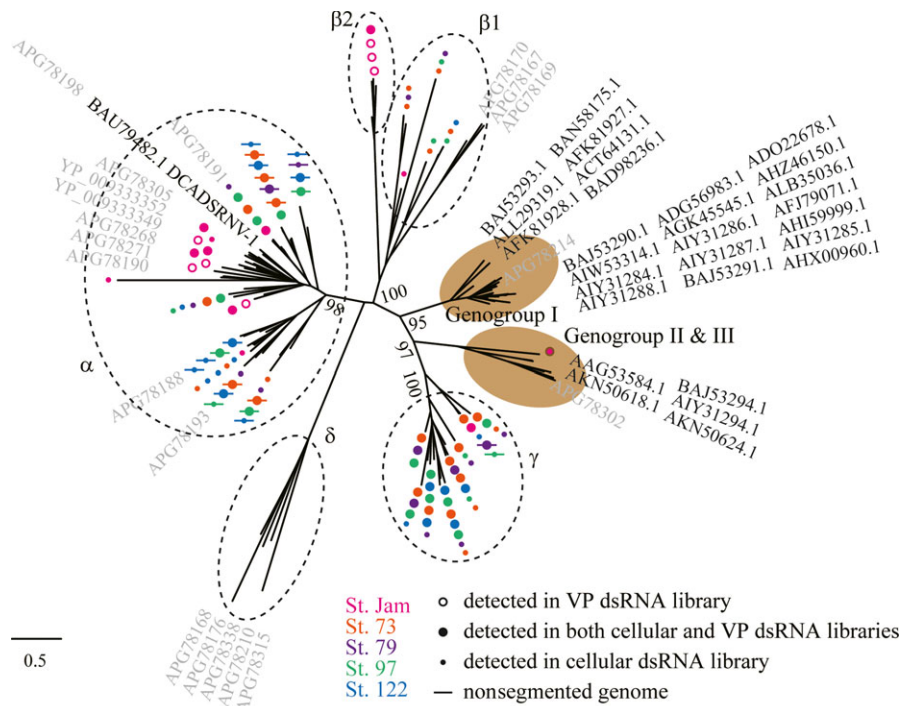
Sequence reads related to picobirnaviruses, which have mostly been identified from faeces of animals including humans (Ganesh, Masachessi, & Mladenova, 2014), predominated in both cellular and VP



**FIGURE 5** Network-based linkage of viral OTUs among stations. OTUs with more than 3× average coverage were used and are presented as small circles. Circle colours indicate the number of connecting nodes: blue, 1; green, 2; yellow, 3; red, 4 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

dsRNA viromes at subarctic pelagic stations (Figures 3 and 4). In fact, we identified a total of 175 contigs/genome segments encoding *Picobirnavirus*-like RdRp domains. Phylogenetic analysis of *Picobirnavirus*-like and known *Picobirnavirus* RdRp amino acid sequences indicated that almost all marine RdRp sequences were phylogenetically distinct from the previously established genogroups I, II and III, which include all picobirnaviruses identified from animal faeces (Smits et al., 2014; Verma, Mor, Erber, & Goyal, 2015) (Figure 6). Thus, we provisionally named the four novel branches as Marine Group (MG)  $\alpha$ ,  $\beta$ 1,  $\beta$ 2 and  $\gamma$  in this study. MG  $\alpha$  includes only one *Picobirnavirus*-like virus identified as being of nonanimal origin, named diatom colony associated dsRNA virus-1 (DCADSRV-1), in addition to the sequences obtained in this study (Urayama et al., 2016). MG  $\beta$ 1 consists of RdRp sequences mainly derived from cellular dsRNA viromes, and MG  $\beta$ 2 contains RdRp sequences mainly derived from the VP dsRNA library in coastal waters (St. Jam). Some of the OTUs identified in this study formed a cluster with the previously characterized genogroups II and III; however, most of them were distinguishable from these genogroups and formed a novel cluster named MG  $\gamma$ .

The genome of the established *Picobirnavirus* is bipartite (King et al., 2012). Genome segment 1 encodes two open reading frames (ORF1 and ORF2), and ORF2 encodes the capsid protein. The smaller genome segment 2 has a single ORF-encoding RdRp (Ganesh et al., 2014). However, we identified two novel types of genome structure encoding *Picobirnavirus*-like RdRp, bisegmented genome



**FIGURE 6** Maximum-likelihood trees of RdRp amino acid sequences from representative members of the family Picobirnaviridae and related sequences obtained in this study. Numbers indicate percentage bootstrap support following 1,000 data resamplings. The best-fitted substitution model was [LG + G + F]. Viruses identified from vertebrates (stool samples) and invertebrates are marked with black and gray, respectively. Colour-coded symbols represent *Picobirnavirus*-like sequences identified in this study. Brown circles indicate the previously established genogroup of picobirnaviruses. Group  $\alpha$  only consists of the *Picobirnavirus*-like virus sequences identified from non-animal origins, named diatom colony-associated dsRNA virus-1 (DCADSRV-1) (Urayama et al., 2016). Group  $\beta$  consists of branches including sequences mainly identified in cellular dsRNA viromes from pelagic stations and in VP dsRNA viromes from coastal waters (St. Jam). Group  $\gamma$  is a novel marine cluster [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(type A) or nonsegmented genome (type B), in the marine virome (Figure 2). Type A genomes were similar to those of previously identified picobirnaviruses, but only one CDS, which has no significant homology with known viral proteins, was identified in the non-RdRp encoding segment. In contrast, MG  $\alpha$  and a branch of MG  $\gamma$  harboured nonsegmented type B genomes that have never been reported. We could not exclude the possibility that a paired segment of type B was overlooked in our analysis; however, no segment with an identical terminal sequence shared with type B segments was found in the viromes.

## 4 | DISCUSSION

### 4.1 | Evaluation of metatranscriptomes and advantages of FLDS version 2

FLDS version 2 developed in this study has significant advantages over conventional shotgun RNA sequencing for nonretro RNA virus surveillance. First, the abundance of RNA viral reads in FLDS version 2 increased more than two orders of magnitude compared with the conventional RNA sequencing method (Supporting information Table S3), although this high abundance of RNA viral sequences is due to the enrichment of viral sequences through dsRNA

purification. Second, FLDS version 2 enabled us to identify complete genome sequences of new RNA viruses, whereas RNA sequencing generally cannot determine the full-length genome of a new virus.

Read assembly and mapping and subsequent comparison of terminal sequences in FLDS version 2 revealed the presence of previously unknown RNA viral lineages that were not identified by conventional RNA sequencing and homology search methods (Figure 2). In the rarefaction analyses, as the number of resampled reads increased, the count of mapped RdRp contigs from the cellular dsRNA became saturated. On the other hand, the unsaturated contig length in this study implies that the sequences include previously unidentified lineages of RdRp and/or the presence of nonviral sequences in the sequence libraries. In addition, previously unknown diversity in genomic structures was also found in RNA viruses phylogenetically related to known viral families. In homology-based analysis, we identified viral genome sequences related to a total of 23 RNA viral families, whereas previous studies on marine RNA viromes using traditional sequencing methods identified sequences related to only nine RNA viral families (Culley et al., 2006, 2014; Steward et al., 2013).

The abundance of extracellular particles with rRNA in pelagic water was comparable or more than that of RNA VPs. Unexpectedly, VP ssRNA transcriptomes from the four subarctic sampling stations

were dominated by rRNA sequences (Supporting information Table S3). VP samples were obtained by the iron precipitation method (Culley et al., 2014; Steward et al., 2013), and a low abundance of rRNA reads occurred in the VP ssRNA metatranscriptome from St. Jam. Thus, the high abundance of rRNA in pelagic VP ssRNA metatranscriptomes was not likely a result of technical issues in sample processing, but suggests the occurrence of rRNA in extracellular membrane particles at these stations.

## 4.2 | Picobirnavirus-like sequences

Our analysis of cell fractions also provides new perspectives into the relationships between virus and host. Unexpectedly, in dsRNA viromes, *Picobirnavirus*-like sequences were detected in cell fractions of marine microorganisms (Figures 3 and 4). *Picobirnavirus* and its relatives are believed to infect vertebrates and have been mainly detected in the faeces of vertebrates (Ganesh et al., 2014) and the bodies of invertebrates (Shi et al., 2016). However, direct infection of animal cells or tissues has not been observed in previous inoculation experiments (Ludert, Abdul-Latiff, Liprandi, & Liprandi, 1995; Malik et al., 2014), and *Picobirnavirus*-like sequences have recently been reported from nonanimal organisms (Urayama et al., 2016). In the phylogenetic analysis of RdRp domain sequences, some *Picobirnavirus*-like sequences identified in this study formed MG  $\alpha$  and  $\beta$  clusters with known invertebrate-derived *Picobirnavirus*-like sequences. In addition, a novel marine cluster, MG  $\gamma$ , was identified (Figure 5). This study revealed that *Picobirnavirus*-like sequences associated with invertebrates are more diverse than those associated with vertebrates. This observation and previous findings from inoculation experiments suggest that some previously reported *Picobirnavirus*-like sequences found in vertebrates likely do not infect vertebrates but originated from microorganisms associated with vertebrates.

## 4.3 | Ecology of RNA viruses in surface seawater and cellular RNA viruses

Although the marine virosphere is one of the most analysed environments explored by metagenomics, this study illuminated a previously unknown RNA virus population associated with diverse planktonic marine microbes in surface seawater. In addition to diverse previously unknown putative viral genome sequences, we found sequences phylogenetically related to more than 20 RNA viral families, most of which have been identified in terrestrial organisms to date, in only a total of 10 L of surface seawater. Previous marine metagenomic investigations of RNA VPs or host organisms (Culley et al., 2006, 2014; Dijkeng et al., 2009; Steward et al., 2013) may have overlooked the diversity of RNA viruses that lack an extracellular phase and/or capsid. We could not exclude the possibility that sampling, purification, and/or library construction processes influenced the RNA virus diversity of the VP fraction. In addition, the actual diversity of RNA viruses in the VP fraction might also have influenced the result. If the VP fraction harboured RNA virus

populations with low abundance and high diversity, the RNA virus diversity in the VP fraction may have been underestimated in the sequence analysis.

The identified RNA virus-like sequences were related to the known RNA viral sequences obtained from well-studied organisms such as humans and mosquitos (animals), soya beans and tomatoes (plants) and *Aspergillus* and *Fusarium* (fungi) (Supporting information Tables S6 and S7), while communities of potential hosts were dominated by monocellular eukaryotes (Supporting information Figure S4). In other words, RNA viruses that likely belong to the same taxonomic level (family or order) were found from phylogenetically distinct host species. For example, endornaviruses are known as plant or fungal viruses (King et al., 2012), however, *Endornavirus*-like sequences were probably detected from monocellular eukaryotes in this study. These data suggest that the host ranges of known RNA viral groups (families or orders) are wider than expected, although this study is not sufficient to define the hosts for each viral sequence. In fact, transmission of an RNA virus from plant to fungi has recently been reported (Andika et al., 2017).

Since RNA viruses cause not only disease but also physiological changes to their hosts, including toxin production (Magliani, Conti, Gerloni, Bertolotti, & Polonelli, 1997), growth rate (Boland, 1992) and stress tolerance (Bottacin, Lévesque, & Punja, 1994), some RNA viruses identified in this study may be detachable genetic elements that modify the functions of host cells without genomic alteration, such as plasmids.

## 5 | CONCLUSION

To date, viral diversity has been assessed through the diversity of extracellular VPs in the marine virosphere. However, the great advantage of the FLDS version 2 method presented in this study is that it targets intracellular dsRNA viral genomes and can provide novel insights into RNA viral diversity. Recent advancements in RNA virology indicate that RNA viruses are not necessarily pathogens but may act as modifiers of cellular function as detachable genetic elements (Roossinck, 2011). The application of FLDS version 2 in diverse environments and organisms, such as soil and gut virome analyses, will expand our knowledge of the previously unseen viral diversity and function in each ecosystem.

## ACKNOWLEDGEMENTS

We thank the captain, crew and science party aboard the *R/V Mirai* on the Japan Agency for Marine-Earth Science & Technology (JAM-STE) MR14-04 cruise. We thank Shinsuke Kawagucci, Norio Miyamoto, Mitsuhiro Yoshida, Akinori Yabuki, Yuji Tomaru, Keizo Nagasaki, Phurt Harnvoravongchai, Seiya Takahashi, Takashi Toyofuku, Katsunori Kimoto, Tomohiko Kikuchi and Shinji Shimode for discussions, suggestions, sample collections and preliminary experiments related to this study. This study was supported by JSPS KAKENHI (Grant No. JP26892031) and by Grants-in-Aid for

Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports, and Technology (MEXT) of Japan (Grant Nos. 16H06429, 16K21723, and 16H06437). JAMSTEC has filed a patent application related to FLDS, of which S.U., T.N. and S.D. are the named inventors.

## AUTHOR CONTRIBUTIONS

S.U. and T.N. designed the research. S.U., Y.T. and T.N. performed the research. S.U. and T.N. collected samples. Y.T., S.N. and S.U. analysed the data. S.U., Y.T., Y.Y., S.D., K.T. and T.N. wrote the manuscript.

## DATA ACCESSIBILITY

Data sets supporting the results of this study are available in the GenBank database repository (Accession Nos. DDBJ: BDQA01000001–BDQA01005007, BDQB01000001–BDQB01001937, BDQC01000001–BDQC01001116, BDQD01000001–BDQD01001458, BDQE01000001–BDQE01001418) and Short Read Archive database (Accession No. DDBJ: DRA005015).

## REFERENCES

- Andika, I. B., Wei, S., Cao, C., Salaipeth, L., Kondo, H., & Sun, L. (2017). Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 12267–12272. <https://doi.org/10.1073/pnas.1714916114>
- Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, *29*, 2607–2618. <https://doi.org/10.1093/nar/29.12.2607>
- Boland, G. J. (1992). Hypovirulence and double-stranded RNA in *Sclerotinia sclerotiorum*. *Canadian Journal of Plant Pathology*, *14*, 10–17.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bottacin, A., Lévesque, C., & Punja, Z. (1994). Characterization of dsRNA in *Chalara elegans* and effects on growth and virulence. *Phytopathology*, *84*, 303–312.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., ... Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 14250–14255. <https://doi.org/10.1073/pnas.202488399>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*, 59–60. <https://doi.org/10.1038/nmeth.3176>
- Bushnell, B. (2016). *BBMap short read aligner*. Berkeley, CA: University of California. Retrieved from <https://sourceforge.net/projects/bbmap>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Culley, A. (2017). New insight into the RNA aquatic virosphere via viromics. *Virus Research*, *244*, 84–89. <https://doi.org/10.1016/j.virusres.2017.11.008>
- Culley, A. I., Lang, A. S., & Suttle, C. A. (2006). Metagenomic analysis of coastal RNA virus communities. *Science*, *312*, 1795–1798. <https://doi.org/10.1126/science.1127404>
- Culley, A. I., Mueller, J. A., Belcaid, M., Wood-Charlson, E. M., Poisson, G., & Steward, G. F. (2014). The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *mBio*, *5*, e01210–e1214. <https://doi.org/10.1128/mBio.01210-14>
- Decker, C. J., & Parker, R. (2014). Analysis of double-stranded RNA from microbial communities identifies double-stranded RNA virus-like elements. *Cell Reports*, *7*, 898–906. <https://doi.org/10.1016/j.celrep.2014.03.049>
- Djikeng, A., Kuzmickas, R., Anderson, N. G., & Spiro, D. J. (2009). Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One*, *4*, e7264. <https://doi.org/10.1371/journal.pone.0007264>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Edwards, R. A., & Rohwer, F. (2005). Viral metagenomics. *Nature Reviews Microbiology*, *3*, 504–510.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*, D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, *399*, 541–548. <https://doi.org/10.1038/21119>
- Ganesh, B., Masachessi, G., & Mladenova, Z. (2014). *Animalpicobimavirus*. *Virus-disease*, *25*, 223–238. <https://doi.org/10.1007/s13337-014-0207-y>
- Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., ... Tappu, R. (2016). MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Computational Biology*, *12*, e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>
- Hutchinson, E. C., von Kirchbach, J. C., Gog, J. R., & Digard, P. (2010). Genome packaging in influenza A virus. *Journal of General Virology*, *91*, 313–328. <https://doi.org/10.1099/vir.0.017608-0>
- John, S. G., Mendez, C. B., Deng, L., Poulos, B., Kauffman, A. K. M., Kern, S., ... Sullivan, M. B. (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environmental Microbiology Reports*, *3*, 195–202. <https://doi.org/10.1111/j.1758-2229.2010.00208.x>
- King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (2012). *Virus taxonomy: Classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. London, UK: Elsevier Academic Press.
- Koch, L. (2016). Metagenomics: Marine genomics goes viral. *Nature Reviews Genetics*, *17*, 660. <https://doi.org/10.1038/nrg.2016.130>
- Koonin, E. (2010). The two empires and three domains of life in the postgenomic age. *Nature Education*, *3*, 27.
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*, 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>
- Koyama, S., Sakai, C., Thomas, C. E., Nunoura, T., & Urayama, S. I. (2016). A new member of the family Totiviridae associated with arboreal ants (*Camponotus nipponicus*). *Archives of Virology*, *161*(7), 2043–2045. <https://doi.org/10.1007/s00705-016-2876-x>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. <https://doi.org/10.1038/nmeth.1923>

- Leininger, S., Urlich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., ... Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, 442, 806.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Ludert, J., Abdul-Latif, L., Liprandi, A., & Liprandi, F. (1995). Identification of picobirnavirus, viruses with bisegmented double stranded RNA, in rabbit faeces. *Research in Veterinary Science*, 59, 222–225. [https://doi.org/10.1016/0034-5288\(95\)90006-3](https://doi.org/10.1016/0034-5288(95)90006-3)
- Magliani, W., Conti, S., Gerloni, M., Bertolotti, D., & Polonelli, L. (1997). Yeast killer systems. *Clinical Microbiology Reviews*, 10, 369–400.
- Malik, Y. S., Kumar, N., Sharma, K., Dhama, K., Shabbir, M. Z., Ganesh, B., ... Banyai, K. (2014). Epidemiology, phylogeny, and evolution of emerging enteric Picobirnaviruses of animal origin and their relationship to human strains. *BioMed Research International*, 2014, 780752.
- Marquez, L. M., Redman, R. S., Rodriguez, R. J., & Roossinck, M. J. (2007). A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science*, 315, 513–515. <https://doi.org/10.1126/science.1136237>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet. Journal*, 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., & Marshall, D. (2010). Tablet—next generation sequence assembly visualization. *Bioinformatics*, 26, 401–402. <https://doi.org/10.1093/bioinformatics/btp666>
- Morris, T., & Dodds, J. (1979). Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology*, 69, 854–858. <https://doi.org/10.1094/Phyto-69-854>
- Okada, R., Kiyota, E., Moriyama, H., Fukuhara, T., & Natsuaki, T. (2015). A simple and rapid method to purify viral dsRNA from plant and fungal tissue. *Journal of General Plant Pathology*, 81, 103–107. <https://doi.org/10.1007/s10327-014-0575-6>
- Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Hunteemann, M., Mikhailova, N., ... Kyrpides, N. C. (2016). Uncovering earth's virome. *Nature*, 536, 425–430. <https://doi.org/10.1038/nature19094>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rambaut, A. (2014). *FigTree 1.4. 2 software*. Institute of Evolutionary Biology, Univ. Edinburgh. Retrieved from <https://tree.bio.ed.ac.uk/software/figtree/>
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Roossinck, M. J. (2011). The good viruses: Viral mutualistic symbioses. *Nature Reviews: Microbiology*, 9, 99–108. <https://doi.org/10.1038/nrmicro2491>
- Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., ... Roe, B. A. (2010). Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology*, 19(Suppl 1), 81–88. <https://doi.org/10.1111/j.1365-294X.2009.04470.x>
- Ruby, J. G., Bellare, P., & Derisi, J. L. (2013). PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)*, 3, 865–880. <https://doi.org/10.1534/g3.113.005967>
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27, 863–864. <https://doi.org/10.1093/bioinformatics/btr026>
- Scotto-Lavino, E., Du, G., & Frohman, M. A. (2006). 5' end cDNA amplification using classic RACE. *Nature Protocols*, 1, 2555–2562. <https://doi.org/10.1038/nprot.2006.480>
- Shi, M., Lin, X. D., Tian, J. H., Chen, L.-J., Chen, X., Li, C.-X., ... Zhang, Y.-Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*. <https://doi.org/10.1038/nature20167>
- Smits, S. L., Schapendonk, C. M., van Beek, J., Vennema, H., Schürch, A. C., Schipper, D., ... Koopmans, M. P. (2014). New viruses in idiopathic human diarrhea cases, the Netherlands. *Emerging Infectious Diseases*, 20, 1218–1222. <https://doi.org/10.3201/eid2007.140190>
- Smoot, M. E., Ono, K., Ruschinski, J., Wang, P. L., & Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, 27, 431–432. <https://doi.org/10.1093/bioinformatics/btq675>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Steward, G. F., Culley, A. I., Mueller, J. A., Wood-Charlson, E. M., Belcaid, M., & Poisson, G. (2013). Are we missing half of the viruses in the ocean? *ISME Journal*, 7, 672–679. <https://doi.org/10.1038/ismej.2012.121>
- 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. *Communications in Statistics-Theory and Methods*, 7, 13–26. <https://doi.org/10.1080/03610927808827599>
- Suttle, C. A. (2005). Viruses in the sea. *Nature*, 437, 356–361. <https://doi.org/10.1038/nature04160>
- Suttle, C. A. (2016). Environmental microbiology: Viral diversity on the global stage. *Nature Microbiology*, 1, 16205. <https://doi.org/10.1038/nmicrobiol.2016.205>
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28, 2731–2739. <https://doi.org/10.1093/molbev/msr121>
- 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. *Molecular Ecology Resources*, 11, 914–921. <https://doi.org/10.1111/j.1755-0998.2011.03021.x>
- Uchida, H., & Doi, T. (2002). *WHP P01 Revisit in 2014 Data Book*. Yokosuka: Jamstec.
- Urayama, S.-i., Takaki, Y., & Nunoura, T. (2016). FLDS: A Comprehensive dsRNA sequencing method for intracellular RNA virus surveillance. *Microbes and Environments*, 31(1), 33–40. <https://doi.org/10.1264/jsme2.ME15171>
- Urayama, S., Yoshida-Takashima, Y., Yoshida, M., Tomaru, Y., Moriyama, H., Takai, K., & Nunoura, T. (2015). A new fractionation and recovery method of viral genomes based on nucleic acid composition and structure using tandem column chromatography. *Microbes and Environments*, 30, 199–203. <https://doi.org/10.1264/jsme2.ME14174>
- Verma, H., Mor, S. K., Erber, J., & Goyal, S. M. (2015). Prevalence and complete genome characterization of turkey picobirnaviruses. *Infection, Genetics and Evolution*, 30, 134–139. <https://doi.org/10.1016/j.meegid.2014.12.014>
- Zeigler Allen, L., McCrow, J. P., Ininbergs, K., et al. (2017). The Baltic Sea virome: Diversity and transcriptional activity of dna and rna viruses. *mSystems*, 2, e00125–16. <https://doi.org/10.1128/mSystems.00125-16>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Urayama S-I, Takaki Y, Nishi S, et al. Unveiling the RNA virosphere associated with marine microorganisms. *Mol Ecol Resour*. 2018;18:1444–1455. <https://doi.org/10.1111/1755-0998.12936>