

## Characterization of the diversity of marine RNA viruses

Alexander I. Culley<sup>1\*</sup>, Curtis A. Suttle<sup>2</sup>, and Grieg F. Steward<sup>1</sup>

<sup>1</sup>Department of Oceanography, University of Hawai'i at Manoa, Honolulu, HI 96822

<sup>2</sup>University of British Columbia, Department of Earth and Ocean Sciences, Botany, and Microbiology & Immunology, 1461 Biosciences-6270 University Blvd., Vancouver, BC, V6T 1Z4

### Abstract

The diversity of ribonucleic acid (RNA) viruses in the ocean and the ongoing isolation and characterization of RNA viruses that infect important primary producers suggests that RNA viruses are active members of the marine microbial assemblage. At this point, little is known about the composition, dynamics, and ecology of the RNA viroplankton. In this chapter, we describe two methods to assess RNA virus diversity from seawater.

Ribonucleic acid (RNA) viruses that infect marine plankton have been poorly studied compared with their deoxyribonucleic acid (DNA)-containing counterparts, but evidence of their importance has been accumulating. RNA-containing viruses have been isolated that infect marine protists, including a noxious dinoflagellate

(Tomaru et al. 2004), a common fungoid protist (Takao et al. 2005), a red tide-forming raphidophyte (Tai et al. 2003), and three types of diatom (Nagasaki et al. 2004; Shirai et al. 2008; Tomaru et al. 2009). There is also at least one report of a marine RNA-containing bacteriophage (Hidaka and Ichida 1976). More detailed descriptions of these viral isolates can be found in reviews by Munn (2006), Brussaard and Martinez (2008), Nagasaki (2008) and Lang et al. (2009).

The isolation and cultivation of a virus from the environment is essential if the virus is to be fully characterized. Nev-

ertheless, cultivation is an unrealistic approach to describing the full range of diversity in natural marine viral assemblages. Cultivation-independent methods have thus been adopted, which include categorization based on morphology (Frank and Moebus 1987), genome size distributions (Steward et al. 2000), the phylogeny or finger print of a specific viral gene (Chen et al. 1996; Fuller et al. 1998; Short and Suttle 1999; Filée et al. 2005; Larsen et al. 2008), or construction of metagenomic libraries (Breitbart et al. 2002, 2004; Angly et al. 2006; Bench et al. 2007; Helton and Wommack 2009).

Characterizing RNA virus diversity based on morphology is the least informative of the above approaches because it is laborious and provides little resolution among viruses. To the best of our knowledge, most RNA viruses are relatively small and untailed with little morphological detail to distinguish them from small, untailed DNA-containing viruses. Moreover, there appears to be no relationship between the morphology of an RNA virus and the type of host that it is able to infect. Characterizing the distribution of genomes sizes within a sample is a more tractable approach to determining RNA virus diversity, because these results are relatively quick and easy to obtain and can provide quantitative data on diversity. Our preliminary results suggest that RNA sizing by denaturing agarose gels electrophoresis can be used to generate "fingerprints" of RNA virus genotypes from marine assemblages. However, this method is limited because relatively large amounts of viral RNA are required, and only significant differences in genome size are discriminated and not differences in sequence.

The superior sensitivity and resolution achievable by sequence analysis has made sequence-based methods the most widely used approach to categorizing microbial diversity, and molecular surveys have made it clear that much of the diversity among RNA viruses has yet to be discovered. Many of the novel gene sequences recovered so far appear to derive from viruses of marine protists. Their high diversity suggests that

\*Corresponding author: E-mail: aculley@hawaii.edu, phone: (808) 956-8629

### Acknowledgments

Publication costs for the Manual of Aquatic Viral Ecology were provided by the Gordon and Betty Moore Foundation. This document is based on work partially supported by the U.S. National Science Foundation (NSF) to the Scientific Committee for Oceanographic Research under Grant OCE-0608600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

The writing of this chapter was supported in part by grants from NSF to the authors (OCE-04-42664 and OCE-0826650) and to the Center for Microbial Ecology Research and Education (EF-0424599). The authors acknowledge the efforts of two anonymous peer reviewers and their suggestions to improve the manuscript.

ISBN 978-0-9845591-0-7, DOI 10.4319/mave.2010.978-0-9845591-0-7.193

Suggested citation format: Culley, A. I., C. A. Suttle, and G. F. Steward. 2010. Characterization of the diversity of marine RNA viruses, p. 193–201. In S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle [eds.], Manual of Aquatic Viral Ecology. ASLO.

viral infections are a persistent force-shaping protistan community composition in the sea.

Because this is a relatively new area of research, the number of methods described in the literature for investigating the diversity of marine RNA viruses is still limited. We present here two protocols that have been used successfully in published articles (Culley et al. 2006; Culley and Steward 2007; Djikeng et al. 2008, 2009), but note that this is a nascent field and there are no routine methods. We hope the protocols described will at least serve as starting point for further improvements.

### **Protocol 1. A degenerate primer reverse transcription-polymerase chain reaction–based protocol to determine the diversity of picorna-like viruses**

This protocol is designed to detect picorna-like viruses from marine samples. It is based on a strategy first reported by Culley et al. (2003) and refined by Culley and Steward (2007). All picorna-like viruses have single-stranded positive-sense RNA genomes and are classified in the order *Picornavirales* (Le Gall et al. 2008). Picorna-like viruses are responsible for several significant human and animal diseases and infect a diversity of marine protists including a diatom (Nagasaki 2008; Lang et al. 2009). The procedure outlined below uses direct filtration of relatively small volumes of water and reverse transcription-polymerase chain reaction (RT-PCR) amplification of an RNA-dependent RNA polymerase (RdRP) gene fragment using degenerate primer sets.

With this method, viral polymerase sequences were amplified in several distinct aquatic environments, including an estuarine urban canal (Ala Wai canal, Waikiki, HI, USA), a tropical bay (Kaneohe Bay, HI, USA), and a temperate bay (Monterey Bay, CA, USA). Amplification occurred in samples from the same site in different seasons and at different depths showing that RNA viruses are widespread and consistently present. Sequencing of the amplified gene fragments revealed novel sequences that are highly divergent from any known isolates.

### **Material and equipment**

**Equipment**—Peristaltic pump and pump heads, thermocycler, heating block, incubator, gel electrophoresis unit, electroporator, gel documentation system

**Supplies**—Sterile, 1.7 mL nucleic acid-free microcentrifuge tubes; 0.2 mL sterile, nucleic acid-free PCR tubes; 0.02  $\mu$ m aluminum oxide filters (Anotop, Whatman); 10 mL sterile syringes; peristaltic pump tubing; sterile razor blades; MinElute Gel Extraction kit (Qiagen); MinElute PCR Purification kit (Qiagen); Masterpure Complete DNA and RNA Purification kit (Epicenter Biotechnologies); PCR Terminator End Repair kit (Lucigen); CloneSmart HCKan Blunt Cloning kit with Ecloni Supreme cells (Lucigen); Turbo DNA-free Kit (Applied Biosystems)

**Solutions, reagents, and media**—0.02-filtered, sterile, nucleic acid-free TE buffer (10 mM Tris, 1 mM EDTA; pH 8); 0.1 M dithiothreitol (DTT); Superscript III Reverse Transcriptase and

buffers (Invitrogen); RNase Out (Invitrogen); RNase H (Invitrogen); Platinum Taq DNA polymerase and buffers (Invitrogen)

Primers in concentrations discussed as follows: 0.5  $\times$  TBE (45 mM Tris-borate, 1 mM EDTA; pH 8.0) electrophoresis buffer; nucleic acid-free, sterile water; 10 mM dNTP mix

### **Steps**

**Collection**—Whole seawater (if intracellular viruses are to be included in the analysis) or 0.22-filtered seawater (if only the free virus community is targeted) is gently filtered (7 mmHg) through a 0.02  $\mu$ m aluminum oxide filter (Anotop, Whatman). Filtration should continue until the rate slows dramatically or drops to zero (note that it is important to measure the total volume of sample that passes through the filter). Use pumped air to remove as much residual sample as possible. Once dry, seal the filter inlets and outlets with parafilm, label the filter, and then flash-freeze the filter in liquid nitrogen and store it at  $-80^{\circ}\text{C}$  until extraction.

**Extraction**—Total nucleic acids are extracted from the aluminum oxide filters using a Masterpure complete DNA and RNA Purification kit (Epicenter, Biotechnologies) with a protocol slightly modified from the manufacturer's instructions. The detailed protocol of this method is presented elsewhere in this book (Steward and Culley 2010, this volume).

**DNase treatment**—To reduce the likelihood of mispriming and increase the sensitivity of the subsequent cDNA synthesis and PCR reactions, we remove contaminating DNA from our RNA preparations with the Turbo DNA-free kit (Applied Biosystems) as described in the protocol provided with the kit. Although the manufacturers suggest that treated RNA template should not exceed 40% of the total cDNA synthesis reaction, using the reaction discussed below, we have used 55% template without inhibition.

**cDNA synthesis**—We have synthesized complementary DNA (cDNA) using reverse transcriptase (Superscript III, Invitrogen) in reactions primed with random hexamers (N6) or the specific primers RdRp2, Mpl.sc1R, Mpl.sc2R, Mpl.sc3R, or Mpl.cdhr (details for these primers are listed in Table 1), and have found no detectable difference in sensitivity between the two approaches. Priming with random hexamers is preferable because the resultant cDNA can be used in a PCR reaction with any of the five primer sets listed in Table 1, eliminating the necessity for a different cDNA reaction for each primer set.

Each cDNA reaction contains 11  $\mu$ L of extracted Turbo DNA-free-treated RNA template, 0.2 mM of each dNTP, and 100 ng of N6 primer in a total volume of 13  $\mu$ L. Denaturation and annealing of the primers to the RNA template occurs by heating the sample to  $65^{\circ}\text{C}$  for 5 min and then cooling it on ice. While still on ice, DTT (0.5 mM final conc.) is added to the reaction as an enzyme stabilization reagent with 40 U RNase OUT (Invitrogen) to protect the sample from RNase activity. The complementary DNA strand is synthesized with 200 U Superscript III (Invitrogen) Reverse Transcriptase. The final reaction volume should be 20  $\mu$ L.

**Table 1.** RT and PCR primer details.

Name	Sequence (5'-3')	°C annealing	~Product (bp)
Mpl.sc1F	TIGCIGGWGAYTWYARM	50	500
Mpl.sc1R	YTCCTTWTCRGSCATKGT		
Mpl.sc2F	ITWGCIGGIGATTWCA	43	500
Mpl.sc2R	CKYTTCARAAWTCAGCATC		
Mpl.sc3F	TIATIGMKGGIGAYTA	49	500
Mpl.sc3R	TTMARGAAIKMAGCATCTT		
Mpl.cdhF	GMIGGTGAYTAYAGCGCTTWYGAY	44	500
Mpl.cdhR	ATACCCAATGCCTYTTIARRAA		
RdRp1	GGRGAYTACASCIRWTTTGAT	50	450
RdRp2	MACCCAACKMCKCTTSARRAA		
SL1	CAGTCCAGTTACGCTGGAGTC	50	NA
SR2	GGTCAGGTATGATTTAAATGGTCAGT		
FR26RV-N	GCCGGAGCTCTGCAGATATCNNNNNN	NA	NA
FR40RV-T	GCCGGAGCTCTGCAGATATC(T)20	NA	NA
FR20RV	GCCGGAGCTCTGCAGATATC	65	NA

NA, not applicable.

The reaction is incubated initially at 25°C for 5 min so that the relatively unstable hexamer primers remain annealed to the template while cDNA synthesis commences. The temperature is then increased to 50°C, the temperature at which Superscript III's processivity is highest, for 60 min. The activity of the enzyme is terminated by incubating the reaction at 85°C for 5 min. After cDNA synthesis, add 1 µL (2 U) of RNase H (Invitrogen) to the reaction and incubate at 37°C for 20 min to digest the RNA template from the cDNA:RNA molecule.

**PCR**—PCR can be performed with the RdRp, Mpl.sc1, Mpl.sc2, Mpl.sc3, and Mpl.cdh primer sets listed in Table 1 and the cDNA synthesized in the previous step. In a 0.2 mL nuclease-free PCR tube, add reaction components to achieve final concentrations of 1× Platinum *Taq* buffer, 3 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 1 µM of each primer, and 1 unit of Platinum *Taq* DNA polymerase (Invitrogen). Incubate the reactions with the following thermal cycling conditions: 94°C for 75 s (this is necessary to activate the enzyme and ensures complete initial denaturation of template), followed by 35 cycles of denaturation at 94°C for 45 s, annealing at a primer-specific temperature (Table 1) for 45 s, and extension at 72°C for 45 s (note that the target product size is approximately 500 bp). Complete the cycle with a single final extension step of 9 min 15 s at 72°C.

Before gel separation, we purify and concentrate the PCR reactions with a MinElute PCR cleanup column (Qiagen) as described by the manufacturer. A smaller volume of product allows one to load the product on a thinner gel, ultimately resulting in a more efficient recovery of target DNA from the excised band. Purified PCR products are loaded onto a 1% agarose gel containing 1× SYBR safe stain (Invitrogen) and 0.5× TBE buffer. Bands of DNA of the appropriate size (approximately 500 bp) are excised and purified with a MinElute Gel Extraction kit (Qiagen) according to the manufacturer's instructions. In the final step of this procedure, we recom-

mend eluting DNA from the column with three washes of 10 µL nuclease-free water in preparation for the end repair reaction described below.

**Cloning and sequencing**—Several commercial kits are available for the efficient cloning of PCR products. We use the CloneSmart Blunt Cloning kit from Lucigen. Some advantages of this cloning kit are the high efficiency of recombinants that eliminate the need for screening (e.g., XGAL/IPTG) and the prevention of transcription of inserts in the vector reduces cloning bias. However, the pSMART vector requires blunt ends with 5' phosphate groups (unlike TOPO-TA [Invitrogen] cloning kits for example). The additional step adds time, increases cost, and increases risk of sample loss, but we have found that it works well.

In preparation for ligation, PCR products are polished and phosphorylated with the PCR Terminator End Repair kit (Lucigen) as described by the manufacturer. After a 15-min incubation at room temperature the reaction is purified and concentrated with a MinElute Reaction Cleanup column (Qiagen). The eluted PCR products are subsequently ligated into the pSMART-HCKan vector (Lucigen) according to the manufacturer. In this reaction, we typically use the maximum amount of template (6.5 µL) and the maximum incubation time (2 h). After terminating the ligation reaction with by incubating for 15 min at 70°C, 2 µL of the ligation reaction is transformed into Ecloni 10G Supreme cells (Lucigen) via electroporation. We do not recommend using greater than 2 µL of the ligation mixture in the transformation reaction because the chances of arcing are greatly increased during electroporation. Before initiating a large-scale sequencing effort, we recommend screening 10 to 20 colonies for inserts by PCR amplification with the primers SL1 and SR2 (Table 1) to assess the quality of the library. Colony PCR products can be visualized on a gel and the products in the correct size range purified and sequenced.

## Assessment

The method described above is based on reverse transcription (RT) of RNA into cDNA and PCR amplification of mixed templates. PCR in combination with cloning, sequencing and bioinformatic analysis, has resulted in the identification and cultivation-independent classification of thousands of microbes (Rappé and Giovannoni 2003). However, bias associated with every step of a PCR-based assay, including sample collection, nucleic acid extraction, PCR amplification, and sequence analysis of PCR amplicons, can contribute to an inaccurate portrayal of the community under examination (von Wintzingerode et al. 1997).

Reverse transcriptase synthesizes DNA from an RNA template. The enzyme lacks 3' to 5' proofreading exonuclease activity, contributing to its relatively high error rate in comparison with other DNA polymerases (Roberts et al. 1988). For example, Superscript II reverse transcriptase (Invitrogen) is approximately 13 times more error prone than Platinum Taq polymerase (Invitrogen) (Roberts et al. 1988). Factors such as the concentration of target RNA, the amount of template secondary structure and priming conditions including annealing temperature can significantly affect the precision, efficiency and production of the RT reaction (Stahlberg et al. 2004).

Amplification of environmental targets with PCR can result in differential amplification, the formation of chimeras and heteroduplexes, and artifacts from DNA polymerase error, among other biases (von Wintzingerode et al. 1997; Kanagwa 2003). Polz and Cavanaugh (1998) found that PCR with degenerate primers did not maintain the original ratio of template after 25 cycles and that templates with GC-rich priming sites were preferentially amplified. Moreover, Suzuki and Giovannoni (1996) observed that in reactions with greater than 35 cycles, a 1:1 ratio of products occurred regardless of the initial ratio of target sequences. PCR can also produce artifacts. Chimeras, molecules formed from parts of two different sequences, can comprise approximately 10% of PCR amplified products and appear to increase in frequency with cycle number (von Wintzingerode et al. 1997). The DNA polymerase in PCR can mis-incorporate bases during amplification resulting in sequencing artifacts. Eckert and Kunkel (1990) calculated Taq error rates as high as  $3 \times 10^{-3}$ , and Acinas et al. (2005) identified Taq DNA polymerase error as the primary cause of artifacts during the construction of rRNA clone libraries.

The cloning of amplified products can be another significant source of bias. Factors that may lead to cloning bias include the expression of deleterious genes (a significant concern when cloning phage genes), a decrease in cloning efficiency with increasing insert size, the formation of heteroduplexes, and inappropriate antibiotic resistance (Kanagwa 2003). For example, Rainey et al. (1994) observed significant differences in community composition between clone libraries constructed with different cloning systems from the same sample. Although PCR-based methods can be effective in

characterizing the richness of a community as well as the identity and phylogeny of its members, estimates of evenness are dubious, because of the biases discussed above.

The advantages of the method described above are that it is relatively simple and inexpensive (compared with a metagenomic approach for example). Using this approach, novel marine RNA viruses are being discovered and characterized at an increasing rate (Lang et al. 2009; Nagasaki 2008). The primers used in the assay can be redesigned to incorporate the most current available sequence information. Moreover, this assay is not limited to picorna-viruses, and can be targeted toward any group of RNA viruses that have an appropriate molecular marker (e.g., RdRP, helicase, or capsid genes).

## Protocol 2. Construction of shotgun libraries from RNA virus assemblages

The following protocol is used to construct shotgun libraries from RNA virus assemblages. This method is designed to detect all RNA viruses, regardless of their genome orientation, and therefore provides a broader assessment of RNA virus diversity compared to the single-gene-based method described in Protocol 1.

Metagenomics is loosely defined as the analysis of genome sequences from a community of organisms. In the field of marine virology, metagenomic approaches have been predominantly used to characterize natural assemblages of marine DNA phages (see Edwards and Rohwer 2005 for a review). These studies have generated an immense quantity of sequence data that has resulted in the discovery of a significant number of novel viral genes and has revealed the extraordinary diversity of DNA viruses in the ocean.

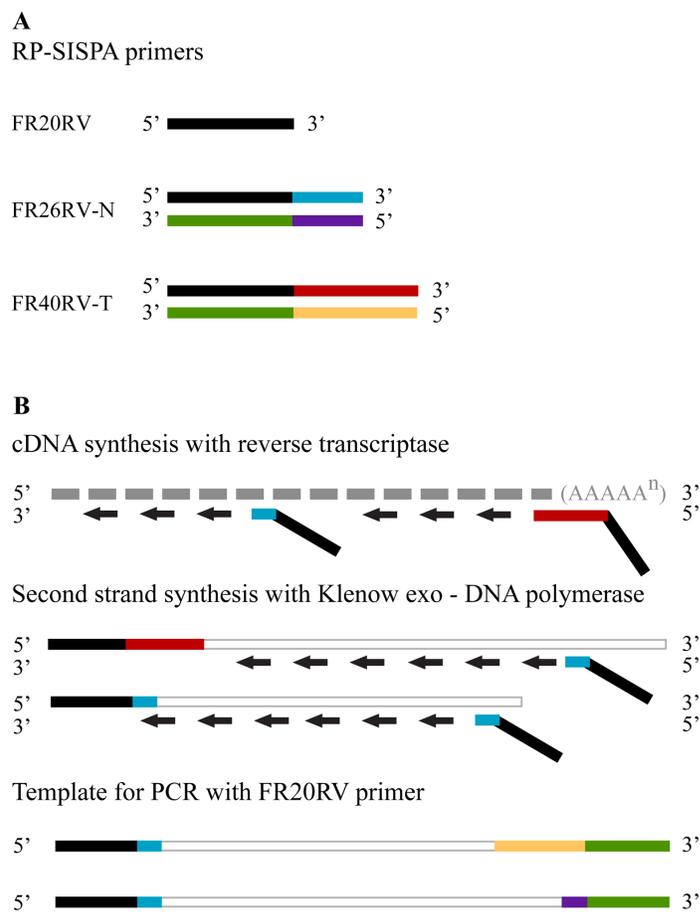
The diversity of RNA virus assemblages in human feces (Zhang et al. 2006b; Victoria et al. 2009), reclaimed water (Rosario et al. 2009), and a freshwater lake (Djikeng et al. 2009) have been characterized with metagenomic methods. These studies detected unexpected viral genomes from each respective sampling environment as well as sundry novel RNA virus genes. The only metagenomic analysis of marine RNA virus assemblages to date characterized two marine RNA virus communities from coastal British Columbia (Culley et al. 2006). In this study, most of the sequences were unrelated to known sequences. For the sequences that were related to known sequences, positive-sense ssRNA viruses that are distant relatives of known RNA viruses dominated the samples. One RNA virus library was characterized by diverse picorna-like viruses, while the second library was dominated by viruses related to members of the family *Tombusviridae* and genus *Umbravirus* (Culley et al. 2006). In the latter library, 59% of the sequence fragments that formed overlapping contiguous sections fell into one segment. Similarly, 66% of JP sequence fragments contributed to only four contigs (Culley et al. 2006). These sequences presumably represent the most abundant viruses in the two marine RNA virus communities sampled. This first "snap shot" suggests that marine RNA viruses have

primarily eukaryotic hosts and are orders of magnitude less diverse than DNA viruses.

There are two primary challenges to constructing shotgun libraries of marine RNA virus communities: producing a purified fraction of viral RNA from seawater and the unbiased production of DNA appropriate for sequencing from a relatively small amount of RNA template. We have constructed libraries with a low percentage of non-viral sequences from concentrated RNA virus assemblages that have simply been filtered and enzymatically treated (Culley et al. 2006). However, we have subsequently found that filtration-only is ineffective in removing contaminating nucleic acid from some samples. To consistently produce the necessary purity of viral RNA for library construction, we recommend that filtration is followed by two rounds of density gradient purification of the viruses (Lawrence and Steward 2010, this volume, Protocol B), followed by DNase treatment in order to remove contaminating DNA. There are several means of producing microgram amounts of double-stranded DNA from sub-nanogram amounts of RNA beyond the method described here. Examples include 1) linear amplification of RNA (Frias-Lopez et al. 2008) and ds cDNA synthesis, 2) RNA-tailing (Botero et al. 2005) and PCR, 3) cDNA tailing and PCR (Iscove et al. 2002), 4) ds cDNA synthesis, primer adapter addition and PCR (Culley et al. 2006; Zhang et al. 2006b), and 5) random ds cDNA synthesis and multiple displacement amplification (Berthet et al. 2008; Rosario et al. 2009). These alternatives may prove to be less biased and more sensitive and merit further investigation. However, to the best of our knowledge, the accuracy, reproducibility, and limitations of these methods have not been evaluated in a systematic manner and/or the protocol has not been used with RNA viruses.

The method described below, based on the work of Reyes and Kim (1991), Froussard (1992), and Allander et al. (2005), is referred to as random priming-mediated sequence-independent single-primer amplification or RP-SISPA, and was developed by Djikeng et al. (2008) to generate whole genome shotgun libraries of virus communities. Victoria et al. (2008) described a similar approach to characterize virus diversity from tissue samples. In RP-SISPA, cDNA is synthesized from a DNA-free viral RNA template with a primer mixture that contains two types of primers. The first primer type, FR26RV-N, has a 5', 20-bp primer sequence (FR20RV), and a 3' random hexamer (N6) segment. The second primer type (FR40RV-T) has the same 5' 20 bp primer (FR20RV) sequence coupled with a 3' poly T tail (Figure 1A). Second strand synthesis is primed by reannealing the cDNA primers (FR26RV-N) remaining from the RT reaction to the newly synthesized cDNA strand and catalyzed by Klenow exo - DNA polymerase (Figure 1B). Single-strand PCR is performed with the double stranded cDNA template and a single 20 bp primer (FR20RV) that is the same sequence as the 5' cDNA primers (Figure 1B). PCR products are visualized on a gel and DNA of the target size can be excised and either cloned or processed directly for pyrosequencing. An evaluation of the accuracy of RP-SISPA by Djikeng et al. (2008) demonstrated

that the method consistently produced libraries with > 90% genome coverage and 15-fold depth of coverage of positive and negative sense RNA virus genomes. Subsequently, RP-SISPA was used to characterize seasonal differences in the diversity of RNA viruses in a freshwater lake (Djikeng et al. 2009). Just as in the marine environment, a majority of sequences did not show significant similarity to known sequences. The sequences that were characterized suggest that the lake RNA virus assemblage includes pathogens of protists, plants, insects, fish, and humans (Djikeng et al. 2009).



**Figure 1.** RP-SISPA Schematic. This figure is based on Djikeng et al. (2008). Figure 1A shows the three primers used in RP-SISPA. FR26RV-N is composed of the 20 bp 5' primer sequence, FR20RV (black bar), and a random hexamer at the 3' end (light blue bar). Primer FR40RV-T is a composite of the FR20RV primer (black bar) and a 20 bp poly T tail (red bar). FR20RV (black bar) is used in PCR. The reverse complement of FR26RV-N (green and orange bar) and FR40RV-T (green and yellow bar) are shown. In Figure 1B, cDNA is synthesized from viral RNA (gray segmented line) with reverse transcriptase primed with FR26RV-N and FR40RV-T. To synthesize the second strand, FR26RV-N is annealed to the newly synthesized cDNA strand where it primes Klenow exo - DNA polymerase "gap filling" activity, resulting in an FR20RV site on both 5' and 3' ends. The DNA templates, which represent random stretches of the initial viral RNA genomes, are then amplified with the FR20RV primer to generate more material for cloning and sequencing.

## Material and equipment

**Equipment**—Thermocycler, heating block, incubator, gel electrophoresis unit, gel documentation system

**Supplies**—1.7 mL sterile, nucleic acid-free microcentrifuge tubes; 0.2 mL sterile, nucleic acid-free PCR tubes; sterile razor blades; QIAamp Viral RNA Mini kit (Qiagen); MinElute Gel Extraction kit (Qiagen); MinElute PCR Purification kit (Qiagen); MinElute Reaction Cleanup kit (Qiagen); PCR Terminator End Repair kit (Lucigen); CloneSmart HCKan Blunt Cloning kit with Ecloni Supreme cells (Lucigen); Turbo DNA-free Kit (Applied Biosystems); electroporation cuvettes

**Solutions, reagents, and media**—Superscript III Reverse Transcriptase and buffers (Invitrogen), 10  $\mu$ M dNTP mix, RNase Out (Invitrogen), Klenow Fragment, 3'-5' exo - (New England Biolabs); primers (Table 1): 0.5  $\times$  TBE (45 mM Tris-borate, 1 mM EDTA [pH 8.0]) electrophoresis buffer; nucleic acid-free, sterile water; Ampligold Taq Polymerase and buffers (Applied Biosystems); 0.02-filtered SM Buffer (100 mM NaCl, 8 mM MgSO<sub>4</sub>, 50 mM Tris pH 7.5)

## Steps

**Sample collection, purification, and extraction**—The ideal starting amount for this method is on the order of nanograms or greater of purified viral RNA, however the sensitivity of RT-PCR suggests that sub-nanogram starting template will be successful. Because of how little RNA is present in a viral genome, the virus community in tens to thousands of liters of seawater must be isolated and concentrated before one can proceed with shotgun library construction. Other chapters in this book (Wommack et al. 2010, this volume; Steward and Culley 2010, this volume) discuss approaches to concentrating viral communities from seawater.

RP-SISPA will not discriminate between viral and non-viral nucleic acids, and it is therefore critical that only pure viral RNA is present as template. We have found that purification through two sequential cesium chloride gradients (Protocol B in Lawrence and Steward 2010, this volume) is effective for removing contaminating cellular and exogenous nucleic acids prior to library construction.

Once the cesium chloride fractions have been collected and concentrated with a centrifugal ultrafiltration unit as described, viruses are recovered by eluting in 3  $\times$  50  $\mu$ L of 0.02-filtered SM buffer. Each density fraction is then extracted with the QIAamp Viral RNA Mini kit (Qiagen) as directed by the manufacturer with the following exception. We do not add carrier RNA as suggested to the "AL" lysis buffer to avoid introducing non-target RNA to the sample. Note that there is no RNase treatment before extraction because exogenous RNA will pellet and is thus removed during the CsCl fractionation procedure. We recommend quantifying the RNA from each density fraction to identify in which fraction the concentration of viral RNA is highest. The template for RP-SISPA can be RNA extracted from multiple

pooled fractions or from a single fraction depending on your research objectives.

**Enzymatic treatment**—As in Protocol 1, we remove contaminating DNA from our RNA preparations with the Turbo DNA-free kit (Applied Biosystems) as described in the protocol provided with the kit.

### RP-SISPA

**cDNA synthesis:** In preparation for cDNA synthesis, 10  $\mu$ L purified RNA viral template is mixed with a final dNTP concentration of 0.2 mM and 1  $\mu$ M and 5 nM final concentrations of FR26RV-N and FR40RV-T primer, respectively (see Table 1 for the sequence of each primer). FR40RV-T is added to take advantage of the fact that a majority of characterized RNA virus genomes have poly(A) tails. The addition of a poly(T) primer may increase the likelihood of the 3' ends being sequenced (Figure 1B). The reaction is heated to 65°C then cooled on ice to allow the primers to anneal. While still on ice, DTT (0.5 mM final conc.) is added to the reaction as an enzyme stabilization reagent with 40 U RNase OUT (Invitrogen) to protect the sample from RNase activity. The complementary DNA strand is synthesized with 200 U of Superscript III (Invitrogen) reverse transcriptase. The final reaction volume should be 20  $\mu$ L. The reaction is incubated initially at 25°C for 10 min so that the hexamer 3' end of primer FR26RV-N and the poly(T)<sub>20</sub> 3' end of primer FR40RV-T remain annealed to the template while cDNA synthesis commences. The temperature is then increased to 50°C, the temperature at which Superscript III's processivity is highest, for 60 min.

**Second strand synthesis:** In a simple and elegant step, the second strand synthesis reaction results in a ds cDNA template with a primer site added to both the 5' and 3' end (Figure 1B). After the hour-long incubation at 50°C, the first strand synthesis reaction is heated immediately to 94°C for 3 min and then rapidly cooled on ice. This step results in the reannealing of excess FR26RV-N primer to the nascent cDNA strand. A complementary second strand is subsequently synthesized at 37°C for 60 min with the addition of 2.5 U of Klenow Fragment, 3'-5' exo - (New England Biolabs). The Klenow reaction is terminated with a final incubation at 75°C for 10 min.

**PCR:** PCR is used to produce a sufficient quantity of DNA from the ds cDNA template from the second strand reaction for sequencing (Figure 1B). One PCR reaction contains 5  $\mu$ L of template taken directly from the second strand synthesis reaction, 40 pM of FR20RV primer (see Table 1), a final dNTP concentration of 0.2 mM, 1  $\times$  Gold buffer, 2.5 mM MgCl<sub>2</sub>, and 2.5 U of Ampligold DNA polymerase (Applied Biosystems) in a final volume of 50  $\mu$ L. The reaction is incubated at 94°C for 10 min to fully denature the template and activate the hot start enzyme (we have found that a hot start is absolutely essential), followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 65°C for 1 min, and extension at 72°C for 2 min and a final extension for 13 min that permits the completion of complementary strand synthesis.

Before gel separation, we purify and concentrate the PCR reactions with a MinElute PCR cleanup column (Qiagen) as described by the manufacturer. This reduces the thickness of the gel, ultimately resulting in a more efficient recovery of target DNA from the excised band. Purified PCR products are loaded onto a 1% agarose gel containing 1 × SYBR safe stain (Invitrogen) and 0.5 × TBE buffer. Bands of DNA of the appropriate size range (the size range excised should be based on what type of sequencing method is being used; for example, we target the 800-2000 bp size range for Sanger sequencing) are excised and purified with a MinElute Gel Extraction kit (Qiagen) according to the manufacturer's instructions. When visualizing the gel, prolonged exposure to UV irradiation can damage the DNA and greatly reduce downstream cloning efficiency. An illuminator with blue light-emitting bulbs is ideal for gel visualization. If this is not available, take steps to reduce the exposure of the gel to ultraviolet irradiation during excision as much as possible. To mitigate the biases associated with high cycle number PCR (see the "Assessment" section of Protocol 1), we recommend pooling the products from multiple PCR reactions using the same ds cDNA template.

If the sample is to be cloned for Sanger sequencing, we recommend eluting DNA from the column with three washes of 10 µL nuclease-free water in preparation for the PCR Terminator (Lucigen) end repair reaction. For the cloning protocol, please refer to the "Cloning and sequencing" section of Protocol 1 in this chapter. The procedure from this point forward is the same. The material purified from the gel may instead be processed for pyrosequencing.

### Assessment

The primary strengths of the RP-SISPA are that it requires no previous knowledge of the composition of the target RNA virus assemblage, it has a low detection limit (nanograms or less) and that it is a relatively straightforward, inexpensive procedure that uses equipment standard to any molecular laboratory. A systematic comparison of metagenomic approaches to characterizing marine RNA virus diversity is not available. However, several different methods have been developed to accurately interrogate the transcriptome of single cells (Peano et al. 2006). These methods share several of the same challenges as the examination of natural communities of RNA viruses, starting with very low concentrations of starting RNA template. Surprisingly, those methods that include an exponential amplification step after the addition of primer sites to the target template (such as RP-SISPA) generally introduced less bias than other approaches such as linear RNA amplification (Iscove et al. 2002; Subkhankulova and Livesey 2006). Presumably those protocols that rely on multiple displacement amplification during RNA virus shotgun library construction are subject to the same high incidence of chimera formation as found in previous studies (Zhang et al. 2006a; Lasken and Stockwell 2007), however this has yet to be demonstrated.

The limitations of the RP-SISPA method included the general biases associated with RT-PCR discussed in the assessment section of Protocol 1. RP-SISPA appears to achieve full coverage of RNA viral genomes less than or equal to 15 kb in size, but only recovered an average of 50% of a 50 kb ds DNA bacteriophage genome. Furthermore sequences from the 3' end of all RNA viral genomes tested were consistently underrepresented, as expected with a method based on randomly primed RT. Gaps in the sequence occurred and may have been due to regions of secondary structure, extreme codon bias, or shearing of the RNA template (Djikeng et al. 2009). Finally, the requirement of having viral RNA free from contamination can be a significant challenge with some samples.

### Conclusions

Study of the ecology of marine RNA viruses is still in its infancy and thus the number of methods available are limited. In this chapter, we have presented two molecular approaches to determining RNA viral diversity from seawater samples. The first is a relatively rapid and inexpensive approach based on the amplification of a single-gene, whereas the second method, based on the sequencing of whole genome shotgun libraries, requires more investment but produces a more comprehensive picture of the RNA virus community. The choice of method is dependent on the research objective and the resources available. We hope that this chapter will kindle interest in marine RNA viruses and serve as an informative starting point for further research and development on this understudied component of the plankton.

### References

- Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. 2005. PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* 71:8966-8969.
- Allander, T., M. T. Tammi, M. Eriksson, A. Bjerkner, A. Tiveljung-Lindell, and B. Andersson. 2005. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc. Natl. Acad. Sci. U.S.A.* 102:12891-12896.
- Angly, F. E., and others. 2006. The marine viromes of four oceanic regions. *PLoS Biol.* 4:2121-2131.
- Bench, S. R., and others. 2007. Metagenomic characterization of Chesapeake bay viroplankton. *Appl. Environ. Microbiol.* 73:7629-7641.
- Berthet, N., and others. 2008. Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. *BMC Mol. Biol.* 9:77.
- Botero, L. M., S. D'imperio, M. Burr, T. R. Mcdermott, M. Young, and D. J. Hassett. 2005. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Appl. Environ. Microbiol.* 71:1267-1275.
- Breitbart, M., and others. 2004. Diversity and population

- structure of a near-shore marine sediment viral community. *Proc. Royal Soc. Biol. Sci. B* 271:565-574.
- , and others. 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U.S.A.* 99:14250-14255.
- Brussaard, C. P. D., and J. M. Martinez. 2008. Algal bloom viruses. *Plant Viruses* 2:1-13.
- Chen, F., C. A. Suttle, and S. M. Short. 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Appl. Environ. Microbiol.* 62:2869-2874.
- Culley, A. I., A. S. Lang, and C. A. Suttle. 2003. High diversity of unknown picorna-like viruses in the sea. *Nature* 424:1054-1057.
- , ———, and ———. 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312:1795-1798.
- , and G. F. Steward. 2007. New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Appl. Environ. Microbiol.* 73:5937-5944.
- Djikeng, A., and others. 2008. Viral genome sequencing by random priming methods. *BMC Genomics* 9:5.
- , R. Kuzmickas, N. G. Anderson, and D. J. Spiro. 2009. Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS ONE* 4:e7264.
- Eckert, K. A., and T. A. Kunkel. 1990. High fidelity DNA-synthesis by the *Thermus-aquaticus* DNA-polymerase. *Nucleic Acids Res.* 18:3739-3744.
- Edwards, R. A., and F. Rohwer. 2005. Viral metagenomics. *Nat. Rev. Microbiol.* 3:504-510.
- Filée, J., F. Tétart, C. A. Suttle, and H. M. Krisch. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc. Natl. Acad. Sci. U.S.A.* 102:12471-12476.
- Frank, H., and K. Moebus. 1987. An electron microscopic study of bacteriophages from marine waters. *Helgol. Mar. Res.* 41:385-414.
- Frias-Lopez, J., and others. 2008. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U.S.A.* 105:3805-3810.
- Froussard, P. 1992. A random-PCR method (RPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Res.* 20:2900.
- Fuller, N. J., W. H. Wilson, I. R. Joint, and N. H. Mann. 1998. Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. *Appl. Environ. Microbiol.* 64:2051-2060.
- Helton, R. R., and K. E. Wommack. 2009. Seasonal dynamics and metagenomic characterization of estuarine virobenthos assemblages by randomly amplified polymorphic DNA PCR. *Appl. Environ. Microbiol.* 75:2259-2265.
- Hidaka, T., and K. Ichida. 1976. Properties of a marine RNA-containing bacteriophage. *Mem. Fac. Fish. Kagoshima Univ.* 25:77-89.
- Iscove, N. N., M. Barbara, M. Gu, M. Gibson, C. Modi, and N. Winegarden. 2002. Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat. Biotechnol.* 20:940-943.
- Kanagwa, T. 2003. Bias and artifacts in multi-template polymerase chain reaction (PCR). *J. Biosci. Bioeng.* 96:317-323.
- Lang, A. S., M. L. Rise, A. I. Culley, and G. F. Steward. 2009. RNA viruses in the sea. *FEMS Microbiol. Rev.* 33:295-323.
- Larsen, J. B., A. Larsen, G. Bratbak, and R. A. Sandaa. 2008. Phylogenetic analysis of members of the *Phycodnaviridae* virus family, using amplified fragments of the major capsid protein gene. *Appl. Environ. Microbiol.* 74:3048-3057.
- Lasken, R. S., and T. B. Stockwell. 2007. Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol.* 7:19.
- Lawrence, J. E., and G. F. Steward. 2010. Purification of viruses by centrifugation, p. 166-181. *In* S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle [eds.], *Manual of Aquatic Viral Ecology*. ASLO.
- Le Gall, O., and others. 2008. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T=3 virion architecture. *Arch. Virol.* 153:715-727.
- Munn, C. B. 2006. Viruses as pathogens of marine organisms - from bacteria to whales. *J. Mar. Biolog. Assoc. U.K.* 86:453-467.
- Nagasaki, K. 2008. Dinoflagellates, diatoms, and their viruses. *J. Microbiol.* 46:235-243.
- , and others. 2004. Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Appl. Environ. Microbiol.* 70:704-711.
- Peano, C., M. Severgnini, I. Cifola, G. De Bellis, and C. Battaglia. 2006. Transcriptome amplification methods in gene expression profiling. *Expert Rev. Mol. Diagn.* 6:465-480.
- Polz, M. F., and C. M. Cavanaugh. 1998. Bias in template-to-product ratios in multi-template PCR. *Appl. Environ. Microbiol.* 64:3724-3730.
- Rainey, F. A., N. Ward, L.I. Sly, and E. Stackerbrandt. 1994. Dependence on the taxon composition of clone libraries for PCR amplified, naturally occurring 16S rDNA, on the primer pair and the cloning system used. *Experientia* 50:796-797.
- Rappé, M. S., and S. J. Giovannoni. 2003. The uncultured microbial majority. *Ann. Rev. Microbiol.* 57:369-394.
- Reyes, G. R., and J. P. Kim. 1991. Sequence-independent, single-primer amplification (sispa) of complex DNA populations. *Mol. Cell. Probes* 5:473-481.
- Roberts, J. D., K. Bebenek, and T. A. Kunkel. 1988. The accuracy of reverse-transcriptase from HIV-1. *Science* 242:1171-1173.
- Rosario, K., C. Nilsson, Y. Lim, Y. Ruan, and M. Breitbart. 2009. Metagenomic analysis of viruses in reclaimed water. *Environ. Microbiol.* 11:2806-2820.

- Shirai, Y., Y. Tomaru, Y. Takao, H. Suzuki, T. Nagumo, and K. Nagasaki. 2008. Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Appl. Environ. Microbiol.* 74:4022-4027.
- Short, S. M., and C. A. Suttle. 2002. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl. Environ. Microbiol.* 68:1290-1296.
- Stahlberg, A., J. Hakansson, X. J. Xian, H. Semb, and M. Kubista. 2004. Properties of the reverse transcription reaction in mRNA quantification. *Clin. Chem.* 50:509-515.
- Steward, G. F., J. L. Montiel, and F. Azam. 2000. Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnol. Oceanogr.* 45:1697-1706.
- , and A. I. Culley. 2010. Extraction and purification of nucleic acids from viruses, p. 154-165. *In* S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle [eds.], *Manual of Aquatic Viral Ecology*. ASLO.
- Subkhankulova, T., and F. J. Livesey. 2006. Comparative evaluation of linear and exponential amplification techniques for expression profiling at the single-cell level. *Genome Biol.* 7: R18.1-R18.16.
- Suzuki, M. T., and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* 62:625-630.
- Tai, V., J. E. Lawrence, A. S. Lang, A. M. Chan, A. I. Culley, and C. A. Suttle. 2003. Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *J. Phycol.* 39:343-352.
- Takao, Y., K. Nagasaki, K. Mise, T. Okuno, and T. Honda. 2005. Isolation and characterization of a novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium* sp. (*Thraustochytriaceae, Labyrinthulea*). *Appl. Environ. Microbiol.* 71:4516-4522.
- Tomaru, Y. and others. 2004. Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. *Aquat. Microb. Ecol.* 34:207-218.
- , Y. Takao, H. Suzuki, T. Nagumo, and K. Nagasaki. 2009. Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming diatom *Chaetoceros socialis*. *Appl. Environ. Microbiol.* 75:2375-2381.
- Victoria, J. G., A. Kapoor, K. Dupuis, D. P. Schnurr, and E. L. Delwart. 2008. Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog.* 4:e1000163.
- , and others. 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J. Virol.* 83:4642-4651.
- Von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* 21:213-229.
- Wommack, K. E., T. Sime-Ngando, D. M. Winget, S. Jamindar, and R. R. Helton. 2010. Filtration-based methods for the collection of viral concentrates from large water samples, p. 110-117. *In* S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle [eds.], *Manual of Aquatic Viral Ecology*. ASLO.
- Zhang, K., and others. 2006a. Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24:680-686.
- Zhang, T., and others. 2006b. RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biol.* 4:108-118.