Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses

Article *in* The ISME Journal · May 2013

DOI: 10.1038/ismej.2013.67 · Source: PubMed

CITATIONS
113

READS
288

3 authors, including:

Ryan Schenck
University of Oxford
15 PUBLICATIONS **212** CITATIONS

SEE PROFILE

Matthew B Sullivan
The Ohio State University
462 PUBLICATIONS **20,626** CITATIONS

SEE PROFILE

Global Morphological Analysis of Marine Viruses Shows Minimal Regional Variation and Dominance of Non-Tailed Viruses

Jennifer R. Brum, Ryan O. Schenck, and Matthew B. Sullivan*

Running Title: Morphological characterization of marine viruses

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

*Corresponding author. Mailing address: 1007 E. Lowell St., Life Sciences South 203, Tucson, Arizona, 85721. Phone: (520) 626-9100. E-mail: mbsulli@email.arizona.edu

Abstract

Viruses influence oceanic ecosystems by causing mortality of microorganisms, altering nutrient and organic matter flux via lysis and auxiliary metabolic gene expression, and changing the trajectory of microbial evolution through horizontal gene transfer. Limited host range and differing genetic potential of individual virus types mean that investigations into the types of viruses that exist in the ocean and their spatial distribution throughout the world's oceans are critical to understanding the global impacts of marine viruses. Here we evaluate viral morphological characteristics (morphotype, capsid diameter, and tail length) using a quantitative transmission electron microscopy (qTEM) method across six of the world's oceans and seas sampled through the Tara Oceans Expedition. Extensive experimental validation of the qTEM method shows that neither sample preservation nor preparation significantly alters natural viral morphological characteristics. The global sampling analysis demonstrated that morphological characteristics did not vary consistently with depth (surface versus deep chlorophyll maximum waters) or oceanic region. Instead, temperature, salinity, and oxygen concentration, but not chlorophyll *a* concentration, were more explanatory in evaluating differences in viral assemblage morphological characteristics. Surprisingly, given that the majority of cultivated bacterial viruses are tailed, non-tailed viruses appear to numerically dominate the upper oceans as they comprised 51–92% of the viral particles observed. Together these results document global marine viral morphological characteristics, show that their minimal variability is more explained by environmental conditions than geography, and suggest that non-tailed viruses might represent the most ecologically important targets for future research.

Introduction

Viruses are key players in the Earth's ecosystem not only because they are the most abundant and diverse biological entities in marine environments (reviewed by Wommack and Colwell, 2000; Breitbart et al., 2007), but also because they have considerable influence on ecological, biogeochemical, and evolutionary processes in the ocean (reviewed by Fuhrman, 1999; Weinbauer, 2004; Suttle, 2007; Breitbart, 2012). Viral-induced mortality of microorganisms in the ocean can affect microbial species composition (Thingstad, 2000) and alter the flux of nutrients and organic matter by increasing recycling of these materials through the microbial loop (reviewed by Fuhrman, 1999). Expression of viral auxiliary metabolic genes (*sensu* Breitbart et al., 2007), such as core photosystem genes, during infection may also substantially impact oceanic productivity (Lindell et al., 2005; Clokie et al., 2006; Lindell et al., 2007; Sharon et al., 2007; Dammeyer et al., 2008; Thompson et al., 2011). In addition, viral-mediated horizontal gene transfer can profoundly alter the evolution of oceanic microorganisms as has been demonstrated in marine cyanobacteria (e.g., Lindell et al., 2004; Sullivan et al., 2006; Ignacio-Espinoza and Sullivan, 2012).

With these significant roles in oceanic ecosystems, it is important to understand the characteristics of marine viruses and their distribution in the oceans. The majority of marine viruses are thought to infect bacteria (Wommack and Colwell, 2000) and taxonomic surveys based on the bacterial 16S rRNA gene have shown that bacterial assemblages vary between oceanic regions (Schattenhofer et al., 2009; Barberan et al., 2012). Thus, one would expect viral assemblages to vary between oceanic regions as well. Viruses do not have a universal marker gene so assessing their diversity across spatial scales is challenging and has resulted in the use of metagenomics to compare viral assemblages from different environments (Breitbart et al., 2004b; Angly et al., 2006; Dinsdale et al., 2008; Hurwitz and Sullivan, 2013). The first study to

compare marine water column viral metagenomes showed that viral assemblage genetic distance increases with geographic distance, but also that there is considerable overlap in viral assemblages across sites even though constituent viral abundances vary (Angly et al., 2006). In fact, one particular podovirus DNA polymerase sequence is present in several aquatic and terrestrial environments (Breitbart et al., 2004a). A much larger-scale Pacific Ocean viral metagenomic dataset (Hurwitz and Sullivan, 2013) employing quantitative methodologies (John et al., 2011; Duhaime and Sullivan, 2012; Duhaime et al., 2012; Hurwitz et al., 2012) is now available to examine biogeography, but such studies have not yet been conducted. This is because the database representation for sequence comparisons are so poor that most ocean viruses are not yet identifiable (e.g., Angly et al., 2006; Hurwitz and Sullivan, 2013). Thus simple questions such as how viral assemblages vary across oceanic regions remain unanswered.

An alternative to metagenomics is comparing viral assemblages throughout the world's oceans using morphology. Viral morphology is central to modern viral taxonomy (King et al., 2012), and commonly correlates with whole-genome-derived taxonomy (Rohwer and Edwards, 2002) and aspects of their biology (reviewed by Ackermann, 2001). Thus, morphological metrics have applications ranging from medical diagnostics (Doane, 1980) to environmental virology (e.g., Bratbak et al., 1990; Weinbauer and Peduzzi, 1994). In aquatic environments, morphological metrics documented spatiotemporal changes in viral assemblages, revealing aquatic viruses as dynamic and varied across large environmental gradients (Bratbak et al., 1990; Auguet et al., 2009; Brum and Steward, 2010; Bettarel et al., 2011b; Bettarel et al., 2011a). Environmental morphological studies also aid viral discovery, finding novel morphologies including large viruses (Bratbak et al., 1992; Gowing, 1993; Sommaruga et al., 1995), spindle-shaped viruses (Oren et al., 1997), and filamentous viruses (Hofer and Sommaruga, 2001).

Finally, morphological analyses are not plagued by the database bias issues (Edwards and Rohwer, 2005) that undermine quantitative viral taxonomic analyses in metagenomic studies.

Sample preparation, however, has only recently been resolved for quantitative viral metagenomic studies (reviewed in Duhaime and Sullivan, 2012), and remains an obstacle to being quantitative in environmental viral morphological studies. Transmission electron microscopy (TEM) sample preparation generally includes one of two approaches: either viruses are concentrated and then adsorbed to TEM grids (e.g., Sommaruga et al., 1995; Stopar et al., 2003), or they are directly deposited onto TEM grids using traditional (e.g., Bergh et al., 1989) or air-driven ultracentrifugation (Maranger et al., 1994; Brum and Steward, 2010). Here, we use an air-driven ultracentrifuge with a rotor designed to quantitatively deposit viruses onto TEM grids (Hammond et al., 1981), resulting in high recovery of viruses (Maranger et al., 1994). We evaluate this quantitative TEM (qTEM) method to determine the best conditions for sample collection and processing, as well as its biases when applied to marine samples. Using qTEM, we then document viral morphological diversity in the upper water column at 14 stations in 6 global ocean regions using highly-contextualized samples collected on the Tara Oceans Expedition (Karsenti et al., 2011).

Materials and Methods

qTEM method

Viruses were deposited onto TEM grids with an air-driven ultracentrifuge (Airfuge CLS, Beckman Coulter, Brea, CA, USA) as previously described (Brum and Steward, 2010) except that grids were rendered hydrophilic using 20s of glow discharge with a sputter coater (Hummer 6.2, Anatech, Union City, CA, USA). A detailed protocol, including suggestions from the

scientific community, is maintained at http://eebweb.arizona.edu/faculty/mbsulli/protocols.htm. Deposited material was then positively stained by immersing the grid into 2% uranyl acetate (Ted Pella, Redding, CA, USA) for 30s followed by three, 10-s washes in ultra-pure water (Milli-Q, Millipore, Billerica, MA, USA), with excess liquid wicked away with filter paper. Grids were then dried at ambient conditions overnight and stored desiccated until analysis. Positive staining was chosen because negative staining results in uneven staining on grids that would introduce observational bias to the analysis and undermine the goal of a quantitative method.

Prepared grids were examined at 65,000-100,000× magnification using a transmission electron microscope (Philips CM12, FEI, Hilsboro, OR, USA) with 100kV accelerating voltage. Micrographs were collected using a Macrofire Monochrome CCD camera (Optronics, Goleta, CA, USA). Viruses were classified as myoviruses, podoviruses, siphoviruses, or icosahedral non-tailed viruses (referred to as non-tailed viruses hereafter) based on their morphology as defined by the International Committee on Taxonomy of Viruses (King et al., 2012). Viral capsid diameters and tail lengths were measured using ImageJ software (Abramoff et al., 2004).

qTEM method evaluation

Several variables were tested to evaluate sample collection, sample processing, and biases inherent in the qTEM method. First, we determined the number of viruses needed per sample to accurately assess morphological characteristics. A 400-µl unfiltered seawater sample from the Biosphere 2 Ocean environment (Oracle, AZ, USA) was deposited onto a grid. Morphotype composition and viral capsid diameter distributions were then compared for the first 50, 100, and 200 viruses observed.

We next evaluated the effects of freezing on viral morphology. Water collected from the Biosphere 2 Ocean was preserved with EM-grade glutaraldehyde (2% final concentration, Sigma). One 400-µl volume was processed immediately (termed 'fresh') using the qTEM method, while another 400µl was flash-frozen in liquid nitrogen (termed 'frozen'), thawed at room temperature, and then similarly processed. Images of 100 viruses per treatment were analyzed to compare morphotype composition and capsid diameter distributions between treatments.

Finally, we evaluated the extent of tail loss resulting from the qTEM method. Water samples (20ml) from Scripps Pier (San Diego, CA, USA), Beaufort Inlet (Beaufort, NC, USA), and Kaneohe Bay (Kaneohe, HI, USA) were filtered through 0.22-µm pore-size filters (Steripak, Millipore), stored in the dark at 4ºC, and concentrated to 250µl with 100kDa cut-off centrifugal filter units (Amicon, Millipore). Triplicate grids were prepared from these concentrated samples using each the qTEM method described above (50-µl volumes), and the adsorption method (Ackermann and Heldal, 2010), where a 10-µl volume was placed on a hydrophilic grid for 10 minutes followed by positive staining of viruses adsorbed to the grid. One hundred viruses per grid were analyzed for viral morphotype composition as described above.

Tara Oceans sample collection

Samples were collected from 14 Tara Oceans Expedition stations in the Mediterranean Sea, Red Sea, Arabian Sea, Indian Ocean, Atlantic Ocean, and Pacific Ocean (Figure S1, Table S1). A rosette equipped with a CTD (Sea-Bird Electronics; SBE 911*plus* with Searam recorder), dissolved oxygen sensor (Sea-Bird Electronics; SBE 43) and fluorometer (WET Labs; ECO-FLrtd) was used to obtain environmental context for each station.

Samples for qTEM analysis were collected from the surface and deep chlorophyll maximum (DCM) using a peristaltic pump, except for DCM samples at stations 30 and 98 where Niskin bottles were used. Samples (2ml) were preserved with EM-grade glutaraldehyde (final concentration 2%), flash-frozen and stored in liquid nitrogen aboard the ship and at -80ºC on land until analysis. Samples (400µl) were thawed at room temperature (ca. 22ºC) and prepared using the qTEM method. Micrographs of 100 viruses per sample were collected and analyzed for viral morphotype, capsid diameter, and tail length.

Statistical analyses

For qTEM method evaluations, upper and lower 95% confidence intervals of viral morphotypes were calculated according to Zar (1996), and binomial regression to compare proportions of viral morphotypes was done with JMP statistical software (SAS, Cary, NC, USA). Morisita's index of similarity (Krebs, 1999), which ranges from zero (no similarity) to slightly greater than 1 (completely similar), was used to compare viral capsid diameter distributions. Sigmaplot (Systat Software, San Jose, CA, USA) was used to perform statistical tests to compare sets of data. Several of the data sets in this study could not be normalized, therefore non-parametric statistics were used in these cases.

Correspondence analysis (CA) was performed using the vegan package (Oksanen et al., 2013) in R version 2.15.2 (R Core Team, 2012) to obtain an ordination plot of viral assemblages based on histograms of viral capsid diameters from each Tara Oceans sample (omitting the station 36 surface sample due to lack of oxygen data). Vectors and response surfaces of environmental variables were fitted to the CA ordination plot using the function 'envfit' in vegan with 10,000 simulations to estimate p-values, and the function 'ordisurf' in vegan, respectively

(Wood, 2011; Oksanen et al., 2013). These analyses were performed using histogram data generated with the average optimal capsid diameter bin size for all samples determined with the 'hist' function in R using the method of Sturges (1926). Sensitivity to bin size was explored by repeating the analyses using the lower and upper limits of the optimal bin size for all samples.

Results

Evaluation of the qTEM method

Several experiments were conducted to rigorously evaluate the qTEM method as follows. First, there was no significant difference when analyzing 50, 100, or 200 viruses per sample by viral morphotype composition (Figure S2A) or capsid diameter distribution (Figure S2B). While more data decreased 95% confidence intervals for morphotype analysis, 100 viruses per sample best balanced accuracy, time and cost to morphologically characterize a viral assemblage and was used for all work presented here. Second, we found no significant difference between samples prepared immediately (fresh) and those prepared after storage in liquid nitrogen (frozen) for either viral morphotype composition (Figure S2C) or capsid diameter distributions (Figure S2D). Third, the percent of each viral morphotype was not significantly different between samples prepared using either the adsorption or qTEM methods with seawater from three marine environments (Figure S3). This suggested that the qTEM method did not cause tail loss. Thus, sample storage and qTEM preparation does not significantly alter morphological characteristics of marine viral assemblages.

Morphological characteristics of oceanic viral assemblages by depth and oceanic region

The Tara Oceans samples were collected from the surface and DCM of 14 stations in 6 oceanic regions with a range of environmental conditions (Table S1). Across 2600 viruses and 26 samples examined, only four viral morphotypes were observed: myoviruses, podoviruses, siphoviruses, and non-tailed viruses (Figure 1). Overall, viral morphotype composition and capsid diameter were remarkably consistent with depth and oceanic region (Figure 2; details for each sample in Figures S4-S9). Non-tailed viruses dominated in each depth and oceanic region (average 66–85%), while myoviruses, podoviruses, and siphoviruses were the next most abundant morphotypes, in that order, except in the Mediterranean Sea where podoviruses exceeded myoviruses (Figure 2A). Regionally, non-tailed viruses were negatively correlated with salinity and podoviruses were positively correlated with salinity (Table S2, Figure S10). For correlations among individual samples, non-tailed viruses and podoviruses were correlated with salinity while myoviruses and podoviruses were correlated with temperature (Table S2). However, these relationships reflected changes in the range of the relative percent of these morphotypes and were often driven by only three to four samples (Figure S10). No morphotype was significantly correlated with oxygen or chlorophyll concentration (Table S2).

With respect to capsid diameters, there was no significant difference between pooled surface and DCM samples (Figure 2B). Regionally, viral capsid diameters in the Mediterranean, Red, and Arabian Seas were significantly larger than those in the Indian, Atlantic, and Pacific Oceans (Figure 2B). These larger overall capsid sizes occurred in the highest salinity oceanic regions (Table S1) with average capsid diameter positively correlated with salinity for individual samples (Table S2, Figure S10). There were no significant relationships between average capsid diameter and environmental parameters when considering pooled data for oceanic regions (Table S2).

Correspondence analysis (CA) to compare sample capsid diameter distributions, as well as capsid diameter bins (Figure 3A, B), was then used to more deeply explore biogeography and the influence of environmental variables on viral assemblage morphological characteristics. Differences between surface and DCM samples were highly variable (Fig. 3A), with some surface samples more similar to the DCM sample at the same station (e.g., station 41) and others much more divergent (e.g., station 34). Further, there was no significant correlation between depth of the DCM and distance between surface and DCM samples at each station on either the CA1 or CA2 axes of the ordination plot (Pearson correlations; $p>0.3$ for both). Biogeographical differences in viral assemblages were also not well-supported, with considerable overlap between samples from each ocean and sea. In fact, the distance between samples on the CA1 or CA2 axis of the plot was not significantly correlated with geographical distance between samples considering either all samples or only surface or DCM samples separately (Pearson correlations, $p>0.4$ for all).

Environmental variables were more explanatory than geography or depth in evaluating viral assemblage morphology in the global oceans. Salinity was the most important environmental variable explaining capsid diameter distributions (CA1 was negatively correlated with salinity and explained the most inertia in the ordination plot; Figure 3A, B). Vectors and response surfaces of environmental variables showed that, while the relationship with temperature was non-linear, temperature, salinity, and oxygen, but not chlorophyll *a*, significantly influenced capsid diameter distributions (Figure 3C–F). For example, samples from the surface at station 23 and the DCM at stations 23 and 30 in the Mediterranean Sea grouped together (Figure 3A), sharing both narrow capsid diameter peaks (49-63nm; Figure S4) and similar environmental conditions (low temperature plus higher salinity and oxygen; Table S1).

In contrast, samples from the DCM at station 41 and surface at stations 34 and 41 from the Red and Arabian Seas grouped together (Figure 3A), sharing wider capsid diameter peaks (49-91nm; Figures S5 and S6) and similar environmental conditions (higher salinity and temperature, lower oxygen; Table S1). However, most samples were closer to the CA plot origin suggesting weaker influences from environmental variables (Figure 3A).

Ordination of capsid diameter bins was also influenced by environmental variables (Figure 3B). However, bins furthest from the origin tended to have the fewest viruses, although this relationship was only significant for the CA2 axis (Pearson correlation, $r=-0.659$, $p=0.004$), suggesting that bins with the most viruses were least influenced by the environmental extremes observed, resulting in relatively consistent abundances across samples. To evaluate the influence of low abundance bins (<5 viruses), the CA was repeated without them, and did not significantly change the analysis results (Table S3).

Similarly, the ordination analyses were relatively insensitive to capsid diameter bin size. Analyses using each the minimum (5nm) and maximum (10nm) optimal bin sizes determined for the samples provided similar results for the influence of environmental parameters on capsid diameters of viral assemblages (Table S3). Exceptions include reduced significance of the temperature vector and oxygen response surface with 10nm bins, most likely because this larger bin size insufficiently resolved capsid diameter distributions in most samples.

Tailed virus sample size was relatively low, reducing statistical power to evaluate spatial differences in their morphological characteristics. With this caveat, morphotype-specific tail lengths were not different between the surface and DCM samples except for siphovirus tails which were longer in surface samples (Figure 4, but note that only six siphoviruses were detected in DCM samples). Among oceanic regions, myovirus tails were longer in the Arabian Sea than

278 Mediterranean Sea, Red Sea, and Atlantic Ocean; siphovirus tails were longer in the Red Sea
279 than Mediterranean Sea; and podovirus tail lengths were not significantly different among the
280 oceanic regions (Figure 4). Correlation analyses between tail lengths and environmental
281 variables were not attempted due to low sample sizes.
282
283 *Global marine viral morphological characteristics*
284 Pooling all sample data allowed examination of overall characteristics of upper water
285 column viruses. Again, non-tailed viruses dominated (averaging 79% of all viruses), followed
286 by myoviruses, podoviruses, and siphoviruses, in that order (Figure 5A). Myoviruses had the
287 largest overall capsid diameters followed by siphoviruses, podoviruses, and non-tailed viruses,
288 with combined tailed viruses having significantly larger capsids than non-tailed viruses (Figure
289 5B). As well, tail lengths statistically differed with siphoviruses having the longest tails,
290 followed by myoviruses, then podoviruses (Figure 5C). In addition, 48% of the 27 observed
291 siphoviruses had prolate capsids and 3% of all observed myoviruses had both capsid diameters
292 and tail lengths either within or smaller than the dimensions described for dwarf myoviruses
293 (Comeau et al., 2012).
294
295 **Discussion**
296 Global ocean qTEM analyses showed that while viral assemblage morphological
297 attributes vary between samples, there is little evidence for consistent variation with depth or
298 oceanic region. The proportion of observed morphotypes (myoviruses, podoviruses,
299 siphoviruses, and non-tailed viruses) was highly similar in each oceanographic region,
300 suggesting that there are controlling factors maintaining their relative abundances in the world's

oceans. Average capsid diameter was significantly greater in the Mediterranean, Red, and Arabian Seas, but neither depth nor inter-sample geographical distance explained variations in sample capsid diameter distributions. Thus, viral morphological attributes in the upper global oceans were not explained by depth or biogeography.

Instead, environmental conditions appear to influence viral morphological characteristics. While no strong relationships between viral morphotype percentages and environmental variables emerged, larger average viral capsid diameters were significantly associated with higher salinity in individual samples. Using capsid diameter distributions as a more refined metric for viral morphology resulted in temperature, salinity, and oxygen concentration, but not chlorophyll *a* concentration, having significant influences on viral assemblages, with salinity as most explanatory. However, this effect was most evident at relative extremes of environmental conditions examined, and most samples lacked such evident environmental influence. This is probably explained by limited variations in surface ocean physico-chemical variables compared with previous studies in which freshwater to saline (Bettarel et al., 2011b) or oxic to anoxic gradients (Brum and Steward, 2010) resulted in very strong changes in viral assemblage morphological characteristics. Linking these global viral morphology data to viral genomic and bacterial taxonomic data will be the logical next step in refining our understanding of marine viral biogeography.

Only four morphotypes were observed in this study, indicating that other morphotypes (e.g., lemon-shaped or filamentous) comprised <1% of these marine viral assemblages (with 100 viruses examined per sample). Additionally, while 100 viruses per sample sufficiently characterized viral assemblages, this resulted in insufficient data to fully investigate spatial variability of tailed viral morphological attributes (e.g., tail length). We estimate that 5 to 100-

fold more viruses per sample (depending upon morphotype) are required to investigate the possible presence of other morphotypes and more robustly evaluate effects of geography and environmental variables on morphological characteristics of tailed virus subgroups.

With the assumption that most marine viruses are phages (viruses that infect bacteria; Wommack and Colwell, 2000) and the knowledge that ca. 96% of all isolated phage are tailed (Ackermann, 2007), one would expect most marine viruses to be tailed. Instead we found that non-tailed icosahedral viruses dominate the upper water column of the global oceans, comprising 51–92% of viral assemblages. This corroborates two previous marine studies and contrasts three in freshwater systems (Table 1). Commonly, however, this high proportion of non-tailed viruses in marine environments is attributed to tail loss during sample preparation (reviewed by Proctor, 1997). The only empirical test of this assertion showed substantial viral tail loss from marine sediment samples (Williamson et al., 2012), but used harsher preparation methods (sonication and/or vortexing) than was used for qTEM in this study. In contrast, qTEM sample preservation and preparation does not cause tail loss or substantially alter other community viral morphological characteristics for water column samples. In addition, not once, in 2600 viruses documented in Tara Oceans samples, were viral tails observed separated from capsids.

It is possible that small podovirus tails may be obscured if these viruses landed directly on their tails when deposited onto the grid and the g-force used (118,000×*g*) was insufficient to force them to a prone position. This would result in erroneous documentation of podoviruses as non-tailed viruses, but would not change our major conclusions. Specifically, even if 50% of podoviruses were recorded as non-tailed, podovirus fractional abundances would double (to 12%) and non-tailed fractional abundances would only decrease to 73% (refer to Figure 5), leaving our concluded relative order of viral morphotypes intact. Further, for non-tailed viruses

to actually be rotated podoviruses would require this scenario to occur at much higher frequency in seawater than freshwater, as non-tailed viruses only comprise 0–30% of investigated freshwater viral assemblages (Table 1).

Marine viruses may lose their tails prior to sample collection through natural decay. In this scenario, one would expect similar capsid diameter distributions for tailed and non-tailed viruses if the 'non-tailed' viruses had lost their tails; instead, tailed viruses had significantly larger capsids than non-tailed viruses. Further, the much lower portion of non-tailed viruses observed in freshwater environments (Table 1) would require vastly different viral decay processes in fresh vs. saltwater, which seems unlikely.

The observation that upper ocean viruses are predominantly non-tailed raises questions regarding what organisms these viruses infect, and whether they contain double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), or RNA genomes. The most abundant potential hosts for viruses in the surface ocean are bacteria (reviewed by Pomeroy et al., 2007), but there are few marine non-tailed phage isolates (Table 2). Early marine phage isolations yielded one non-tailed dsDNA phage in 1968 and one non-tailed RNA phage in 1976, and more recent efforts have added nine ssDNA phages and a phage of unknown nucleic acid type (Table 2). Notably, two of these non-tailed phages were isolated using the cyanobacterium *Synechococcus* sp. WH7803 (McDaniel et al., 2006; Kuznetsov et al., 2012) from which a decade of viral isolations had previously resulted in only tailed phages (Waterbury and Valois, 1993; Wilson et al., 1993; Fuller et al., 1998; Lu et al., 2001; Chen and Lu, 2002; Marston and Sallee, 2003; Sullivan et al., 2003). Collectively, this suggests that the relative dearth of non-tailed phage isolates (Ackermann, 2007) may result from ascertainment bias derived from a combination of limited

host diversity and non-tailed phages being less easily propagated or recognized than their tailed counterparts.

The upper ocean, while dominated by bacteria, contains other potential microbial hosts for viruses including archaea and eukaryotes. Marine archaea numerically dominate the mesopelagic oceans (Karner et al., 2001), with increased abundance in some surface waters (e.g., the Southern Ocean; DeLong et al., 1994), yet their viruses are represented by a single isolate – a lemon-shaped virus from a hydrothermal deep-sea environment that infects *Pyrococcus abyssi* (Geslin et al., 2007). We observed no lemon-shaped viruses, nor any of the myriad 'exceptional' morphotypes isolated from archaeal extremophiles (reviewed by Prangishvili et al., 2006). This is likely because physico-chemical variables in the oceanic samples did not approach the 'extreme' conditions from which these exceptional morphotypes have been isolated. However, there are non-marine archaeal viral isolates with icosahedral non-tailed morphology (Bamford et al., 2005; Atanasova et al., 2012; Jaakkola et al., 2012) and further exploration of marine archaeal virus-host systems may yield more examples.

To date, the majority of isolated marine non-tailed viruses derive from eukaryotes including 28 dsDNA viruses isolated from marine alga; 3 ssDNA viruses isolated from marine diatoms; and 6 RNA viruses isolated from diatoms, a fungoid protist, and picophytoplankton (Table 2). While less abundant than prokaryotes, the relatively high number of viruses released per eukaryotic cell (reviewed by Lang et al., 2009) may increase representation of their viruses in the oceans (Steward et al., 2013) such that they could comprise a significant portion of non-tailed viruses.

Capsid diameters of marine non-tailed viral isolates (Table 2), while admittedly limited, may be useful in hypothesizing potential hosts for the observed non-tailed viruses. The range of

capsid diameters for isolated eukaryotic dsDNA viruses (115–220nm), smaller eukaryotic RNA viruses (22–32nm), larger eukaryotic RNA viruses (90–95nm), and smaller ssDNA phages (30–32nm) each comprised <1% of non-tailed viruses in the Tara Oceans samples, while eukaryotic ssDNA viruses (30–38nm) and larger ssDNA phages (72–77nm) only comprised 3% and 5%, respectively. However, the lone dsDNA and RNA non-tailed phages isolated from marine bacteria had 60nm capsids, which most closely represented the mean capsid diameter for Tara Oceans non-tailed viruses (54±12nm). Assuming that these trends from so few cultivated non-tailed viruses are robust, this suggests that most non-tailed marine viruses may infect the numerically dominant bacteria. However, the primary conclusion from comparing capsid diameters is that most observed non-tailed viruses have no cultivated representatives.

Cultivation-independent approaches also provide information about marine non-tailed viruses. First, marine viral metagenomes have yielded assembled genomes with similarity to non-tailed ssDNA *Microviridae* phages (Tucker et al., 2011; Roux et al., 2012), and to several families of eukaryotic non-tailed RNA viruses (Culley et al., 2006), providing genomic information about uncultured groups. Second, recent work suggests that RNA viruses are nearly as abundant as dsDNA viruses, comprising 15-77% of total viruses at one coastal Hawaii location (Steward et al., 2013). Extrapolating this to the global oceans where 51–92% of viruses were non-tailed, and assuming all RNA viruses are non-tailed, suggests that RNA viruses could comprise 16–100% of the non-tailed viruses observed.

Finally, 65–93% (reviewed by Hurwitz and Sullivan, 2013) and 41–81% (Culley et al., 2006; Steward et al., 2013) of sequences in marine DNA and RNA viral metagenomes, respectively, are not represented in existing genomic databases. Given that observed non-tailed virus capsid diameters were largely inconsistent with those from cultivated marine non-tailed

viruses, we posit that non-tailed viruses may comprise the majority of this vast 'unknown' marine viral metagenomic sequence space. Several existing and emerging approaches will likely help identify and characterize non-tailed marine viruses. These include culture-based approaches (e.g., targeted isolations with existing and new marine bacterial, archaeal, and eukaryotic cultures), as well as new methods that either require only the host to be in culture (e.g., viral tagging; Deng et al., 2012) or are completely cultivation-independent (e.g., physical fractionation of viral assemblages; Bergeron et al., 2007; Steward and Rappé, 2007; Brum and Steward, 2011; Brum et al., 2013). The abundance and distribution of genetically-characterized, non-tailed viruses could also be explored using phageFISH (Allers et al., 2013). As well, viruses with particular nucleic acid types can be examined by enriching for ssDNA (Kim and Bae, 2011) or specifically targeting dsDNA, ssDNA and RNA pools (Andrews-Pfannkoch et al., 2010).

In summary, morphological analysis was fundamental to the origin of modern aquatic viral research (e.g., Bergh et al., 1989; Borsheim et al., 1990; Bratbak et al., 1990; Borsheim, 1993) and, with careful methodological evaluation, it continues to be a valuable tool to understand the ecology and diversity of aquatic viruses. This use of qTEM to assess marine viruses across six ocean regions shifts the paradigm to non-tailed viruses as dominant, which should guide future work towards characterizing these abundant and nearly unexplored viruses.

Acknowledgements

We thank the Tucson Marine Phage Lab for manuscript review; Tony Day for electron microscopy assistance; Stefanie Kandels and John Adams for logistical support; the Tara Oceans Consortium for sample collection; Céline Dimier and Marc Picheral for assistance with environmental data acquisition; Jesse Czekanski-Moir for suggesting, and assistance with,

correspondence analysis; and Dana Hunt, Grieg Steward, and Eric Allen for collecting samples for methods testing. Funding was provided by the Gordon and Betty Moore Foundation to MBS.

Figure Captions:

Figure 1. Examples of the four viral morphotypes observed in this study (A, myovirus; B, podovirus; C, siphovirus; D, non-tailed virus).

Figure 2. (A) Percent of viral morphotypes in all surface samples combined, all DCM samples combined, and each oceanic region. Error bars represent standard deviations of the means of all samples. Letters indicate significant differences between depths or oceanic regions while numbers indicate significant differences within depths or oceanic regions (ANOVA with Tukey post-hoc test, $p<0.001$ for all). (B) Box and whisker plots of viral capsid diameters in all surface samples combined, all DCM samples combined, and each oceanic region. Top, middle, and bottom lines of each box correspond to the 75th, 50th (median), and 25th percentiles, respectively. Whiskers extending from the top and bottom of each box correspond to the 90th and 10th percentiles, respectively. Circles represent capsid diameters that are outside of the 90th and 10th percentiles (outliers). Letters indicate significant differences between depths or oceanic regions (ANOVA with Tukey post-hoc test, $p<0.001$ for all). The number of viruses used for each dataset is given in parentheses.

Figure 3. Ordination of Tara Oceans samples (A) and capsid diameter bins in nm (B) using correspondence analysis (CA) based on distribution of viral capsid diameters with 7nm bins (s, surface sample; d, DCM sample; surface sample from station 36 is omitted due to missing

461 oxygen data; percent of total inertia explained by CA1 and CA2 is reported on the axes).
462 Lengths of vectors overlaid on the sample ordination plot correspond to the strength of influence
463 for each environmental variable, with r^2 and p-values reported for each vector (C–F). Response
464 surfaces for each environmental variable are also overlaid on the sample ordination plot to assess
465 linearity of the relationship, with r^2(adjusted), p-values, and the percent of deviance explained
466 reported for each response surface (C–F). CA1 was negatively correlated with salinity (Pearson
467 correlation, r=-0.486, p=0.014) while CA2 was negatively correlated with temperature (Pearson
468 correlation, r=-0.623, p<0.001) and positively correlated with oxygen (Pearson correlation,
469 r=0.646, p<0.001).
470
471 Figure 4. Box and whisker plots of myovirus, siphovirus, and podovirus tail lengths in all
472 surface samples combined, all DCM samples combined, and each oceanic region. Refer to
473 Figure 2 for a description of box and whisker plot construction. The number of viruses used for
474 each dataset is given in parentheses. Letters indicate significant differences between depths (t-
475 test, p=0.001) or oceanic regions (ANOVA on ranks with Dunn's post-hoc test, p<0.05 for all).
476
477 Figure 5. Morphological results for all viruses in this study including the percent of each
478 morphotype (A), as well as capsid diameters (B) and tail lengths (C) of all viruses and each
479 morphotype. The average and standard deviation are given for each set of viruses, with ranges
480 reported in parentheses, and the number of viruses analyzed (N) is given for capsid diameters
481 and tail lengths. Refer to Figure 2 for a description of box and whisker plot construction. Letters
482 indicate significant differences between morphotypes (ANOVA on ranks, p<0.001 for all) and

numbers indicate significant differences between capsid diameters of non-tailed and all tailed viruses combined (B; Mann-Whitney rank sum test, $p<0.001$).

Figure S1. Location and station numbers of the Tara Oceans samples used in this study. Map created in Ocean Data View (Schlitzer, 2011).

Figure S2. Viral morphotypes and size-frequency histograms of viral capsid diameters in the Biosphere 2 Ocean after analysis of 50, 100, and 200 viruses from the sample (A and B, respectively) and after analysis of 100 viruses each from fresh and frozen samples (C and D, respectively). Error bars for viral morphotype percentages represent 95% confidence intervals. There was no significant difference in the percent of each morphotype when comparing analysis of 50, 100, and 200 viruses (A; binomial regression, $p>0.1$ for all) or fresh versus frozen samples (C; binomial regression, $p>0.1$ for all). Capsid diameter distributions were also highly similar between 50, 100, and 200 analyzed viruses (B; Morisita's index 1.03–1.06 for all) and between fresh and frozen samples (D; Morisita's index of 0.92).

Figure S3. Percent of viral morphotypes observed after deposition of samples from Scripps Pier (SP), Beaufort Inlet (BI), and Kaneohe Bay (KB) onto grids using the adsorption and qTEM methods. Error bars are standard deviations of the mean of triplicate samples. There were no significant differences in the percent of each morphotype between the adsorption and qTEM methods for any sample (t-tests, $p>0.1$ for all).

505 Figure S4. Viral morphotype compositions and size-frequency histograms of viral capsid
506 diameters and tail lengths in the Mediterranean Sea.
507
508 Figure S5. Viral morphotype compositions and size-frequency histograms of viral capsid
509 diameters and tail lengths in the Red Sea.
510
511 Figure S6. Viral morphotype compositions and size-frequency histograms of viral capsid
512 diameters and tail lengths in the Arabian Sea.
513
514 Figure S7. Viral morphotype compositions and size-frequency histograms of viral capsid
515 diameters and tail lengths in the Indian Ocean.
516
517 Figure S8. Viral morphotype compositions and size-frequency histograms of viral capsid
518 diameters and tail lengths in the Atlantic Ocean.
519
520 Figure S9. Viral morphotype compositions and size-frequency histograms of viral capsid
521 diameters and tail lengths in the Pacific Ocean.
522
523 Figure S10. Relationships between viral morphological characteristics and environmental
524 parameters with significant correlations (Table S2). Error bars are standard deviations of the
525 means for oceanic regions.
526
527 **References**

528 Abramoff MD, Magalhaes PJ, Ram SJ. (2004). Image processing with ImageJ. Biophotonics
529 International 11:36-42.
530
531 Ackermann HW. (2007). 5500 phages examined in the electron microscope. Arch Virol 152:227-
532 243.
533
534 Ackermann HW. (2001). Frequency of morphological phage descriptions in the year 2000. Arch
535 Virol 146:843-857.
536
537 Ackermann H-W, Heldal M. (2010). Basic electron microscopy of aquatic viruses. In: Wilhelm
538 SW, Weinbauer MG, Suttle CA (eds). Manual of Aquatic Viral Ecology. ASLO, pp 182-192.
539
540 Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J, *et al.* Single-cell
541 and population level viral infection dynamics revealed by phageFISH, a method to visualize
542 intracellular and free viruses. Environ Microbiol 2013; e-pub ahead of print 14 March 2013, doi:
543 10.1111/1462-2920.12100.
544
545 Andrews-Pfannkoch C, Fadrosh DW, Thorpe J, Williamson SJ. (2010). Hydroxyapatite-
546 mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from
547 natural viral assemblages. Appl Environ Microbiol 76:5039-5045.
548
549 Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, *et al.* (2006). The marine
550 viromes of four oceanic regions. PLoS Biol 4:2121-2131.

551

552 Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM. (2012). Global network of specific
553 virus-host interactions in hypersaline environments. Environ Microbiol 14:426-440.

554

555 Attoui H, Jaafar FM, Belhouchet M, de Micco P, de Lamballerie X, Brussaard CPD. (2006).
556 *Micromonas pusilla* reovirus: a new member of the family *Reoviridae* assigned to a novel
557 proposed genus (*Mimoreovirus*). J Gen Virol 87:1375-1383.

558

559 Auguet JC, Montanie H, Lebaron P. (2006). Structure of virioplankton in the Charente Estuary
560 (France): transmission electron microscopy *versus* pulsed field gel electrophoresis. Microb Ecol
561 2006:197-208.

562

563 Auguet JC, Montanie H, Hartmann HJ, Lebaron P, Casamayor EO, Catala P, *et al.* (2009).
564 Potential effects of freshwater virus on the structure and activity of bacterial communities in the
565 Marennes-Oleron Bay (France). Microb Ecol 57:295-306.

566

567 Bamford DH, Ravantti JJ, Ronnholm G, Laurinavicius S, Kukkaro P, Dyall-Smith M, *et al.*
568 (2005). Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon
569 *Haloarcula hispanica*. J Virol 79:9097-9107.

570

571 Barberan A, Fernandez-Guerra A, Bohannan BJM, Casamayor EO. (2012). Exploration of
572 community traits as ecological markers in microbial metagenomes. Mol Ecol 21:1909-1917.

573 Bergeron A, Belcaid M, Steward GF, Poisson G. (2007). Divide and conquer: enriching
574 environmental sequencing data. PLoS ONE 2:e830.
575
576 Bergh O, Borsheim KY, Bratbak G, Heldal M. (1989). High abundance of viruses found in
577 aquatic environments. Nature 340:467-468.
578
579 Bettarel Y, Bouvier T, Bouvier C, Carre C, Desnues A, Domaizon I, *et al.* (2011a). Ecological
580 traits of planktonic viruses and prokaryotes along a full-salinity gradient. FEMS Microbiol Ecol
581 76:360-372.
582
583 Bettarel Y, Bouvier T, Agis M, Bouvier C, Van Chu T, Combe M, *et al.* (2011b). Viral
584 distribution and life strategies in the Bach Dang Estuary, Vietnam. Microb Ecol 62:143-154.
585
586 Borsheim KY. (1993). Native marine bacteriophages. FEMS Microbiol Ecol 102:141-159.
587
588 Borsheim KY, Bratbak G, Heldal M. (1990). Enumeration and biomass estimation of planktonic
589 bacteria and viruses by transmission electron microscopy. Appl Environ Microbiol 56:352-356.
590
591 Bratbak G, Heldal M, Norland S, Thingstad TF. (1990). Viruses as partners in spring bloom
592 microbial trophodynamics. Appl Environ Microbiol 56:1400-1405.
593
594 Bratbak G, Haslund OH, Heldal M, Naess A, Roeggen T. (1992). Giant marine viruses? Mar
595 Ecol Prog Ser 85:202-202.

596

597 Breitbart M. (2012). Marine Viruses: Truth or Dare. Ann Rev Mar Sci 4:425-448.

598

599 Breitbart M, Miyake JH, Rohwer F. (2004a). Global distribution of nearly identical phage-

600 encoded DNA sequences. FEMS Microbiol Lett 236:249-256.

601

602 Breitbart M, Thompson LR, Suttle CA, Sullivan MB. (2007). Exploring the vast diversity of

603 marine viruses. Oceanography 20:135-139.

604

605 Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, *et al.* (2004b). Diversity and

606 population structure of a near-shore marine-sediment viral community. Proc Roy Soc B 271:565-

607 574.

608

609 Brum JR, Steward GF. (2010). Morphological characterization of viruses in the stratified water

610 column of alkaline, hypersaline Mono Lake. Microb Ecol 60:636-643.

611

612 Brum JR, Steward GF. (2011). Physical fractionation of aquatic viral assemblages. Limnol

613 Oceanogr Methods 9:150-163.

614

615 Brum JR, Culley AI, Steward GF. (2013). Assembly of a marine viral metagenome after

616 fractionation. PLoS ONE (in press).

617

618 Castberg T, Thyrhaug R, Larsen A, Sandaa R-A, Heldal M, Van Etten JL, *et al.* (2002). Isolation
619 and characterization of a virus that infects *Emiliania huxleyi* (Haptophyta). J Phycol 38:767-774.
620
621 Chen F, Lu J. (2002). Genomic sequence and evolution of marine cyanophage P60: a new insight
622 on lytic and lysogenic phages. Appl Environ Microbiol 68:2589-2594.
623
624 Clokie M, Shan J, Bailey S, Jia Y, Krisch HM, West S, *et al.* (2006). Transcription of a
625 "photosynthetic" T4-type phage during infection of a marine cyanobacterium. Environ Microbiol
626 8:827-835.
627
628 Colombet J, Sime-Ngando T, Cauchie HM, Fonty G, Hoffmann L, Demeure G. (2006). Depth-
629 related gradients of viral activity in Lake Pavin. Appl Environ Microbiol 72:4440-4445.
630
631 Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, Tovkach FI, *et al.* (2012). Phage
632 morphology recapitulates phylogeny: The comparative genomics of a new group of myoviruses.
633 PLoS ONE 7:e40102.
634
635 Cottrell MT, Suttle CA. (1991). Wide-spread occurrence and clonal variation in viruses which
636 cause lysis of a cosmopolitan eukaryotic marine phytoplankter, *Micromonas pusilla*. Mar Ecol
637 Prog Ser 78:1-9.
638
639 Culley AI, Lang AS, Suttle CA. (2006). Metagenomic analysis of coastal RNA virus
640 communities. Science 312:1795-1798.

Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW, Frankenberg-Dinkel N. (2008). Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. Curr Biol 18:442-448.

DeLong EF, Wu KY, Prezelin BB, Jovine RVM. (1994). High abundance of Archaea in Antarctic marine picoplankton. Nature 371:695-697.

Demuth J, Neve H, Witzel K-P. (1993). Direct electron microscopy study on the morphological diversity of bacteriophage populations in Lake Plussee. Appl Environ Microbiol 59:3378-3384.

Deng L, Gregory A, Yilmaz S, Poulos BT, Hugenholtz P, Sullivan MB. (2012). Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. mBio 3:e00373-12.

Derelle E, Ferraz C, Escande M-L, Eychenie S, Cooke R, Piganeau G, *et al.* (2008). Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. PLoS ONE 3:e2250.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, *et al.* (2008). Functional metagenomic profiling of nine biomes. Nature 452:629-632.

Doane FW. (1980). Viral morphology as an aid for rapid diagnosis. Yale J. Biol. Med. 53:19-25.

664 Duhaime MB, Sullivan MB. (2012). Ocean viruses: Rigorously evaluating the metagenomic
665 sample-to-sequence pipeline. Virology 434:181-186.
666
667 Duhaime MBD, Deng L, Poulos BT, Sullivan MB. (2012). Towards quantitative metagenomics
668 of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and
669 optimization of the linker amplification method. Environ Microbiol 14:2526-2537.
670
671 Edwards RA, Rohwer F. (2005). Viral metagenomics. Nat Rev Microbiol 3:504-510.
672
673 Espejo RT, Canelo ES. (1968). Properties of bacteriophage PM2: a lipid-containing bacterial
674 virus. Virology 34:738-747.
675
676 Fuhrman JA. (1999). Marine viruses and their biogeochemical and ecological effects. Nature
677 399:541-548.
678
679 Fuller NJ, Wilson WH, Joint IR, Mann NH. (1998). Occurrence of a sequence in marine
680 cyanophages similar to that of T4 g20 and its application to PCR-based detection and
681 quantification techniques. Appl Environ Microbiol 64:2051-2060.
682
683 Geslin C, Gaillard M, Flament D, Rouault K, Le Romancer M, Prieur D, *et al.* (2007). Analysis
684 of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from
685 *Pyrococcus abyssi*. J Bacteriol 189:4510-4519.
686

687 Gowing MM. (1993). Large virus-like particles from vacuoles of phaeodarian radiolarians and
688 from other marine samples. Mar Ecol Prog Ser 101:33-43.
689
690 Hammond GW, Hazelton PR, Chuang I, Klisko B. (1981). Improved detection of viruses by
691 electron microscopy after direct ultracentrifuge preparation of specimens. J Clin Microbiol
692 14:210-221.
693
694 Hidaka T, Ichida K-i. (1976). Properties of a marine RNA-containing bacteriophage. Mem Fac
695 Fish Kagoshima Univ 25:77-89.
696
697 Hofer JS, Sommaruga R. (2001). Seasonal dynamics of viruses in an alpine lake: importance of
698 filamentous forms. Aquat Microb Ecol 26:1-11.
699
700 Holmfeldt K, Odic D, Sullivan MB, Middelboe M, Riemann L. (2012). Cultivated single
701 stranded DNA phages that infect marine *Bacteroidetes* prove difficult to detect with DNA
702 binding stains. Appl Environ Microbiol 78:892-894.
703
704 Hurwitz BL, Sullivan MB. (2013). The Pacific Ocean Virome (POV): A marine viral
705 metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS ONE
706 8:e57355.
707

708 Hurwitz BL, Deng L, Poulos BT, Sullivan MB. Evaluation of methods to concentrate and purify
709 ocean virus communities through comparative, replicated metagenomics. Environ Microbiol; e-
710 pub ahead of print 9 July 2012, doi: 10.1111/j.1462-2920.2012.02836.x.
711
712 Ignacio-Espinoza JC, Sullivan MB. (2012). Phylogenomics of T4 cyanophages: lateral gene
713 transfer in the 'core' and origins of host genes. Environ Microbiol 14:2113-2126.
714
715 Jaakkola ST, Penttinen RK, Vilen ST, Jalasvuori M, Ronnholm G, Bamford JKH, *et al.* (2012).
716 Closely related archaeal *Haloarcula hispanica* icosahedral viruses HHIV-2 and SH1 have
717 nonhomologous genes encoding host recognition functions. J Virol 86:4734-4742.
718
719 Jacobsen A, Bratbak G, Heldal M. (1996). Isolation and characterization of a virus infecting
720 *Phaeocystis pouchetii* (Prymnesiophyceae). J Phycol 32:923-927.
721
722 John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, *et al.* (2011). A simple and
723 efficient method for concentration of ocean viruses by chemical flocculation. Environmental
724 Microbiology Reports 3:195-202.
725
726 Karner MB, Delong EF, Karl DM. (2001). Archaeal dominance in the mesopelagic zone of the
727 Pacific Ocean. Nature 409:507-510.
728
729 Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J, *et al.* (2011). A holistic
730 approach to marine eco-systems biology. PLoS Biol 9:e1001177.

Kim K-H, Bae J-W. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol 77:7663-7668.

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. (2012). Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Academic Press: San Diego.

Krebs CJ. (1999). Ecological Methodology, Second Edition. Addison-Welsey Educational Publishers, Inc.: Menlo Park, California.

Kuznetsov YG, Chang S-C, Credaroli A, Martiny J, McPherson A. (2012). An atomic force microscopy investigation of cyanophage structure. Micron 43:1336-1342.

Lang AS, Rise ML, Culley AI, Steward GF. (2009). RNA viruses in the sea. FEMS Microbiol Rev 33:295-323.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438:86-89.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc Natl Acad Sci USA 101:11013-11018.

754 Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, *et al.* (2007). Genome-
755 side expression dynamics of a marine virus and host reveal features of co-evolution. Nature
756 449:83-86.
757
758 Lu J, Chen F, Hodson RE. (2001). Distribution, isolation, host specificity, and diversity of
759 cyanophages infecting marine *Synechococcus* spp. in river estuaries. Appl Environ Microbiol
760 67:3285-3290.
761
762 Maranger R, Bird DF, Juniper SK. (1994). Viral and bacterial dynamics in Arctic sea ice during
763 the spring algal bloom near Resolute, N.W.T., Canada. Mar Ecol Prog Ser 111:121-127.
764
765 Marston MF, Sallee JL. (2003). Genetic diversity and temporal variation in the cyanophage
766 community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Appl
767 Environ Microbiol 69:4639-4647.
768
769 McDaniel LD, DelaRosa M, Paul JH. (2006). Temperate and lytic cyanophages from the Gulf of
770 Mexico. J Mar Biol Assoc UK 86:517-527.
771
772 Nagasaki K, Yamaguchi M. (1997). Isolation of a virus infectious to the harmful bloom causing
773 microalga *Heterosigma akashiwo* (Raphidophyceae). Aquat Microb Ecol 13:135-140.
774

775 Nagasaki K, Tomaru Y, Katanozake N, Shirai Y, Nishida K, Itakura S, *et al.* (2004). Isoloation
776 and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom
777 *Rhizosolenia setigera*. Appl Environ Microbiol 70:704-711.
778
779 Nagasaki K, Tomaru Y, Takao Y, Nishida K, Shirai Y, Suzuki H, *et al.* (2005). Previously
780 unknown virus infects marine diatom. Appl Environ Microbiol 71:3528-3535.
781
782 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al.* (2013) vegan:
783 Community Ecology Package. R package version 2.1-27/r2451. http://R-Forge.R-
784 project.org/projects/vegan/.
785
786 Oren A, Bratbak G, Heldal M. (1997). Occurrence of virus-like particles in the Dead Sea.
787 Extremophiles 1:143-149.
788
789 Pomeroy LR, Williams PJl, Azam F, Hobbie JE. (2007). The microbial loop. Oceanography
790 20:28-33.
791
792 Prangishvili D, Forterre P, Garrett RA. (2006). Viruses of the Archaea: a unifying view. Nat Rev
793 Microbiol 4:837-848.
794
795 Proctor LM. (1997). Advances in the study of marine viruses. Microsc Res Tech 37:136-161.
796

R Core Team. (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rohwer F, Edwards R. (2002). The phage proteomic tree: a genome-based taxonomy for phage. J Bacteriol 184:4529-4535.

Roux S, Krupovic M, Poulet A, Debroas D, Enault F. (2012). Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS ONE 7:e40418.

Sandaa R-A, Heldal M, Castberg T, Thyrhaug R, Bratbak G. (2001). Isolation and characterization of two viruses with large genome size infection *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). Virology 290:272-280.

Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environ Microbiol 11:2078-2093.

Schlitzer R. (2011). Ocean Data View. http://odv.awi.de.

Schroeder DC, Oke J, Malin G, Wilson WH. (2002). Coccolithovirus (Phycodnaviridae): characterization of a new large dsDNA algal virus that infects *Emiliania huxleyi*. Arch Virol 147:1685-1698.

Sharon I, Tzahor S, Williamson S, Shmoish M, Man-Aharonovich D, Rusch DB, *et al.* (2007). Viral photosynthetic reaction center genes and transcripts in the marine environment. ISME J 1:492-501.

Sommaruga R, Krossbacher M, Salvenmoser W, Catalan J, Psenner R. (1995). Presence of large virus-like particles in a eutrophic reservoir. Aquat Microb Ecol 9:305-308.

Steward GF, Rappé MS. (2007). What's the 'meta' with metagenomics? ISME J 1:100-102.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G. (2013). Are we missing half of the viruses in the ocean? ISME J 7:672-679.

Stopar D, Cerne A, Zigman M, Poljsak-Prijatelj M, Turk V. (2003). Viral abundance and a high proportion of lysogens suggests that viruses are important members of the microbial community in the Gulf of Trieste. Microb Ecol 46:249-256.

Sturges HA. (1926). The choice of a class interval. J Am Stat Assoc 21:65-66.

Sullivan MB, Waterbury JB, Chisholm SW. (2003). Cyanophage infecting the oceanic cyanobacterium *Prochlorococcus*. Nature 424:1047-1051.

842 Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. (2006).
843 Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their
844 hosts. PLoS Biol 4:1344-1357.
845
846 Suttle CA. (2007). Marine viruses - major players in the global ecosystem. Nat Rev Microbiol
847 5:801-812.
848
849 Suttle CA, Chan AM. (1995). Viruses infecting the marine Prymnesiophyte *Chrysochromulina*
850 spp.: isolation, preliminary charactrization and natural abundance. Mar Ecol Prog Ser 118:275-
851 282.
852
853 Tai V, Lawrence JE, Lang AS, Chan AM, Culley AI, Suttle CA. (2003). Charactrization of
854 HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo*
855 (Raphidophyceae). J Phycol 39:343-352.
856
857 Takao Y, Nagasaki K, Mise K, Okuno T, Honda D. (2005). Isolation and characterization of a
858 novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium* sp.
859 (Thraustochytriaceae, Labyrinthulea). Appl Environ Microbiol 71:4516-4522.
860
861 Tapper MA, Hicks RE. (1998). Temperate viruses and lysogeny in Lake Superior
862 bacterioplankton. Limnol Oceanogr 43:95-103.
863

Tarutani K, Nagasaki K, Itakura S, Yamaguchi M. (2001). Isolation of a virus infecting the novel shellfish-killing dinoflagellate *Heterocapsa circularisquama*. Aquat Microb Ecol 23:103-111.

Thingstad TF. (2000). Elements of a theory for the mechanisms coltrolling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45:1320-1328.

Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, *et al.* (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc Natl Acad Sci USA 108:E757-E764.

Tomaru Y, Shirai Y, Suzuki H, Nagumo T, Nagasaki K. (2008). Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. Aquat Microb Ecol 50:103-112.

Tomaru Y, Takao Y, Suzuki H, Nagumo T, Nagasaki K. (2009). Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming diatom *Chaetoceros socialis*. Appl Environ Microbiol 75:2375-2381.

Tomaru Y, Takao Y, Suzuki H, Nagumo T, Koike K, Nagasaki K. (2011). Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* Grunow. Appl Environ Microbiol 77:5285-5293.

887 Tomaru Y, Katanozake N, Nishida K, Shirai Y, Tarutani K, Yamaguchi M, *et al.* (2004).
888 Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus
889 infecting the bivalve-killing microalga *Heterocapsa circularisquama*. Aquat Microb Ecol
890 34:207-218.
891
892 Tucker KP, Parsons R, Symonds EM, Breitbart M. (2011). Diversity and distribution of single-
893 stranded DNA phages in the North Atlantic Ocean. ISME J 5:822-830.
894
895 Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine
896 *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl Environ
897 Microbiol 59:3393-3399.
898
899 Weinbauer MG. (2004). Ecology of prokaryotic viruses. FEMS Microbiol Rev 28:127-181.
900
901 Weinbauer MG, Peduzzi P. (1994). Frequency, size and distribution of bacteriophages in
902 different marine bacterial morphotypes. Mar Ecol Prog Ser 108:11-20.
903
904 Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH. (2009). From small hosts come
905 big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. Environ
906 Microbiol 11:2821-2839.
907

908 Weynberg KD, Allen MJ, Gilg IC, Scanlan DJ, Wilson WH. (2011). Genome sequence of
909 *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the
910 ocean. J Virol 85:4520-4529.
911
912 Williamson KE, Helton RR, Wommack KE. (2012). Bias in bacteriophage morphological
913 classification by transmission electron microscopy due to breakage or loss of tail structures.
914 Microsc Res Tech 75:452-457.
915
916 Wilson WH, Joint IR, Carr NG, Mann NH. (1993). Isolation and molecular characterization of
917 five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. Appl Environ
918 Microbiol 59:3736-3743.
919
920 Wolf S, Muller D, Maier I. (2000). Assembly of a large icosahedral DNA virus, MclaV-1, in the
921 marine alga *Myriotrichia clavaeformis* (Dictyosiphonales, Phaeophyceae). Eur J Phycol 35:163-
922 171.
923
924 Wommack KE, Colwell RR. (2000). Virioplankton: viruses in aquatic ecosystems. Microbiol
925 Mol Biol Rev 64:69-114.
926
927 Wood SN. (2011). Fast stable restricted maximum likelihood estimation of semiparametric
928 generalized linear models. Journal of the Royal Statistical Society (B) 73:3-36.
929
930 Zar J. (1996). Biostatistical Analysis. Prentice Hall: Upper Saddle River.

931

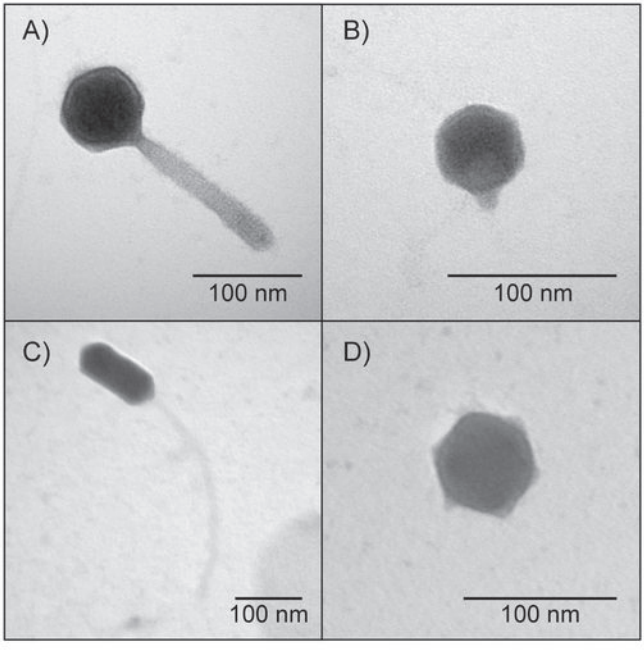
Table 1. Percent of non-tailed viruses in viral assemblages from freshwater lakes and marine environments.

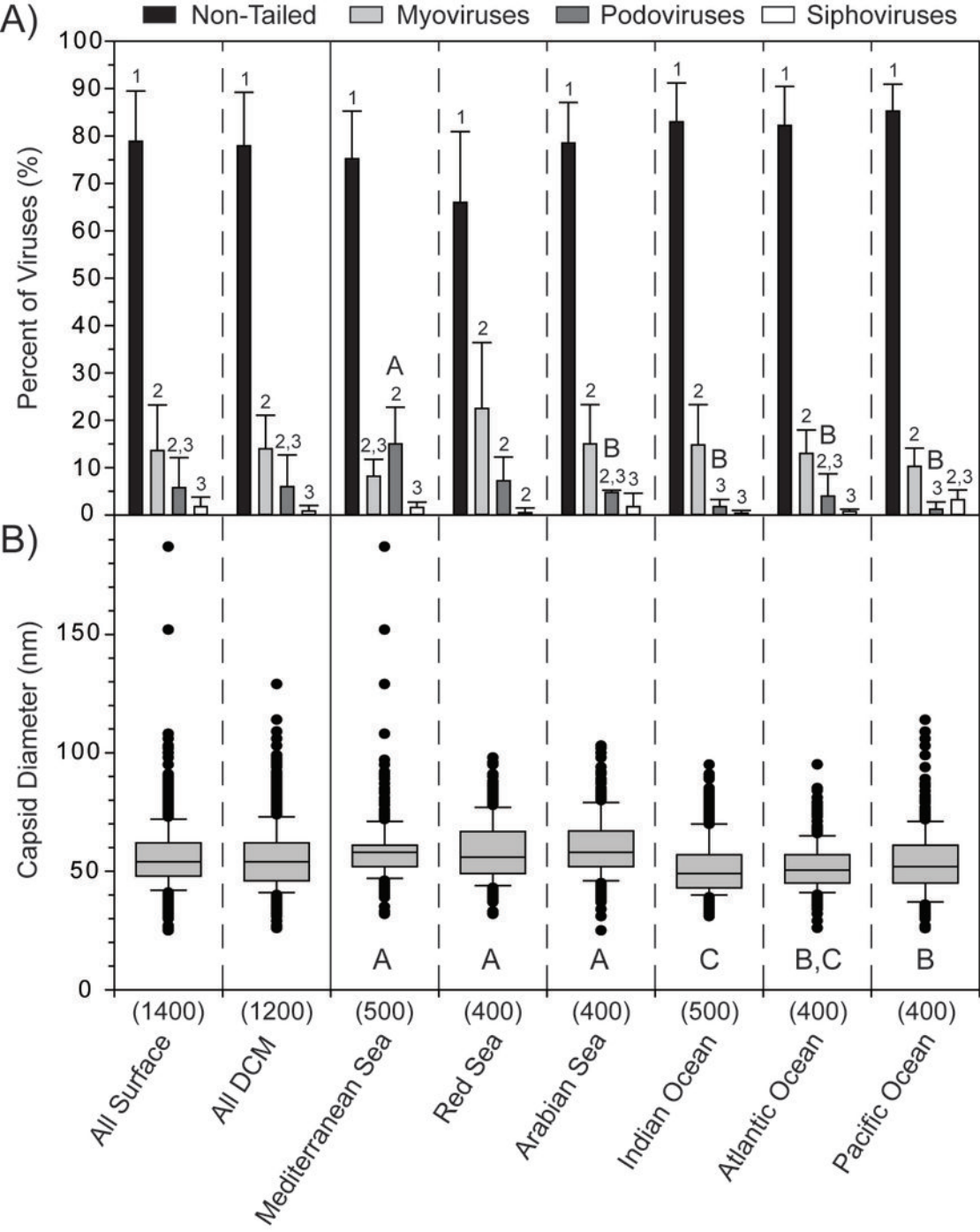
Environment	Non-Tailed Viruses	Location	Reference
Freshwater Lakes	0%	Lake Plußsee, Germany	Demuth et al., 1993
	30%	Lake Superior, USA	Tapper and Hicks, 1998
	25%	Lake Pavin, France	Colombet et al., 2006
Marine Environments[1]	91±4%	Pertuis d'Antioche, French Atlantic coast	Auguet et al., 2006
	74%	Gulf of Trieste, Adriatic Sea	Stopar et al., 2003
	79% (51–92%)	Global survey	This study

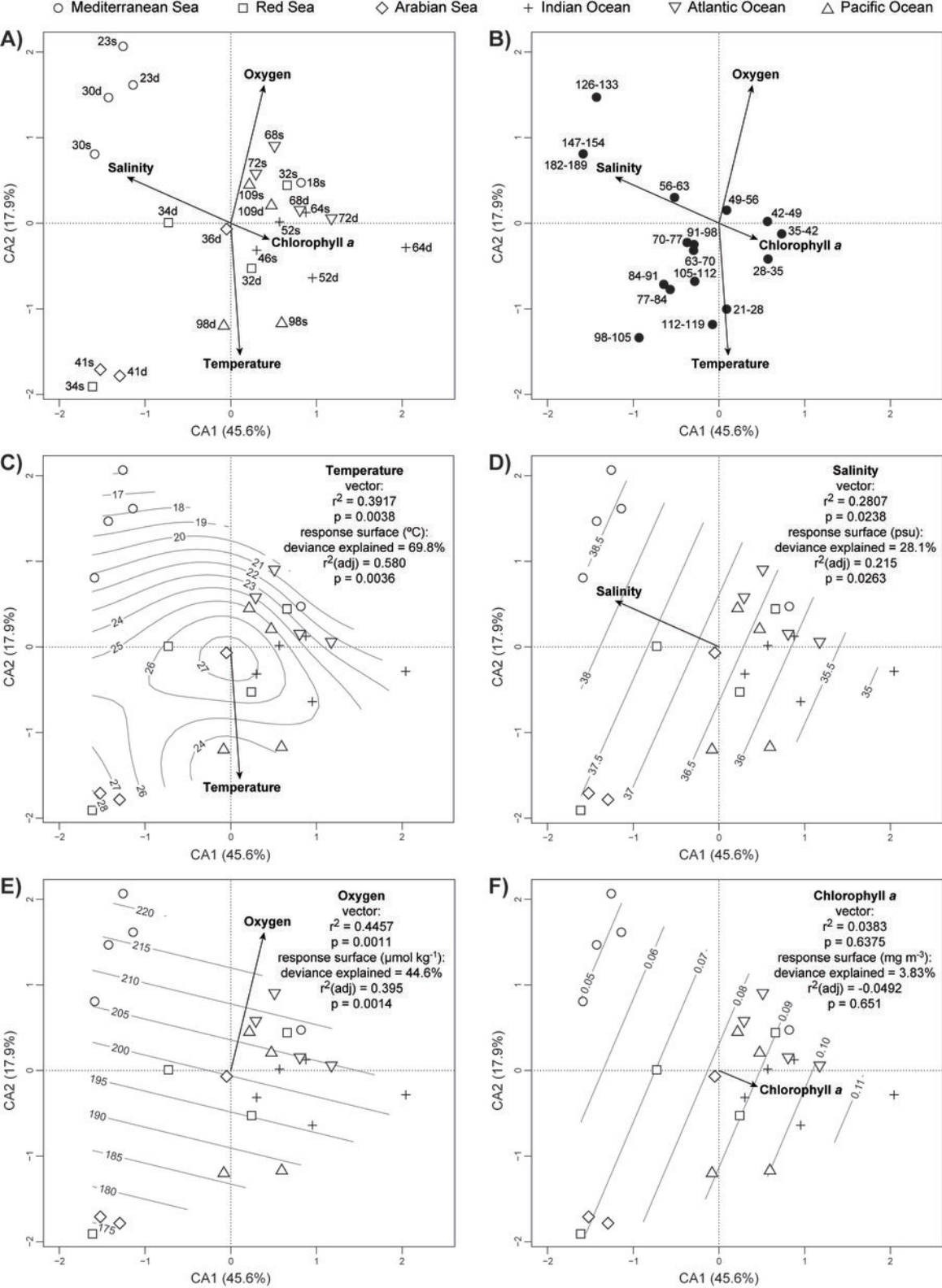
[1]Bratbak et al. (1990) also reported that non-tailed viruses were 'dominating' in coastal waters of Norway, but did not quantify their contribution to the viral assemblage.

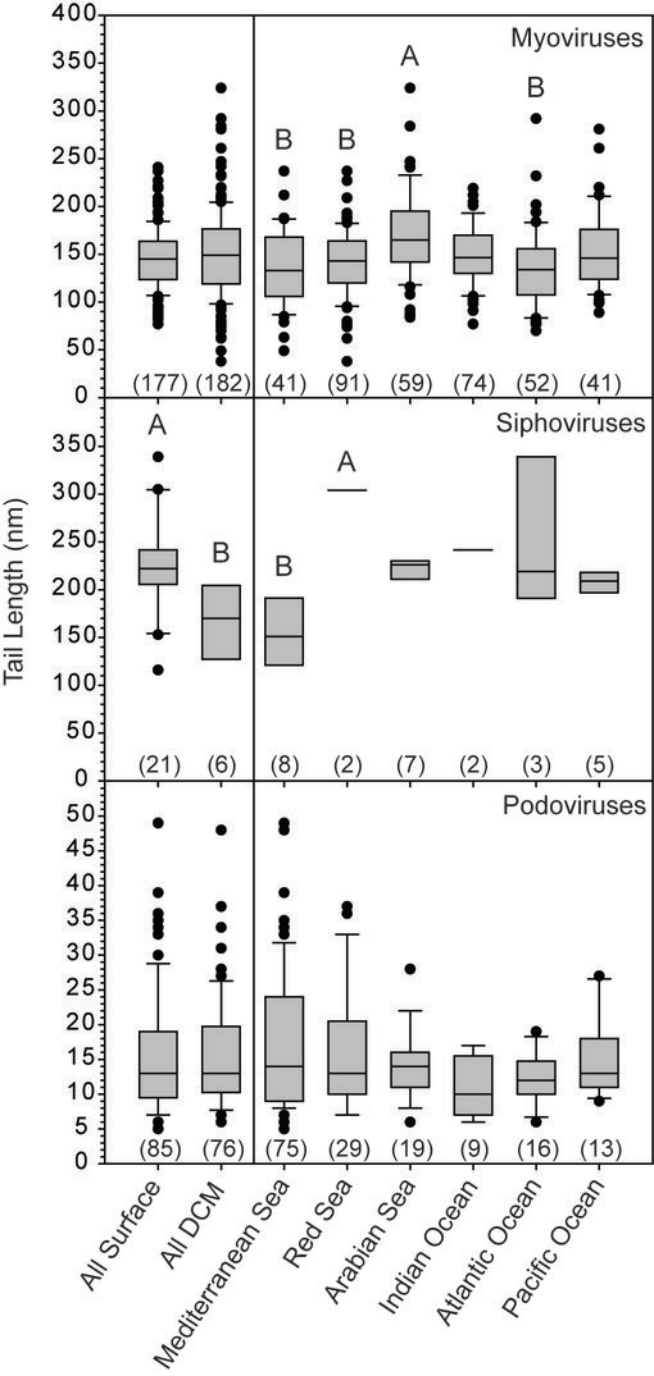
Table 2. Published non-tailed viruses isolated from marine bacteria and single-celled eukaryotes. NR = not reported

	Host Organism	Virus	Nucleic Acid	Capsid Diameter (nm)	Reference
Bacteria	*Pseudoalteromonas* sp.	PM2	dsDNA	60	Espejo and Canelo, 1968
	Synechococcus sp. WH7803	cyanophage N	ssDNA	ca. 77	McDaniel et al., 2006
	Cellulophaga baltica	Φ3:2	ssDNA	73 ± 0.5	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ46:2	ssDNA	72 ± 1.9	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ48:2	ssDNA	72 ± 1.1	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ12:2	ssDNA	31 ± 2.1	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ12a:1	ssDNA	30 ± 1.8	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ18:4	ssDNA	32 ± 2.6	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ14:1	ssDNA	NR	Holmfeldt et al., 2012
	Cellulophaga baltica	Φ48:1	ssDNA	NR	Holmfeldt et al., 2012
	06N-58	06N-58P	RNA	60	Hidaka and Ichida, 1976
	Synechococcus sp. WH7803	NR	NR	125	Kuznetsov et al., 2012
Single-Celled Eukaryotes	*Micromonas pusilla*	MPV-PB5	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Micromonas pusilla	MPV-PB7	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Micromonas pusilla	MPV-PB8	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Micromonas pusilla	MPV-GM1	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Micromonas pusilla	MPV-PL1	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Micromonas pusilla	MPV-SP1	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Micromonas pusilla	MPV-SG1	dsDNA	ca. 115	Cottrell and Suttle, 1991
	Chrysochromulina brevifilum	CbV-PW1	dsDNA	145 – 170	Suttle and Chan, 1995
	Phaeocystis pouchetii	PpV01	dsDNA	130 – 160	Jacobsen et al., 1996
	Heterosigma akashiwo	HaV	dsDNA	202 ± 6	Nagasaki and Yamaguchi, 1997
	Myriotrichia clavaeformis	MclaV-1	dsDNA	195 ± 5	Wolf et al., 2000
	Chrysochromulina ericina	CeV-01B	dsDNA	160	Sandaa et al., 2001
	Pyramimonas orientalis	PoV-01B	dsDNA	220 × 180	Sandaa et al., 2001
	Heterocapsa circularisquama	HcV	dsDNA	197 ± 8	Tarutani et al., 2001
	Emiliania huxleyi	EhV	dsDNA	160 – 180	Castberg et al., 2002
	Emiliania huxleyi	EhV (10 isolates)	dsDNA	170 – 200	Schroeder et al., 2002
	Ostreococcus tauri	OtV5	dsDNA	122 ± 9	Derelle et al., 2008
	Ostreococcus tauri	OtV-1	dsDNA	100 – 120	Weynberg et al., 2009
	Ostreococcus tauri	OtV-2	dsDNA	NR	Weynberg et al., 2011
	Chaetoceros salsugineum	CsNIV	ssDNA	38	Nagasaki et al., 2005
	Chaetoceros debilis	CdebDNAV	ssDNA	30	Tomaru et al., 2008
	Chaetoceros lorenzianus	ClorDNAV	ssDNA	34	Tomaru et al., 2011
	Heterosigma akashiwo	HaRNAV	RNA	25	Tai et al., 2003
	Rhizosolenia setigera	RsRNAV	RNA	32	Nagasaki et al., 2004
	Heterocapsa circularisquama	HcRNAV	RNA	30	Tomaru et al., 2004
	Schizochytrium sp.	SssRNAV	RNA	25	Takao et al., 2005
	Micromonas pusilla	MpRV	RNA	90 – 95	Attoui et al., 2006
	Chaetoceros socialis	CsfrRNAV	RNA	22	Tomaru et al., 2009


A)
100 nm
B)
100 nm
C)
100 nm
D)
100 nm


A)
Non-Tailed
Myoviruses
Podoviruses
Siphoviruses
Percent of Viruses (%)
B)
Capsid Diameter (nm)
A
A
A
C
B,C
B
(1400)
(1200)
(500)
(400)
(400)
(500)
(400)
(400)
All Surface
All DCM
Mediterranean Sea
Red Sea
Arabian Sea
Indian Ocean
Atlantic Ocean
Pacific Ocean


Mediterranean Sea
Red Sea
Arabian Sea
Indian Ocean
Atlantic Ocean
Pacific Ocean
A)
B)
C)
D)
E)
F)
CA1 (45.6%)
CA2 (17.9%)
Oxygen
Salinity
Chlorophyll a
Temperature
Temperature
vector:
r2 = 0.3917
p = 0.0038
response surface (ºC):
deviance explained = 69.8%
r2(adj) = 0.580
p = 0.0036
Salinity
vector:
r2 = 0.2807
p = 0.0238
response surface (psu):
deviance explained = 28.1%
r2(adj) = 0.215
p = 0.0263
Oxygen
vector:
r2 = 0.4457
p = 0.0011
response surface (µmol kg-1):
deviance explained = 44.6%
r2(adj) = 0.395
p = 0.0014
Chlorophyll a
vector:
r2 = 0.0383
p = 0.6375
response surface (mg m-3):
deviance explained = 3.83%
r2(adj) = -0.0492
p = 0.651


Myoviruses
Siphoviruses
Podoviruses
Tail Length (nm)
All Surface
All DCM
Mediterranean Sea
Red Sea
Arabian Sea
Indian Ocean
Atlantic Ocean
Pacific Ocean
(177) (182) (41) (91) (59) (74) (52) (41)
(21) (6) (8) (2) (7) (2) (3) (5)
(85) (76) (75) (29) (19) (9) (16) (13)