The ISME Journal (2013) 7, 1738–1751

www.nature.com/ismej

ORIGINAL ARTICLE

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

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.
The ISME Journal (2013) **7,** 1738–1751; doi:10.1038/ismej.2013.67; published online 2 May 2013
Subject Category: microbial population and community ecology
Keywords: biogeography; diversity; marine viruses; seawater; viral morphology

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 (for example, 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

Correspondence: MB Sullivan, Department of Ecology and Evolutionary Biology, University of Arizona, 1007 E Lowell Street, Life Sciences South 203, Tucson, AZ 85721, USA.
E-mail: mbsulli@email.arizona.edu
Received 10 December 2012; revised 16 March 2013; accepted 19 March 2013; published online 2 May 2013

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 not only increases with geographical 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 data set (Hurwitz and Sullivan, 2013) employing quantitative methodologies (John *et al.*, 2011; Duhaime and Sullivan, 2012; Duhaime *et al.*, 2012; Hurwitz *et al.*, 2013; Solonenko *et al.*, in press) 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 (for example, 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 (for example, 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.*, 2011a, b). 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 the two approaches: either viruses are concentrated and then adsorbed to TEM grids (for example, Sommaruga *et al.*, 1995; Stopar *et al.*, 2003) or they are directly deposited onto TEM grids using traditional (for example, 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 six 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 20 s 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 in 2% uranyl acetate (Ted Pella, Redding, CA, USA) for 30 s followed by three 10-s washes in ultra-pure water (Milli-Q, Millipore, Billerica, MA, USA), with excess liquid wicked away by 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 $\times 65\,000$–$100\,000$ magnification using a transmission electron microscope (Philips CM12, FEI, Hilsboro, OR, USA) with 100 kV 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 (US National Institutes of Health, Bethesda, MD, USA; 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-Aldrich, St. Louis, MO, USA). One 400-µl volume was processed immediately (termed 'fresh') using the qTEM method, while another 400 µl 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 (20 ml) 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 4 °C and concentrated to 250 µl with 100 kDa cutoff centrifugal filter units (Amicon, Millipore). Triplicate grids were prepared from these concentrated samples using 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 min 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 (Supplementary Figure S1, Supplementary Table S1). A rosette equipped with a CTD (Sea-Bird Electronics, Bellevue, WA, USA; SBE 911*plus* with Searam recorder), dissolved oxygen sensor (Sea-Bird Electronics; SBE 43) and fluorometer (WET Labs, Philomath, OR, USA; 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 (2 ml) were preserved with EM-grade glutaraldehyde (final concentration 2%), flash-frozen and stored in liquid nitrogen aboard the ship and at −80 °C on land until analysis. Samples (400 µl) were thawed at room temperature (ca. 22 °C) 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 >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 the 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 the 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 (Supplementary Figure S2A) or capsid diameter distribution (Supplementary Figure S2B). Although 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 the work presented here. Second, we found no significant difference between the samples prepared immediately (fresh) and those prepared after storage in liquid nitrogen (frozen) for either viral morphotype composition (Supplementary Figure S2C) or capsid diameter distributions (Supplementary Figure S2D). Third, the percentage of each viral morphotype was not significantly different between samples prepared using either the adsorption or the qTEM method with seawater from three marine environments (Supplementary 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 six oceanic regions with a range of environmental conditions (Supplementary 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 Supplementary 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 (Supplementary Table S2, Supplementary Figure S10). For correlations among individual samples, non-tailed viruses and podoviruses were correlated with salinity while myoviruses and podoviruses were correlated with temperature (Supplementary Table S2). However, these relationships reflected changes in the range of the relative percentage of these morphotypes and were often driven by only 3–4 samples (Supplementary Figure S10). No morphotype was significantly correlated with oxygen or chlorophyll concentration (Supplementary Table S2).

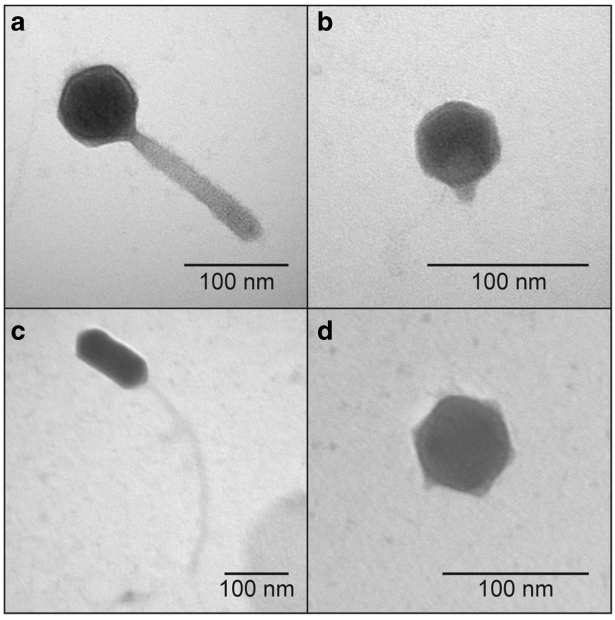

Figure 1 Examples of the four viral morphotypes observed in this study ((**a**), myovirus; (**b**), podovirus; (**c**), siphovirus; (**d**), non-tailed virus).

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 (Supplementary Table S1) with average capsid diameter positively correlated with salinity for individual samples (Supplementary Table S2, Supplementary Figure S10). There were no significant relationships between average capsid diameter and environmental parameters when considering pooled data for oceanic regions (Supplementary Table S2).

CA to compare sample capsid diameter distributions, as well as capsid diameter bins (Figures 3a and 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 (Figure 3a), with some surface samples more similar to the DCM sample at the same station (for example, station 41) and others much more divergent (for example, 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's 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's 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; Figures 3a). 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 (Figures 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 (49–63 nm; Supplementary Figure S4) and similar environmental conditions (low temperature plus higher salinity and oxygen; Supplementary Table S1). By contrast, samples from the DCM at station 41 and surface at stations 34 and 41 from the

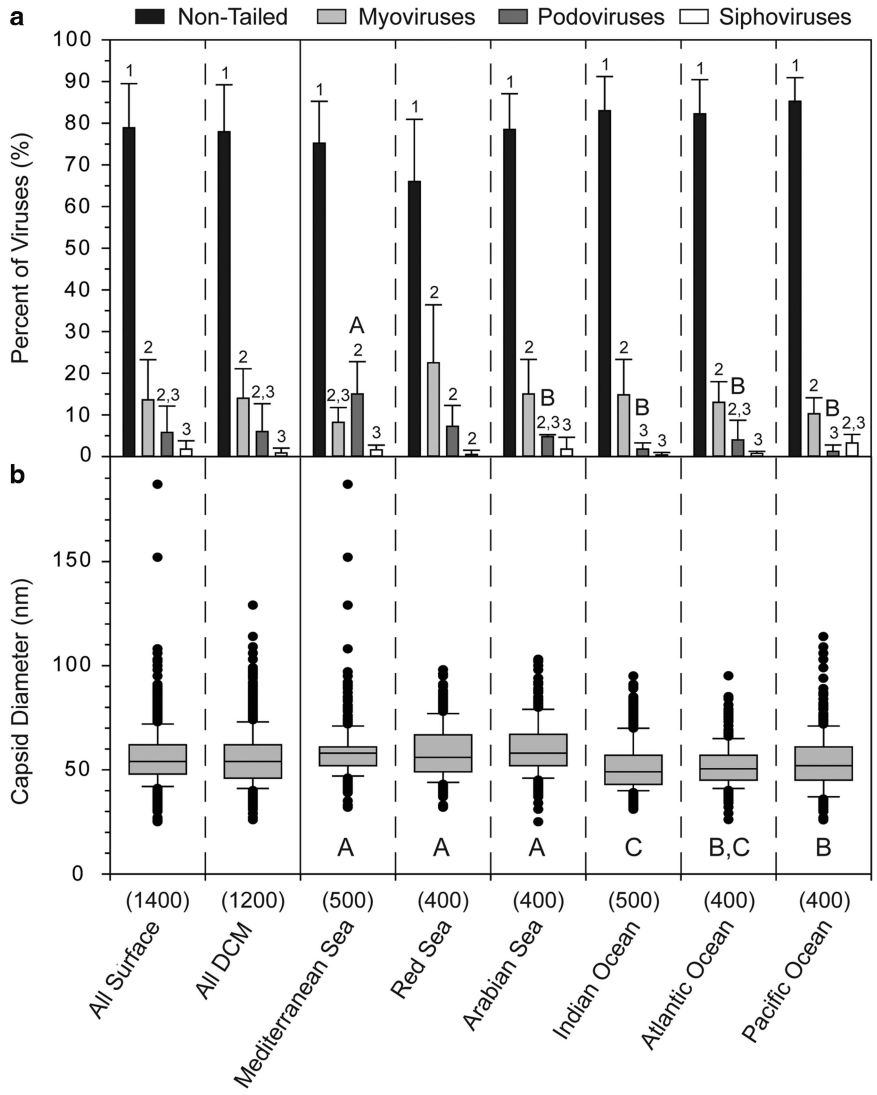

Figure 2 (**a**) Percentage of viral morphotypes in all the surface samples combined, all the DCM samples combined and each oceanic region. Error bars represent s.ds. of the means of all the samples. Letters indicate significant differences between depths or oceanic regions while numbers indicate significant differences within depths or oceanic regions (ANOVA with Tukey's *post-hoc* test, $P<0.001$ for all). (**b**) Box and whisker plots of viral capsid diameters in all the surface samples combined, all the 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's *post-hoc* test, $P<0.001$ for all). The number of viruses used for each data set is given in parentheses.

Red and Arabian Seas grouped together (Figure 3a), sharing wider capsid diameter peaks (49–91 nm; Supplementary Figures S5 and S6) and similar environmental conditions (higher salinity and temperature, lower oxygen; Supplementary 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's 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 (Supplementary Table S3).

Similarly, the ordination analyses were relatively insensitive to capsid diameter bin size. Analyses using each of the minimum (5 nm) and maximum (10 nm) optimal bin sizes determined for the samples provided similar results for the influence of environmental parameters on capsid diameters of viral assemblages (Supplementary Table S3).

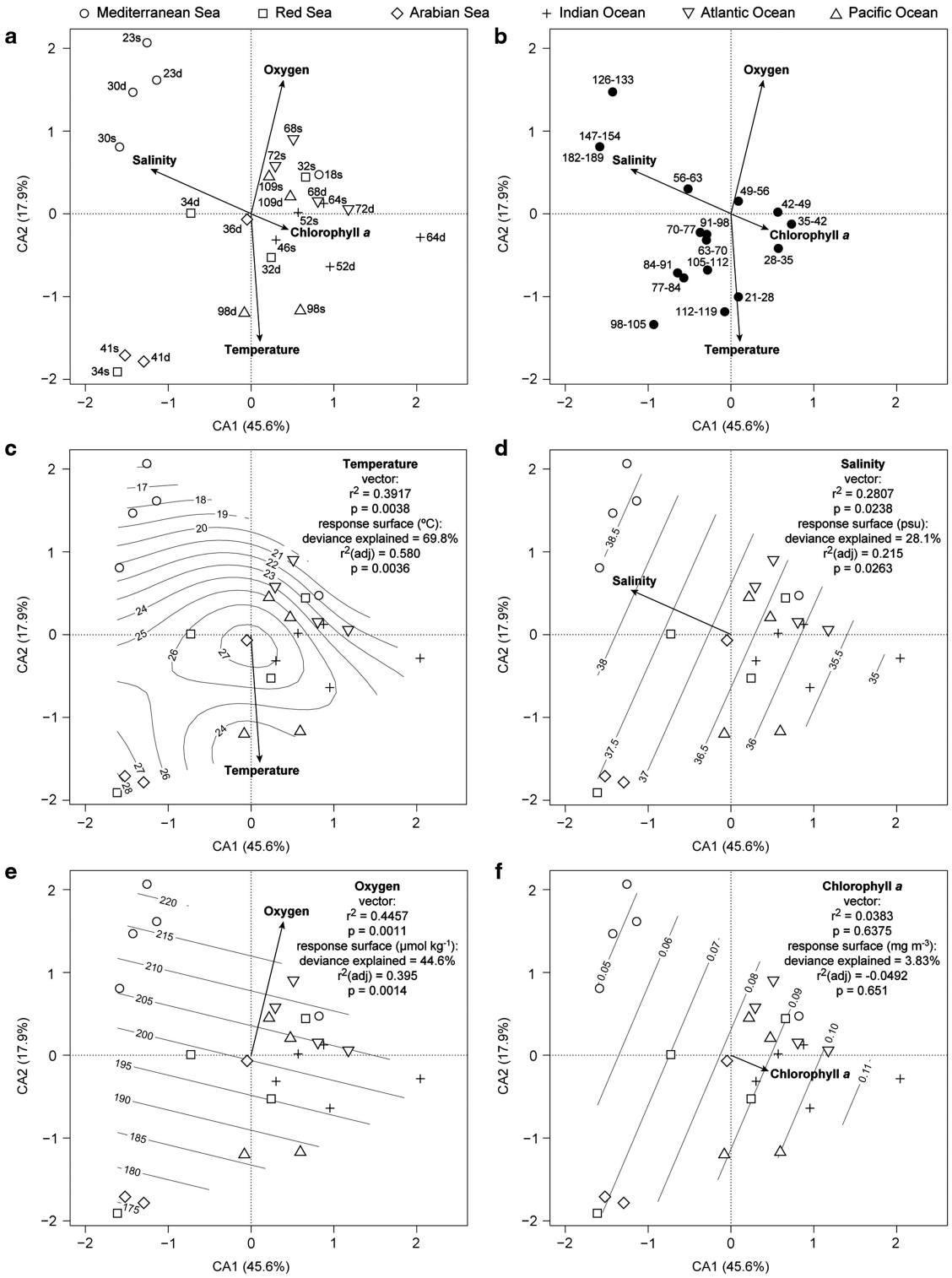

Figure 3 Ordination of Tara Oceans samples (**a**) and capsid diameter bins in nm (**b**) using CA based on distribution of viral capsid diameters with 7 nm bins (s, surface sample; d, DCM sample; surface sample from station 36 is omitted due to missing oxygen data; percentage of total inertia explained by CA1 and CA2 is reported on the axes). Lengths of vectors overlaid on the sample ordination plot correspond to the strength of influence for each environmental variable, with r^2 and *P*-values reported for each vector (**c–f**). Response surfaces for each environmental variable are also overlaid on the sample ordination plot to assess linearity of the relationship, with r^2 (adjusted), *P*-values and the percentage of deviance explained reported for each response surface (**c–f**). CA1 was negatively correlated with salinity (Pearson's correlation, $r = -0.486$, $P = 0.014$) while CA2 was negatively correlated with temperature (Pearson's correlation, $r = -0.623$, $P < 0.001$) and positively correlated with oxygen (Pearson's correlation, $r = 0.646$, $P < 0.001$).

Exceptions include reduced significance of the temperature vector and oxygen response surface with 10 nm 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 Mediterranean Sea, Red Sea and Atlantic Ocean; siphovirus tails were longer in the Red Sea than Mediterranean Sea; and podovirus tail lengths were not significantly different among the oceanic regions (Figure 4). Correlation analyses between tail lengths and environmental variables were not attempted owing to low sample sizes.

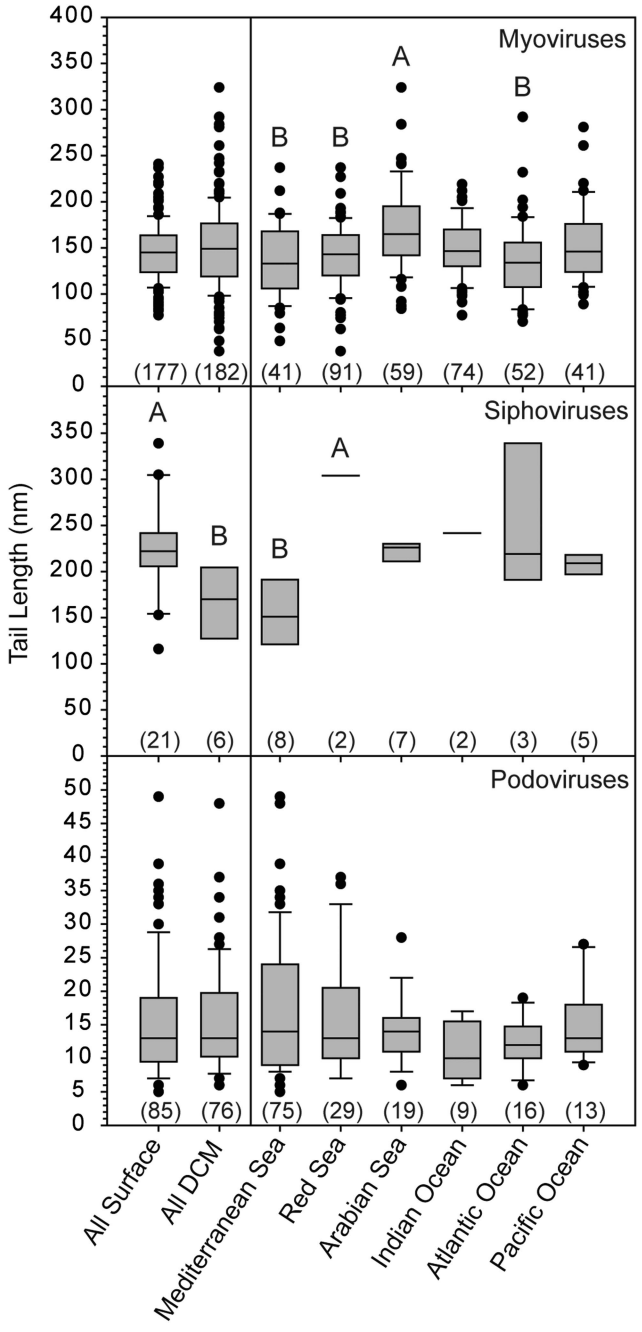

Figure 4 Box and whisker plots of myovirus, siphovirus and podovirus tail lengths in all the surface samples combined, all the DCM samples combined and each oceanic region. Refer to Figure 2 for a description of box and whisker plot construction. The number of viruses used for each data set is given in parentheses. Letters indicate significant differences between depths (*t*-test, $P=0.001$) or oceanic regions (ANOVA on ranks with Dunn's *post-hoc* test, $P<0.05$ for all).

Global marine viral morphological characteristics

Pooling all the sample data allowed examination of overall characteristics of upper water column viruses. Again, non-tailed viruses dominated (averaging 79% of all the viruses), followed by myoviruses, podoviruses and siphoviruses, in that order (Figure 5a). Myoviruses had the largest overall capsid diameters followed by siphoviruses, podoviruses and non-tailed viruses, with combined tailed viruses having significantly larger capsids than non-tailed viruses (Figure 5b). Also, tail lengths statistically differed with siphoviruses having the longest tails, followed by myoviruses, then podoviruses (Figure 5c). In addition, 48% of the 27 observed siphoviruses had prolate capsids and 3% of all the observed myoviruses had both capsid diameters and tail lengths either within or smaller than the dimensions described for dwarf myoviruses (Comeau *et al.*, 2012).

Discussion

Global ocean qTEM analyses showed that while viral assemblage morphological attributes vary between samples, there is little evidence for consistent variation with depth or oceanic region. The proportion of observed morphotypes (myoviruses, podoviruses, siphoviruses and non-tailed viruses) was highly similar in each oceanographic region, 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. Although no strong relationships between viral morphotype percentages and environmental variables emerged, larger average viral capsid diameters were significantly associated with higher salinity in

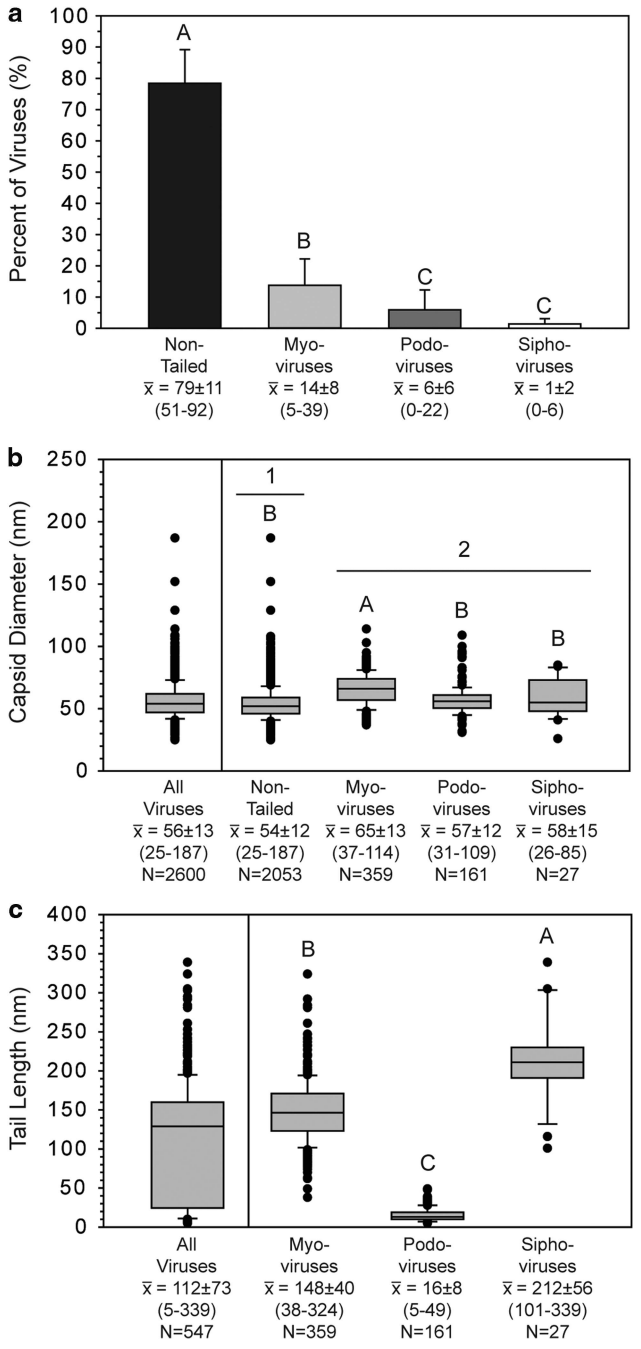

Figure 5 Morphological results of all the viruses in this study, including the percentage of each morphotype (**a**), as well as capsid diameters (**b**) and tail lengths (**c**) of all the viruses and each morphotype. The average and s.d. are given for each set of viruses, with ranges reported in parentheses, and the number of viruses analyzed (*N*) is given for capsid diameters and tail lengths. Refer to Figure 2 for a description of box and whisker plot construction. Letters 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, $P<0.001$).

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 the 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 next logical step in refining our understanding of marine viral biogeography.

Only four morphotypes were observed in this study, indicating that other morphotypes (for example, 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 (for example, tail length). We estimate that 5–100-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. By 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 ($118\,000 \times g$) was insufficient to force them to a prone position. This would result in erroneous documentation of podoviruses as non-tailed viruses

Table 1 Percentage 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[a]	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

[a]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.

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 before 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 versus 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, although 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 (for example, 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 are derived from eukaryotes, including 28 dsDNA viruses isolated from marine algae; three ssDNA viruses isolated from marine diatoms; and six RNA viruses isolated from diatoms, a fungoid protist and picophytoplankton (Table 2). Although 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 (115–220 nm), smaller eukaryotic RNA viruses (22–32 nm), larger eukaryotic RNA viruses (90–95 nm) and smaller ssDNA phages (30–32 nm) each comprised $<1\%$ of non-tailed viruses in the Tara Oceans samples, while eukaryotic ssDNA viruses

Table 2 Published non-tailed viruses isolated from marine bacteria and single-celled eukaryotes

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
	06 N-58	06 N-58 P	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

Abbreviation: NR, not reported.

(30–38 nm) and larger ssDNA phages (72–77 nm) only comprised 3 and 5%, respectively. However, the lone dsDNA and RNA non-tailed phages isolated from marine bacteria had 60 nm capsids, which most closely represented the mean capsid diameter for Tara Oceans non-tailed viruses (54 ± 12 nm). 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 the 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 (for example, 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 (for example, viral tagging; Deng *et al.*, 2012) or are completely cultivation-independent (for example, 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). Also, 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 (for example, 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.

Conflict of Interest

The authors declare no conflict of interest.

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; Céline Dimier and Marc Picheral for assistance with environmental data acquisition; Jesse Czekanski-Moir for suggesting, and assistance with, correspondence analysis; Brian Enquist for 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. We thank the coordinators and members of the Tara Oceans consortium (http://www.embl.de/tara-oceans/start/) for organizing sampling and data analysis. We thank the commitment of the following people and sponsors who made this singular expedition possible: CNRS, EMBL, Genoscope/CEA, VIB, Stazione Zoologica Anton Dohrn, UNIMIB, ANR (projects POSEIDON/ANR-09-BLAN-0348, BIOMARKS/ANR-08-BDVA-003, PROMETHEUS/ANR-09-GENM-031 and TARA-GIRUS/ANR-09-PCS-GENM-218), EU FP7 (MicroB3/No. 287589), FWO, BIO5, Biosphere 2, agnès b., the Veolia Environment Foundation, Region Bretagne, World Courier, Illumina, Cap L'Orient, the EDF Foundation EDF Diversiterre, FRB, the Prince Albert II de Monaco Foundation, Etienne Bourgois, the Tara schooner and its captain and crew. Tara Oceans would not exist without continuous support from 23 institutes (http://oceans.taraexpeditions.org). This article is contribution number 0004 of the Tara Oceans Expedition 2009–2012.

References

Abramoff MD, Magalhaes PJ, Ram SJ. (2004). Image processing with ImageJ. *Biophotonics Int* **11**: 36–42.

Ackermann HW. (2007). 5500 phages examined in the electron microscope. *Arch Virol* **152**: 227–243.

Ackermann HW. (2001). Frequency of morphological phage descriptions in the year 2000. *Arch Virol* **146**: 843–857.

Ackermann H-W, Heldal M. (2010). Basic electron microscopy of aquatic viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds). *Manual of Aquatic Viral Ecology*. ASLO: Waco, pp 182–192.

Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J *et al.* (2013). Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol; e-pub ahead of print 14 March 2013; Doi:10.1111/1462-2920.12100.*

Andrews-Pfannkoch C, Fadrosh DW, Thorpe J, Williamson SJ. (2010). Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. *Appl Environ Microbiol* **76**: 5039–5045.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C *et al.* (2006). The marine viromes of four oceanic regions. *PLoS Biol* **4**: 2121–2131.

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

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

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

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

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

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

Bergeron A, Belcaid M, Steward GF, Poisson G. (2007). Divide and conquer: enriching environmental sequencing data. *PLoS ONE* **2**: e830.

Bergh O, Borsheim KY, Bratbak G, Heldal M. (1989). High abundance of viruses found in aquatic environments. *Nature* **340**: 467–468.

Bettarel Y, Bouvier T, Bouvier C, Carre C, Desnues A, Domaizon I *et al.* (2011a). Ecological traits of planktonic viruses and prokaryotes along a full-salinity gradient. *FEMS Microbiol Ecol* **76**: 360–372.

Bettarel Y, Bouvier T, Agis M, Bouvier C, Van Chu T, Combe M *et al.* (2011b). Viral distribution and life strategies in the Bach Dang Estuary, Vietnam. *Microb Ecol* **62**: 143–154.

Borsheim KY. (1993). Native marine bacteriophages. *FEMS Microbiol Ecol* **102**: 141–159.
Borsheim KY, Bratbak G, Heldal M. (1990). Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. *Appl Environ Microbiol* **56**: 352–356.
Bratbak G, Heldal M, Norland S, Thingstad TF. (1990). Viruses as partners in spring bloom microbial trophodynamics. *Appl Environ Microbiol* **56**: 1400–1405.
Bratbak G, Haslund OH, Heldal M, Naess A, Roeggen T. (1992). Giant marine viruses? *Mar Ecol Prog Ser* **85**: 202–202.
Breitbart M. (2012). Marine viruses: truth or dare. *Ann Rev Mar Sci* **4**: 425–448.
Breitbart M, Miyake JH, Rohwer F. (2004a). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* **236**: 249–256.
Breitbart M, Thompson LR, Suttle CA, Sullivan MB. (2007). Exploring the vast diversity of marine viruses. *Oceanography* **20**: 135–139.
Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P *et al.* (2004b). Diversity and population structure of a near-shore marine-sediment viral community. *Proc R Soc B* **271**: 565–574.
Brum JR, Steward GF. (2010). Morphological characterization of viruses in the stratified water column of alkaline, hypersaline Mono Lake. *Microb Ecol* **60**: 636–643.
Brum JR, Steward GF. (2011). Physical fractionation of aquatic viral assemblages. *Limnol Oceanogr Methods* **9**: 150–163.
Brum JR, Culley AI, Steward GF. (2013). Assembly of a marine viral metagenome after fractionation. *PLoS ONE* **8**: e60604.
Castberg T, Thyrhaug R, Larsen A, Sandaa R-A, Heldal M, Van Etten JL *et al.* (2002). Isolation and characterization of a virus that infects *Emiliania huxleyi* (Haptophyta). *J Phycol* **38**: 767–774.
Chen F, Lu J. (2002). Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol* **68**: 2589–2594.
Clokie M, Shan J, Bailey S, Jia Y, Krisch HM, West S *et al.* (2006). Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol* **8**: 827–835.
Colombet J, Sime-Ngando T, Cauchie HM, Fonty G, Hoffmann L, Demeure G. (2006). Depth-related gradients of viral activity in Lake Pavin. *Appl Environ Microbiol* **72**: 4440–4445.
Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, Tovkach FI *et al.* (2012). Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. *PLoS ONE* **7**: e40102.
Cottrell MT, Suttle CA. (1991). Wide-spread occurrence and clonal variation in viruses which cause lysis of a cosmopolitan eukaryotic marine phytoplankter, *Micromonas pusilla*. *Mar Ecol Prog Ser* **78**: 1–9.
Culley AI, Lang AS, Suttle CA. (2006). Metagenomic analysis of coastal RNA virus 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.
Duhaime MB, Sullivan MB. (2012). Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **434**: 181–186.
Duhaime MBD, Deng L, Poulos BT, Sullivan MB. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ Microbiol* **14**: 2526–2537.
Edwards RA, Rohwer F. (2005). Viral metagenomics. *Nat Rev Microbiol* **3**: 504–510.
Espejo RT, Canelo ES. (1968). Properties of bacteriophage PM2: a lipid-containing bacterial virus. *Virology* **34**: 738–747.
Fuhrman JA. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548.
Fuller NJ, Wilson WH, Joint IR, Mann NH. (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.
Geslin C, Gaillard M, Flament D, Rouault K, Le Romancer M, Prieur D *et al.* (2007). Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi*. *J Bacteriol* **189**: 4510–4519.
Gowing MM. (1993). Large virus-like particles from vacuoles of phaeodarian radiolarians and from other marine samples. *Mar Ecol Prog Ser* **101**: 33–43.
Hammond GW, Hazelton PR, Chuang I, Klisko B. (1981). Improved detection of viruses by electron microscopy after direct ultracentrifuge preparation of specimens. *J Clin Microbiol* **14**: 210–221.
Hidaka T, Ichida K-i. (1976). Properties of a marine RNA-containing bacteriophage. *Mem Fac Fish Kagoshima Univ* **25**: 77–89.
Hofer JS, Sommaruga R. (2001). Seasonal dynamics of viruses in an alpine lake: importance of filamentous forms. *Aquat Microb Ecol* **26**: 1–11.
Holmfeldt K, Odic D, Sullivan MB, Middelboe M, Riemann L. (2012). Cultivated single stranded DNA phages that infect marine *Bacteroidetes* prove difficult to detect with DNA binding stains. *Appl Environ Microbiol* **78**: 892–894.
Hurwitz BL, Sullivan MB. (2013). The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **8**: e57355.

Hurwitz BL, Deng L, Poulos BT, Sullivan MB. (2013). Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol; e-pub ahead of print 9 July 2013; Doi:10.1111/j.1462-2920.2012.02836.x*.

Ignacio-Espinoza JC, Sullivan MB. (2012). Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ Microbiol* **14**: 2113–2126.

Jaakkola ST, Penttinen RK, Vilen ST, Jalasvuori M, Ronnholm G, Bamford JKH *et al.* (2012). Closely related archaeal *Haloarcula hispanica* icosahedral viruses HHIV-2 and SH1 have nonhomologous genes encoding host recognition functions. *J Virol* **86**: 4734–4742.

Jacobsen A, Bratbak G, Heldal M. (1996). Isolation and characterization of a virus infecting *Phaeocystis pouchetii* (Prymnesiophyceae). *J Phycol* **32**: 923–927.

John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S *et al.* (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* **3**: 195–202.

Karner MB, Delong EF, Karl DM. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507–510.

Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J *et al.* (2011). A holistic approach to marine ecosystems 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, CA, USA.

Krebs CJ. (1999). *Ecological Methodology*, 2nd edn Addison-Welsey Educational Publishers, Inc.: Menlo Park, CA, USA.

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.

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T *et al.* (2007). Genome-side expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83–86.

Lu J, Chen F, Hodson RE. (2001). Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Appl Environ Microbiol* **67**: 3285–3290.

Maranger R, Bird DF, Juniper SK. (1994). Viral and bacterial dynamics in Arctic sea ice during the spring algal bloom near Resolute, N.W.T., Canada. *Mar Ecol Prog Ser* **111**: 121–127.

Marston MF, Sallee JL. (2003). Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* **69**: 4639–4647.

McDaniel LD, DelaRosa M, Paul JH. (2006). Temperate and lytic cyanophages from the Gulf of Mexico. *J Mar Biol Assoc UK* **86**: 517–527.

Nagasaki K, Yamaguchi M. (1997). Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). *Aquat Microb Ecol* **13**: 135–140.

Nagasaki K, Tomaru Y, Katanozake N, Shirai Y, Nishida K, Itakura S *et al.* (2004). Isoloation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Appl Environ Microbiol* **70**: 704–711.

Nagasaki K, Tomaru Y, Takao Y, Nishida K, Shirai Y, Suzuki H *et al.* (2005). Previously unknown virus infects marine diatom. *Appl Environ Microbiol* **71**: 3528–3535.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2013). vegan: Community Ecology Package. R package version 2.1-27/r2451. http://R-Forge.R-project.org/projects/vegan/.

Oren A, Bratbak G, Heldal M. (1997). Occurrence of virus-like particles in the Dead Sea. *Extremophiles* **1**: 143–149.

Pomeroy LR, Williams PJl, Azam F, Hobbie JE. (2007). The microbial loop. *Oceanography* **20**: 28–33.

Prangishvili D, Forterre P, Garrett RA. (2006). Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**: 837–848.

Proctor LM. (1997). Advances in the study of marine viruses. *Microsc Res Tech* **37**: 136–161.

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.

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.

Solonenko SA, Ignacio-Espinoza JC, Alberti A, Cruaud C, Hallam S, Konstantinidis K *et al.* (2013). Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* (in press).

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.

Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: 1344–1357.

Suttle CA. (2007). Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.

Suttle CA, Chan AM. (1995). Viruses infecting the marine Prymnesiophyte *Chrysochromulina* spp.: isolation, preliminary charactrization and natural abundance. *Mar Ecol Prog Ser* **118**: 275–282.

Tai V, Lawrence JE, Lang AS, Chan AM, Culley AI, Suttle CA. (2003). Charactrization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *J Phycol* **39**: 343–352.

Takao Y, Nagasaki K, Mise K, Okuno T, Honda D. (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.

Tapper MA, Hicks RE. (1998). Temperate viruses and lysogeny in Lake Superior bacterioplankton. *Limnol Oceanogr* **43**: 95–103.

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.

Tomaru Y, Katanozake N, Nishida K, Shirai Y, Tarutani K, Yamaguchi M *et al.* (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.

Tucker KP, Parsons R, Symonds EM, Breitbart M. (2011). Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J* **5**: 822–830.

Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**: 3393–3399.

Weinbauer MG. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**: 127–181.

Weinbauer MG, Peduzzi P. (1994). Frequency, size and distribution of bacteriophages in different marine bacterial morphotypes. *Mar Ecol Prog Ser* **108**: 11–20.

Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH. (2009). From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ Microbiol* **11**: 2821–2839.

Weynberg KD, Allen MJ, Gilg IC, Scanlan DJ, Wilson WH. (2011). Genome sequence of *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. *J Virol* **85**: 4520–4529.

Williamson KE, Helton RR, Wommack KE. (2012). Bias in bacteriophage morphological classification by transmission electron microscopy due to breakage or loss of tail structures. *Microsc Res Tech* **75**: 452–457.

Wilson WH, Joint IR, Carr NG, Mann NH. (1993). Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol* **59**: 3736–3743.

Wolf S, Muller D, Maier I. (2000). Assembly of a large icosahedral DNA virus, MclaV-1, in the marine alga *Myriotrichia clavaeformis* (Dictyosiphonales, Phaeophyceae). *Eur J Phycol* **35**: 163–171.

Wommack KE, Colwell RR. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69–114.

Wood SN. (2011). Fast stable restricted maximum likelihood estimation of semiparametric generalized linear models. *J R Stat Soc B* **73**: 3–36.

Zar J. (1996). *Biostatistical Analysis*. Prentice Hall: Upper Saddle River, NJ, USA.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)