

# Re-examination of the relationship between marine virus and microbial cell abundances

Charles H. Wigington<sup>1</sup>, Derek Sonderegger<sup>2</sup>, Corina P. D. Brussaard<sup>3,4</sup>, Alison Buchan<sup>5</sup>, Jan F. Finke<sup>6</sup>, Jed A. Fuhrman<sup>7</sup>, Jay T. Lennon<sup>8</sup>, Mathias Middelboe<sup>9</sup>, Curtis A. Suttle<sup>6,10,11,12</sup>, Charles Stock<sup>13</sup>, William H. Wilson<sup>14</sup>, K. Eric Wommack<sup>15</sup>, Steven W. Wilhelm<sup>5\*</sup> and Joshua S. Weitz<sup>1,16\*</sup>

**Marine viruses are critical drivers of ocean biogeochemistry, and their abundances vary spatiotemporally in the global oceans, with upper estimates exceeding  $10^8$  per ml. Over many years, a consensus has emerged that virus abundances are typically tenfold higher than microbial cell abundances. However, the true explanatory power of a linear relationship and its robustness across diverse ocean environments is unclear. Here, we compile 5,671 microbial cell and virus abundance estimates from 25 distinct marine surveys and find substantial variation in the virus-to-microbial cell ratio, in which a 10:1 model has either limited or no explanatory power. Instead, virus abundances are better described as nonlinear, power-law functions of microbial cell abundances. The fitted scaling exponents are typically less than 1, implying that the virus-to-microbial cell ratio decreases with microbial cell density, rather than remaining fixed. The observed scaling also implies that viral effect sizes derived from 'representative' abundances require substantial refinement to be extrapolated to regional or global scales.**

Viruses of microbes have been linked to central processes across the global oceans, including biogeochemical cycling<sup>1–6</sup> and the maintenance and generation of microbial diversity<sup>1,4,7–9</sup>. Virus propagation requires that virus particles both contact and subsequently infect cells. The per cell rate at which microbial cells—including bacteria, archaea and microeukaryotes—are contacted by viruses is assumed to be proportional to the product of virus and microbial abundances<sup>10</sup>. If virus and microbe abundances were related in a predictable way it would be possible to infer the rate of virus–cell contacts from estimates of microbial abundance alone.

Virus ecology underwent a transformation in the late 1980s with the recognition that virus abundances, as estimated using culture-independent methods, were orders of magnitude higher than estimates via culture-based methods<sup>11</sup>. Soon thereafter, researchers began to report the 'virus to bacterium ratio' (VBR) as a statistical proxy for the strength of the relationship between viruses and their potential hosts in both freshwater and marine systems<sup>12</sup>. This ratio is more appropriately termed the 'virus-to-microbial cell ratio' (VMR), a convention that we use here (Supplementary Section 1).

Observations accumulated over the past 25 years have noted a wide variation in the VMR, yet there is a consensus that a suitable first approximation is that the VMR is equal to 10 (Supplementary Table 1). This ratio also reflects a consensus that typical microbial

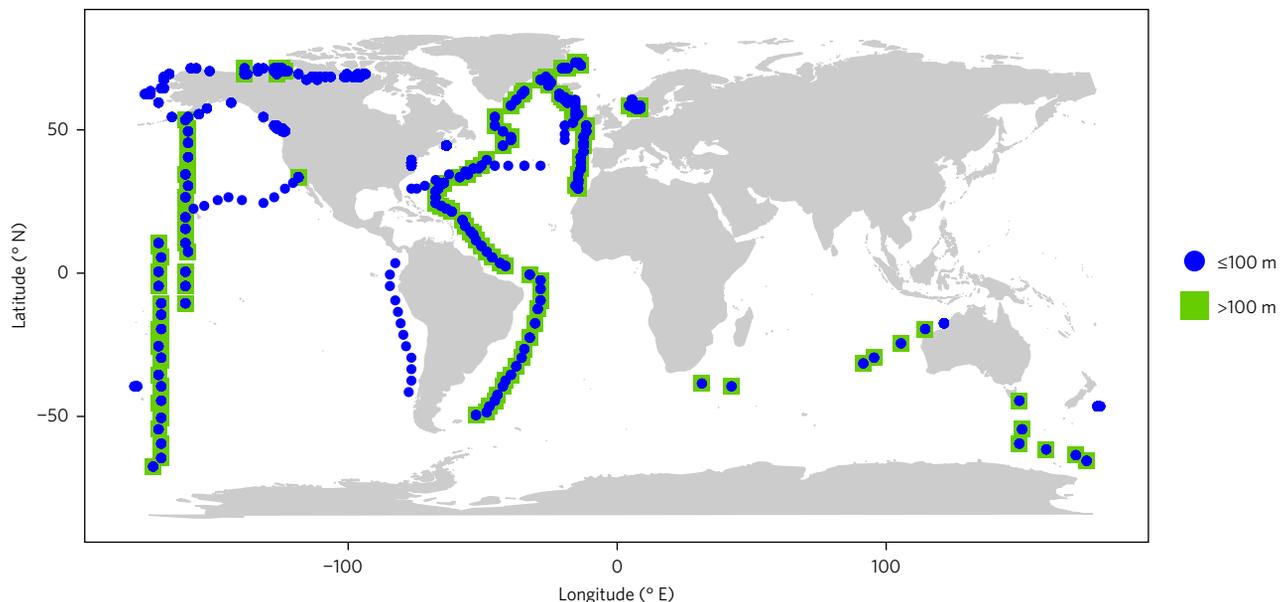
abundances are approximately  $10^6$  per ml and typical virus abundances are approximately  $10^7$  per ml<sup>13,14</sup>. Yet, the use of a fixed ratio carries with it another assumption—that of linearity—that is, if microbial abundance were to double, then viruses are expected to double as well. An alternative is that the relationship between virus and microbial abundances is better described in terms of a nonlinear relationship, for example, a power law.

In practice, efforts to predict the regional or global-scale effects of viruses on marine microbial mortality, turnover and even biogeochemical cycles depend critically on the predictability of the relative density of viruses and microbial cells. The expected community-scale contact rate, as inferred from the product of virus and microbial abundances, is a key factor for inferring virus-induced cell lysis rates at a site or sites (see, for example, ref. 15), which also depend on diversity<sup>16</sup>, latent infections<sup>17</sup> and virus–microbe infection networks<sup>18</sup>. Here, we directly query the nature of the relationship between virus and microbial densities via a large-scale compilation and re-analysis of abundance data across marine environments.

## Results

**VMR exhibits substantial variation in the global oceans.** In the compiled marine survey data (Fig. 1, Table 1 and the 'Materials and methods'), 95% of microbial abundances range from  $5.0 \times 10^3$

<sup>1</sup>School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. <sup>2</sup>Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, Arizona 86011, USA. <sup>3</sup>Department of Biological Oceanography, Royal Netherlands Institute for Sea Research (NIOZ), 1790 AB Den Burg, Texel, The Netherlands. <sup>4</sup>Department of Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, 1090 GE, Amsterdam, The Netherlands. <sup>5</sup>Department of Microbiology, The University of Tennessee, Knoxville, Tennessee 37996, USA. <sup>6</sup>Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. <sup>7</sup>Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA. <sup>8</sup>Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. <sup>9</sup>Marine Biological Section, Department of Biology, University of Copenhagen, DK-3000, Helsingør, Denmark. <sup>10</sup>Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. <sup>11</sup>Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. <sup>12</sup>Program in Integrated Microbial Diversity, Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8, Canada. <sup>13</sup>Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey 08540, USA. <sup>14</sup>Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK. <sup>15</sup>Plant and Soil Sciences, Delaware Biotechnology Institute, Delaware Technology Park, Newark, Delaware 19711, USA. <sup>16</sup>School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. \*e-mail: wilhelm@utk.edu; jsweitz@gatech.edu



**Figure 1 | Global distribution and information regarding sample sites.** Each point denotes a location from which one or more samples were taken. Sampling depth ranged from 5,500 m below sea level up to the surface, with 2,921 taken near the surface ( $\leq 100$  m, noted as circles) and 2,750 below that ( $>100$  m, noted as squares). The numbers of points for each study—the ‘frequency’—are provided in Supplementary Table 2. Full access to data is provided as Supplementary Data 1.

to  $4.1 \times 10^6$  per ml, and 95% of virus abundances range from roughly  $3.7 \times 10^5$  to  $6.4 \times 10^7$  per ml (Fig. 2a). Both microbial and virus concentrations generally decrease with depth, as reported previously (for example, see ref. 19). We separated the samples according to depth using an operational definition of the near-surface and sub-surface, corresponding to samples taken at depths  $\leq 100$  m and  $>100$  m, respectively. The cutoff of 100 m was chosen as a typical depth scale for the euphotic zone in systems with low to moderate chlorophyll<sup>20</sup>. The precise depth varies spatiotemporally. Our intent was to distinguish zones strongly shaped by active planktonic foodweb dynamics in well-lit waters (the ‘near-surface’) from dark mesopelagic waters shaped primarily by decaying particle fluxes with greater depth (the ‘sub-surface’). The median VMR for the near-surface samples ( $\leq 100$  m) is 10.5, and the median VMR for the sub-surface samples ( $>100$  m) is 16.0. In that sense, the consensus 10:1 ratio does accurately represent the median VMR for the surface data. We also observe substantial variation in the VMR, as has been noted in earlier surveys and reviews (Supplementary Table 1). Figure 2b shows that 95% of the variation in VMR in the near-surface ocean lies between 1.4 and 160, and between 3.9 and 74 in the sub-surface ocean. For the near-surface ocean, 44% of the VMR values are between 5 and 15, 16% are less than 5 and 40% exceed 15. This wide distribution, both for the near-surface and the sub-surface, demonstrates the potential limitations in using the 10:1 VMR, or any fixed ratio, as the basis for a predictive model of virus abundance derived from estimates of microbial abundance.

**Virus abundance does not vary linearly with microbial abundance.** Figure 3 shows two alternative predictive models of the relationship between logarithmically scaled virus and microbial abundances for water column samples. The models correspond to a fixed-ratio model and a power-law model. To clarify the interpretation of fitting in log–log space, consider a fixed-ratio model with a 12:1 ratio between virus and microbial abundance,  $V = 12 \times B$ . Then, in log–log space the relationship is

$$\log_{10}(V) = \log_{10}12 + \log_{10}B \quad (1)$$

which we interpret as a line with a  $y$  intercept of  $\log_{10}12 = 1.08$  and a slope (change in  $\log_{10}V$  for a one-unit change in  $\log_{10}B$ ) of 1. By the same logic, any fixed-ratio model will result in a line with slope 1 in the log–log plot, and the  $y$  intercept will vary logarithmically with VMR. The alternative predictive model is that of a power law:  $V = cB^{\alpha_1}$ . In log–log space, the relationship is

$$\log_{10}V = \log_{10}c + \alpha_1 \log_{10}B \quad (2)$$

$$\log_{10}V = \alpha_0 + \alpha_1 \log_{10}B \quad (3)$$

The slope  $\alpha_1$  of a fitted line on log-transformed data denotes the power-law exponent that best describes the relationship between the variables. The intercept  $\alpha_0$  of a fitted line on log-transformed data denotes the logarithmically transformed pre-factor.

The 10:1 line has residual squared errors of  $-16\%$  and  $-25\%$  in the surface and deep samples, respectively (Supplementary Table 3). In both cases, this result means that a 10:1 line explains less of the variation in virus abundance compared with a model in which virus abundance is predicted by its mean value across the data. To evaluate the generality of this result, we considered an ensemble of fixed-ratio models, each with a different VMR. In the near-surface samples, we find that all fixed-ratio models explain less of the variation (that is, have negative values of  $R^2$ ) than a ‘model’ in which virus abundance is predicted to be the global mean in the data set (Supplementary Fig. 1). This reflects the failure of constant-ratio (that is, linear) models to capture the cluster of high VMRs at low microbial density apparent in the density contours of Fig. 2a and the shoulder of elevated high VMR frequency in Fig. 2b. The largest contributor to this cluster of points is the ARCTICSBI study (Table 1). In the sub-surface samples, fixed-ratio models in which the VMR varies between 12 and 22 do have positive explanatory power, but all perform worse than the power-law model (Supplementary Fig. 1). In contrast, the best-fitting power-law model explains 15 and 64% of the variation in the data for near- and sub-surface samples, respectively (Supplementary Table 3). The best-fit power-law scaling exponents are 0.42 (with 95% confidence intervals (CIs) of (0.39,0.46)) for near-surface samples and 0.53 (with 95% CIs of (0.52,0.55)) for sub-surface samples.

**Table 1 | Virus and microbial abundance data from 25 different marine virus abundance studies from 11 different laboratory groups.**

Study name	Laboratory	Study type	Location	Regime	Citation
NORTHSEA2001	Bratbak	Spatial	North Sea	Coastal	Supplementary Data 1
RAUNEFJORD2000	Bratbak	Temporal	North Sea	Coastal	Supplementary Data 1
BATS	Breitbart	Temporal	Sargasso Sea	Non-coastal	Ref. 21
STRATIPHYT1	Brussaard	Spatial	North Atlantic Transect	Non-coastal	Ref. 15
STRATIPHYT2	Brussaard	Spatial	North Atlantic Transect	Non-coastal	Supplementary Data 1
USC MO	Fuhrman	Temporal	Santa Barbara Channel	Non-coastal	Ref. 22
GEOTRACES	Herndl	Spatial	Atlantic Transect	Non-coastal	Ref. 23
GEOTRACES_LEG3	Herndl	Spatial	Atlantic Transect	Non-coastal	Supplementary Data 1
BEDFORDBASIN	Li	Temporal	North Atlantic Ocean	Coastal	Ref. 24
GREENLAND 2012	Middelboe	Spatial	Greenland Sea	Non-coastal	Supplementary Data 1
INDIANOCEAN2006	Middelboe	Spatial	Indian Ocean	Non-coastal	Supplementary Data 1
KH04-5	Nagata	Spatial	Southern Pacific Ocean	Non-coastal	Ref. 25
KH05-2	Nagata	Spatial	Northern Pacific Ocean	Non-coastal	Ref. 25
CASES03-04	Suttle	Spatial	Arctic Ocean	Non-coastal	Ref. 9
SOG/ELA/TROUT/SWAT	Suttle	Temporal	Pacific Ocean—Strait of Georgia	Coastal	Ref. 26
ARCTICSB1	Wilhelm	Spatial	Gulf of Alaska	Coastal	Ref. 27
FECYCLE1	Wilhelm	Spatial	South Pacific Ocean	Non-coastal	Ref. 28
FECYCLE2	Wilhelm	Spatial	South Pacific Ocean	Non-coastal	Ref. 29
NASB2005	Wilhelm	Spatial	North Atlantic Ocean	Non-coastal	Ref. 30
POWOW	Wilhelm	Spatial	Pacific Ocean	Non-coastal	Supplementary Data 1
TABASCO	Wilhelm	Spatial	South Pacific Ocean	Non-coastal	Ref. 31
MOVE	Wommack	Temporal	Atlantic—Chesapeake	Coastal	Ref. 32

A total of 5,671 data points were aggregated. The data collection dates range from 2000 to 2011. For sampling convenience, data primarily comes from coastal waters in the northern hemisphere and were collected predominately during the summer months, with the notable exceptions of long-term coastal monthly monitoring sites (USC MO, BATS, Chesapeake Bay).

The difference between a linear and a power-law model can be understood, in part, by comparing predictions of viral abundances as a function of variation in microbial abundances. For example, doubling the microbial abundance along either regression line is not expected to lead to a doubling in virus abundance, but rather a  $2^{0.42} = 1.3$ - and  $2^{0.53} = 1.4$ -fold increase, respectively. The difference between models becomes more apparent with scale, for example, 10- and 100-fold increases in near-surface microbial abundances are predicted to be associated with  $10^{0.53} = 3.4$ - and  $100^{0.53} = 11$ -fold increases in viral abundances, respectively, given a power-law model. The power-law model is an improvement over the fixed-ratio model in both the near- and sub-surface, even when accounting for the increase in parameters (Supplementary Table 3). In the near-surface, refitting the surface data without outliers improves the explanatory power to approximately  $R^2 = 0.3$  from an  $R^2 = 0.65$  for the sub-surface (Methods and Supplementary Fig. 2). The power-law exponents in the near- and sub-surface are qualitatively robust to variation in the choice of depth threshold, for example, as explored over the range between 50 and 150 m (Supplementary Fig. 3). In summary, the predictive value of a power-law model is much stronger in the sub-surface than in the near-surface, where confidence in the interpretation of power-law exponents is limited.

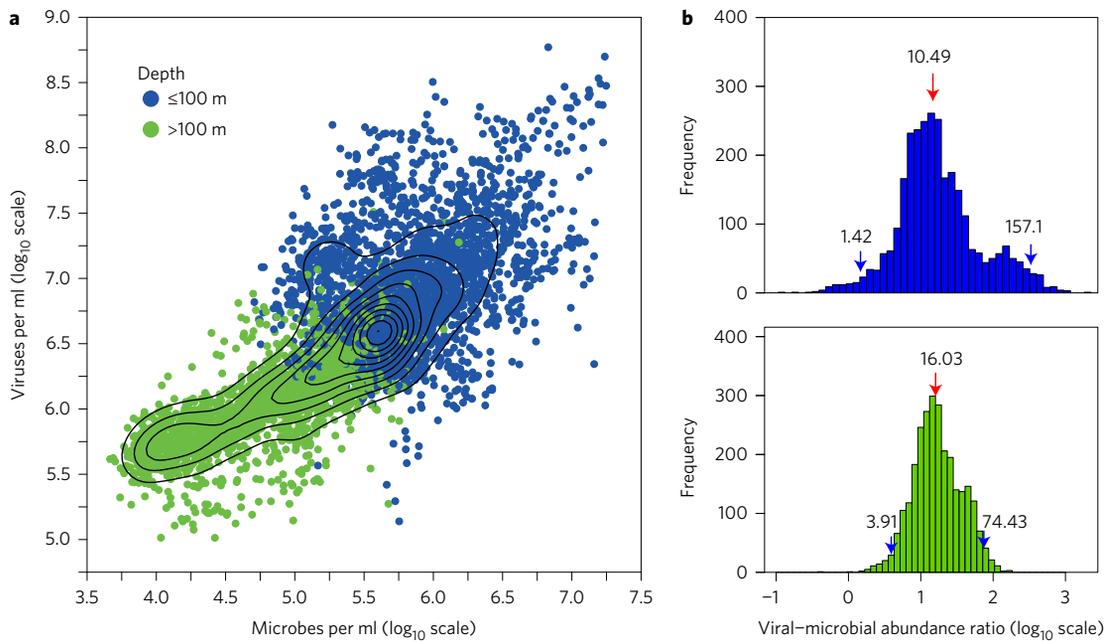
#### Study-to-study measurement variation is unlikely to explain the intrinsic variability of virus abundances in the surface ocean.

Next, we explored the possibility that the variation in methodologies affected the baseline offset of virus abundance measurements and thereby decreased the explanatory power of predicting virus abundances based on microbial abundances. That is, if  $V^*$  is the true and unknown abundance of viruses, then it is possible that two studies would estimate  $\hat{V}_1 = V^*(1 + \epsilon_1)$  and  $\hat{V}_2 = V^*(1 + \epsilon_2)$ , where  $|\epsilon_1|$  and  $|\epsilon_2|$  denote the relative magnitude of study-specific shifts. We constrain the relative variation in measurement, such that the measurement uncertainty is 50% or less (see 'Materials and methods'). The constrained regression model improves the explanatory power of the model (Supplementary Table 3), but, in doing so, the model forces 18 of the 25 studies to the maximum level of measurement variation

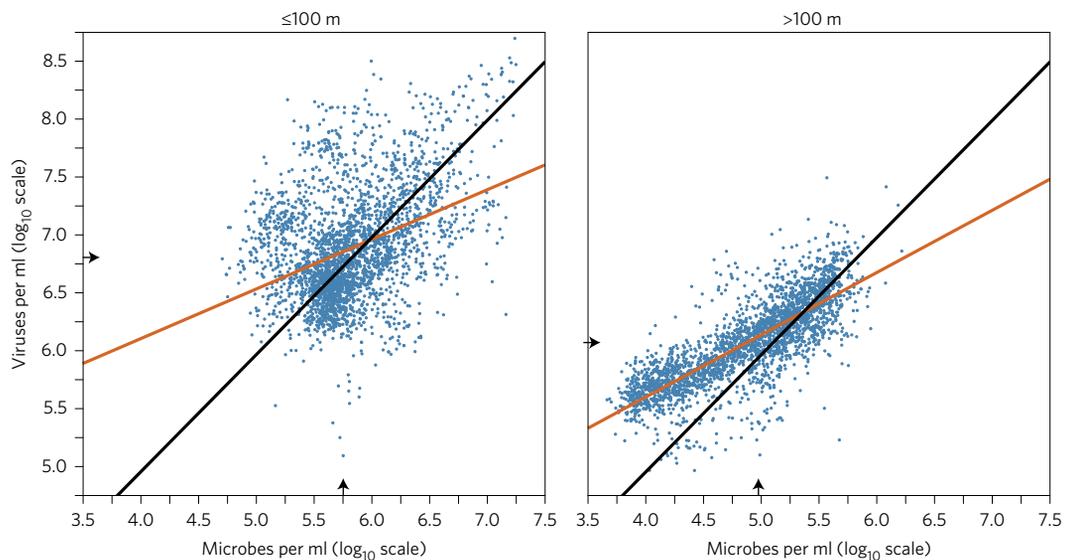
permitted (Supplementary Fig. 4). We do not expect differences in measurement protocols to explain the nearly two orders of magnitude variation in estimating virus abundance, given the same true virus abundance in a sample. Note that when sub-surface samples were analysed through the constrained power-law model, there was only a marginal increase of 2% in  $R^2$  and, moreover, 9 of the 12 studies were fit given the maximum level of measurement variation permitted (Supplementary Fig. 4). The constrained intercept model results suggest that the observed variation in virus abundance in the surface oceans is not well explained strictly by the variation in measurement protocol between studies.

#### VMR decreases with increasing microbial abundance, a hallmark of power-law relationships.

We next evaluate an ensemble of power-law models,  $V_i = c_i N^{\alpha_i}$ , where index  $i$  denotes the use of distinct intercepts and power-law exponents for each survey. The interpretation of this model is that the nonlinear nature of the virus-to-microbial relationship may differ in distinct oceanic realms or due to underlying differences in sites or systems, rather than due to measurement differences. Figure 4 presents the results of fitting using the study-specific power-law model in the surface ocean samples. Study-specific power-law fits are significant in 18 of 25 cases in the surface ocean. The median power-law exponent for studies in the surface ocean is 0.50. Furthermore, of those significant power-law fits, the 95% distribution of the power-law exponent excludes a slope of one and is entirely less than one in 11 of 18 cases (Fig. 5). This model, in which the power-law exponent varies with study, is a significant improvement in terms of  $R^2$  (Supplementary Table 3). For sub-surface samples, study-specific power-law fits are significant in 10 of 12 cases in the sub-surface (Supplementary Fig. 5). The median power-law exponent for studies in the sub-surface is 0.67. Of those significant power-law fits, the central 95% distribution of the power-law exponent is less than one in 6 of 10 cases (Supplementary Fig. 6). A power-law exponent of less than one means that the virus abundance increases less than proportionately given increases in microbial abundance. This study-specific analysis extends the findings that



**Figure 2 | Variation in virus and microbial abundances and the VMR.** **a**, Microbial abundance versus virus abundance coloured by depth. Each point represents a biological sample. Contours denote regions of equal probability distribution in the data set. **b**, Histogram of the logarithm of VMR. Top and bottom panels correspond to near- and sub-surface water column samples, respectively. The red arrow denotes the median value and the blue arrows denote the central 95% range of values. The numbers associated with each arrow denote the non-transformed value of VMR.

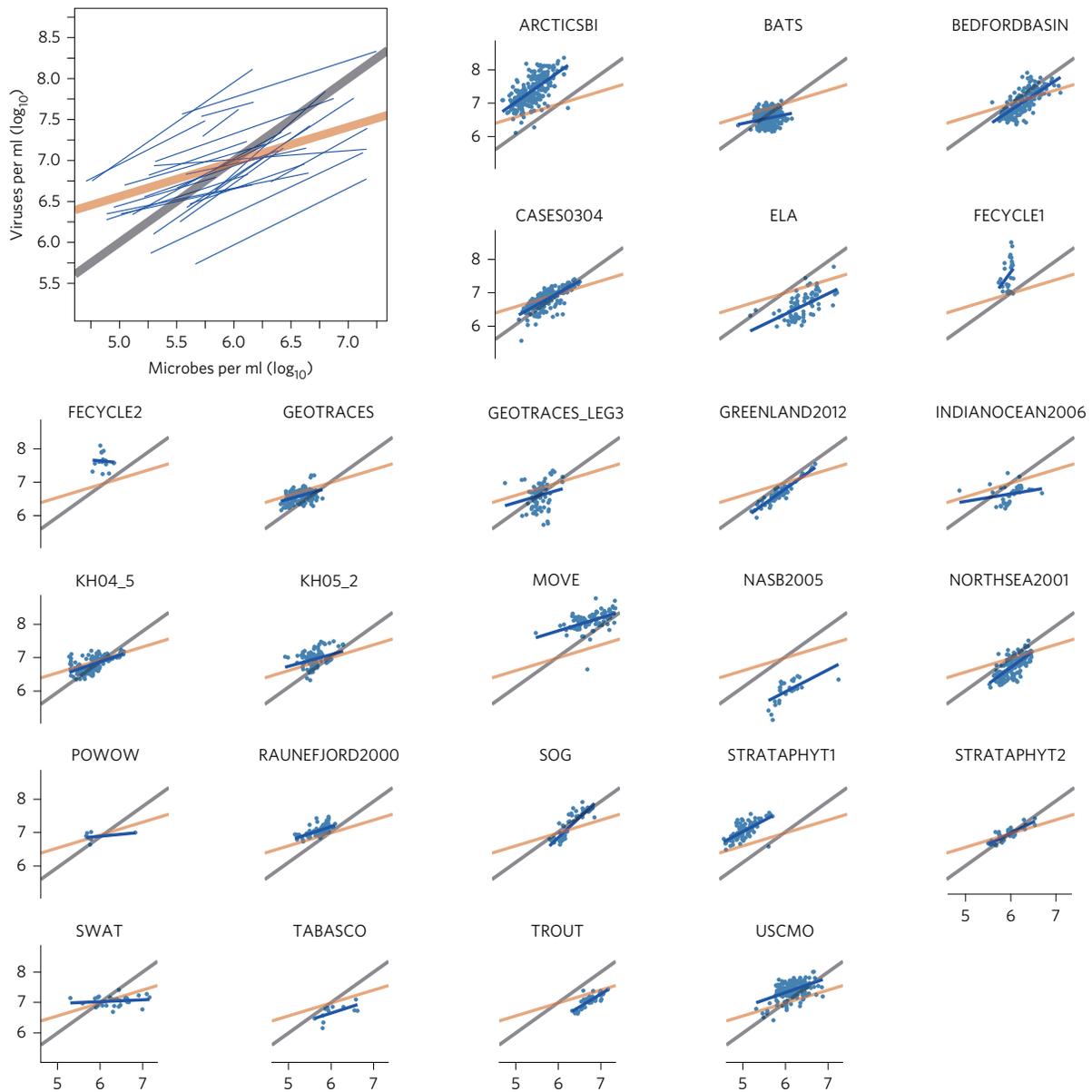


**Figure 3 | Virus abundance is poorly fit by a model of a tenfold increase relative to microbial abundance.** Left: surface ocean. The orange line denotes the best-fit power law with an exponent of 0.42 and the black line denotes the 10:1 curve. The best-fit power law explains 15% of the variation and the 10:1 line explains ~16% of the variation. See main text for the interpretation of negative  $R^2$  and the importance of outliers in these fits. Right: deeper water column. The orange line denotes the best-fit power law with an exponent of 0.53 and the black line denotes the 10:1 curve. The best-fit power law explains 64% of the variation and the 10:1 line explains ~26% of the variation. In both cases the arrows on the axes denote the median of the respective abundances.

nonlinear, rather than linear, models are more suited to describing the relationship between virus and microbial abundances. We find that the dominant trend in both near-surface and sub-surface samples is that the VMR decreases as microbial abundance increases. The increased explanatory power by study is stronger for near-surface than for sub-surface samples. This increase in  $R^2$  comes with a caveat: study-specific models do not enable *a priori* predictions of virus abundance given a new environment or sample site, without further effort to disentangle the biotic and abiotic factors underlying the different scaling relationships.

## Discussion

Viruses are increasingly considered in efforts to describe the factors controlling marine microbial mortality, productivity and biogeochemical cycles<sup>3,4,33–36</sup>. Quantitative estimates of virus-induced effects can be measured directly, but are often inferred indirectly using the relative abundance of viruses to microbial cells. To do so, there is a consensus that assuming the VMR is 10 in the global oceans—despite the observed variation—is a reasonable starting point. Here, we have re-analysed the relationship of virus to microbial abundances in 25 marine survey data sets. We find that

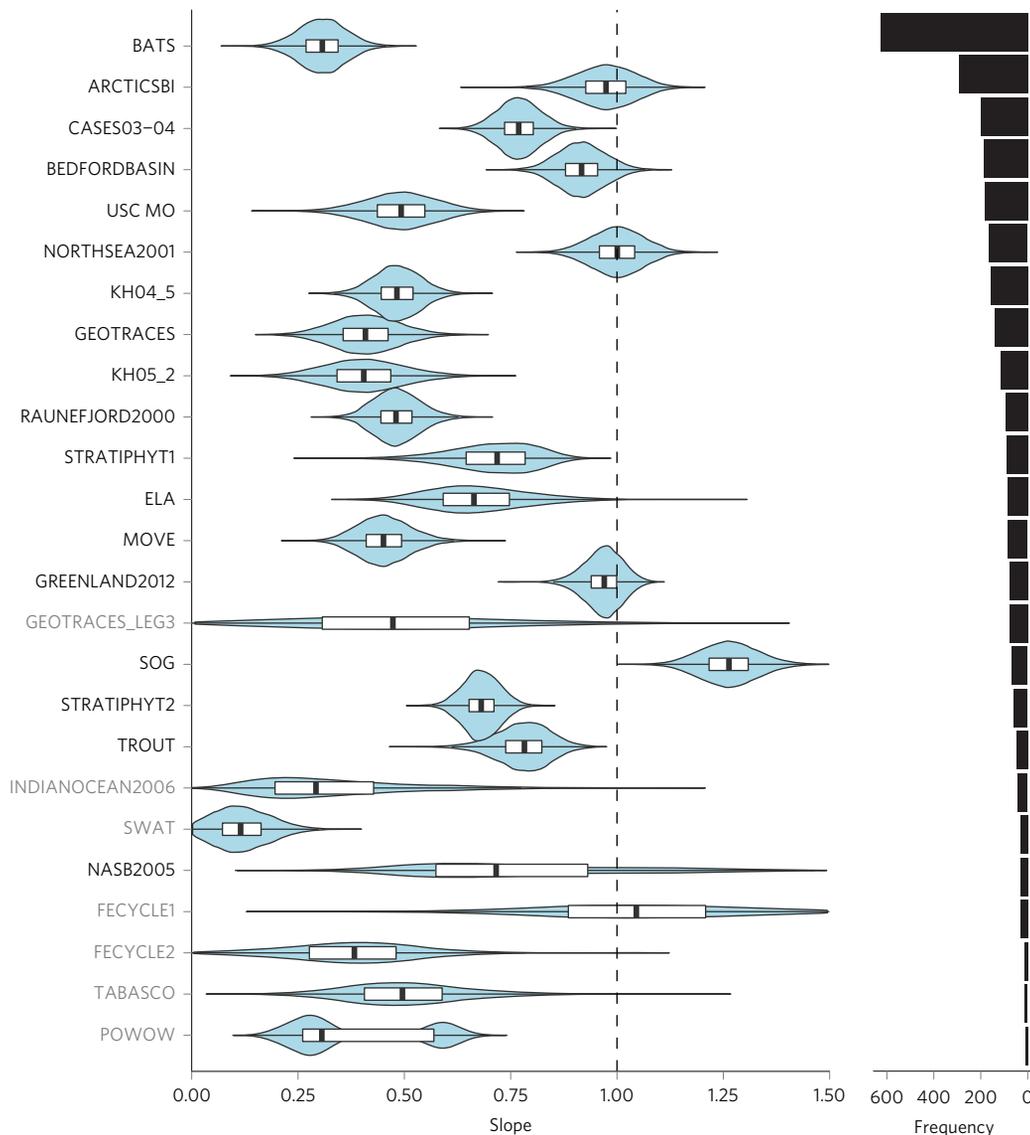


**Figure 4 | Virus-microbial relationships given the variable slope and intercept mixed-effects model.** Upper left: best-fit power law for each study (blue lines) plotted together with the best-fit power law of the entire data set (orange line) and the 10:1 line (grey line). Individual panels: best-fit power-law model (blue line) on log-transformed data (blue points) for each study, with the power-law model regression (orange) and 10:1 line (grey) as reference. The power-law exponents and associated confidence intervals are shown in Fig. 5.

95% of the variation in VMR ranges from 1.4 to 160 in the near-surface ocean and from 3.9 to 74 in the sub-surface. Although the 10:1 ratio accurately describes the median of the VMR in the surface ocean, the broad distribution of VMR implies that microbial abundance is a poor quantitative predictor of virus abundance. Moreover, increases in microbial abundance do not lead to proportionate increases in virus abundance. Instead, we propose that the virus to microbial abundance relationship is nonlinear and that the degree of nonlinearity—as quantified via a power-law exponent—is typically less than 1. This sublinear relationship can be interpreted to mean that the VMR decreases as an increasing function of microbial abundance and generalizes earlier observations<sup>13</sup>. Power-law relationships between virus and microbial abundance emerge from complex feedbacks involving both exogenous and endogenous factors. The question of exogenous factors could be addressed, in part, by examining environmental covariates at survey sites. For example, if microbial and virus abundances

varied systematically with another environmental co-factor during a transect, then this would potentially influence the inferred relationship between virus and microbial abundances. In that same way, variation in environmental correlates, including temperature and incident radiation, may directly modify virus life history traits<sup>37,38</sup>. Also, some of the marine survey data sets examined here constitute repeated measurements at the same location (for example, in the Bermuda Atlantic Time-series Study (BATS)). Time-varying environmental factors could influence the relative abundance of microbes and viruses. It is also interesting to note that viruses-induced mortality is considered to be more important at eutrophic sites<sup>13</sup>, where microbial abundance is higher, yet the observed decline in VMR with microbial abundance would suggest the opposite.

It could also be the case that variation in endogenous factors determines total abundances. Endogenous factors can include the life history traits of viruses and microbes that determine which



**Figure 5 | Study-specific 95% confidence intervals of power-law exponents for relationships between virus and microbial abundances in the surface.**

The confidence intervals are plotted using 'violin' plots that include the median (centre black line), 75% distribution (white bars) and 95% distribution (black line), with the distribution overlaid (blue shaded area). The numbers of points included as part of each study are displayed on the rightmost bar plots. Study labels in black indicate those studies for which the regression fit had a  $P$  value less than  $0.002 = 0.05/25$  (accounting for a multiple comparison correction given the analysis of 25 studies). Study labels in grey indicate a  $P$  value above this threshold.

hosts are infected by which viruses<sup>18</sup>, as well as the quantitative rates of growth, defence and infection. For example, relative strain abundances are predicted to depend on niche differences according to the 'kill-the-winner' theory, which presupposes tradeoffs between growth and defence<sup>1,39</sup>. Similarly, the recent hypothesis of a complementary 'king-of-the-mountain' mechanism suggests that relative abundance relationships may depend on life history trait differences, even when tradeoffs are not strict<sup>40</sup>. In both examples, total abundances may nonetheless depend on other factors, including the strength of grazing.

The analysis of abundance relationships also requires a consideration of variation in time. As is well known, virus-microbe interactions can lead to intrinsic oscillatory dynamics. Indeed, previous observations of a declining relationship between VMR and microbial abundance have been attributed to changing ratios across phytoplankton bloom events, including possible virus-induced termination of blooms<sup>13</sup>. Similar arguments have been proposed in the analysis of tidal sediments<sup>41</sup>. Alternatively, observations

of declining VMR with microbial density have been attributed to a variation in underlying diversity<sup>42</sup>. Another factor potentially complicating abundance predictions is that episodic events, including the induction of lysogenic populations, influence total microbial and viral counts. Varying degrees of lysogenic and co-infection relationships have been measured in marine virus-host systems<sup>14,17,43</sup>, the consequences of which may differ from those given interactions with lytic viruses, as is commonly the focus of model- and empirical-based studies. Whatever the mechanism(s), it is striking that virus abundances in some surveys can be strongly predicted via alternative power-law functions of microbial abundances. Mechanistic models are needed to further elucidate these emergent macroecological patterns and relationships, akin to recent efforts to explain emergent power laws between terrestrial predators and prey<sup>44</sup>.

The present analysis first separated the abundance data according to depth and then according to survey as a means to identify different relationships between virus and microbial abundances in

the global oceans. The predictive value of total microbial abundance is strong when considering sub-surface samples. In contrast, microbial abundance is not a strong predictor of virus abundance in near-surface samples, when using linear or nonlinear models. The predictive power of nonlinear models improved substantially in the near-surface when evaluating each marine survey separately. The minimal predictive value of microbial cell abundances for inferring viral abundances in the near-surface when aggregating across all surveys is problematic given that virus–microbe interactions have significant roles in driving microbial mortality and ecosystem functioning<sup>3,5,33</sup>. Indeed the aggregation of abundance measurements in terms of total microbial abundances may represent part of the problem.

At a given site and time of sampling, each microbial cell in the community is potentially targeted by a subset of the total viral pool. In moving forward, understanding the variation in virus abundance and its relationship to microbial abundance requires a critical examination of correlations at functionally relevant temporal and spatial scales, that is, at the scale of interacting pairs of viruses and microbes. These scales will help inform comparisons of virus–microbe contact rates with viral-induced lysis rates, thereby linking abundance and process measurements. We encourage the research community to prioritize examination of these scales of interaction as part of efforts to understand the mechanisms underlying nonlinear virus–microbe abundance relationships in the global oceans.

## Methods

**Data source.** Marine virus abundance data were aggregated from 25 studies (Table 1). A total of 5,671 data points were aggregated. The data collection dates ranged from 1996 to 2012. Data were primarily collected from coastal waters in the northern hemisphere, predominately during the summer months, with the notable exceptions of long-term coastal monthly monitoring sites, that is, the studies USC MO, BATS and MOVE.

**Data processing.** Analyses of the data were performed using R version 3.1.1. Scripts and original data are provided at [https://github.com/WeitzGroup/Virus\\_Microbe\\_Abundance](https://github.com/WeitzGroup/Virus_Microbe_Abundance).

**Power-law model.** A power-law regression model used the  $\log_{10}$  of the predictor variable, microbial abundance per ml  $N$  and the  $\log_{10}$  of the outcome variable, virus abundance per ml  $V$ . The power-law regression was calculated using the equation  $\log_{10} V = \alpha_0 + \alpha_1 \log_{10} N$ . The  $\alpha_0$  and  $\alpha_1$  parameters were fit via ordinary least squares (OLS) regression to minimize the sum of square error.

**Constrained variable-intercept model.** The constrained model is a ‘mixed-effects’ regression model using the same predictor and outcome variables,  $\log_{10}$  of microbial abundance per ml and the  $\log_{10}$  virus abundance per ml, respectively. This model includes study-specific intercepts, which were constrained such that the values for any of the intercepts were restricted to one standard error above or below the intercept value taken from the power-law model. The standard error value for this model came from the power-law model. The equation for this model is  $V = \alpha_0^{(i)} + \alpha_1 N$ , where  $\alpha_0^{(i)}$  is the study-specific intercept and  $\alpha_1$  is the slope common to all studies,  $N$  is the predictor variable and  $V$  is the outcome variable.

**Variable slope and variable intercept model.** A power-law model where the exponent and intercept varied with each study was evaluated using the same predictor variable,  $\log_{10}$  microbial abundance per ml, and the same outcome variable,  $\log_{10}$  virus abundance per ml. In this model, there was a study-specific  $\alpha_0$  and  $\alpha_1$  and an OLS regression calculated using the equation  $V = \alpha_0^{(i)} + \alpha_1^{(i)} N$ .

**Bootstrapping model CIs.** Bootstrap analyses of the power-law model and mixed-effects models were conducted to derive 95% CIs surrounding the parameters estimated by the models. For all models the original data set was sampled with replacement, by study, to arrive at a bootstrap sample data set; this process was repeated 10,000 times. Distributions for all parameters were generated and the 2.5, 50 and 97.5% points were identified from among the 10,000 parameter estimates.

**Outlier identification.** Outliers in the data were identified by calculating the top and bottom 2% of estimated VMRs amongst the entire 5,671 samples. The outliers corresponded to ratios below 1.81 and above 128. Those samples with VMRs that fell outside these bounds were considered outliers. There were 218 outlier samples taken at depths of  $\leq 100$  m and 10 outlier samples taken at depths of  $> 100$  m.

**Depth cutoff robustness.** The cutoff point for which data were partitioned into either the near-surface or the sub-surface was varied from 50 m to 150 m in 1 m increments. For each step, a power-law model was evaluated for both the near-surface and the sub-surface.

Received 15 September 2015; accepted 7 December 2015;  
published 25 January 2016

## References

- Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**, 1320–1328 (2000).
- Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
- Brussaard, C. P. D. *et al.* Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J.* **20**, 575–578 (2008).
- Rohwer, F. & Thurber, R. V. Viruses manipulate the marine environment. *Nature* **459**, 207–212 (2009).
- Weitz, J. S. & Wilhelm, S. W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Reports* **4**, 17 (2012).
- Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W. & Weitz, J. S. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nature Rev. Microbiol.* **12**, 519–528 (2014).
- Weitz, J. S., Hartman, H. & Levin, S. A. Coevolutionary arms races between bacteria and bacteriophage. *Proc. Natl Acad. Sci. USA* **102**, 9535–9540 (2005).
- Avrani, S., Schwartz, D. A. & Lindell, D. Virus-host swinging party in the oceans: incorporating biological complexity into paradigms of antagonistic coexistence. *Mobile Genet. Elements* **2**, 88–95 (2012).
- Payet, J. P. & Suttle, C. A. To kill or not to kill: the balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol. Oceanogr.* **58**, 465–474 (2013).
- Murray, A. G. & Jackson, G. A. Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. *Marine Ecol. Progr. Ser.* **89**, 103–116 (1992).
- Bergh, O., Borhseim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
- Maranger, R. & Bird, D. F. Viral abundance in aquatic systems—a comparison between marine and fresh-waters. *Marine Ecol. Progr. Ser.* **1210**, 217–226 (1995).
- Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
- Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
- Mojica, K. D. A. *et al.* Phytoplankton community structure in relation to vertical stratification along a north–south gradient in the Northeast Atlantic Ocean. *Limnol. Oceanogr.* **60**, 1498–1521 (2015).
- Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nature Rev. Microbiol.* **3**, 504–510 (2005).
- Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W. & Sullivan, M. B. Seasonal time bombs: dominant temperate viruses affect southern ocean microbial dynamics. *ISME J.* <http://dx.doi.org/10.1038/ismej.2015.125> (2015).
- Weitz, J. S. *et al.* Phage–bacteria infection networks. *Trends Microbiol.* **21**, 82–91 (2013).
- Danovaro, R. *et al.* Marine viruses and global climate change. *FEMS Microbiol. Rev.* **35**, 993–1034 (2011).
- Morel, A. & Berthon, J. Surface pigments, algal biomass profiles, and potential production of the euphotic layer: relationships reinvestigated in view of remote-sensing applications. *Limnol. Oceanogr.* **340**, 1545–1562 (1989).
- Parsons, R. J., Breitbart, M., Lomas, M. W. & Carlson, C. A. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J.* **6**, 273–284 (2011).
- Fuhrman, J. A. *et al.* Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103**, 13104–13109 (2006).
- De Corte, D., Sintès, E., Yokokawa, T., Reinthaler, T. & Herndl, G. J. Links between viruses and prokaryotes throughout the water column along a north Atlantic latitudinal transect. *ISME J.* **60**, 1566–1577 (2012).
- Li, W. K. W. & Dickie, P. M. Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry* **440**, 236–246 (2001).
- Yang, Y., Yokokawa, T., Motegi, C. & Nagata, T. Large-scale distribution of viruses in deep waters of the Pacific and southern oceans. *Aquatic Microbial Ecol.* **710**, 193–202 (2013).
- Clasen, J. L., Brigden, S. M., Payet, J. P. & Suttle, C. A. Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshwater Biol.* **53**, 1090–1100 (2008).
- Balsom, A. L. *Macrofaunal Community Composition and Biomass, and Bacterial and Viral Abundances from the Gulf of Alaska to the Canadian Archipelago: a Biodiversity Study.* MSc thesis, Univ. Tennessee (2003).

28. Strzepek, R. F. *et al.* Spinning the 'ferrous wheel': the importance of the microbial community in an iron budget during the FeCycle experiment. *Global Biogeochem. Cycles* **19**, GB4S26 (2005).
29. Matteson, A. R. *et al.* Production of viruses during a spring phytoplankton bloom in the South Pacific Ocean near New Zealand. *FEMS Microbiol. Ecol.* **79**, 709–719 (2012).
30. Rowe, J. M. *et al.* Constraints on viral production in the Sargasso Sea and North Atlantic. *Aquatic Microbial Ecol.* **520**, 233–244 (2008).
31. Wilhelm, S. W. *et al.* UV radiation induced DNA damage in marine viruses along a latitudinal gradient in the southeastern Pacific Ocean. *Aquatic Microbial Ecol.* **310**, 1–8 (2003).
32. Wang, K., Eric Wommack, K. & Chen, F. Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl. Environ. Microbiol.* **770**, 7459–7468 (2011).
33. Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nature Rev. Microbiol.* **5**, 801–812 (2007).
34. Danovaro, R. *et al.* Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* **454**, 1084–1U27 (2008).
35. Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Rev. Microbiol.* **130**, 147–159 (2015).
36. Weitz, J. S. *et al.* A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J.* **90**, 1352–1364 (2015).
37. Suttle, C. A. & Chan, A. M. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Marine Ecol. Progr. Ser.* **92**, 99–109 (1993).
38. De Paepe, M. & Taddei, F. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol.* **4**, e193 (2006).
39. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecol.* **13**, 19–27 (1997).
40. Giovannoni, S., Temperton, B. & Zhao, Y. *Nature* **499**, E4–E5 (2013).
41. Carreira, C., Larsen, M., Glud, R. N., Brussaard, C. P. D. & Middelboe, M. Heterogeneous distribution of prokaryotes and viruses at the microscale in a tidal sediment. *Aquatic Microbial Ecol.* **69**, 183–192 (2013).
42. Bratbak, G. & Heldal, M. in *Molecular Ecology of Aquatic Microbes* Vol. 238 (ed. Joint, I.) 249–264 (Springer-Verlag, 1995).
43. Williamson, S. J., Houchin, L. A., McDaniel, L. & Paul, J. H. Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. *Appl. Environ. Microbiol.* **68**, 4307–4314 (2002).
44. Hatton, I. A. *et al.* The predator–prey power law: biomass scaling across terrestrial and aquatic biomes. *Science* **349**, aac6284 (2015).

### Acknowledgements

This work was supported by National Science Foundation (NSF) grants OCE-1233760 (to J.S.W.) and OCE-1061352 (to A.B. and S.W.W.), a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (to J.S.W.) and a Simons Foundation SCOPE grant (to J.S.W.). This work was conducted as part of the Ocean Viral Dynamics Working Group at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award DBI-1300426, with additional support from The University of Tennessee, Knoxville.

### Author contributions

C.H.W. developed the code and implemented all scripts, analysed data, performed statistical analysis and contributed to writing the manuscript. D.S. developed and reviewed the code, analysed data, performed statistical analysis and provided feedback on the manuscript. A.B., J.F., J.T.L., M.M., C.A.S., C.S., W.H.W. and K.E.W. contributed to the design and implementation of the study, the assessment and collection of data sets, and provided feedback on the manuscript. C.P.D.B. and J.F.F. contributed to the assessment and collection of data sets and provided feedback on the manuscript. S.W.W. co-led the design of the study, led the data collection and assessment component, and contributed to writing the manuscript. J.S.W. co-led the design of the study, led the code and statistical analysis component, and wrote the manuscript.

### Additional information

Supplementary information is available [online](http://www.nature.com/online). Reprints and permissions information is available online at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.W.W. and J.S.W.

### Competing interests

The authors declare no competing financial interests.