https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Open Access Earth System Science Data Discussions

A global viral oceanography database (gVOD)

Le Xie[1], Wei Wei[1], Lanlan Cai[1], Xiaowei Chen[1], Yuhong Huang[1], Nianzhi Jiao[1], Rui Zhang[1, *], Ya-Wei Luo[1, *]

[1]State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen, Fujian 361102, China

Correspondence to: Rui Zhang (ruizhang@xmu.edu.cn) and Ya-Wei Luo (ywluo@xmu.edu.cn)

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Open Access Earth System Science Data Discussions

Abstract. Virioplankton is a key component of marine biosphere in maintaining diversity of microorganisms and stabilizing ecosystems. They also greatly contribute to nutrient recycles by releasing organic matter after lysis of hosts. In this study, we constructed the first global viral oceanography database (gVOD) by collecting 10,931 viral abundance (VA) data and 727 viral production (VP) data, along with host and other oceanographic parameters when available. Most VA data were obtained in the North Atlantic (32%) and North Pacific Oceans (29%), while the Southeast Pacific and Indian Oceans were quite under sampled. The VA in the global ocean was 1.17(±3.31) $\times 10^7$ particles ml^{-1}. The lytic and lysogenic VP in the global ocean was 9.87(±24.16) $\times 10^5$ and 2.53(±8.64) $\times 10^5$ particles ml^{-1} h^{-1}, respectively. Average VA in coastal oceans was higher than that in surface open oceans [3.61(±6.30) $\times 10^7$ versus 0.73(±1.24) $\times 10^7$ particles ml^{-1}], while average VP in coastal and surface open oceans was close. Vertically, VA, lytic and lysogenic VP deceased from surface to deep ocean by about one order of magnitude. The total number of viruses in the global ocean estimated by bin average and random forest methods was 1.4 $\times 10^{30}$ particles and 1.39 $\times 10^{30}$ particles, leading to an estimate of global ocean viral biomass at 32.3 and 32.2 Tg C, respectively. We expect that the gVOD will be a fundamental and very useful database for laboratory, field and modeling studies in marine ecology and biogeochemistry. The full gVOD database (Xie et al., 2020) is stored at PANGAEA (a temporary link: https://www.pangaea.de/tok/19f9d7b496a00f57f491e639440708aee00b6a49).

1 Introduction

Virioplankton are the most abundant biological entities and one of the largest genetic reservoirs in the ocean (Breitbart, 2012; Fuhrman, 1999). With an estimation of ~10^{23} marine microbes being infected every second, viruses play important roles in affecting microbial mortality, regulating community composition and impacting biogeochemical cycles (Suttle, 2005; Zhang et al., 2007). Viruses were estimated to kill ~20–40% of marine bacterioplankton every day, similar to the microbial mortality caused by zooplankton grazing (Fuhrman, 1999). In particular, virus-mediated cell lysis effectively 'shunts' approximately 25% of the photosynthetically fixed carbon, which otherwise is transferred to higher trophic levels, to dissolved organic matter (DOM) pool, partly forming to the basis of the microbial loop and leading to the recycling of nutrients (Suttle, 2007; Wilhelm and Suttle, 1999). Furthermore, viral lysis can contribute to biological pump through the release of sticky lysates, accelerating the aggregation and sink of carbon into the deep sea (Suttle, 2005).

Compilation of the observations of viral abundance and activity in the global ocean is very necessary and urgent in understanding spatiotemporal distributions of viruses, exploring the controlling factors of viral processes, qualitatively and quantitatively assessing virus-host interactions and viral functioning in marine ecosystems, and even improving predictions of large-scale marine ecosystem and Earth system models. Previous two studies (Bar-On and Milo, 2019; Wigington et al., 2016) summarized viral abundance data in the ocean and estimated viral biomass as well as virus-to-prokaryote ratio. However, the lacking of host parameters such as bacterial production, and oceanographic parameters such as temperature, salinity, nutrient concentrations, limits the usage of these datasets in broader oceanographic contexts. More importantly, there is no public database of viral activity in the global ocean, which substantially hinders our understanding of the ecological and

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020


Open Access Earth System Science Data Discussions

biogeochemical role of virioplankton on global scale. In addition, the ecological function of viruses in ecosystem is tightly linked to their life strategies, mainly including the lytic and the lysogenic infection (Wommack and Colwell, 2000). The significance of viruses in oceanic biogeochemistry is mainly reflected through the lytic infection, which results in cell lysis and the release of DOM. In contrast, those temperate viruses choosing the lysogenic infection can influence microbial diversity and metabolism by transferring new genes to their hosts, altering the expression of host genes, not killing hosts for many generations until an environmental or cellular trigger causes them to enter the lytic cycle and hence serving as molecular "time bomb" (Paul, 2008). Therefore, it is necessary to include the quantity and quality data of viral life strategy in a viral oceanographic database.

In this study, we construct the first global viral oceanography database, namely gVOD, by collecting data of viral abundance (VA), lytic and lysogenic viral production (VP), as well as other related viral, host and oceanographic metadata when available. Based on the database, we estimate the total viral number and biomass in the global ocean. In addition, the data of VA and VP generated with different techniques were compared to provide references for evaluating the possible technical bias.

2 Data and methods

2.1 Database summary

In the gVOD, direct measurements of three core parameters (VA, lytic and lysogenic VP), as well as accessary viral, prokaryotic and oceanographic parameters when available, were collected from published papers or acquired from lead authors or principal investigators (Table 1). Sampling information including date, latitude, longitude, depth and methods was included for each data record. We used ocean depth shallower or deeper than 200 m as a criterion to identify coastal or open ocean samples. The open ocean samples were further separated into surface and deep samples that collected in 0–200 m and >200 m, respectively.

The quality-controlled database consists of 10,931 VA data points (Table S1), 608 lytic VP data points and 119 lysogenic VP data points (Table S2). Most of VA (99.2%) and lytic VP (98.4%) and all lysogenic VP data have accompanying data of prokaryotic abundance (Table 1). For some samples, the abundances of flagellate, picoeukaryotes, *Synechococcus* and *Prochlorococcus* are also available. Prokaryotic productivity measurements cover 22.1% of VA, 57.7% of lytic VP and 76.5% of lysogenic VP data. The most available environmental parameters are salinity and temperature, providing oceanographic information for about half of VA, two-thirds of lytic VP and nearly all lysogenic VP data. Oxygen and chlorophyll *a* concentration data are also adequate particularly for VA. The concentrations of different types of nutrient, including nitrate, silicate and phosphate, are available for many samples. Other environmental parameters (pH, light intensity, dissolved organic carbon concentration) are relatively scarce. Moreover, given that the frequency of viral infected cells was usually calculated, independently or together with VP, to quantify the impact of viral infection within the microbial community (Chen et al., 2019; Payet and Suttle, 2013; Weinbauer et al., 2003), the reported frequencies of lytic infection (n = 438) and lysogenic infection (n = 266) in the literature were also collected into the database to facilitate the future exploration of marine viral activities.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020


Open Access Earth System Science Data Discussions

Lastly, we collected 83 viral decay rate data, 206 viral burst size data and 111 virus-mediated mortality data, which can be useful for certain studies.

2.2 Viral abundance

The measurements of viral abundance in this database are based on three methods described below. In the first method, viruses were harvested by ultracentrifuging onto copper grids and stained with uranyl acetate, and then enumerated using transmission election microscopy (TEM) (Akaike, 1974). In the second method, viruses were collected onto 0.02-μm filters and stained with a nucleic acid-specific fluorescent dye (e.g., SYBR Green I), and then were counted under an epifluorescence microscope (EFM) (Noble and Fuhrman, 1998). The third method counted viruses by using flow cytometer (FCM), before which viruses were stained with fluorescent dye (e.g., SYBR Green I and SYBR Gold), and identified on the basis of the green fluorescence vs. side scatter signal (Brussaard, 2004; Marie et al., 1999). The details of these three approaches have been described elsewhere (Weinbauer, 2004).

2.3 Lytic viral activity

Lytic VP is paramount and widely employed to assess the activity of lytic viruses in community-level and the roles of viruses in marine ecosystems. In this database, the lytic VP was estimated by one of following five methods. The first method estimated VP by calculating expected viral release rates by multiplying fraction of viral infected cells (mainly prokaryotes), prokaryotic productivity (assuming equal prokaryotic mortality rate) and burst size obtained from TEM studies (Noble and Steward, 2001) or virus-dilution approach (Weinbauer et al., 2002). For simplifying reason, in this paper we label this method as FPB to represent the three variables listed above and used in the estimation. In the second method, called as radioactive incorporation approach (RIA), lytic VP was estimated by determining viral DNA synthesis rates using labeled radiotracer (e.g., ^{3}H-, ^{32}P-, or ^{14}C-labeled thymidine or leucine) and a conversion factor to quantify the incorporated radiotracer into viral particles (Noble and Steward, 2001; Steward et al., 1992). The third method was estimated by viral decay rates (VDR), assuming that the abundance of virus particles is in steady state, then the loss rate of virus particles should be balanced by the production rate (Heldal and Bratbak, 1991). The fourth method used fluorescently labeled viral tracers (FLVT) to measure the dilution rates from the decay of labeled viruses and net changes of the non-labeled viruses in natural viral community (Noble and Fuhrman, 2000). The fifth method quantified the increase of viral abundance during time course incubation using a virus dilution or virus reduction approach (VRA) (Weinbauer et al., 2010; Winget et al., 2005), which effectively avoided new viral infection by reducing viral abundances using pore-size filter or tangential flow filtration system.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020


Open Access Earth System Science Data Discussions

2.4 Lysogenic viral activity

Lysogenic VP is generally measured to detect the proviruses (temperate viruses) that choose lysogenic infection in the environment. Lysogenic VP in this database was estimated using VRA described above after the provirus induction by Mitomycin C (Weinbauer et al., 2002). Hence, provirus induction was defined as the difference in viral abundance between Mitomycin C treated and control samples.

2.5 Quality control

VA, lytic and lysogenic VP in the ocean can vary greatly. We conducted quality control for the database. A negative value lytic VP (Wells and Deming, 2006) was removed. All zero-value (below detection limit) data were kept in the database but were not included in the following analyses. For positive-value data, we applied the Chauvenet's criterion to their log-transformed values to identify outliers (Glover et al., 2011): A datum was treated as an outlier when its probability of deviating from the observed mean was less than 1/(2n), where n was the number of data samples. Outliers were marked in the database and not included in the following analyses.

2.6 Total number and biomass of viruses in global ocean

Based on the VA data of our database, we estimate the total number of viruses in the global ocean using two methods. In the first method, we separately estimated the total number of viruses in different sea areas, including the North and South Atlantic, North and South Pacific, Indian Oceans, Mediterranean and Baltic Sea and Arctic Ocean. For each ocean or sea, virus numbers were calculated for coastal and open oceans separately. The mean VA in several depth bins (coastal oceans: bins separated at 5, 10, 25, 50 and 100 m; open oceans: bins separated at 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500, 1000 and 2000 m), were multiplied by water volume of each bin to calculate total number of viruses. The total number of viruses in global ocean was then calculated by summing up the estimates of the coastal and open ocean regions of the three oceans. Due to the insufficient data, mean VA in deep waters of Arctic Ocean, Mediterranean Sea and Baltic Sea were substituted by the value of North Atlantic.

The second method used the random forest (MATLAB machine learning toolbox) (Breiman, 2001) to construct a model of VA based on sampling latitude, longitude, months and depths. VA data were binned to $1° \times 1°$ with 44 vertical layers and the mean VA of each bin, if data available, was fed into the random forest. When implementing the random forest, 75% randomly selected data were used for training the model while the rest data are used for model validation. The trained model was then used to predict VA for each bin and then to estimate the total viral number in the global ocean.

The viral biomass of the global ocean was calculated from the virus numbers using a conversion factor of 0.023 fg C per viral particle, which was based on a relationship between carbon contents in heads of marine viruses (C_{head}) and their sizes (Jover et al., 2014):

$$C_{head} = 41(r - 2.5)^3 + 130(7.5r^2 - 18.75r + 15.63), \quad (1)$$

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020


Open Access Earth System Science Data Discussions

where r is radium of viral head for which an average of 26.3 nm from the Tara Ocean expedition was used (Brum et al., 2015).

In this paper, all the uncertainties reported in parentheses after means are standard deviations, except that the standard errors of the mean are reported for the estimates of total viral particles and biomass for the global ocean, because the mean values are used in the estimates and therefore the uncertainties of the means were most interested.

3 Results and discussion

3.1 Data distribution

Most VA data were collected in the north hemisphere (particularly in tropical and subtropical regions), while less data in the southern hemisphere (Figs. 1a−1c). In total, nearly two-thirds of the VA data were sampled in the North Atlantic Ocean (32%) and North Pacific Ocean (29%) (Fig. 2a). In addition, 6 long-term time-series observations of VA were included in the compilation (Figs. 1a & 1b): Bermuda Atlantic Time-series Study (BATS) in 2000–2009, San Pedro Ocean Time Series (SPOTS) Microbial Observatory in 2000–2011, the Bedford Basin Monitor (BBM) in 1996–2000, the Rivers Inlet (RI) in 2008–2010, the Saanich Inlet (SI) in 2010–2012 and 2014–2015, and the Guanabara Bay (GB) in 2011–2014. Weekly VA were measured at BBM, approximately monthly samples were collected at BATS, SPOTS, SI, and GB year round, and monthly samples were collected at RI only in spring and summer. Vertically, most VA data were sampled in surface ocean (≤ 200 m, 71%) while less data in deep ocean (> 200 m, 29%), particularly depth below 1,000 m (Fig. 3a). Summer VA samples were most abundant while the fewest data in winter (Fig. 4a).

Lytic VP data in the north hemisphere are much more than in south hemisphere (Figs. 1d−1f), with almost half of lytic VP data were sampled in the North Pacific Ocean (31%) and North Atlantic Ocean (18%) (Fig. 2b). A majority of lytic VP data (86%) was collected in the surface ocean (Fig. 3b). Most deep data were sampled in North Atlantic Ocean and western and northeastern Pacific Ocean (Fig. 1e). There were seasonal biases in lytic VP data, most of which were sampled in summer while rarely sampled in autumn (Fig. 4b). More lytic VP data were sampled in open oceans (63%) than in coastal waters (37%) (Fig. 5). Almost every lytic and lysogenic VP data accompanied with VA measurements.

There are very limited number of lysogenic VP data in both surface and deep oceans (Figs. 1g & 1h), while those deep samples were even much less than the already scarce surface samples (Fig. 3c). Lysogenic VP data in the northern hemisphere are slightly more than the southern hemisphere (Fig. 1i), with most lysogenic VP data sampled in the North Pacific Ocean (29%), North Atlantic Ocean (24%) and South Atlantic Ocean (23%) (Fig. 2c). Similar to lytic VP data, lysogenic VP data were tended to be collected in spring and summer than in other seasons particularly winter (Fig. 4c). Lysogenic VP data in the open ocean (77%) were also much more than those in coastal waters (23%) (Fig. 5).

In summary, most viral data were sampled in North Atlantic and Northeast Pacific Oceans (Figs. 1 & 2), and more data in surface than in the deep oceans (Fig. 3). Viral data also tended to be sampled in summer (Fig. 4). Although the total coastal samples were less than the open ocean samples (Fig. 5), considering the far less proportion of coastal oceans than open oceans, viral samples were more concentrated in coastal areas.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020


Open Access Earth System Science Data Discussions

3.2 Viral Abundance in global ocean

In surface oceans, VA (n=7,768) mostly varies in the order of 10^6 to 10^8 particles ml^{-1}, with mean VA in coastal waters [3.61(±6.3) × 10^7 particles ml^{-1}] about 5 times higher than that in surface open oceans [7.3(±12.4) × 10^6 particles ml^{-1}] (Figs. 6a & 6b). VA in coastal South Atlantic Ocean and Mediterranean and Baltic Seas is higher than that in other coastal oceans (Fig. 6a). Although VA across different surface open oceans distributes in similar ranges, the average VA in Pacific (particularly the southern portion) is higher than other basins (Fig. 6b), a pattern previously found in another study (Lara et al., 2017) using less data than this study. VA decreased with depth, with those in the global deep ocean [1.26(±2.44) × 10^6 particles ml^{-1}, n=3,164] about one order of magnitude lower than those in the surface (Figs. 7a & 7b). The vertical profiles in different open ocean basins more clearly show that VA in Pacific is higher than Atlantic in surface 1,000 m, while the difference does not exist in deeper oceans (Fig. 7b).

Previous studies have showed that the VA counted using FCM, which become more popular in studies after 2014 (Table S1), has strong correlation with those using EFM (Brussaard et al., 2010; Marie et al., 1999; Payet and Suttle, 2008). In our database, most VA samples were measured using FCM (7,353, 67.26%) and EFM (3,465, 31.71%), while only 112 (1.03%) VA samples were counted using TEM. VA counted by the three methods distribute in similar ranges and do not show systematic difference, except for deep open ocean samples in which VA using TEM are substantially lower than those using the other two methods (Fig. 8). But considering much fewer VA data points using TEM than others (Fig. 8 & Table S1), we cannot conclude TEM substantially underestimates VA in deep water samples. Herein, our database provides references for methodological comparison on a global scale.

The total number of global ocean viruses estimated by binning the VA data (Figs. 7a & 7b) is 1.4(±0.02) × 10^{30} particles (mean ±s.e.), which is very close to the estimate of 1.39(±0.03) × 10^{30} particles (mean ±s.e.) using the random forest method (Fig. S1). Both values are consistent to the previous estimates of 10^{30} (Suttle, 2007), 1.29 × 10^{30} (Cobian Guemes et al., 2016) and 1.5 × 10^{30} (Bar-On and Milo, 2019) viral particles for the global ocean. Using a conversion factor of 0.023 fg C per viral particle (see Methods), these two values of total viral number give estimates of total viral biomass in the global ocean at 32.3 ± 0.05 and 32.2 ± 0.06 Tg C, respectively, confirming a recent estimate of 30 Tg C (Bar-On and Milo, 2019).

3.3 Viral production

In surface ocean, lytic VP (n=523) varies greatly from 10^3 to 10^7 particles ml^{-1} h^{-1} in different ocean basins (Figs. 6c & 6d). The overall mean and standard deviation of lytic VP in the global ocean were 9.87(±24.16) × 10^5 particles ml^{-1} h^{-1}. Lytic VP in surface open Pacific Ocean is about one order of magnitude higher than that in surface open Atlantic Ocean (Fig. 6d), which is a consistent pattern as VA (Fig. 6d). Lytic VP in surface Arctic Ocean is much lower than other basins, which is expected considering its much lower biological productivity (Figs. 6c & 6d). Although insufficient lytic VP data (n=82) are available for meaningful statistical analyses in deep waters (Fig. 9), the existing data do show a general trend that VP decreases at one

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

order of magnitude from surface to deep open oceans (Fig. 7c). Unlike VA, average lytic VP in coastal water is close to that in the surface open ocean (Fig. 6c).

Several studies have tried to compare different methodological approaches to estimate the lytic VP rate, revealing that the VRA method is more reliable and less laborious, compared to the probable overestimation by FLVT approach and the potential underestimation by RIA method, though such comparisons were mainly constrained to coastal ocean (Helton et al., 2005; Karuza et al., 2010; Rastelli et al., 2016; Winget et al., 2005). Indeed, most of the lytic VP (84.4%) in this database was estimated by VRA, suggesting that VRA was widely utilized in literature and increasingly became a standard method to estimate VP across different marine environments. Additionally, although a meaningful comparison of reported lytic VP values between disparate marine ecosystems is complicated by the inherent variability among approaches, the lytic VP rates in this database might provide a rough global-scale insight into methodological comparison. The result shows that in both the surface and deep open oceans, the VRA method generally obtains much higher rates than other methods, whereas such difference between methods is not obvious for coastal samples (Fig. 9). However, due to limited number of samples using the methods other than VRA (Fig. 9 and Table S2), we do not have adequate data to tell if the difference in VA is caused by the measurement methods, or the randomness of the samples.

The lysogenic VP data are too few (surface n=85, deep ocean n= 34) to conduct comparison across different ocean basins or between surface and deep oceans, although the results are plotted for readers' reference (Figs. 6e & 6f). The overall lysogenic VP in the global ocean is $2.53(\pm 8.64) \times 10^5$ particles $ml^{-1}\,h^{-1}$, which is about one third of the level of lytic VP, although more data are certainly needed to better compare two different types of VP.

4 Data availability

The gVOD database (Xie et al., 2020) can be downloaded from PANGAEA and released under a temporary link (https://www.pangaea.de/tok/19f9d7b496a00f57f491e639440708aee00b6a49).

5 Conclusion

We constructed a global ocean viral database (gVOD) by compiling 10,931 VA data, 608 lytic VP data and 119 lysogenic VP data. This database may be useful for global-scale studies of viral processes and their roles in marine ecosystems and biogeochemical cycles. The VA, lytic and lysogenic VP data are significantly variable. Most VA were counted using flow cytometers and epifluorescence microscopes, while the virus reduction approach was the most popular method in estimating VP. The lytic VP is about 3 times higher than the lysogenic VP. The calculation using the database also confirms the previous estimates of viral numbers and biomass.

Our database shows that the current investigations have the limitation in spatiotemporal coverage. VA dataset has a poor coverage in South Pacific and Indian Ocean. Lytic VP dataset does not have good coverage in South Pacific, Northwest Pacific,

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Indian Ocean and South Atlantic. Lysogenic VP data are very few across the whole global ocean. Vertically, all viral data were sampled much less in mesopelagic and deep ocean than in the surface oceans. Thus, measurements of viral parameters in these regions and depths should be given high priority. In addition, more viral data should be sampled in winter to avoid seasonal biases of virus-related properties.

The database is stored in a public data repository (PANGAEA), and will be updated regularly when new data become available. We hope that the database will be valuable for field and modeling studies in marine ecology, biogeochemistry and other areas of oceanography.

Author contributions.

RZ and YWL conceived and designed structure of database and mathematical analyses of the data. LX, WW, LC, XC and YH collected the data and described the metadata. LX, NJ, RZ and YWL conducted quality control and analyses of the data. LX, RZ and YWL led the writing of the paper, with contribution from all the co-authors.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgements.

We would like to thank all the scientists and crews who collected the historical data. This work was funded by National Natural Science Foundation of China (91951209, 41890802, 41476093).

Reference

Akaike, H.: A new look at the statistical model identification, IEEE transactions on automatic control, 19, 716–723, https://doi.org/10.1007/978-1-4612-1694-0_16, 1974.

Bar-On, Y. M. and Milo, R.: The biomass composition of the oceans: a blueprint of our blue planet, Cell, 179, 1451–1454, https://doi.org/10.1016/j.cell.2019.11.018, 2019.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/a:1010933404324, 2001.

Breitbart, M.: Marine viruses: truth or dare, Ann Rev Mar Sci, 4, 425–448, https://doi.org/10.1146/annurev-marine-120709-142805, 2012.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J. M., Gorsky, G., Gregory, A. C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B. T., Schwenck, S. M., Speich, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Bork, P., Bowler, C., Sunagawa, S., Wincker, P., Karsenti, E., and Sullivan, M. B.: Patterns and ecological drivers of ocean viral communities, Science, 348, 1261498, https://doi.org/10.1126/science.1261498, 2015.

Brussaard, C. P. D.: Optimization of procedures for counting viruses by flow cytometry, Appl Environ Microbiol, 70, 1506−1513, https://doi.org/10.1128/AEM.70.3.1506-1513.2004, 2004.

Brussaard, C. P. D., Payet, J. P., Winter, C., and Weinbauer, M. G.: Quantification of aquatic viruses by flow cytometry. In: Manual of aquatic viral ecology, https://doi.org/10.4319/mave.2010.978-0-9845591-0-7.102, 2010.

Chen, X., Wei, W., Wang, J., Li, H., Sun, J., Ma, R., Jiao, N., and Zhang, R.: Tide driven microbial dynamics through virus-host interactions in the estuarine ecosystem, Water Res, 160, 118−129, https://doi.org/10.1016/j.watres.2019.05.051, 2019.

Cobian Guemes, A. G., Youle, M., Cantu, V. A., Felts, B., Nulton, J., and Rohwer, F.: Viruses as winners in the game of life, Annu Rev Virol, 3, 197−214, https://doi.org/10.1146/annurev-virology-100114-054952, 2016.

Fuhrman, J. A.: Marine viruses and their biogeochemical and ecological effects, Nature, 399, 541−548, https://doi.org/10.1038/21119, 1999.

Glover, D. M., Jenkins, W. J., and Doney, S. C.: Modeling methods for marine science, Cambridge University Press, https://doi.org/10.1017/CBO9780511975721, 2011.

Heldal, M. and Bratbak, G.: Production and decay of viruses in aquatic environments, Mar Ecol Prog Ser, 72, 205−212, https://doi.org/10.3354/meps072205, 1991.

Helton, R. R., Cottrell, M. T., Kirchman, D. L., and Wommack, K. E.: Evaluation of incubation-based methods for estimating virioplankton production in estuaries, Aquat Microb Ecol, 41, 209−219, https://doi.org/10.3354/ame041209, 2005.

Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W., and Weitz, J. S.: The elemental composition of virus particles: implications for marine biogeochemical cycles, Nat Rev Microbiol, 12, 519−528, https://doi.org/10.1038/nrmicro3289, 2014.

Karuza, A., Del Negro, P., Crevatin, E., and Fonda Umani, S.: Viral production in the Gulf of Trieste (Northern Adriatic Sea): Preliminary results using different methodological approaches, J Exp Mar Biol Ecol, 383, 96−104, https://doi.org/10.1016/j.jembe.2009.12.003, 2010.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Lara, E., Vaqué, D., Sà, E. L., Boras, J. A., Gomes, A., Borrull, E., Díez-Vives, C., Teira, E., Pernice, M. C., Garcia, F. C., Forn, I., Castillo, Y. M., Peiró, A., Salazar, G., Morán, X. A. G., Massana, R., Catalá, T. S., Luna, G. M., Agustí, S., Estrada, M., Gasol, J. M., and Duarte, C. M.: Unveiling the role and life strategies of viruses from the surface to the dark ocean, Sci Adv, 3, e1602565, https://doi.org/10.1126/sciadv.1602565, 2017.

Marie, D., Brussaard, C. P. D., Thyrhaug, R., Bratbak, G., and Vaulot, D.: Enumeration of marine viruses in culture and natural samples by flow cytometry, Appl Environ Microbiol, 65, 45–52, https://doi.org/10.1128/AEM.65.1.45-52.1999, 1999.

Noble, R. T. and Fuhrman, J. A.: Rapid virus production and removal as measured with fluorescently labeled viruses as tracers., Appl Environ Microbiol, 66, 3790–3797, https://doi.org/10.1128/Aem.66.9.3790-3797.2000, 2000.

Noble, R. T. and Fuhrman, J. A.: Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria, Aquat Microb Ecol, 14, 113–118, https://doi.org/10.3354/ame014113, 1998.

Noble, R. T. and Steward, G. F.: Estimating viral proliferation in aquatic samples, Method Microbiol, 30, 67–84, https://doi.org/10.1016/S0580-9517(01)30040-5, 2001.

Paul, J. H.: Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas?, ISME J, 2, 579–589, https://doi.org/10.1038/ismej.2008.35, 2008.

Payet, J. P. and Suttle, C. A.: Physical and biological correlates of virus dynamics in the southern Beaufort Sea and Amundsen Gulf, J Marine Syst, 74, 933–945, https://doi.org/10.1016/j.jmarsys.2007.11.002, 2008.

Payet, J. P. and 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, https://doi.org/10.4319/lo.2013.58.2.0465, 2013.

Rastelli, E., Dell'Anno, A., Corinaldesi, C., Middelboe, M., Noble, R. T., and Danovaro, R.: Quantification of viral and prokaryotic production rates in benthic ecosystems: A methods comparison, Front Microbiol, 7, 1501, https://doi.org/10.3389/fmicb.2016.01501, 2016.

Steward, G. F., Wikner, J., Smith, D. C., Cochlan, W. P., and Azam, F.: Estimation of virus production in the sea. I: Method development, Marine Microbial Food Webs 6, 57–78, https://doi.org/10.1007/BF00614106, 1992.

Suttle, C. A.: Marine viruses – major players in the global ecosystem, Nat Rev Microbiol, 5, 801–812, https://doi.org/10.1038/nrmicro1750, 2007.

Suttle, C. A.: Viruses in the sea, Nature, 437, 356–361, https://doi.org/10.1038/nature04160, 2005.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Weinbauer, M., Rowe, J., and Wilhelm, S.: Determining rates of virus production in aquatic systems by the virus reduction approach. In: Manual of Aquatic Viral Ecology, https://doi.org/10.4319/mave.2010.978-0-9845591-0-7.1, 2010.

Weinbauer, M. G.: Ecology of prokaryotic viruses, FEMS Microbiol Rev, 28, 127–181, https://doi.org/10.1016/j.femsre.2003.08.001, 2004.

Weinbauer, M. G., Brettar, I., and Hofle, M. G.: Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters, Limnol Oceanogr, 48, 1457–1465, https://doi.org/10.4319/lo.2003.48.4.1457, 2003.

Weinbauer, M. G., Winter, C., and Höfle, M. G.: Reconsidering transmission electron microscopy based estimates of viral infection of bacterio-plankton using conversion factors derived from natural communities, Aquat Microb Ecol, 27, 103–110, https://doi.org/10.3354/ame027103, 2002.

Wells, L. E. and Deming, J. W.: Significance of bacterivory and viral lysis in bottom waters of Franklin Bay, Canadian Arctic, during winter, Aquat Microb Ecol, 43, 209–221, https://doi.org/10.3354/ame043209, 2006.

Wigington, C. H., Sonderegger, D., Brussaard, C. P. D., Buchan, A., Finke, J. F., Fuhrman, J. A., Lennon, J. T., Middelboe, M., Suttle, C. A., Stock, C., Wilson, W. H., Wommack, K. E., Wilhelm, S. W., and Weitz, J. S.: Re-examination of the relationship between marine virus and microbial cell abundances, Nat Microbiol, 1, 15024, https://doi.org/10.1038/nmicrobiol.2015.24, 2016.

Wilhelm, S. W. and Suttle, C. A.: Viruses and nutrient cycles in the sea, BioScience, 49, 781–788, https://doi.org/10.2307/1313569, 1999.

Winget, D. M., Williamson, K. E., Helton, R. R., and Wommack, K. E.: Tangential flow diafiltration: an improved technique for estimation of virioplankton production, Aquat Microb Ecol, 41, 221–232, https://doi.org/10.3354/ame041221, 2005.

Wommack, K. E. and Colwell, R. R.: Virioplankton viruses in aquatic ecosystems, Microb Mol Biol Rev, 64, 64–114, https://doi.org/10.1128/mmbr.64.1.69-114.2000, 2000.

Xie, L., Wei, W., Cai, L., Chen, X., Huang, Y., Zhang, R., and Luo, Y.-W.: A global viral oceanography database (gVOD) from 1987 to 2018, PANGAEA, https://www.pangaea.de/tok/19f9d7b496a00f57f491e639440708aee00b6a49, 2020.

Zhang, R., Weinbauer, M. G., and Qian, P. Y.: Viruses and flagellates sustain apparent richness and reduce biomass accumulation of bacterioplankton in coastal marine waters, Environ Microbiol, 9, 3008–3018, https://doi.org/10.1111/j.1462-2920.2007.01410.x, 2007.

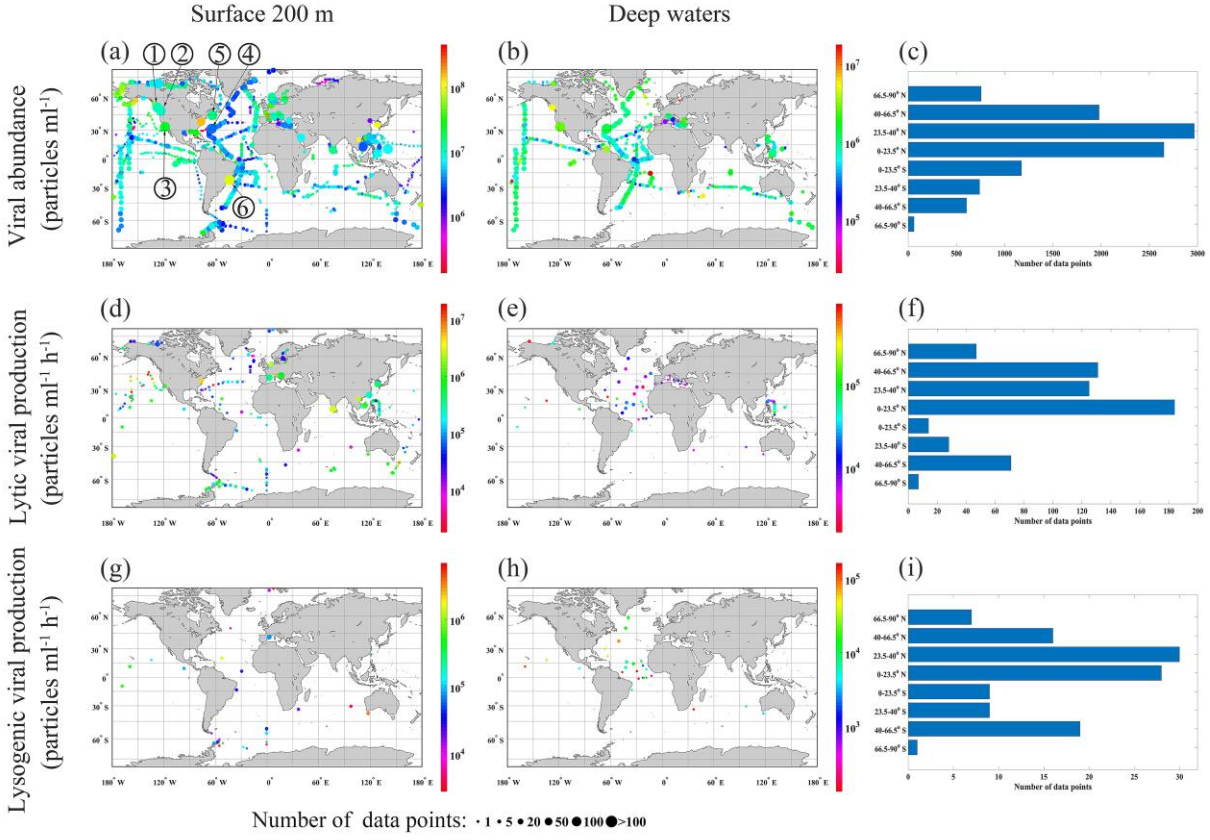
https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

Open Access Earth System Science Data Discussions

Table 1. Number of accessory viral, host and oceanographic parameters associated with each of the core viral parameters (VA, lytic and lysogenic VP).

	VA (*n*=10,931)	Lytic VP (*n*=608)	Lysogenic VP (*n*=119)
Accessory viral parameters			
Frequency of lytic infection	405	142	36
Frequency of lysogenic infection	227	96	36
Viral decay rate	83	65	27
Burst size	206	96	-
Virus-mediated bacterial mortality	46	46	19
% cells lysed	55	53	17
Accessory host parameters			
Prokaryotic abundance	10,846	598	119
Prokaryotic productivity	2,425	352	91
Flagellate abundance	411	44	7
Picoeukaryotic abundance	1,554	68	15
Synechococcus abundance	1,700	80	42
Prochlorococcus abundance	1,567	73	42
Accessory oceanographic parameters			
Chlorophyll *a*	3,949	244	71
Temperature	6,253	399	119
Salinity	6,360	370	85
Oxygen	4,930	82	46
Nitrate	2,707	91	46
Phosphate	3,153	144	51
Silicate	2,638	106	36
pH	96	47	-
Light intensity	35	35	-
Dissolved organic carbon	97	13	-

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020



330 **Figure 1. Collected viral abundance (a, b), lytic viral production (d, f) and lysogenic viral production (g, h) in surface (≤200 m) (a, d and g) and in deep (>200 m) waters (b, e and h), binned on 1 °×1 °grids. Color of each grid codes the mean value of the parameters, and the size of the circles represents number of samples in each bin. Also shown the number of samples in latitudinal bands (c, f and i). Numbers in (a) represent long-term time-series observations of viral abundance: 1, Rivers Inlet; 2: Saanich Inlet; 3: San Pedro Ocean Time Series Microbial Observatory; 4: Bedford Basin Monitor; 5: Bermuda Atlantic Time-series Study**
335 **and 6: Guanabara Bay.**

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

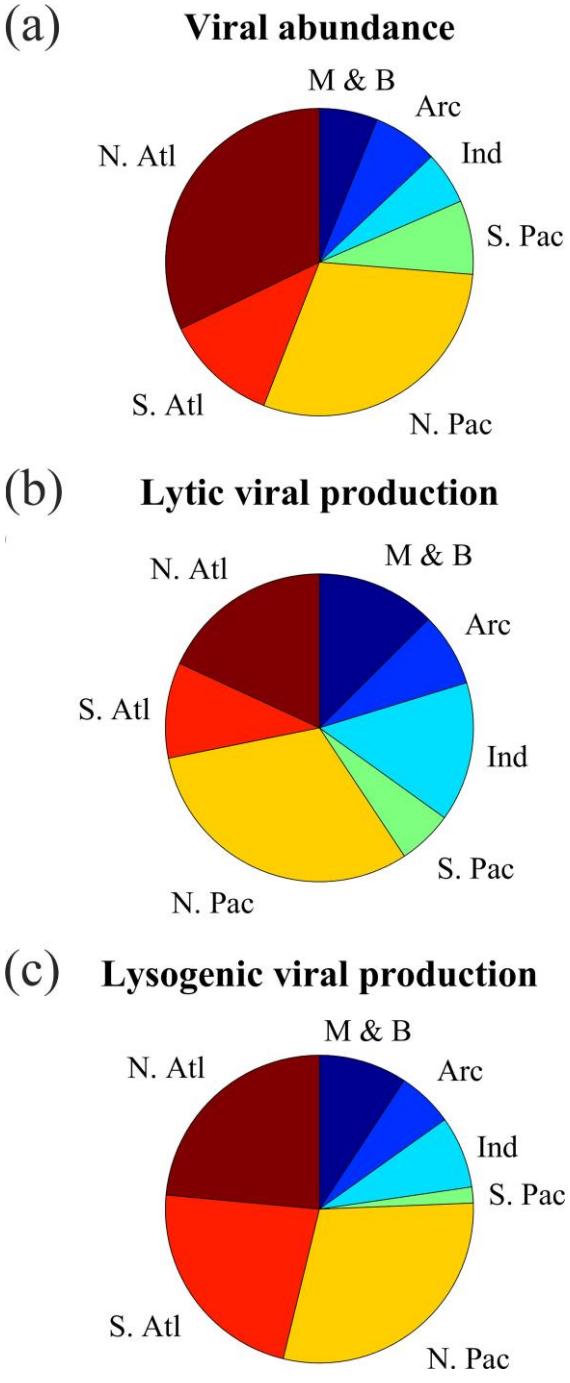


Figure 2. The fraction of (a) viral abundance, (b) lytic viral production and (c) lysogenic viral production data points in different oceans (N. Atl: North Atlantic, S. Atl: South Atlantic, N. Pac: North Pacific, S. Pac: South Pacific, Ind: Indian ocean, Arc: Arctic ocean, M & B: Mediterranean Sea and Baltic Sea).

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

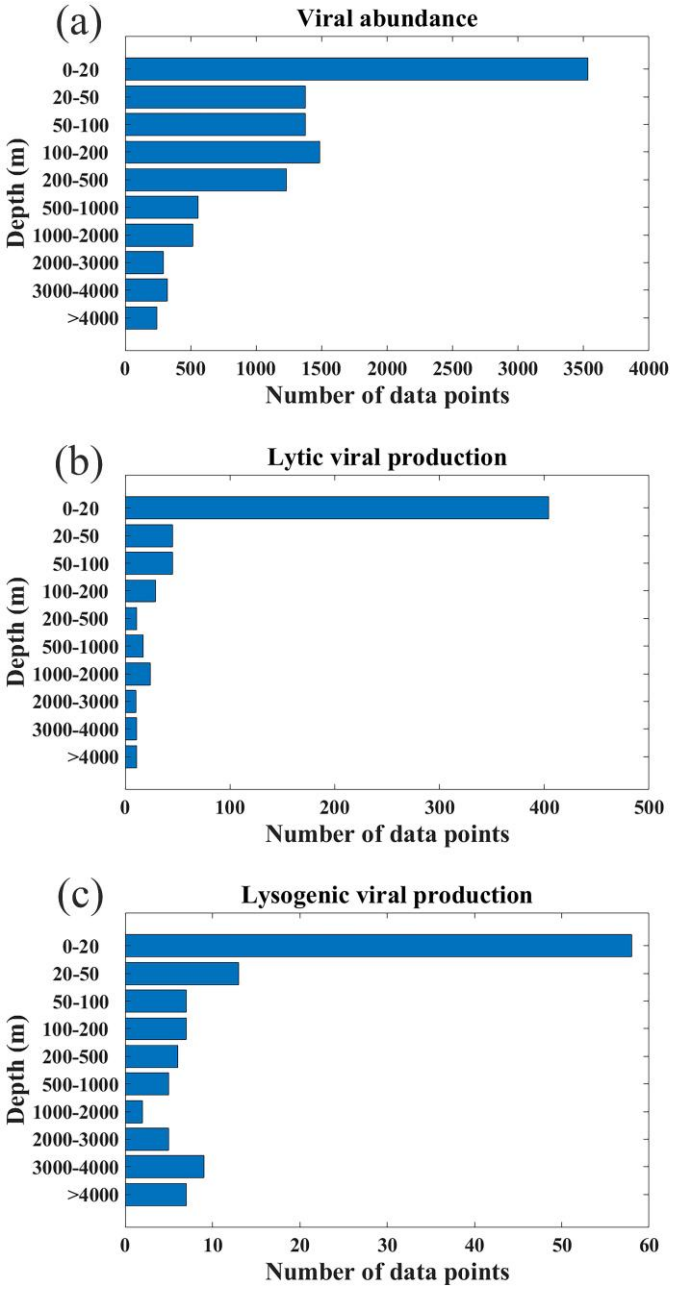


Figure 3. Depth distribution of the data for (a) viral abundance, (b) lytic viral production, and (c) lysogenic viral production.

345

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

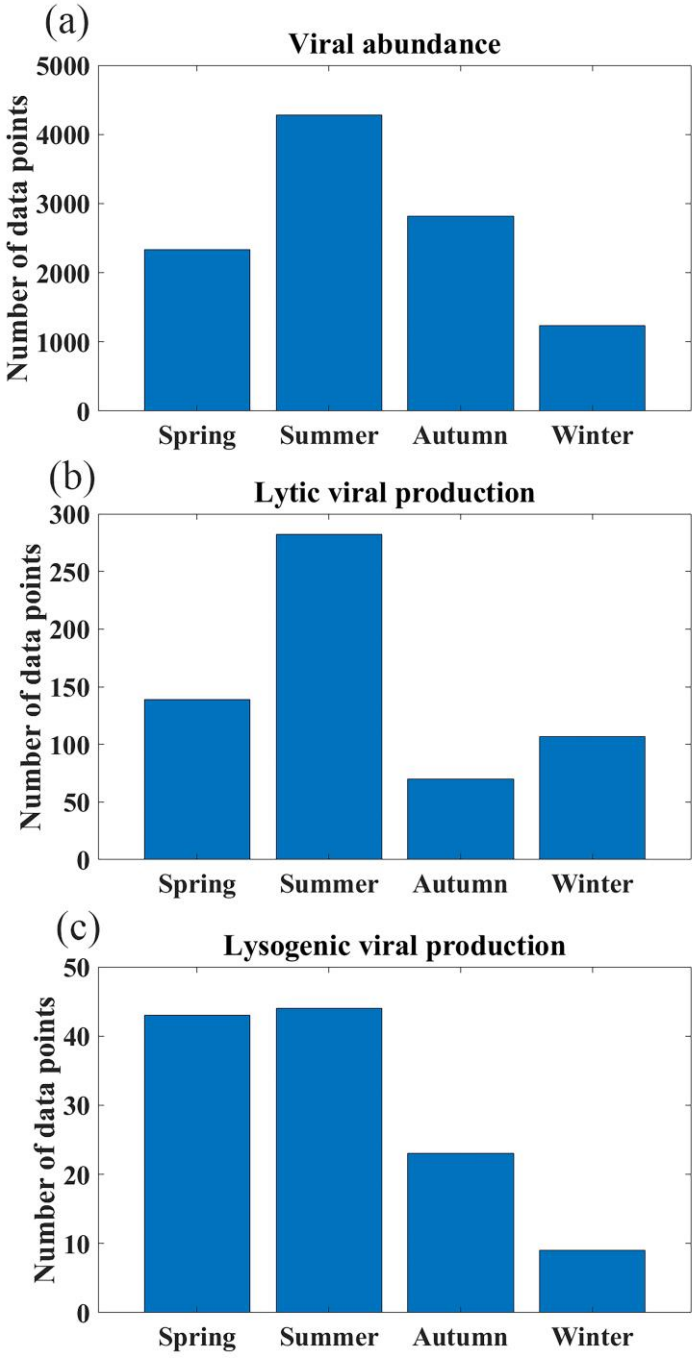
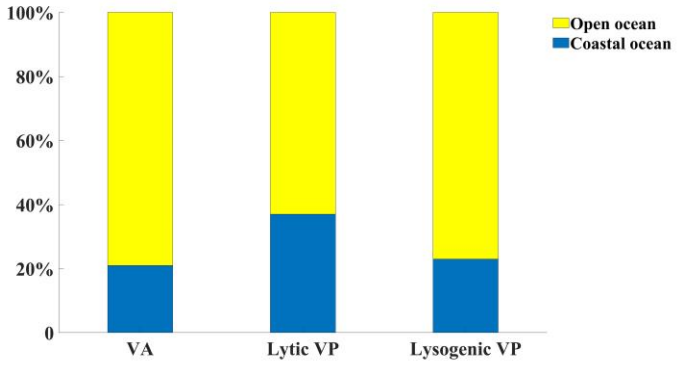


Figure 4. Seasonal distributions of number of samples for (a) viral abundance, (b) lytic viral production and (c) lysogenic viral production.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020



350 **Figure 5. The fraction of viral abundance (VA), lytic viral production (VP) and lysogenic VP data points in coastal versus in open oceans.**

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

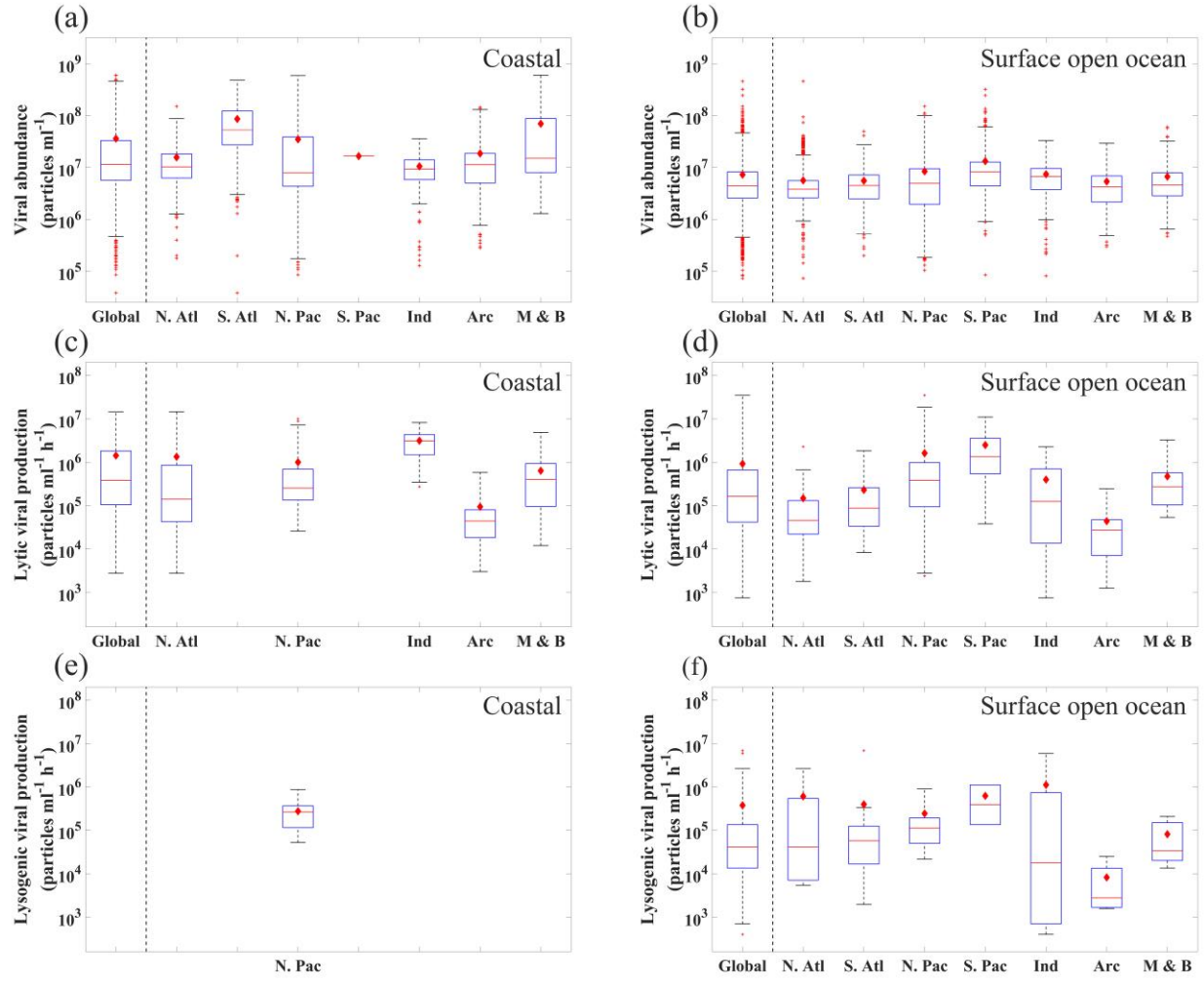
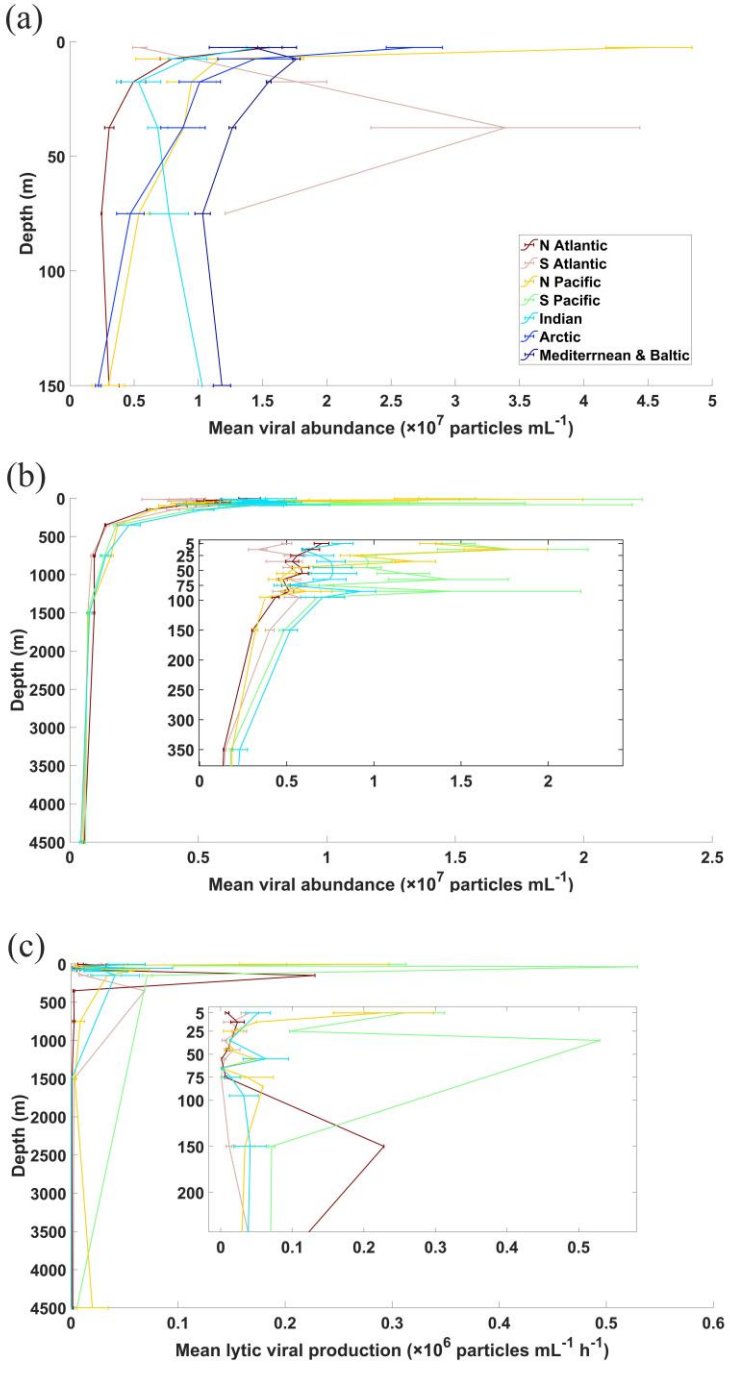


Figure 6. The range of (a, d) viral abundance, (b, e) lytic viral production, and (c, f) lysogenic viral production in the different oceans (N. Atl: North Atlantic, S. Atl: South Atlantic, N. Pac: North Pacific, S. Pac: South Pacific, Ind: Indian ocean, Arc: Arctic ocean, M & B: Mediterranean Sea and Baltic Sea), grouping in coastal ocean (a–c) and open ocean samples in surface 200 m (d–f). All data are shown in logarithmic scales. The red diamonds mark the mean value. The central red lines indicate the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles of the data. Error bars extend to the 5th and 95th percentiles and the remaining outliers are marked with red plus signs.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020

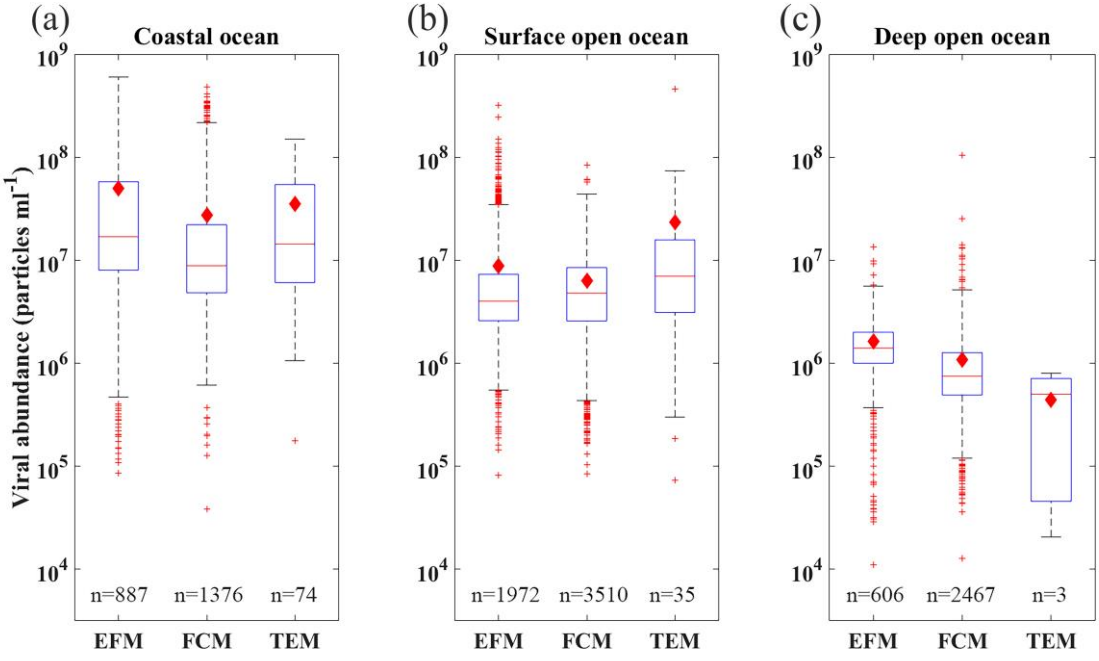
Open Access Earth System Science Data Discussions



365

Figure 7. Vertical profiles of average viral abundance of each ocean basin in (a) coastal and (b) open ocean waters. Also shown the vertical profiles for open water lytic viral production (c), while those for coastal samples were not constructed because of limited data points. Error bars represent one standard error of the mean.

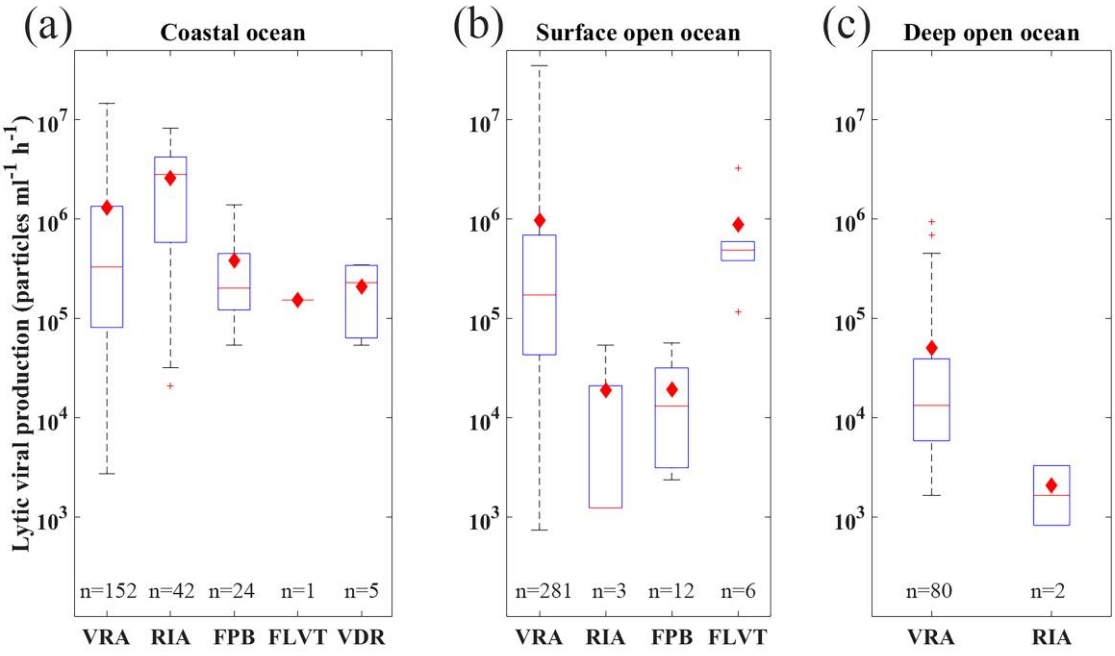
https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020



370

Figure 8. Box plots of distribution of viral abundance using different measurement methods (EFM, FCM and TEM, see text for more details) in coastal (a), surface open ocean (b) and deep open ocean (c) waters. See caption of Figure 6 for details of lines and symbols of the box plots.

https://doi.org/10.5194/essd-2020-120
Preprint. Discussion started: 30 July 2020



375

Figure 9. Comparison of lytic viral production using different measurement methods including VRA, RIA, FPB, FLVT and VDR (FPB: calculated by multiplying fraction of viral infected cells, prokaryotic production and burst size; RIA: radioactive incorporation approach; FLVT: fluorescently labelled viral tracers method; VRA: virus reduction approach; VDR: estimated by viral decay rates. See text for details) in coastal (a), surface open ocean (b) and deep open ocean (c) waters. See caption of Figure 6
380 **for details of lines and symbols of the box plots.**