



TITLE:

# Prevalence of Viral Frequency-Dependent Infection in Coastal Marine Prokaryotes Revealed Using Monthly Time Series Virome Analysis

AUTHOR(S):

Tominaga, Kento; Ogawa-Haruki, Nana; Nishimura, Yosuke; Watai, Hiroyasu; Yamamoto, Keigo; Ogata, Hiroyuki; Yoshida, Takashi

---

CITATION:

Tominaga, Kento ...[et al]. Prevalence of Viral Frequency-Dependent Infection in Coastal Marine Prokaryotes Revealed Using Monthly Time Series Virome Analysis. *mSystems* 2023, 8(1): e00931-22.

ISSUE DATE:

2023-02-23

URL:








<http://hdl.handle.net/2433/279505>

RIGHT:

© 2023 Tominaga et al.; This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



# Prevalence of Viral Frequency-Dependent Infection in Coastal Marine Prokaryotes Revealed Using Monthly Time Series Virome Analysis

 Kento Tominaga,<sup>a,b</sup>  Nana Ogawa-Haruki,<sup>a</sup>  Yosuke Nishimura,<sup>c</sup>  Hiroyasu Watai,<sup>a</sup>  Keigo Yamamoto,<sup>d</sup>  Hiroyuki Ogata,<sup>e</sup>  
 Takashi Yoshida<sup>a</sup>

<sup>a</sup>Graduate School of Agriculture, Kyoto University, Kyoto, Japan

<sup>b</sup>Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

<sup>c</sup>Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa, Japan

<sup>d</sup>Research Institute of Environment, Agriculture and Fisheries, Osaka Prefecture, Osaka, Japan

<sup>e</sup>Institute for Chemical Research, Kyoto University, Kyoto, Japan

**ABSTRACT** Viruses infecting marine prokaryotes have a large impact on the diversity and dynamics of their hosts. Model systems suggest that viral infection is frequency dependent and constrained by the virus-host encounter rate. However, it is unclear whether frequency-dependent infection is pervasive among the abundant prokaryotic populations with different temporal dynamics. To address this question, we performed a comparison of prokaryotic and viral communities using 16S rRNA amplicon and virome sequencing based on samples collected monthly for 2 years at a Japanese coastal site, Osaka Bay. Concurrent seasonal shifts observed in prokaryotic and viral community dynamics indicated that the abundance of viruses correlated with that of their predicted host phyla (or classes). Cooccurrence network analysis between abundant prokaryotes and viruses revealed 6,423 cooccurring pairs, suggesting a tight coupling of host and viral abundances and their “one-to-many” correspondence. Although stable dominant species, such as SAR11, showed few cooccurring viruses, a fast succession of their viruses suggests that viruses infecting these populations changed continuously. Our results suggest that frequency-dependent viral infection prevails in coastal marine prokaryotes regardless of host taxa and temporal dynamics.

**IMPORTANCE** There is little room for doubt that viral infection is prevalent among abundant marine prokaryotes regardless of their taxa or growth strategy. However, comprehensive evaluations of viral infections in natural prokaryotic communities are still technically difficult. In this study, we examined viral infection in abundant prokaryotes by monitoring the monthly dynamics of prokaryotic and viral communities at a eutrophic coastal site, Osaka Bay. We compared the community dynamics of viruses with those of their putative hosts based on genome-based *in silico* host prediction. We observed frequent cooccurrence among the predicted virus-host pairs, suggesting that viral infection is prevalent in abundant prokaryotes regardless of their taxa or temporal dynamics. This likely indicates that frequent lysis of the abundant prokaryotes via viral infection has a considerable contribution to the biogeochemical cycling and maintenance of prokaryotic community diversity.

**KEYWORDS** frequency-dependent selection, *K/r* strategy, marine prokaryotes, marine viruses, virome

Marine prokaryotes are ubiquitous in the ocean and play key roles in global biogeochemical processes (1). Most of the observed species (>35,000 species-level operational taxonomic units [OTUs] based on 97% 16S rRNA sequence identity) fall into several

**Editor** Michael S. Rappe, University of Hawaii at Manoa

**Copyright** © 2023 Tominaga et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Takashi Yoshida, [yoshida.takashi.7a@kyoto-u.ac.jp](mailto:yoshida.takashi.7a@kyoto-u.ac.jp), or Hiroyuki Ogata, [ogata@kuicr.kyoto-u.ac.jp](mailto:ogata@kuicr.kyoto-u.ac.jp).

The authors declare no conflict of interest.

**Received** 26 September 2022

**Accepted** 29 December 2022

**Published** 1 February 2023

major taxa (phyla or classes for *Proteobacteria*), such as *Alphaproteobacteria* (e.g., SAR11), *Bacteroidetes* (e.g., *Flavobacteriaceae*), and *Cyanobacteria* (e.g., *Synechococcus* and *Prochlorococcus*) (2, 3). Although individual species have distinct ecological niches, they are often classified into one of two growth strategists based on their potential growth rate and temporal dynamics: (i) *K*-strategists (slow growing and persistently dominant; e.g., SAR11) and (ii) *r*-strategists (fast growing and opportunistic; e.g., *Flavobacteriaceae*) (4). However, recent high-frequency sampling schemes (e.g., daily) uncovered that species not recognized as *r*-strategists exhibit drastic fluctuations (e.g., Marine Group II [MGI] *Euryarchaeota*) (5, 6). Further, finely resolved populations (genotypes or strains) within a species-level OTU often show distinct temporal dynamics (7–11), indicating that species described as *K*-strategists can show frequent fluctuation.

Viruses infecting prokaryotes are abundantly present in the ocean and are estimated to lyse 20 to 40% of the prokaryotic cells each day (4, 12, 13). Viruses are thought to infect their specific hosts (often restricted to strains within a species) in a frequency-dependent manner, in which the encounter rate between the viruses and their hosts is a determinant for the infection rate (14, 15). Thus, viruses infect host populations that become abundant, and frequencies of host and viruses oscillate over time, leading to maintenance of the diversity of the host community (16, 17). Moreover, mathematical models have demonstrated that a prokaryotic species with a higher growth rate can be more susceptible to viral infection (17). This trend allows *K*-strategists to reach a higher abundance than *r*-strategists because of their higher resistance against viral infection by cryptic escape through reduced cell size and/or specialized defense mechanisms (4, 18). However, the discovery of SAR11 viruses questions this prediction (19). Given the dominance of SAR11 viruses in the ocean (19–21), there is little room for doubt that *K*-strategists are also targeted by viruses. Therefore, comprehensive surveys to examine the prevalence of viral infections in abundant prokaryotes with different growth strategies are still required.

Previous monthly observations of microbial communities revealed that seasonal oceanographic features have a strong influence on the prokaryotic community (22, 23). Seasonal variability of the viral community has also been reported using PCR-based analysis (24, 25) and viral metagenomics (viromics) (26–28). Although viruses are obligate parasites, viral seasonality was often discussed independently from the seasonality of their hosts except for a few well-cultivated prokaryote-virus pairs (e.g., *Synechococcus/Prochlorococcus* and their viruses) (25, 29). Otherwise, the interactions are described solely based on the dynamics of individual viruses and prokaryotes because of the difficulty in connecting uncultured viruses and their hosts (13, 30).

In this study, we aimed to solve the fundamental question of whether viral infection is prevalent among abundant prokaryotic populations or the way viruses infect differs depending on the taxa and/or growth strategies of their hosts. For this purpose, we monitored prokaryotic and viral communities at Osaka Bay, a eutrophic coastal site with input of nutrients from rivers that is affected by an oligotrophic warm current Kuroshio (31), for 2 years at a monthly time interval. In the sampling site, monitoring of microbial communities (32) and characterization of full-length double-stranded DNA (dsDNA) viral genomes assembled from virome sequences (33) and their diurnal transcriptomic dynamics (34, 35) have been reported. However, cooccurrence dynamics alone have been suggested to be a poor predictor of interactions in the context of viral infection of microbial hosts (36). To overcome the limitations of correlation-based inference of virus-host pairs, we compare the community dynamics of viruses with those of their putative hosts using genome-based *in silico* host prediction analysis (29, 30). The prevalence of viral infection is discussed with cooccurrence dynamics among the potential virus-host pairs, which fulfilled the genome-based prediction.

## RESULTS AND DISCUSSION

**Overview of prokaryotic and viral communities in Osaka Bay.** We collected more than 1 year of time series (total of 18 months) seawater samples during the daytime (3 h before or after high tide) of each month. We obtained 2.8 million paired-end

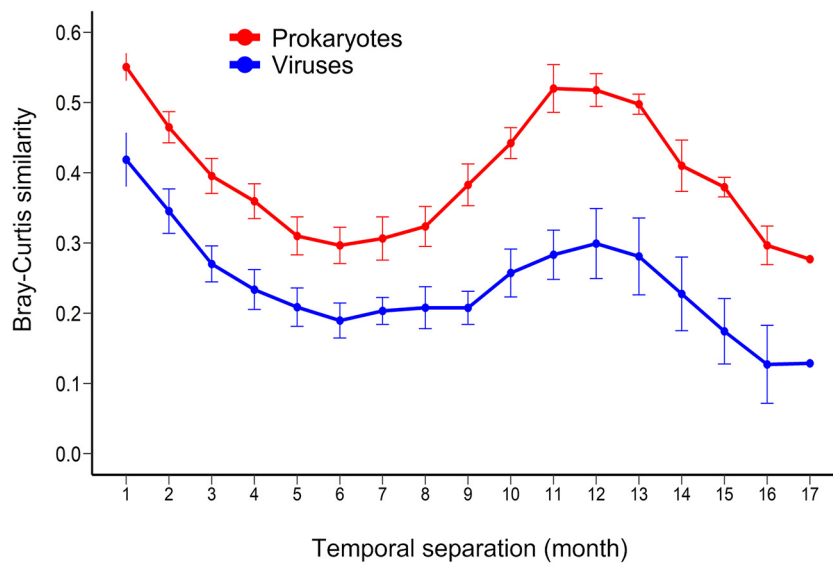
reads (24,168 to 846,565 reads per sample) from the 16S rRNA gene V3-V4 region amplicon sequencing libraries, and these sequences were clustered into 35,191 OTUs (1,462 to 18,268 OTUs per month, median of 3,274) with a sequence identity threshold of 99% (species-level populations are presented in Table S1 in the supplemental material). The prokaryotic community was dominated by *Alphaproteobacteria* (41%), *Gammaproteobacteria* (21%), *Bacteroidetes* (19%), and *Cyanobacteria* (7%) at the phylum level (class level for *Proteobacteria*).

To explore viral community composition, we obtained 60 million paired-end reads of viromes (929,884 to 8,124,354 sequences per sample), which were generated from the virus-size fraction of 17 samples that were concomitantly collected with the prokaryotic-size fractions (Table S1). After decontamination of prokaryotic sequences and dereplication, 5,226 virus-like large contigs (>10 kb), including 202 circularly assembled viral genomes, were obtained (Table S1). Here, we call the 5,226 contigs assembled in this study monthly time series Osaka Bay viral (mts-OBV) contigs. We refer to these contigs (or genomes) operationally as species-level viral populations, according to the previous proposal in viral ecology (37). The majority (~75%) of mts-OBV contigs showed high genomic similarity (genomic similarity score ( $S_G$ ) of >0.15; see reference 38 for the definition of  $S_G$ ) with one of the previously reported viral complete genomes (33) or the 202 circular genomes assembled in this study. Based on the  $S_G$ , these mts-OBV contigs were classified into 314 genus-level taxonomic groups (Table S1).

On average, 40% of virome reads (29 to 53% per sample) were mapped on the mts-OBV contigs and previously reported viral genomes (33). According to the fragments per kilobase per million (FPKM) values calculated from the read counts, mts-OBV contigs (assembled in this study) occupied, on average, 96% of the relative abundance of analyzed viral genomes, suggesting mts-OBVs well represent the viral community observed in the samples. Other relatively abundant genomes included genomes assembled in a previous study at Osaka Bay (e.g., OBV\_N00129 and OBV\_N00081) (33) and *Pelagibacter* virus HTVC010P (19). Further, we confirmed whether the mts-OBV contigs can cover the trend of the whole viral community by the following procedure according to the previous study (34). The abundances of mts-OBV and other smaller contigs (1 to 10 kb) assembled in this study were compared based on the abundances of all detected terminase large subunit genes (*terL*). mts-OBV contigs covered a wide range of abundance, suggesting that they can capture the trend of the whole viral community. In addition, all mts-OBV contigs ranked at the top (>30%) of the community in at least one sample (Fig. S1), indicating that most abundant viruses were captured within mts-OBVs. Thus, in the following sections, only the 5,226 mts-OBV contigs (including 202 complete genomes) and 4,240 previously reported complete viral genomes (33) were considered viral community assessment for read mapping.

Alpha-diversity (Shannon index) of the viral community was higher than that of the prokaryotic community (Fig. S2A and B). Both richness and evenness were also higher in the viral community than in the prokaryotic community (Fig. S2C to F). It should be noted that prokaryotic diversity was evaluated via single marker gene analysis (i.e., 16S rRNA), but viral diversity was evaluated via whole-genome sequencing. Thus, the methodological difference could have caused the relatively higher diversity of the viral community, such as double counting of partial contigs derived from a viral genome. Another possible explanation for the higher viral diversity is that a prokaryotic species can be infected by more than one viral species at each time point (discussed below).

**Seasonal dynamics of prokaryotic and viral communities.** We investigated seasonal dynamics of prokaryotic and viral communities using the Bray-Curtis similarity index between all pairwise combinations of samples (136 pairs, 1- to 17-month intervals), following the previous time series observations in different locations (22, 25, 26, 39, 40). Both prokaryotic and viral communities showed clear seasonal patterns, with a peak average similarity at an interval of about 12 months, representing the same seasons, and the bottom of average similarity at an interval of 6 months, representing opposite seasons (Fig. 1). Similar seasonal peaks have been observed in previous studies,



**FIG 1** Seasonality of the prokaryotes and viruses at Osaka Bay (OB) during observation. The Bray-Curtis community similarity index was calculated among all of the possible sample pairs from normalized abundances of prokaryotic OTUs and OBV contigs and was plotted as a function of the number of months separating their sampling.

such as the San-Pedro Ocean Time (SPOT) series conducted in San Pedro Channel (22, 25, 26, 39–41). However, the lower community similarity observed between opposite seasons in both prokaryotic and viral communities at Osaka Bay than observed at SPOT using a similar methodology (26, 41) suggests more pronounced seasonality at Osaka Bay. Prokaryotic community dynamics were concordant with seasonal environmental variables, such as water temperature and inorganic nutrients, which increased in summer (June to September) presumably because of the increasing river inflow during the rainy season (Table 1; Table S1). The community variation is generally larger (low similarity) between the autumn and spring samples than between the summer and winter samples in both communities (Fig. S3). The similarity between samples was systematically lower for the viral community than for the prokaryotic community (Fig. 1, discussed below). The viral community composition was significantly correlated with the prokaryotic community composition as well as the seasonal environmental variables (Mantel,  $\rho = 0.504$ ,  $P < 0.01$ ; Table 1).

Given that most viruses can only propagate in their specific host and thereby the viral community composition is shaped by the prokaryotic community composition, the

**TABLE 1** Rho values of partial Mantel tests for prokaryotic and viral communities and environmental parameters

	Prokaryotes <sup>a</sup>	Viruses	Environmental combined	Nutrient combined	Temp	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	PO <sub>4</sub>	Salinity
Prokaryotes										
Viruses	0.504 <sup>c</sup>									
Environmental combined	0.308 <sup>c</sup>	0.227 <sup>b</sup>								
Nutrient combined	0.344 <sup>c</sup>	0.288 <sup>c</sup>	0.871 <sup>c</sup>							
Temp	0.505 <sup>c</sup>	0.137	0.37 <sup>c</sup>	0.284 <sup>c</sup>						
NH <sub>4</sub>	-0.097	-0.109	0.328 <sup>c</sup>	0.458 <sup>c</sup>	0.01					
NO <sub>2</sub>	0.037	-0.015	0.379 <sup>c</sup>	0.53 <sup>c</sup>	0.007	0.404 <sup>c</sup>				
NO <sub>3</sub>	0.239 <sup>b</sup>	0.3 <sup>c</sup>	0.206 <sup>b</sup>	0.34 <sup>c</sup>	-0.044	-0.113	-0.06			
PO <sub>4</sub>	0.511 <sup>c</sup>	0.443 <sup>c</sup>	0.487 <sup>c</sup>	0.56 <sup>c</sup>	0.154	0.035	0.108	0.218 <sup>b</sup>		
Salinity	-0.159	-0.037	0.502 <sup>c</sup>	0.191	-0.028	0.078	0.01	-0.16	0.115	
SiO <sub>2</sub>	0.138	0.078	0.565 <sup>c</sup>	0.329 <sup>c</sup>	0.447 <sup>c</sup>	0.028	-0.037	-0.063	0.042	0.459 <sup>c</sup>

<sup>a</sup>The value in each box is the Rho value.

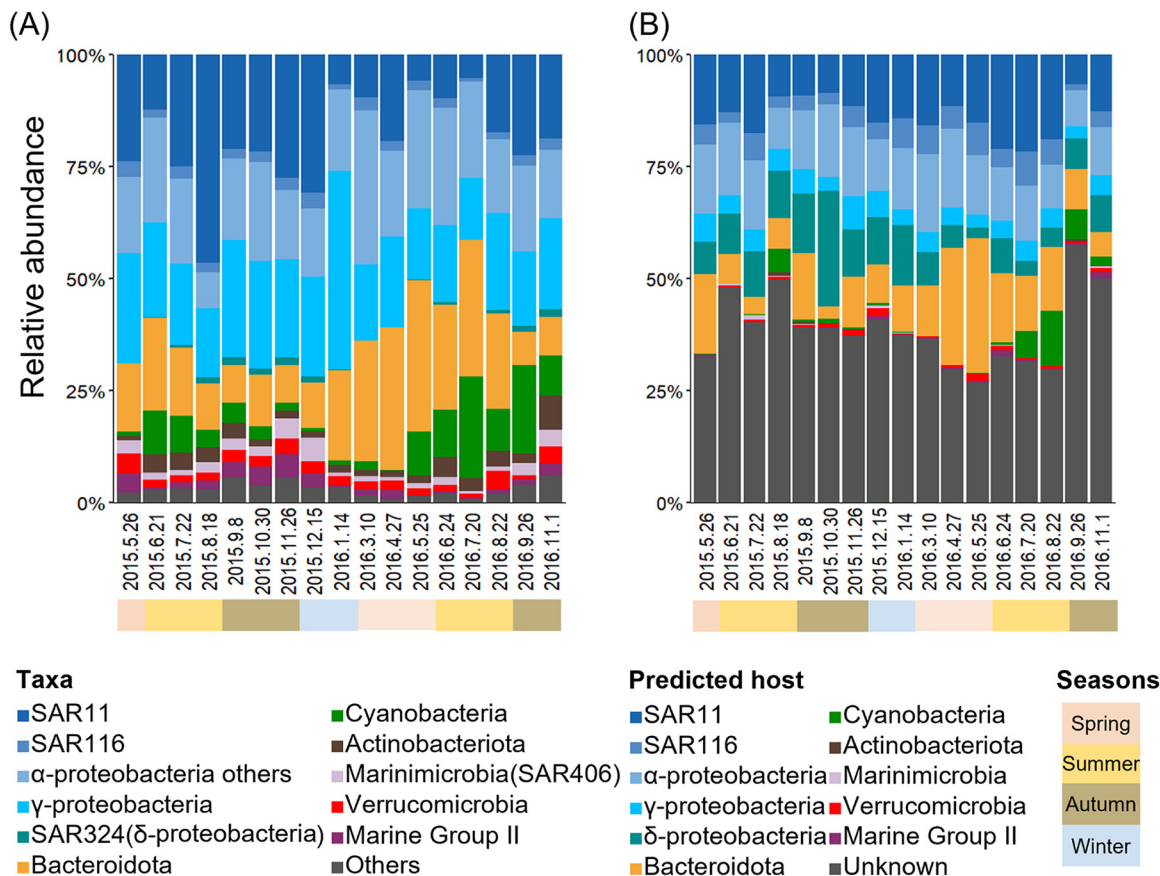
<sup>b</sup> $P < 0.05$ .

<sup>c</sup> $P < 0.01$ .

abundance of each virus might reflect the abundance of its host. To test this hypothesis, compositions of prokaryotic and viral communities were compared using the information of predicted viral hosts (mostly host phylum- or class-level composition). Putative host groups of viruses were predicted using four commonly used genome-based *in silico* prediction methods (similarity with known viruses, CRISPR-spacer match, tRNA match, and genome homology). First, based on the similarity with cultured viruses, putative host groups of 951 mts-OBV contigs (22 genomic OTUs [gOTUs]) were predicted (*Synechococcus/Prochlorococcus*, 182 contigs; SAR11, 501 contigs; SAR116, 214 contigs; *Roseobacter*, 31 contigs; others, 23 contigs; Table S1). Similarly, putative host groups of 504 mts-OBV contigs (39 genera) were predicted based on the genome-wide sequence similarities with uncultured viral genomes according to the previous studies (*Bacteroidetes*, 468 contigs; MGII, 36 contigs [33, 42]; Table S1). For the other 1,460 mts-OBV contigs (*Alphaproteobacteria*, 35 gOTUs, 621 contigs; *Bacteroidetes*, 80 contigs; *Gammaproteobacteria*, 236 contigs; *Deltaproteobacteria*, 326 contigs; others, 53 contigs; Table S1), putative host groups were predicted via the sequence similarity (i.e., CRISPR-spacer matching, tRNA matching, and genome homology) between viral (mts-OBVs with >200,000 previously reported marine viral genomes [26, 33, 38, 43, 44]) and prokaryotic genomic data sets (>8,000 marine prokaryotic metagenome-assembled genomes in previous studies [45–49] and the genomes in the NCBI RefSeq database). Altogether, we assigned potential host groups for 2,844 mts-OBV contigs (*Alphaproteobacteria*, 1,375 contigs; *Bacteroidetes*, 548 contigs; *Deltaproteobacteria*, 326 contigs; *Gammaproteobacteria*, 250 contigs; *Cyanobacteria*, 190 contigs; Table S1).

Major phyla (or classes for *Proteobacteria*) in the prokaryotic community did not change drastically, but the relative abundance of several phyla (classes) exhibited seasonal dynamics (Fig. 2). The seasonal dynamics of the predicted hosts resembled the seasonal dynamics of prokaryotes (Fig. 2). For example, *Cyanobacteria* (79% of reads were assigned to OTU\_8, *Synechococcus*) dominated in summer (up to 9.6% and 22.6% of the community in June 2015 and July 2016, respectively) (Fig. 2), and *Synechococcus* virus abundance also increased in summer (up to 5.3% and 12.1% of the community in August 2015 and August 2016, respectively) (Fig. 2). Similarly, the relative abundance of *Bacteroidetes* increased from winter to spring (up to 33.7% of the community in May 2016) (Fig. 2), and *Bacteroidetes* virus abundance also increased during spring (up to 30.2% of the community in May 2016) (Fig. 2). Relative abundances of both SAR11 (from 5% to 47% of the community) (Fig. 2) and SAR11 viruses (from 9% to 22% of the community) (Fig. 2) showed changes over time, but they were always abundant throughout the observed period. Therefore, the viral community appears to generally follow the dynamics of their host.

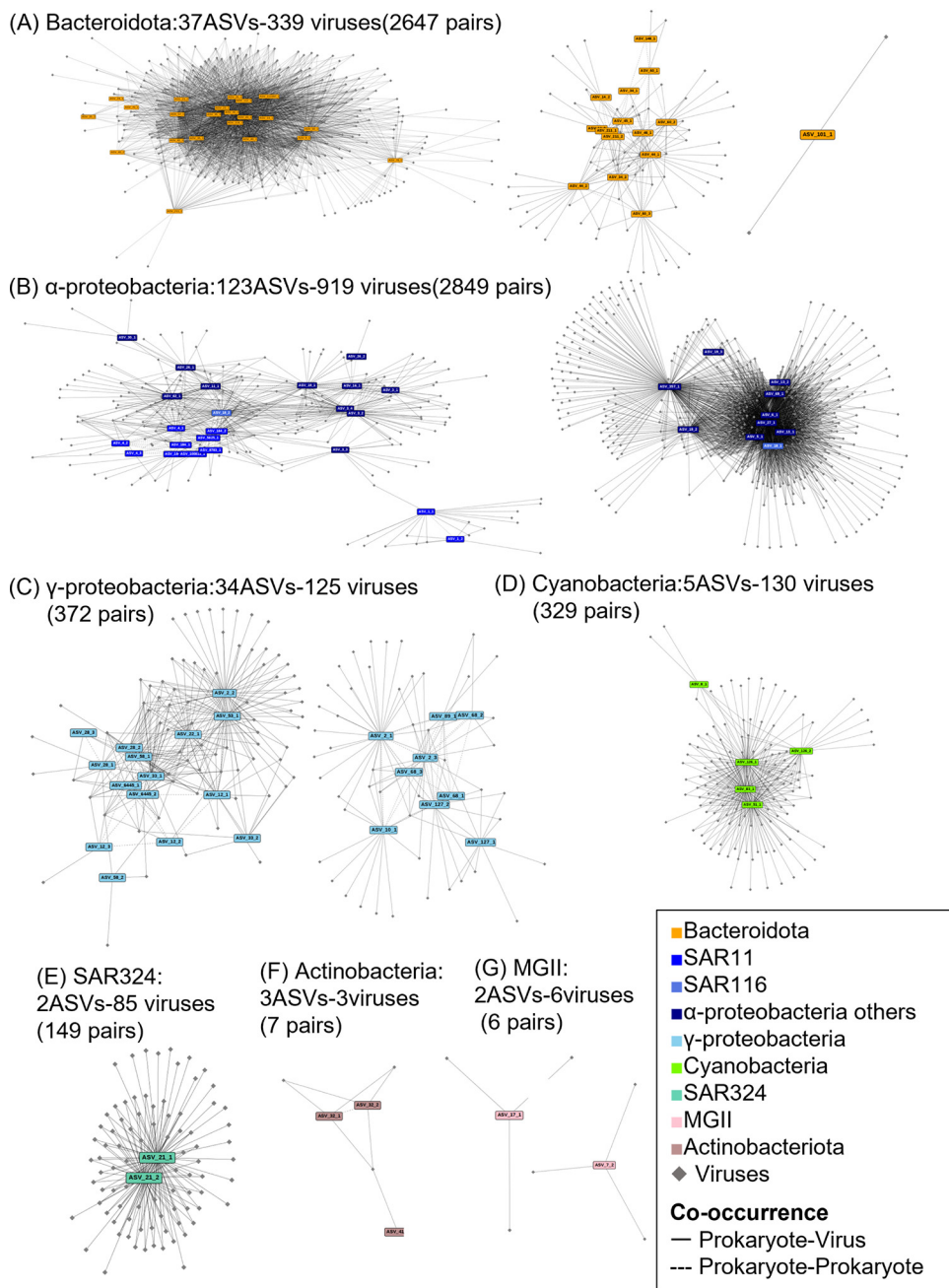
However, viral abundance did not always match with their putative host abundance (Fig. S4). For example, the proportion of putative *Gammaproteobacteria* viruses was lower than that of *Gammaproteobacteria*, and the proportion of putative *Deltaproteobacteria* viruses was much higher than that of *Deltaproteobacteria* (Fig. 2). The lack of a tight correlation between viral and host abundance may not be surprising. Host prediction based on genome analysis in this study was mostly at the phylum or class level except for contigs showing similarity with cultured viruses, such as *Synechococcus/Prochlorococcus* cyanoviruses, while typical prokaryotic viruses could only infect specific host species or strains. Further, although our analysis annotated putative hosts at nearly 60% of the viral community, remaining populations without host prediction may lead to the underestimation of viruses infecting some taxa. The differences in burst sizes among viruses, which have been estimated to range from 6 to 300 in the marine environment (50), can also influence the estimation of viral abundance. Alternatively, as previously observed and discussed in Hevroni et al. (28), differences in viral infection cycles (e.g., consistently low production of virions [51]) or short time delay of peaks of virus and host abundance (52) might cause the mismatch between host abundance and extracellular viral abundance. Next, to investigate whether viral abundance increased according to specific host abundance, we statistically examined associations (i.e., cooccurrence) between the viruses and amplicon sequence variants (ASVs) extracted from the abundant 73 prokaryotic OTUs.



**FIG 2** Comparison of prokaryotic and viral taxonomic community composition based on host prediction. (A) Relative abundance of phylogenetic groups of prokaryotic communities. Quality-controlled reads were clustered into OTUs with a sequence identity of 99% using VSEARCH (73). These OTUs were classified at the phylum level (class level for *Proteobacteria*) using SINA (75). (B) Relative abundance of viruses based on their putative hosts assigned by host prediction. Normalized abundances of viral contigs were calculated from fragments per kilobase per million (FPKM) values and were converted to relative abundance.

### Cooccurrence network analysis between the abundant prokaryotes and viruses.

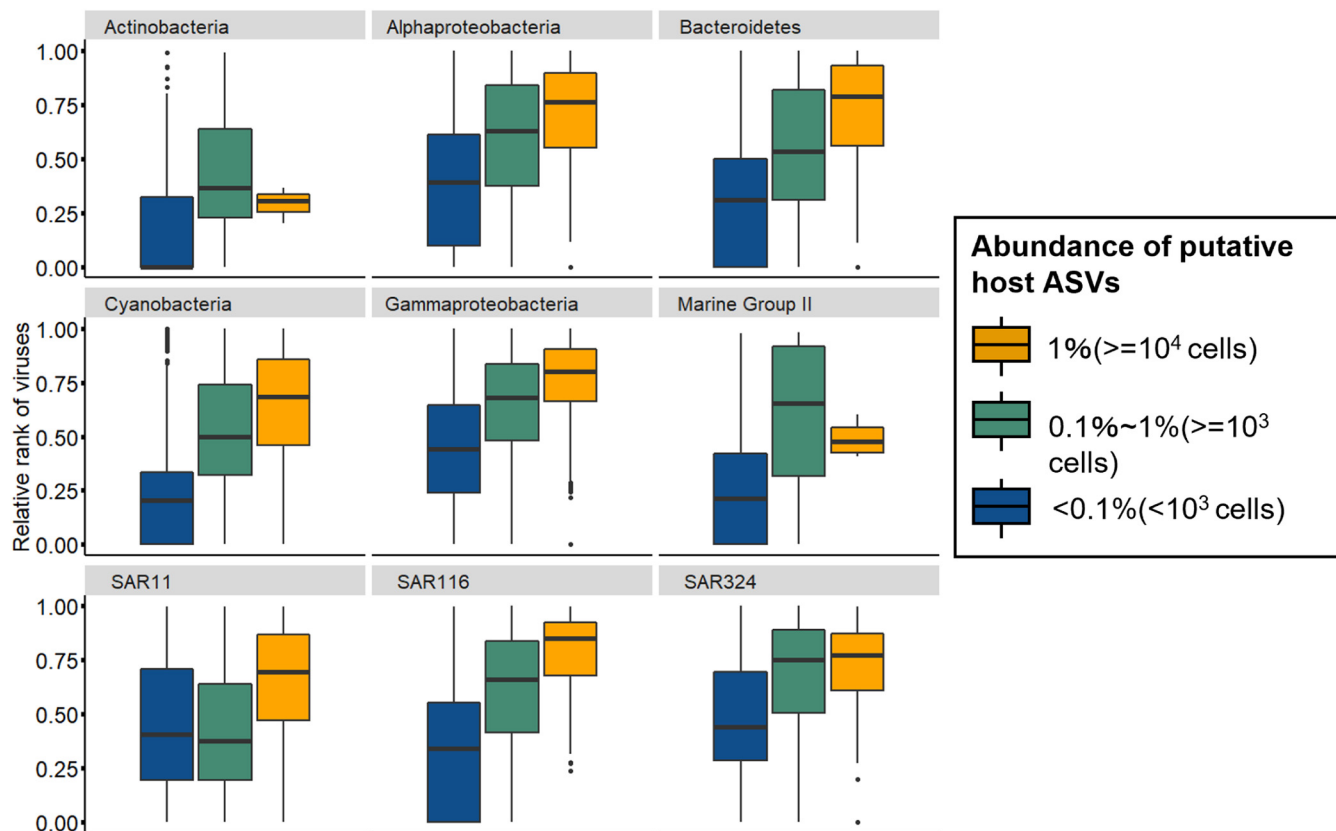
To examine the dynamics of closely related (nearly strain-level) variants within each OTU, 114 ASVs (1~4 ASVs per OTU) (Fig. S5) were extracted from the abundant 74 OTUs via minimum entropy decomposition (MED), according to previous studies (7, 10, 11). Although the majority of ASVs derived from the same abundant OTU showed similar dynamics, several ASVs showed distinct seasonal patterns (e.g., ASVs from SAR86 OTU2 and NS4 marine group *Flavobacteria* OTU14) (Fig. S5). These ASVs likely reflect ecologically meaningful seasonal subpopulations in an OTU that dominates throughout the sampling period. Then, pairwise correlations (cooccurrence network) between the 114 prokaryotic ASVs and the viral species that were predicted to infect the prokaryotic ASVs via host prediction (e.g., 37 *Bacteroidetes* ASVs and 548 mts-OBV contigs predicted as *Bacteroidetes* virus) were determined via Spearman's correlations. In total, 6,423 significant correlations between 104 prokaryotic ASVs and 1,366 viral species were detected (Fig. 3; Fig. S6A). The majority (88.6%) of prokaryotic ASVs correlated with at least one viral species. In contrast, only 34% and 31% of prokaryotic ASVs positively and negatively correlated with environmental variables, respectively (Spearman correlations,  $r > 0.6$ ,  $P < 0.01$ ,  $q < 0.05$ ) (Table S2). The number of cooccurring viral species ranged from 0 (13 ASVs) to 359 (ASV6-1, classified into *Planktomarina*), and the median value was 16. We verified that the correlation is not only because the pattern of viral and host dynamics was similar but also because the virus ranked high in the viral community when the putative host was abundant. We therefore compared the relative rank among the viral community and the host prokaryotic abundance in the detected 6,423



**FIG 3** A broad overview of detected positive correlations between prokaryotic ASVs and viral populations that potentially infect each prokaryotic taxa based on host prediction analysis. (A) *Bacteroidota* and their viruses. (B) *Alphaproteobacteria* and their viruses. (C) *Gammaproteobacteria* and their viruses. (D) *Cyanobacteria* and their viruses. (E to G) Other major groups, such as SAR324 (E), *Actinobacteria* (F), and Marine Group II (G), and their viruses. Prokaryotic nodes are circles, and viral nodes are V shapes. Node color indicates prokaryotic taxa. Solid lines are positive correlations.

putative virus-host pairs. First, four cyanobacterial ASVs and cooccurring 130 cyanovirus species were examined. Because substantial numbers of *Synechococcus/Prochlorococcus*-virus pairs have been reported in culture-based studies (53–56), host prediction for cyanoviruses is likely to be reliable. These cyanoviral species were more dominant in the viral community when their cooccurring ASVs exceeded the predicted minimum host cell density for effective propagation of prokaryotic viruses ( $10^3$  cells/mL [57] or  $10^4$  cells/mL [58]) (Fig. 4; Fig. S6B and C). Thus, cyanobacterial viral species were not abundant or were often undetectable when their putative hosts were less abundant, but they became dominant





**FIG 4** Increase of viral abundance according to host cell density between cooccurring host-virus pairs. The normalized relative rank of each virus in the community (0 to 1) was plotted when their putative host relative abundance exceeded 1% ( $\approx 10^4$  cells/mL, yellow) and 0.1% ( $\approx 10^3$  cells/mL, green) and below 0.1% (blue). Box plots are constructed with the upper and lower lines corresponding to the 25th and 75th percentiles; outliers are displayed as points.

when putative host abundance increased. This viral increase with host abundance was also observed in 98 other prokaryotic ASVs and their cooccurring viral species (Fig. 4; Fig. S6B and C). This result clearly indicates that frequency-dependent viral infection is prevalent in abundant prokaryotes at least between the detected virus-host pairs.

**Characterization of the virus-host interaction by host taxa.** The community of viruses showed a higher alpha-diversity than the community of prokaryotes (Fig. S2), and the cooccurrence analysis indicated one-to-many associations between the host and virus populations (median of 16 viral species per prokaryotic ASV). This suggests that one abundant prokaryotic ASV can interact with multiple viral species. Note that the numbers of cooccurring viral species were overestimated because each contig could be a partial genome fragment derived from the same viral genome (average completeness of mts-OBV contigs was 39%) (Table S1). However, the contigs classified into different genera (average of 8 gOTUs) often cooccurred with an ASV. Next, we characterized the “one-to-many” virus-host interaction network (i.e., how many viruses cooccurred with each ASV) with respect to their host taxa and host temporal dynamics.

The number of cooccurring viral species for prokaryotic ASVs was generally dependent on the predicted number of associated viruses determined via host prediction (Fig. S6D). For example, *Bacteroidetes* viruses (548 viruses) were the second most frequently observed viruses, and an average of 71.5 viruses cooccurred with *Bacteroidetes* ASVs (1 to 208 viruses per ASV, between 37 *Bacteroidetes* ASVs and 339 *Bacteroidetes* viruses). The number of cooccurring viruses could be overestimated because of the double count of cooccurring viruses between two cooccurring ASVs. If ASV-A and ASV-B cooccurred, the viruses cooccurring with ASV-A can also be included in the viruses cooccurring with ASV-B and vice versa. In fact, up to 16 ASV-ASV cooccurring pairs were detected for *Bacteroidetes*. In contrast, the taxa with less frequently detected

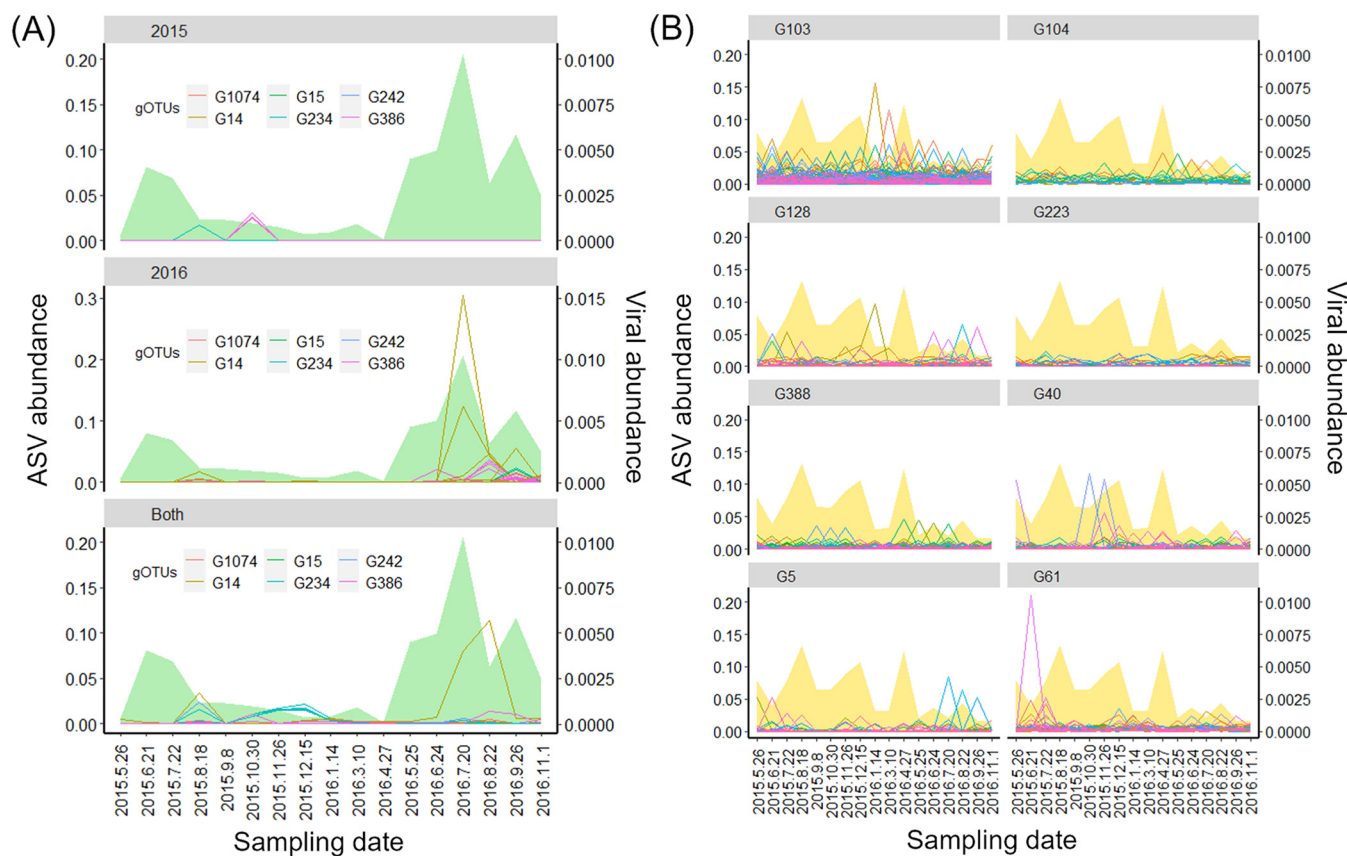
viruses (e.g., MGII, 38 viruses) had a smaller number of cooccurring populations (0 to 3 viruses per ASV) (Fig. S6D). Thus, the number of cooccurring viral species might be underestimated in these taxa because of host prediction limitations. In addition, it should be noted that there are two assumptions for our virus-host cooccurrence analysis: (i) viral host range is narrow as is common in isolated strains (59), and (ii) infectious virions do not persist a long time with low external virion inflow. The latter assumption is following the diel cycling and locality of viral production (34, 60, 61). Thus, it should be noted that exceptions for these assumptions, such as broad-host-range viruses (62), may cause misinterpretation from the cooccurrence analysis. Further studies are required for the quantification of viruses with such exceptional characteristics in the environment.

To confirm the validity of the cooccurrence analysis using the putative virus-host pairs derived from our genome-based host prediction, we compared the number of cooccurring pairs between them and nonhost pairs. In prokaryotic taxa that found many viruses (e.g., *Bacteroidetes* and *Alphaproteobacteria*), more frequent cooccurrences were observed between putative host taxa and their viruses than nonhost-like pairs (Fig. S6E to G). This result suggests that the genome-based host prediction assisted the connection of more plausible pairs. Further, to examine the robustness of our cooccurrence analysis, we converted the relative abundance to the absolute abundance for the samples with cell and virus-like particle (VLP) counts. In this analysis, at least 97 virus-host pairs were confirmed in absolute abundance levels (fig. S6H and I).

**Characterization of the virus-host interaction by host temporal dynamics.** Of note, SAR11 had relatively few cooccurring viral species even though there were more than 500 putative SAR11 viral species (Fig. S6D). SAR11 is often regarded as a *K*-strategist, which is believed to be resistant to viral infection (4), and the growth strategy may influence the cooccurrence dynamics with viruses. Next, we examined the number of cooccurring viruses among prokaryotic ASVs classified in the same taxa depending on the temporal dynamics to solve this issue.

It is difficult to determine the growth strategy of each prokaryotic ASV because maximum growth rate (*r*) and carry capacity (*K*) cannot be directly inferred from their monthly temporal dynamics (63). Thus, instead of assigning strategies to each ASV, we classified the ASVs into two groups according to their temporal dynamics (temporarily abundant or persistently abundant). The temporal dynamics of each prokaryotic ASV was assessed by the indexes that we introduced (see Materials and Methods). According to these, 13 ASVs were the most persistently dominant (i.e., *K*-like index of  $>12$  and *r*-like index of  $<0.1$ ). Among the 13 ASVs, 7 were classified as SAR11 (Fig. S7). Twenty-two of 57 ASVs belonging to the taxa previously predicted as *r*-strategists (i.e., *Flavobacteriaceae*, *Rhodobacteraceae*, *Vibrio*, and Marine Group II) were classified as temporarily abundant ASVs (*K*-like index of  $<3$  and *r*-like index of  $>0.5$ ; total of 33 ASVs) (Fig. S7). Generally, temporarily abundant ASVs, such as members of *Bacteroidetes*, showed many cooccurring viral species (Fig. S7). In contrast, persistently dominant ASVs of *Synechococcus* and SAR11 showed relatively few cooccurring viral species (Fig. S7). The most abundant ASV of *Synechococcus* (ASV8-1, making up 76.7% of all cyanobacterial reads) and SAR11 (ASV1-1, occupying 7 to 64% of all SAR11 reads of each month) showed 7 and 16 cooccurring viruses, respectively, even though 183 cyanoviruses and 500 SAR11 viruses were detected during the observation (Fig. S7).

The cyanoviruses and SAR11 viruses were as abundant as their putative host taxa (e.g., both cyanoviruses and cyanobacteria were dominant during the summer) (Fig. 2). The few cooccurrences of virus and host seem to be not reasonable. However, if a temporal shift of virus-host pairs occurred, cooccurrence analysis may fail to detect virus-host associations. Therefore, we compared the dynamics of the two dominant prokaryotic ASVs and viral species that did not cooccur with their predicted hosts. Representative sequences of ASV8-1 matched with the members of *Synechococcus* subcluster 5.1a with 100% identity. Among the 53 cyanoviral species that did not cooccur with any cyanobacterial ASV, 41 species were classified into two gOTUs (G14 [T7-like cyanosiphovirus] and G386 [T4-like cyanomyovirus]), which are known to infect subcluster 5.1a (e.g., *Synechococcus* sp. WH 8103, clade II), suggesting a plausible interaction between ASV8-1 and these



**FIG 5** Dynamics of the most dominant prokaryotic population (ASV1-1 and ASV8-1) and viruses that were predicted to infect these host taxa by host prediction analysis but did not cooccur with any ASV. (A) Dynamics of ASV8-1 that was classified into *Synechococcus* and 53 cyanoviruses that did not cooccur with cyanobacterial ASVs. The area chart represents the relative abundance of ASV8-1, and the lines represent viral contigs over time. The panels were separated by viral annual pattern (2015 type, 2016 type, and both years type; if the virus was more than five times more abundant in one year than in another year, the virus was defined as a year-specific virus). Colors represent the gOTU (genus) of the virus. (B) Dynamics of ASV1-1 that was classified into the SAR11 clade and 309 putative SAR11 viruses that did not cooccur with any SAR11 ASVs. The area chart represents the relative abundance of ASV1-1, and the lines represent viral contigs over time. The panels were separated based on the classified gOTUs of each virus.

viruses. ASV8-1 especially dominated during the summer (maximum of 8% and 21% of the prokaryotic community in June 2015 and July 2016, respectively) (Fig. 5A). Of these 53 viral species, the abundances of which also increased in summer, 4 were abundant only in 2015 (from 5 to greater than 170 times more abundant in 2015 than in 2016), and another 38 species were more abundant in 2016 (from 5 to greater than 300 times more abundant in 2016 than in 2015) (Fig. 5A). Similarly, ASV1-1 of SAR11 was always abundant (Fig. 5B), and SAR11 viruses occupied a major fraction of the viral community. However, abundant members of SAR11 viruses (309 contigs) were replaced in a relatively short time (a few months) (Fig. 5B). These results suggest that the host-virus interaction might have been underestimated in the cooccurrence analysis, and persistently dominant ASVs can also interact with multiple viruses based on their cell density. Considering a recent observation of infection dynamics of a lysogenic SAR11 virus, which did not show decreased host abundance even as virus abundance increased (51), the continuous dominance of the host with infection likely occurs versus assuming rare hosts produce a large amount of viruses. Although we cannot directly connect each ASV and its growth strategy, the prevalence of interactions between viruses and prokaryotes showing different temporal dynamics suggests that abundant prokaryotes are potentially infected by viruses according to their cell density regardless of their growth strategy.

Finally, we investigated whether the observed viruses, including those not statistically detected as cooccurring viruses with hosts (e.g., 53 cyanoviruses and 309 SAR11 viruses in Fig. 5), were also produced via increased contact frequency with hosts. To infer the contact frequency, we focused on single-nucleotide polymorphisms (SNPs) in viral

genomes. SNPs of closely related viral populations were previously observed in abundant viral populations, such as freshwater cyanoviruses (64) and marine viruses in other coastal areas (26). Because recent studies suggested that the majority of viruses observed in the virome were produced via diel and local viral-host interactions (28, 34, 38, 61), it likely indicates that multiple infection events may lead to the generation of mutations through DNA replication. We thus hypothesized frequent reproduction and mutations for abundant viruses with an increased contact frequency with their hosts. Therefore, SNPs from mts-OBV contigs with more than  $10\times$  coverage depth (2,356 contigs) were calculated. We observed increasing intrapopulation genetic diversity (SNPs quantified by average genomic entropy) as a function of overall viral population abundance regardless of their host taxa (Fig. S8). This suggests that the increase in contact frequency occurred not only in viruses showing clear cooccurrence with hosts (e.g., *Bacteroidetes* viruses) but also in viruses that showed no statistical detection of cooccurrence with hosts (e.g., SAR11 viruses). This result corroborates the notion that contact rate is the key parameter for viral reproduction regardless of whether they show a long-term cooccurrence pattern with their hosts.

**Ecological interpretation inferred from virus-host dynamics.** There are at least three possible mechanisms that could lead to shifts of the dominant viral species in the same host groups (Fig. 5). First, more closely related prokaryotic populations that cannot be differentiated by the 16S rRNA gene polymorphism could cooccur with viruses. Previous studies focusing on the polymorphism of internal transcribed spacer (ITS) sequences (ITS-ASV) in SAR11 and *Cyanobacteria* reported that ITS-ASV dynamics correlate more with viral dynamics (as inferred from T4-like viral marker genes) than 16S-ASV dynamics of these taxa (7, 29). Therefore, dynamics of more highly resolved populations (e.g., ITS-ASVs or whole-genome-sequence-based populations) might have synchronized with observed viral dynamics. Second, the temporal acquisition of host resistance or viral counterresistance, as often observed in culture model systems (65), may cause a shift of the dominant viral species. Third, the shifts of viruses can be interpreted as a result of the founder effect following host fluctuation via genetic drift (66). Seasonal fluctuation of the host population causes a bottleneck effect, and, therefore, the founder effect following the bottleneck effect enables the abundances of several viral species to equally increase. This was suggested as a mechanism of an incomplete selective sweep in the freshwater *Cyanobacteria* populations having different CRISPR-spacer genotypes (67). This scenario is more plausible between ASV8-1 and their viruses because ASV8-1 experienced clear seasonal fluctuation (Fig. 5A).

Altogether, we revealed that frequency-dependent infection occurs in abundant prokaryotic populations according to cell density via “one-to-many” host-virus correspondences regardless of the host temporal dynamics. Although more direct evaluation of growth capacity is required for validation, this suggests “one-to-many” frequency-dependent infection can occur regardless of host growth strategy. Such “one-to-many” cooccurrences seem to support the validity of the previous study, which used cooccurrences between viruses for the host prediction method (68). One-to-many host-virus correspondences may suggest that a prokaryotic species can be attacked by multiple viruses with different infection strategies (e.g., different cell surface targets). This can cause difficulties in establishing complete resistance toward multiple coexisting viruses and sustaining continuous virus-host interaction in the environment. The inability to evolve complete resistance to many viral species in complex virus-host systems in the environment can contribute to the prevalence of frequency-dependent selection in abundant marine prokaryotes.

The comparison of monthly dynamics between prokaryotic and viral communities indicated concurrent seasonal shifts at the whole-community level. Concurrent seasonal shifts were also broadly observed between the corresponding virus and host pairs at the phylum or class level based on the host prediction analysis. We further statistically confirmed their cooccurrence via network analysis among abundant prokaryotic populations and their viruses regardless of the host taxa or temporal dynamics. These results likely suggest that abundant prokaryotes are broadly exposed to frequent viral infection

regardless of their taxa and growth strategy and indicate that lysis of the abundant prokaryotes via viral infection has a considerable impact on biogeochemical cycling and maintenance of prokaryotic community diversity. Further, these abundant prokaryotic populations should reflect actively growing members of the community because they became dominant even though they suffered frequent loss by viral lysis.

## MATERIALS AND METHODS

**Sampling and processing.** Seawater samples (4 l) were collected at a 5-m depth at the entrance of Osaka Bay (34°19'28"N, 135°7'15"E), Japan, between March 2015 and November 2016 at monthly intervals. Because the tide level and diel cycle can influence community composition, the sampling time was unified at 3 h before or after high tide in the daytime. Considering the climate of the sampling area, the seasons were defined as follows: March to May = spring, June to August = summer, September to November = autumn, and December to February = winter. Additionally, from June to July was considered the rainy season. Seawater was filtered through a 142-mm-diameter (3.0- $\mu$ m-pore-size) polycarbonate membrane (Millipore, Billerica, MA) and then sequentially through 0.22- $\mu$ m-pore-size Sterivex filtration units (SVG010RS, EMD Millipore). After filtration, filtration units were directly stored at  $-80^{\circ}\text{C}$  for subsequent DNA extraction. The filtrates were stored at  $4^{\circ}\text{C}$  before treatments. Water temperature and salinity were monitored using fixed water intake systems of the Research Institute of Environment, Agriculture, and Fisheries, Osaka prefecture. Nutrient concentrations ( $\text{NO}_3\text{-N}$ ,  $\text{NO}_2\text{-N}$ ,  $\text{NH}_4\text{-N}$ ,  $\text{PO}_4\text{-P}$ , and  $\text{SiO}_2\text{-Si}$ ) were measured by continuous flow analysis (BL TEC K.K., Japan). Prokaryotic cell and virus-like particles were enumerated using SYBR green and SYBR gold epifluorescence microscopy (69, 70).

**rRNA gene amplicon sequencing analysis.** For prokaryotic community analysis, DNA was extracted from the stored filtration units, as previously described (34, 71). Total 16S rRNA genes were amplified using a primer set based on the V3-V4 hypervariable region of prokaryotic 16S rRNA genes (72) with added overhang adapter sequences at each 5' end according to the sample preparation guide ([https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)). Amplicons were sequenced using a MiSeq sequencing system and MiSeq V3 (2  $\times$  300 bp) reagent kits (Illumina, San Diego, CA).

Paired-end 16S rRNA gene amplicon sequences were merged using VSEARCH with the "-M 1000" option (73). Merged reads containing ambiguous nucleotides (i.e., "N") were discarded. The remaining merged reads were clustered using VSEARCH to form operational taxonomic units (OTUs) at a 99% sequence identity threshold. Singleton OTUs were discarded. The representative sequences of the remaining OTUs were searched against the SILVA rRNA gene database (release 138) (74) to taxonomically annotate OTUs using SINA (75) at a 99% sequence identity threshold. To define abundant OTUs, we considered the reported minimum host cell density for effective viral infection ( $\approx 10^4$  cells/mL) (58). Following typical coastal marine prokaryotic cell density ( $\approx 10^6$  cells/mL) (76), we assumed prokaryotic cell density uniformity as  $10^6$  cells/mL during the whole sampling period, and the cutoff value for abundant OTUs was set as 1% relative abundance ( $10^6 \times 0.01$  [1%] =  $10^4$  cells/mL).

To identify statistically relevant variants within abundant OTUs, we applied minimum entropy decomposition (MED) (11) as previously reported. All the sequences from each 99% OTU were aligned using MAFFT v7.123b (-retree 1 -maxiterate 0 -nofft -parttree) (77). Alignments of sequences containing positions with an entropy of  $>0.25$  were decomposed, and decomposition continued until all positions had an entropy of  $<0.25$ . The minimum number of the most abundant sequence within each amplicon sequence variant (ASV) needed to exceed 50, and ASVs that did not exceed 1% of the parent OTU composition were discarded (7).

**Virome sequencing, assembly, classification, and calculation of relative abundance.** The filtrate containing viruses was concentrated via  $\text{FeCl}_3$  precipitation (78) and purified using DNase and a CsCl density centrifugation step (79). The DNA was then extracted as previously described (80). We failed to obtain enough DNA for virome sequencing for one sample (February 2016), so the sample was removed from the analysis. Libraries were prepared using a Nextera XT DNA sample preparation kit (Illumina, San Diego, CA), according to the manufacturer's protocol, using 0.25 ng of viral DNA. Samples were sequenced using a MiSeq sequencing system and MiSeq V3 (2  $\times$  300 bp) reagent kits (Illumina, San Diego, CA).

Viromes were individually assembled using SPAdes 3.9.1 with default  $k$ -mer lengths (81). Additionally, we used scaffolds of these assemblies (here referred to as contigs for simplicity). Only the virus-like contigs were extracted using VirSorter (categories 1, 2, and 3) (82). Circular contigs were determined as previously described (33). Contig sequences were clustered at 95% global average nucleotide identity with cd-hit-est (options -c 0.95 -G 1 -n 10 -mask NX; 549 redundant contigs were discarded) (83). A total of 5,226 monthly time series Osaka Bay viral (mts-OBV) contigs ( $>10$  kb, 62 to 926 contigs/sample, including 202 circular contigs) were obtained. Genome completeness and quality of mts-OBV contigs were evaluated using checkV (v0.7.0) (84).

In addition, this assembly generated 181,131 short contigs (i.e., from 1 kb to 10 kb). Because the reliability of virus prediction tools decreases in shorter contigs, the abundance of these contigs was assessed based on the relative abundance of terminase large subunit genes (*terL*), as previously described (34). In total, 4,666 genes were detected as putative *terL* genes (i.e., genes with the best hits to PF03354.14, PF04466.12, PF03237.14, and PF05876.11). Fragments per kilobase per million (FPKM) values for putative *terL* genes were calculated using in-house ruby scripts (<https://github.com/yosuken/CountMappedReads2>).

The mts-OBV contigs with complete viral genomic sequence set collected in a previous study (33) were used for viral abundance estimation based on the read mapping. The complete viral genomic

sequence belonged to one of the following two categories: (i) 1,811 environmental viral genomes (EVGs; all are circularly assembled genomes, 45 were assembled in Osaka Bay in a previous study [33]) derived from marine virome studies and (ii) 2,429 reference viral genomes (RVGs) of cultured dsDNA viruses, which were collected previously (33). Genus-level genomic OTUs (gOTUs) were previously assigned for complete genomes based on the genomic similarity score ( $S_g$ ) using ViPTree (85). For the mts-OBV contigs, if a sequence showed a high similarity to one of the complete genomes (with an  $S_g$  of  $>0.15$ ), the sequence was assigned to the gOTU of the most similar genome, as previously described (33, 34). Quality-controlled virome reads were obtained through quality control steps as previously described (33). These reads were mapped against the viral genomic sequence set using Bowtie2 software with the “-score-min L,0,-0.3” parameter (86). FPKM values were calculated using in-house ruby scripts, and relative abundances of each virus among the analyzed viruses were measured.

**Viral host prediction.** First, we assigned putative host groups based on genomic similarity with the viral genomic sequence set collected in a previous study (33). Among the 487 genera (gOTUs) of 2,429 cultured prokaryotic viral genomes in the previous classification, there were only two gOTUs that included viruses infecting two different host phyla (33). The exceptional gOTUs are G617 and G1038. G617 includes viruses that infect *Firmicutes* and *Deinococcus-Thermus*. Most of the members of G1038 infect *Enterobacteriales* in *Proteobacteria*, but a virus (*Staphylococcus* phage SA1) infects *Staphylococcus aureus* in *Firmicutes*. Thus, members in the same gOTU likely infect similar host groups, at least at a relatively broad taxonomic level as predicted in this study (phylum or class level in *Proteobacteria*). If mts-OBV contigs were classified into the same gOTU as the viruses with a known (via cultivation) or predicted (by genomic content [33]) host group, the host group was assigned to the contigs. We also compared similarity between mts-OBV contigs, the viral genomes deposited in Virus-Host DB (<https://www.genome.jp/virushostdb/>; as of October 2018), and recently reported isolates (87, 88).

For the viruses without assigned host groups via genomic similarity, we performed *in silico* host prediction based on the nucleotide sequence similarity between viruses and prokaryotes as follows according to previous studies (42, 89, 90). First, a total of 220,103 viral genomes/contigs derived from marine viromes were collected and used for the analysis (26, 33, 38, 43, 44) (Table S1 in the supplemental material). For the putative host genomes, we collected a total of 8,016 metagenome-assembled genomes (MAGs)/single amplified genomes (SAGs) from marine metagenomic or single-cell genomic studies (45–49). From Pachiadaki et al., we used only 1,040 high-quality SAG assemblies with  $\geq 80\%$  completion (49). To remove the contamination of virus-like contigs from the MAGs/SAGs, 14,967 contigs classified as viral-like sequences using VirSorter (categories 1, 2, and 3) (82) were discarded (Table S1). Details of each prediction method were reviewed previously (91).

**(i) CRISPR-spacer matching.** CRISPR-spacer sequences were predicted using the CRISPR Recognition Tool (92), and a total of 13,305 sequences were extracted from the analyzed 8,016 MAGs/SAGs. Detected spacer sequences and spacer sequences deposited in CRISPRdb (93) were queried against viral genomes using the BLASTn-short function (94), where at least 95% identity over the whole spacer length and only 1 to 2 SNPs at the 5' end of the sequence were allowed.

**(ii) tRNA matching.** tRNAs were recovered from the 8,016 MAGs/SAGs and viral genomes using ARAGORN with the “-t” option (95). A total of 213,939 and 31,439 tRNAs were recovered from MAGs/SAGs and viral genomes, respectively. The recovered prokaryotic and viral tRNAs with 111,385 tRNAs deposited in GtRNAdb (96) were compared using BLASTn (94), and only a perfect match (100% length and 100% sequence identity) was considered indicative of putative host-virus pairs.

**(iii) Nucleotide sequence homology of prokaryotic and viral genomes.** Viral genomes/contigs were queried against prokaryotic MAGs/SAGs and prokaryotic genomes in NCBI RefSeq (as of December 2019) using BLASTn (94). Only the best hits above 80% identity across the alignment with a length of  $\geq 1,500$  bp were considered indicative of host-virus pairs. For the prediction based on MAG/SAG contigs, we performed taxonomic validation of the matching contigs in MAGs/SAGs as previously described (42).

No inconsistent host prediction for each viral contig was observed between these three different methods. Viruses belonging to the same gOTU were assigned consistent host groups (33), with three exceptional gOTUs (G404 including putative alphaproteobacterial viruses and putative deltaproteobacterial viruses, G405 including putative *Bacteroidetes* viruses and putative *Marinimicrobia* viruses, and G495 including putative alphaproteobacterial viruses and putative *Bacteroidetes* viruses), which included viruses predicted to infect multiple host lineages. For the contigs assigned to the three gOTUs, genomic similarity scores among the same gOTU members were calculated, and the potential host of each contig was assigned based on the most similar genomes/contigs, which was annotated via host prediction.

**Statistical analyses.** Before statistical analyses, 16S rRNA amplicon reads were rarefied using the “vegan” package in R (20,803 reads per sample based on minimum sample size) (97). To examine within-sample alpha-diversity (Shannon diversity, evenness, and richness) and beta-diversity (Bray-Curtis similarity:  $1 - \text{Bray-Curtis dissimilarity}$  for all possible pairwise combinations among all of sampling points), we used the vegan package in R (98). Mantel tests were performed using R and the vegan package (98) only on fully overlapping sets of data. Pairwise correlations between the estimated abundance of prokaryotic ASVs and viral contigs (with putative host information and exceeding FPKM values of  $>10$  at least a month, 2,735 contigs) on fully overlapping sets of data were then determined via Spearman correlation ( $\rho > 0.6$ ,  $P < 0.01$ ,  $q < 0.05$ ), as implemented in the local similarity analysis program. (99, 100). Only the pairs supported by host prediction (e.g., *Bacteroidetes* and *Bacteroidetes* viruses) were considered cooccurring virus-host pairs. In the sanity check of the analysis (Figure S6E-G), pairs that were not supported by the host prediction were included. However, even though virus\_A cooccurred with ASV\_X, which was not supported by host prediction, virus\_A cooccurred with another ASV\_Y, which was supported by host prediction, and ASV\_X and ASV\_Y cooccurred; the pair between virus\_A and ASV\_X were

removed as a false positive. In absolute abundance-based analysis, the count data of prokaryotic cells and virus-like particles for the first 7 months were used because of the sample loss of other months. The absolute abundances of each ASV and virus were calculated as the multiplication of cell/virus-like particle counts and relative abundance. Network visualizations of correlation matrices were generated using igraph and ggnetwork (101, 102).

**Classification of temporal dynamics of ASVs.** We established indexes for the classification of temporal dynamics of each ASV by their monthly dynamics. For the analogy of the  $r$  (intrinsic rate of natural increase) of each ASV, the maximum increase of the normalized relative rank (0 to 1) per month of each ASV was applied. Similarly, for the analogy of  $K$  (carrying capacity) for each ASV, the length of the continuously abundant month ( $>0.1\%$  relative abundance, 1 to 18 months) of each ASV was applied. Although these indices are not the same as the original definitions of  $r$  and  $K$ , we refer to these two indices as an  $r$ -like index and a  $K$ -like index, respectively, for simplicity.

**Detection of SNPs.** Reads were mapped to viral contigs using Bowtie2 with a “-score-min L,0,-0.3” (86), and the resulting alignment files were converted to BAM format and sorted using SAMtools (103). The average genome entropy of the contigs, which exceeded more than  $10\times$  coverage each month, was computed using DiversiTools (<http://josephhughes.github.io/DiversiTools/>).

**Data availability.** Sequences obtained from the observations were deposited at the DNA Data Bank of Japan (DDBJ) under project number [PRJDB10879](https://www.ncbi.nlm.nih.gov/PRJDB10879). Raw sequence reads can be found under accession numbers [DRX260081](https://www.ncbi.nlm.nih.gov/DRX260081) to [DRX260115](https://www.ncbi.nlm.nih.gov/DRX260115), and assemblies of viromes can be found under BioSample ID [SAMD00279559](https://www.ncbi.nlm.nih.gov/SAMD00279559).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, PNG file, 0.2 MB.

**FIG S2**, PNG file, 0.8 MB.

**FIG S3**, PNG file, 0.1 MB.

**FIG S4**, PNG file, 0.5 MB.

**FIG S5**, PNG file, 0.6 MB.

**FIG S6**, PDF file, 0.5 MB.

**FIG S7**, PNG file, 0.6 MB.

**FIG S8**, PNG file, 0.3 MB.

**TABLE S1**, XLSX file, 7.8 MB.

**TABLE S2**, XLSX file, 0.01 MB.

## ACKNOWLEDGMENTS

K.T. would like to thank Ryoma Kamikawa, Shigeki Sawayama, and Daichi Morimoto at Kyoto University for their technical comments and preparation of the manuscript. K.T. would also like to thank Hisashi Endo, Florian Prodingler, Hiroaki Takebe, Kentaro Fujiwara, and Tatsuhiro Isozaki at Kyoto University and Keizo Nagasaki at Kochi University for very useful discussions. We thank the Research Institute of Environment, Agriculture, and Fisheries, Osaka prefecture, for supporting the sampling and Sakie Tanaka for continuous flow analysis. Computational work was supported by the Super Computer System, Institute for Chemical Research, Kyoto University. This study was supported by Grants-in-Aid for Scientific Research KAKENHI (numbers 17H03850 and 21H05057) and Challenging Exploratory Research (number 26660171) from the Japan Society for the Promotion of Science (JSPS), Canon Foundation (number 203143100025), JSPS Scientific Research on Innovative Areas (number 16H06437), and the Bilateral Open Partnership Joint Research Project (Japan-Lithuania Research Cooperative Program) “Research on prediction of environmental change in Baltic Sea based on comprehensive metagenomic analysis of microbial viruses.”

K.T. performed the experiments, analysis, and preparation of the manuscript. N.O.-H. and H.W. performed sampling, experiments, and analysis. K.Y. contributed to sampling and experiments. Y.N. and H.O. contributed to the analysis, discussion, and preparation of the manuscript. T.Y. contributed to the research design, discussion, manuscript revision, and overall support for this study.

We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## REFERENCES

- Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034–1039. <https://doi.org/10.1126/science.1153213>.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Ludicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Tara Oceans Coordinators. 2015. Structure and function of the global ocean microbiome. *Science* 348:1261359. <https://doi.org/10.1126/science.1261359>.
- Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P, Tunlid A, Hagström A. 2007. Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* 16:867–880. <https://doi.org/10.1111/j.1365-294X.2006.03189.x>.
- Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>.
- Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bönke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdt G, Wichels A, Wiltshire KH, Glockner FO, Schweder T, Amann R. 2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* 336:608–611. <https://doi.org/10.1126/science.1218344>.
- Needham DM, Fuhrman JA, Cram JA, Fuhrman JA, Sun F. 2016. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiol* 1:16005. <https://doi.org/10.1038/nmicrobiol.2016.5>.
- Needham DM, Sachdeva R, Fuhrman JA. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* 11:1614–1629. <https://doi.org/10.1038/ismej.2017.29>.
- Chafee M, Fernández-Guerra A, Buttigieg PL, Gerdt G, Eren AM, Teeling H, Amann RL. 2018. Recurrent patterns of microdiversity in a temperate coastal marine environment. *ISME J* 12:237–252. <https://doi.org/10.1038/ismej.2017.165>.
- Tikhonov M, Leach RW, Wingreen NS. 2015. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J* 9:68–80. <https://doi.org/10.1038/ismej.2014.117>.
- Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML. 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 4:1111–1119. <https://doi.org/10.1111/2041-210X.12114>.
- Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. 2015. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* 9:968–979. <https://doi.org/10.1038/ismej.2014.195>.
- Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–361. <https://doi.org/10.1038/nature04160>.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nat Microbiol* 3:754–766. <https://doi.org/10.1038/s41564-018-0166-y>.
- Fuhrman J, Suttle C. 1993. Viruses in marine planktonic systems. *Oceanography* 6:51–63. <https://doi.org/10.5670/oceanog.1993.14>.
- Winter C, Bouvier T, Weinbauer MG, Thingstad TF. 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the “killing the winner” hypothesis revisited. *Microbiol Mol Biol Rev* 74:42–57. <https://doi.org/10.1128/MMBR.00034-09>.
- Rodriguez-Valera F, Martín-Cuadrado A-B, Rodríguez-Brito B, Pasić L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* 7:828–836. <https://doi.org/10.1038/nrmicro2235>.
- Thingstad TF. 2000. 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. <https://doi.org/10.4319/lo.2000.45.6.1320>.
- Våge S, Storesund JE, Thingstad TF. 2013. Adding a cost of resistance description extends the ability of virus-host model to explain observed patterns in structure and function of pelagic microbial communities. *Environ Microbiol* 15:1842–1852. <https://doi.org/10.1111/1462-2920.12077>.
- Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB, Giovannoni SJ. 2013. Abundant SAR11 viruses in the ocean. *Nature* 494:357–360. <https://doi.org/10.1038/nature11921>.
- Martínez-Hernández F, Fornas O, Lluésma Gómez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, Anton J, Gasol JM, Rosselli R, Rodríguez-Valera F, Sullivan MB, Acinas SG, Martínez-García M. 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 8:15892. <https://doi.org/10.1038/ncomms15892>.
- Martínez-Hernández F, Fornas O, Lluésma Gómez M, García-Heredia I, Maestre-Carballe L, López-Pérez M, Haro-Moreno JM, Rodríguez-Valera F, Martínez-García M. 2018. Single-cell genomics uncover *Pelagibacter* as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *ISME J* 13:232–236. <https://doi.org/10.1038/s41396-018-0278-7>.
- Fuhrman JA, Cram JA, Needham DM. 2015. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* 13:133–146. <https://doi.org/10.1038/nrmicro3417>.
- Bunse C, Pinhassi J. 2017. Marine bacterioplankton seasonal succession dynamics. *Trends Microbiol* 25:494–505. <https://doi.org/10.1016/j.tim.2016.12.013>.
- Pagarete A, Chow C-ET, Johannessen T, Fuhrman JA, Thingstad TF, Sandaa RA. 2013. Strong seasonality and interannual recurrence in marine myovirus communities. *Appl Environ Microbiol* 79:6253–6259. <https://doi.org/10.1128/AEM.01075-13>.
- Chow CET, Fuhrman JA. 2012. Seasonality and monthly dynamics of marine myovirus communities. *Environ Microbiol* 14:2171–2183. <https://doi.org/10.1111/j.1462-2920.2012.02744.x>.
- Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. 2020. Long-term stability and Red Queen-like strain dynamics in marine viruses. *Nat Microbiol* 5:265–271. <https://doi.org/10.1038/s41564-019-0628-x>.
- Hwang J, Park SY, Park M, Lee S, Lee T-K. 2017. Seasonal dynamics and metagenomic characterization of marine viruses in Goseong Bay, Korea. *PLoS One* 12:e0169841. <https://doi.org/10.1371/journal.pone.0169841>.
- Havroni G, Flores-Urbe J, Béja O, Philofof A. 2020. Seasonal and diel patterns of abundance and activity of viruses in the Red Sea. *Proc Natl Acad Sci U S A* 117:29738–29747. <https://doi.org/10.1073/pnas.2010783117>.
- Ahlgren NA, Perelman JN, Yeh Y, Fuhrman JA. 2019. Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors. *Environ Microbiol* 21:2948–2963. <https://doi.org/10.1111/1462-2920.14687>.
- Brum JR, Sullivan MB. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Microbiol* 13:147–159. <https://doi.org/10.1038/nrmicro3404>.
- Ozaki K, Uye S-I, Kusumoto T, Hagino T. 2004. Interannual variability of the ecosystem of the Kii Channel, the Inland Sea of Japan, as influenced by bottom intrusion of cold and nutrient-rich water from the Pacific Ocean, and a recent trend of warming and oligotrophication. *Fish Oceanogr* 13:65–79. <https://doi.org/10.1046/j.1365-2419.2003.00274.x>.
- Kopf A, Bica M, Kottmann R, Schnetzer J, Kostadinov I, Lehmann K, Fernández-Guerra A, Jeanthon C, Rahav E, Ullrich M, Wichels A, Gerdt G, Polymenakou P, Kotoulas G, Siam R, Abdallah RZ, Sonnenschein EC, Cariou T, O'Gara F, Jackson S, Orlic S, Steinke M, Busch J, Duarte B, Caçador I, Canning-Clode J, Bobrova O, Marteinsson V, Reynisson E, Loureiro CM, Luna GM, Quero GM, Löscher CR, Kremp A, DeLorenzo ME, Øvreås L, Tolman J, LaRoche J, Penna A, Frischer M, Davis T, Katherine B, Meyer CP, Ramos S, Magalhães C, Jude-Lemelleur F, Aguirre-Macedo ML, Wang S, Poulton N, Jones S, et al. 2015. The ocean sampling day consortium. *GigaScience* 4:27. <https://doi.org/10.1186/s13742-015-0066-5>.
- Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, Blanc-Mathieu R, Yamamoto K, Hingamp P, Sako Y, Sullivan MB, Goto S, Ogata H, Yoshida T. 2017. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* 2:e00359-16. <https://doi.org/10.1128/mSphere.00359-16>.
- Yoshida T, Nishimura Y, Watai H, Haruki N, Morimoto D, Kaneko H, Honda T, Yamamoto K, Hingamp P, Sako Y, Goto S, Ogata H. 2018. Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME J* 12:1287–1295. <https://doi.org/10.1038/s41396-018-0052-x>.
- Martínez-Hernández F, Luo E, Tominaga K, Ogata H, Yoshida T, DeLong EF, Martínez-García M. 2020. Diel cycling of the cosmopolitan abundant *Pelagibacter* virus 37-F6: one of the most abundant viruses on earth. *Environ Microbiol Rep* 12:214–219. <https://doi.org/10.1111/1758-2229.12825>.



36. Coenen AR, Weitz JS. 2018. Limitations of correlation-based inference in complex virus-microbe communities. *mSystems* 3:e00084-18. <https://doi.org/10.1128/mSystems.00084-18>.
37. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee KB, Malmstrom RR, Martinez-Garcia M, Mizrahi IK, Ogata H, Páez-Espino D, Petit MA, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, et al. 2019. Minimum information about an uncultivated virus genome (MIUVIG). *Nat Biotechnol* 37:29–37. <https://doi.org/10.1038/nbt.4306>.
38. Luo E, Aylward FO, Mende DR, DeLong EF. 2017. Bacteriophage distributions and temporal variability in the ocean's interior. *mBio* 8:e01903-17. <https://doi.org/10.1128/mBio.01903-17>.
39. Prodinge F, Endo H, Takano Y, Li Y, Tominaga K, Iozaki T, Blanc-Mathieu R, Gotoh Y, Hayashi T, Taniguchi E, Nagasaki K, Yoshida T, Ogata H. 2021. Year-round dynamics of amplicon sequence variant communities differ among eukaryotes, *Imitervirales* and prokaryotes in a coastal ecosystem. *FEMS Microbiol Ecol* 97:fiab167. <https://doi.org/10.1093/femsec/fiab167>.
40. Cram JA, Chow C-ET, Sachdeva R, Needham DM, Parada AE, Steele JA, Fuhrman JA. 2015. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J* 9:563–580. <https://doi.org/10.1038/ismej.2014.153>.
41. Yeh Y-C, Fuhrman JA. 2022. Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time. *ISME Commun* 2:36. <https://doi.org/10.1038/s43705-022-00121-8>.
42. Tominaga K, Morimoto D, Nishimura Y, Ogata H, Yoshida T. 2020. *In silico* prediction of virus-host interactions for marine *Bacteroidetes* with the use of metagenome-assembled genomes. *Front Microbiol* 11:738. <https://doi.org/10.3389/fmicb.2020.00738>.
43. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Domínguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay JÉ, Vik D, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, de Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jallion O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Not F, Ogata H, Poulton N, Raes J, Sardet C, Speich S, Stemann L, Sullivan MB, Sunagawa S, Wincker P, Culley AI, Tara Oceans Coordinators, et al. 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177:1109–1123. <https://doi.org/10.1016/j.cell.2019.03.040>.
44. Mizuno CM, Ghai R, Saghaï A, López-García P, Rodriguez-Valera F. 2016. Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7:e00805-16. <https://doi.org/10.1128/mBio.00805-16>.
45. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* 5:170203. <https://doi.org/10.1038/sdata.2017.203>.
46. Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 2017. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* 5:e3558. <https://doi.org/10.7717/peerj.3558>.
47. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lückner S, Eren AM. 2018. Nitrogen-fixing populations of *Planctomycetes* and *Proteobacteria* are abundant in surface ocean metagenomes. *Nat Microbiol* 3:804–813. <https://doi.org/10.1038/s41564-018-0176-9>.
48. Krüger K, Chafee M, Ben Francis T, Glavina del Rio T, Becher D, Schweder T, Amann RL, Teeling H. 2019. In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. *ISME J* 13:2800–2816. <https://doi.org/10.1038/s41396-019-0476-y>.
49. Pachiadaki MG, Brown JM, Brown J, Bezuidt O, Berube PM, Biller SJ, Poulton NJ, Burkart MD, La Clair JJ, Chisholm SW, Stepanauskas R. 2019. Charting the complexity of the marine microbiome through single-cell genomics. *Cell* 179:1623–1635. <https://doi.org/10.1016/j.cell.2019.11.017>.
50. Parada V, Herndl GJ, Weinbauer MG. 2006. Viral burst size of heterotrophic prokaryotes in aquatic systems. *J Mar Biol Assoc U K* 86:613–621. <https://doi.org/10.1017/S002531540601352X>.
51. Morris RM, Cain KR, Hvorecny KL, Kollman JM. 2020. Lysogenic host-virus interactions in SAR11 marine bacteria. *Nat Microbiol* 5:1011–1015. <https://doi.org/10.1038/s41564-020-0725-x>.
52. Needham DM, Chow C-ET, Cram JA, Sachdeva R, Parada A, Fuhrman JA. 2013. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J* 7:1274–1285. <https://doi.org/10.1038/ismej.2013.19>.
53. Sullivan MB, Waterbury JB, Chisholm SW. 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* 424:1047–1051. <https://doi.org/10.1038/nature01929>.
54. 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. <https://doi.org/10.1128/aem.59.10.3393-3399.1993>.
55. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* 3:e144. <https://doi.org/10.1371/journal.pbio.0030144>.
56. Suttle CA, Chan AM. 1993. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser* 92:99–109. <https://doi.org/10.3354/meps092099>.
57. Suttle CA, Chan AM. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol* 60:3167–3174. <https://doi.org/10.1128/aem.60.9.3167-3174.1994>.
58. Wiggins BA, Alexander M. 1985. Minimum bacterial density for bacteriophage replication: implications for significance of bacteriophages in natural ecosystems. *Appl Environ Microbiol* 49:19–23. <https://doi.org/10.1128/aem.49.1.19-23.1985>.
59. Weinbauer MG. 2004. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28:127–181. <https://doi.org/10.1016/j.femsrev.2003.08.001>.
60. Aylward FO, Boeuf D, Mende DR, Wood-Charlson EM, Vislova A, Eppley JM, Romano AE, DeLong EF. 2017. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc Natl Acad Sci U S A* 114:11446–11451. <https://doi.org/10.1073/pnas.1714821114>.
61. Brum JR, Ignacio-Espinoza JC, Roux S, Doullier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB, Tara Oceans Coordinators. 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. <https://doi.org/10.1126/science.1261498>.
62. Zborowsky S, Lindell D. 2019. Resistance in marine cyanobacteria differs against specialist and generalist cyanophages. *Proc Natl Acad Sci U S A* 116:16899–16908. <https://doi.org/10.1073/pnas.1906897116>.
63. Andrews JH, Harris RF. 1986. *r*- and *K*-selection and microbial ecology, p 99–147. *In* Marshall KC (ed), *Advances in microbial ecology*. Springer, Boston, MA.
64. Kimura S, Sako Y, Yoshida T. 2013. Rapid *Microcystis* cyanophage gene diversification revealed by long- and short-term genetic analyses of the tail sheath gene in a natural pond. *Appl Environ Microbiol* 79:2789–2795. <https://doi.org/10.1128/AEM.03751-12>.
65. Koskella B, Brockhurst MA. 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* 38:916–931. <https://doi.org/10.1111/1574-6976.12072>.
66. Cohan FM, Perry EB. 2007. A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* 17:R373–R386. <https://doi.org/10.1016/j.cub.2007.03.032>.
67. Kimura S, Uehara M, Morimoto D, Yamanaka M, Sako Y, Yoshida T. 2018. Incomplete selective sweeps of *Microcystis* population detected by the leader-end CRISPR fragment analysis in a natural pond. *Front Microbiol* 9:425. <https://doi.org/10.3389/fmicb.2018.00425>.
68. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL. 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun* 8:15955. <https://doi.org/10.1038/ncomms15955>.
69. Chen F, Lu J, Binder BJ, Liu Y, Hodson RE. 2001. Application of digital image analysis and flow cytometry to enumerate marine viruses stained with SYBR gold. *Appl Environ Microbiol* 67:539–545. <https://doi.org/10.1128/AEM.67.2.539-545.2001>.
70. Noble R, Fuhrman J. 1998. Use of SYBR green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microb Ecol* 14:113–118. <https://doi.org/10.3354/ame014113>.
71. Takebe H, Tominaga K, Fujiwara K, Yamamoto K, Yoshida T. 2020. Differential responses of a coastal prokaryotic community to phytoplanktonic organic matter derived from cellular components and exudates. *Microbes and Environments* 35:ME20033. <https://doi.org/10.1264/jmsme2.ME20033>.
72. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. 2014. Development of a prokaryotic universal primer for simultaneous analysis of

- Bacteria* and *Archaea* using next-generation sequencing. *PLoS One* 9: e105592. <https://doi.org/10.1371/journal.pone.0105592>.
73. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>.
  74. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
  75. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829. <https://doi.org/10.1093/bioinformatics/bts252>.
  76. Whitman WB, Coleman DC, Wiebe WJ, Schwalbach MS, Brown MV, Green JL, Brown JH. 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95:6578–6583. <https://doi.org/10.1073/pnas.95.12.6578>.
  77. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
  78. John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* 3:195–202. <https://doi.org/10.1111/j.1758-2229.2010.00208.x>.
  79. 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* 15:1428–1440. <https://doi.org/10.1111/j.1462-2920.2012.02836.x>.
  80. Kimura S, Yoshida T, Hosoda N, Honda T, Kuno S, Kamiji R, Hashimoto R, Sako Y. 2012. Diurnal infection patterns and impact of *Microcystis* cyanophages in a Japanese pond. *Appl Environ Microbiol* 78:5805–5811. <https://doi.org/10.1128/AEM.00571-12>.
  81. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
  82. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.
  83. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
  84. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2020. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 39:578–585. <https://doi.org/10.1038/s41587-020-00774-7>.
  85. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. 2017. ViPTree: the viral proteomic tree server. *Bioinformatics* 33:2379–2380. <https://doi.org/10.1093/bioinformatics/btx157>.
  86. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  87. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H. 2016. Linking virus genomes with host taxonomy. *Viruses* 8:66. <https://doi.org/10.3390/v8030066>.
  88. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, Du S, Tian Z, Zhao Y. 2020. Culturing novel and abundant pelagiphages in the ocean. *Environ Microbiol* 23:1145–1161. <https://doi.org/10.1111/1462-2920.15272>.
  89. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. *Nature* 536:425–430. <https://doi.org/10.1038/nature19094>.
  90. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB, Sullivan MB, Tara Oceans Coordinators. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. <https://doi.org/10.1038/nature19366>.
  91. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiol Rev* 40:258–272. <https://doi.org/10.1093/femsre/fuv048>.
  92. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. <https://doi.org/10.1186/1471-2105-8-209>.
  93. Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. <https://doi.org/10.1186/1471-2105-8-172>.
  94. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
  95. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
  96. Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* 44:D184–D189. <https://doi.org/10.1093/nar/gkv1309>.
  97. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcginn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Maintainer HW. 2020. vegan: community ecology package (version 2.6.2). <https://github.com/vegandevs/vegan>.
  98. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
  99. Xia LC, Steele JA, Cram JA, Cardon ZG, Simmons SL, Vallino JJ, Fuhrman JA, Sun F. 2011. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst Biol* 5: S15. <https://doi.org/10.1186/1752-0509-5-S2-S15>.
  100. Xia LC, Ai D, Cram J, Fuhrman JA, Sun F. 2013. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics* 29:230–237. <https://doi.org/10.1093/bioinformatics/bts668>.
  101. Csardi G, Nepusz T. 2006. The igraph software package (version 1.3.0) for complex network research. *InterJournal, Complex Systems* 1695. <https://igraph.org/>.
  102. Tyner S, Briatte F, Hofmann H. 2017. Network visualization with ggplot2. *R Journal* 9:27–59.
  103. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>.