Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan

Takashi Yoshida, Yosuke Nishimura, Hiroyasu Watai, Nana Haruki, Daichi Morimoto, Hiroto Kaneko, Takashi Honda, Keigo Yamamoto, Pascal Hingamp, Yoshihiko Sako, et al.

The ISME Journal
https://doi.org/10.1038/s41396-018-0052-x

ARTICLE

Corrected: Correction

Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan

Takashi Yoshida[1] · Yosuke Nishimura[1,2] · Hiroyasu Watai[1] · Nana Haruki[1] · Daichi Morimoto[1] · Hiroto Kaneko[3] · Takashi Honda[1] · Keigo Yamamoto[4] · Pascal Hingamp[5] · Yoshihiko Sako[1] · Susumu Goto[6] · Hiroyuki Ogata[3]

Received: 9 August 2017 / Revised: 10 November 2017 / Accepted: 20 December 2017

Abstract

Viruses infecting microorganisms are ubiquitous and abundant in the ocean. However, it is unclear when and where the numerous viral particles we observe in the sea are produced and whether they are active. To address these questions, we performed time-series analyses of viral metagenomes and microbial metatranscriptomes collected over a period of 24 h at a Japanese coastal site. Through mapping the metatranscriptomic reads on three sets of viral genomes ((i) 878 contigs of Osaka Bay viromes (OBV), (ii) 1766 environmental viral genomes from marine viromes, and (iii) 2429 reference viral genomes), we revealed that all the local OBV contigs were transcribed in the host fraction. This indicates that the majority of viral populations detected in viromes are active, and suggests that virions are rapidly diluted as a result of diffusion, currents, and mixing. Our data further revealed a peak of cyanophage gene expression in the afternoon/dusk followed by an increase of genomes from their virions at night and less-coherent infectious patterns for viruses putatively infecting various groups of heterotrophs. This suggests that cyanophages drive the diel release of cyanobacteria-derived organic matter into the environment and viruses of heterotrophic bacteria might have adapted to the population-specific life cycles of hosts.

Introduction

Viruses infecting microorganisms ubiquitously and abundantly present in the ocean [1] and virus-mediated microbial lysis has a considerable impact on the marine biogeochemical cycles [2–4]. As the discovery of highly abundant small particles in the sea [1, 4], initially referred to as "virus like particles (VLPs)", reflecting their unknown status of activity, numerous viruses have been isolated from marine environments. Various approaches have subsequently been undertaken to characterize the biogeography of viral populations and to reveal ecological and evolutionary forces that viruses exert on their hosts [5]. Of the different approaches, viral metagenome (i.e., "virome") sequencing has substantially expanded our knowledge of the diversity of marine viruses [6, 7]. However, a virome alone does not offer any direct evidence of viral infection, though it does reveal the presence of virions and associated nucleic acids. Viral infection can be detected by monitoring viral gene transcription in the host by either reverse transcription polymerase chain reactions targeting the marker genes of a specific group of viruses [8] or metatranscriptomics uncovering transcripts of a whole range of viruses [9, 10].

Takashi Yoshida and Yosuke Nishimura are contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41396-018-0052-x) contains supplementary material, which is available to authorized users.

✉ Takashi Yoshida
yoshiten@kais.kyoto-u.ac.jp
Hiroyuki Ogata
ogata@kuicr.kyoto-u.ac.jp

[1] Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan

[2] School of Life Science and Technology, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152-8550, Japan

[3] Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

[4] Research Institute of Environment, Agriculture and Fisheries, Osaka Prefecture, 442, Shakudo, Habikino, Osaka, Japan

[5] Aix Marseille Université, Université de Toulon, CNRS, IRD, MIO UM 110, 13288 Marseille, France

[6] Database Center for Life Science, Joint-Support Center for Data Science Research, Research Organization of Information and Systems, Wakashiba, Kashiwa, Chiba 277-0871, Japan

Published online: 30 January 2018

Osaka Bay, Japan, is an enclosed gulf with an area of 1500 square kilometers surrounded by densely populated districts and forests. Osaka Bay is a typical eutrophic environment owing to the input of nutrients from rivers, although it is also affected by Kuroshio, an oligotrophic warm current, intruding into the bay through the Kii Channel [11]. We previously analyzed a set of time-series viromes collected over a period of 24 h (nine time points) at Osaka Bay (Osaka Bay viromes, OBVs [5]) and successfully obtained complete genomes of not only well-documented viral lineages (e.g., cyanophages, SAR11 phages) but also novel viral lineages infecting heterotrophic bacteria such as marine group II Euryarchaeota, Bacteroidetes, SAR86, and SAR116 members. Photoautotrophic cyanobacteria as well as heterotrophic bacteria display population-specific diel cycling in many of their transcripts [12], suggesting that coordinated infection cycles of viruses infecting these bacteria also occur in this enclosed gulf. Here, we generated a time-series of host metatranscriptomes from samples concomitantly collected at the same sampling site as the viromes. We addressed the activities and diel infection dynamics of marine viruses through cross-comparisons of the coupled virome and metatranscriptomic data.

Materials and methods

Sampling

Prokaryotic fractions were prepared from a time-series of nine seawater samples from a 5 m depth at the entrance of Osaka Bay, Japan, every 3 h over a period of 24 h on August 25 and 26, 2014. On these days, the sunrise and sunset times were 5:26 and 18:35, respectively. During the sampling, the times and heights (cm) for high tide and low tide were as follows: high tide, 6:58/159, 19:12/160; low tide, 12:55/89, 0:30/44. For metatranscriptomic analysis, 1 l of seawater was prefiltered through a 142 mm 3.0 µm-pore polycarbonate membrane (EMD Millipore, USA) and then filtered sequentially through 0.22 µm-pore Sterivex filtration units (SVGV010RS, EMD Millipore). After filtration, the 0.22 µm filtration units were either directly transferred to −80 °C (for subsequent DNA extraction) or immediately filled with RNA*later* (Thermo Fisher Scientific, USA) and stored at −80 °C (for subsequent RNA extraction).

DNA/RNA extraction and sequencing

For prokaryotic host community analysis, DNA was extracted from the stored filtration units using a xanthogenate–sodium dodecyl sulphate method according to a previously published protocol [7]. A total of 16 S rDNA was amplified using a primer set based on the V3–V4 hypervariable region of prokaryotic 16 S rRNA genes for the simultaneous detection of Bacteria and Archaea [13] with added overhang adapter sequences at each 5′-end according to the 16 S sample preparation guide (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 commercial sequencing kit, version 3 (2 × 300 bp paired-end) (Illumina, USA). Virome library preparation, DNA extraction, and sequencing were performed as previously described [5]. For metatranscriptomic analysis, total RNA was extracted from the stored filters using mirVana (Thermo Fisher Scientific). After removal of rRNA with a Ribo-Zero rRNA removal kit (bacteria) (Illumina), the RNAs were converted to double-stranded cDNA with PrimeScript (TaKaRa, Japan). One ng of cDNA was segmented and tagged with NexteraXT (Illumina) following a 12-cycle amplification as described by the Illumina protocol and sequenced using sequencing kit version 3 (2 × 75 bp paired-end) (Illumina).

16 S rRNA gene amplicon sequence analysis

Paired-end sequences of the 16 S rDNA amplicon were merged using FLASH software (version 1.2.11) with option "-M 1000" [14]. Merged reads containing ambiguous nucleotides (i.e., "N") were discarded. The remaining merged reads were clustered using CD-HIT-EST (version 4.7) [15] to form operational taxonomic units (OTUs) with a sequence identity threshold of 97%. Singleton OTUs were discarded at this stage. The representative sequences of the remaining OTUs were searched against the SILVA ribosomal RNA gene database (release 128) [16] to taxonomically annotate OTUs using CD-HIT-EST-2D with a threshold of 97% sequence identity.

Virome and metatranscriptome analysis

Co-assembly of sequence reads from the nine OBV samples generated, after decontamination of prokaryotic sequences, 878 viral contigs longer than 10 kb (OBV long contigs; including 45 circularly assembled genomes and 833 non-circular genomes) [5]. In addition, this assembly generated 27,551 short contigs (i.e., longer than 1 kb but no longer than 10 kb). The abundance of these contigs was assessed based on the relative abundance of terminase large subunit genes (*terL*). Putative *terL* genes were detected as follows. Gene predictions were computed using MetaGeneMark [17]. HHsearch (probability > 85% and *p*-value < 1e-5) [18] was performed for each predicted gene against the Pfam database (version 31.0) after constructing query hidden Markov models using jackhammer (included in the

HMMER package) [19] with default settings against genes included in 5095 viral genomic sequences (described below; Supplementary Table S1). As a result, 1348 genes (from 1330 contigs) were detected as putative *terL* genes (i.e., genes with the best hit to PF03354.14, PF04466.12, PF03237.14, and PF05876.11). The FPKM (fragments per kilobase per mapped million reads) for putative *terL* genes were calculated by in-house ruby scripts. A longer *terL* was selected if two putative *terL* genes were detected in one contig.

The viral genomic sequence set (likely originating from prokaryotic dsDNA viruses; 5095 sequences; Supplementary Table S1) was collected in a previous study [5]. These sequences belong to one of the following three categories: (i) 878 viral contigs (> 10 kb, including 45 circularly assembled genomes and 833 non-circular genomes) obtained from OBVs, (ii) 1766 environmental viral genomes (EVGs; all are circularly assembled genomes) derived from marine virome studies (not including OBVs), and (iii) 2429 reference viral genomes (RVGs) of cultured dsDNA viruses. Genus-level genomic OTUs (gOTUs) were assigned for complete genomes (i.e., 45 circularly assembled genomes in (i) and all genomes in (ii), and (iii)) as previously described [5]. In addition, for the OBV non-circular (i.e., possibly incomplete) genome sequences, if the sequence showed similarity to one of the complete genomes (with genomic similarity $S_G > 0.15$), the sequence was assigned to the gOTU of the most similar complete genome. Putative host groups (Cyanobacteria, Bacteroidetes, SAR11, SAR116, and SAR86) were predicted from genomic similarity to known viruses and/or gene contents of complete genomes [5]. As viruses belonging to the same gOTU were predicted to have consistent host groups (i.e., each gOTU was not assigned to more than one host group), putative host groups were also applied to the other members of each gOTU.

Quality controlled virome and metatranscriptome sequence reads (Supplementary Table S2) were obtained through quality control steps as previously described [5]. These reads were mapped against the viral genomic sequence set using Bowtie 2 software with a parameter "--score-min L,0,-0.3" [20]. FPKM values were calculated by in-house ruby scripts. Hierarchical clustering of OBV contigs based on Pearson correlation of time-series virome abundance was performed using the average linkage method implemented on Multiple Experiment Viewer version 4.9.0. The gap statistics calculated by the R package clusGap [21] were used to assess the number of clusters obtained from the hierarchical clustering with parameters 'Kmax = 10, bootstrap = 10000'. In this analysis, the optimum number of clusters was estimated to be seven. All statistical tests were performed with R software (version 3.3.1).

Results and discussion

Sequence reads from the virion-derived metagenomes (OBV; 8.5 M read pairs; 2.4 Gbp generated from the virus enriched <0.22 μm size fraction; Supplementary Table S2) were assembled into 878 contigs (>10 kb), which included 45 putatively complete viral genomes (28.5–192 kb, average 54.2 kb) [5]. Such metagenomic data provide near-quantitative estimates of dsDNA genomes packed in viral particles [22]. These 878 contigs recruited 29.0% of the quality controlled OBV metagenomic reads (Supplementary Table S2). Our analysis indicated that these 878 contigs represent a wide range of viral populations with varying relative abundances. Based on the relative abundances of the terminase large subunit genes (*terL*), one of the core genes of tailed phages [23], in the whole set of viromes, we estimated that the relative abundances of viral populations represented by the long contigs could be as low as 0.0193% (OBV_N00282), whereas the highest was 0.779% (OBV_N00192) (Supplementary Figure S1).

Next, we prepared a set of viral genomic data that was composed of three groups: (i) the 878 OBV contigs, (ii) 1766 EVGs, and (iii) 2429 RVGs from cultured dsDNA viruses. The OBV reads were preferentially recruited on the OBV contigs (Fig. 1a). The proportion of genomic sequences covered by reads was higher for the OBV contigs (median 99.86%) compared with other viral genomes (EVGs: 1.54%; RVGs: 0.89%; Fig. 1b), demonstrating once again the locality of viral communities as previously suggested by a global ocean virome study [24].

Despite the locality of viral communities, some viral genomes abundant in our samples were also observed in other viromes. Of the top 50 most abundant completely sequenced viral genomes, 45 were from OBVs (OBV complete genomes; Supplementary Figure S2). The other five abundant genomes were those of SAR11 phage HTVC010P, *Synechococcus* phage S-RIP1, *Synechococcus* phage S-SM2, and two EVGs (uvMED-CGR-U-MedDCM-OCT-S38-C40 (AP013455) and TARA_ERS488757_N000101). SAR11 phage HTVC010P was previously proposed as the most abundant marine virus [25], and its abundance was ranked 44th in the Osaka Bay data set. We also noted that OBV-N00073 (ranked 35th) was related to SAR11 phage HTVC019, one of the four previously sequenced SAR11 phages (Supplementary Figure S2). Among the EVGs and RVGs, nine genomes (two EVGs and seven RVGs) showed high abundances (i.e., average FPKM, >10; %-virome mapped region, >60%; Fig. 1a, b). Of the nine genomes, six belonged to the set of 38 globally abundant virus clusters identified in a recent study [6]. These globally abundant viruses may have some common characteristics, such as a broad host range, allowing their wide geographic distribution.

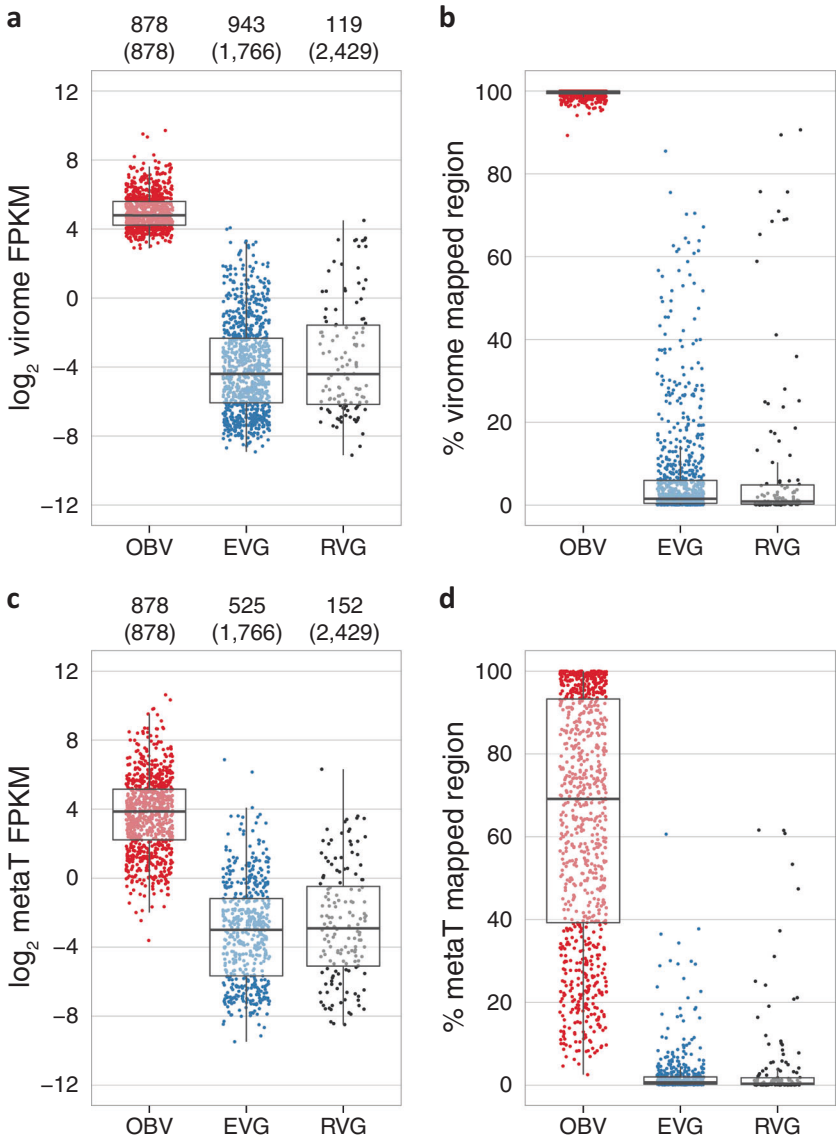

Fig. 1 Virome abundance and viral transcript abundance in Osaka Bay. Sequence reads obtained from nine viromes **a**, **b** and nine metatranscriptomes **c**, **d** were mapped on OBV contigs (red), EVGs (blue), and RVGs (black). *y*-axis represents log2-scaled average FPKM (fragments per kilobase per mapped million reads) **a**, **c** and % of the read mapped region **b**, **d**. For each panel, only sequences with ≥ 1 mapped reads are shown. The numbers of the read mapped sequences are indicated on the top of panels **a**, **c**, with the total numbers of sequences in parentheses. The boxes represent the first quartile, median, and third quartile. metaT: metatranscriptome

Given that viromes were derived from virions (i.e., 0.22 µm filtered seawater excluding host cells), a fundamental question is when and where these virions were generated; were the viral genomes generated through local infection at the time of sampling (i.e., "local production") or did they come from further afield through transport by sea currents or water mixing?

To address this question, we compared the viromes with the metatranscriptomes derived from host size fractions (0.22–3 µm) that were concomitantly collected with the virome fractions. Mapping the metatranscriptomic reads on the set of viral genomes confirmed that all viral genomes represented by the 878 OBV contigs were transcribed in the host cells in the same water (Fig. 1c, d). Both the depth and %-region of the metatranscriptome mapping were substantially higher for the OBV contigs (median FPKM, 14.5; median %-region, 69.14%) than for the EVGs (0.125; 0.67%) and RVGs (0.133; 0.34%). This result supports the hypothesis of overwhelming local production of viral particles in our samples. Namely, virion-associated genomes we captured were derived from virions produced by viral infection that occurred at the time of sampling during the period of 24 h. Local production of viruses readily explains the locality of viral ("particle") communities observed in larger scale virome studies [24]. The fact that all of the viral contigs (>10 kb) were transcriptionally active with no exception further suggests that the concentration of these viral particles rapidly decreases, if there is no successive infection, to levels undetectable by metagenomic approaches upon their diffusion owing to water movement.

Local production of viruses implies that viral community structures revealed by viromes are possibly linked with viral activities (i.e., viral infection) in the same water. However, the relative abundances of virion genomes were poorly

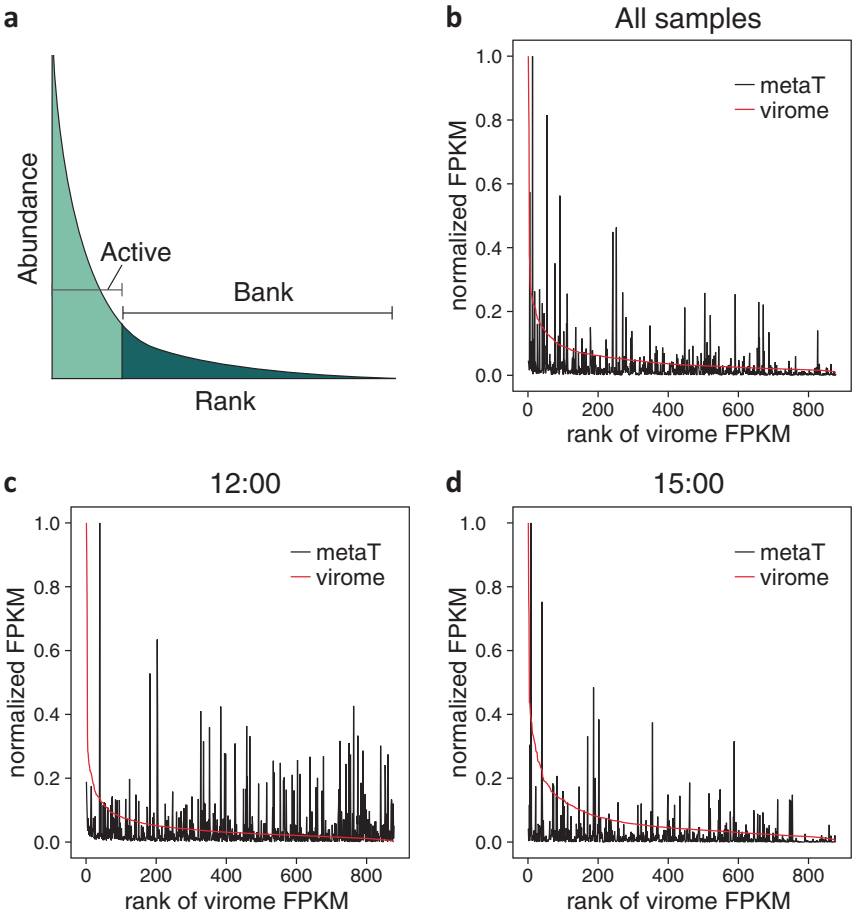

Fig. 2 Rank-abundance curves of viruses with transcriptomic abundance. **a** Illustration of the Bank model proposed by [26]. **b** Average abundance of the nine time points for 878 OBV viral populations (contigs). *y*-axis represents normalized virome and metatranscriptome FPKM (i.e., each FPKM is divided by the FPKM of the most abundant sequence). **c** Abundance at 12:00 (the first sample). **d** Abundance at 15:00. metaT: metatranscriptome

correlated with those of viral transcripts (i.e., the maximum of the coefficient of determination $r^2 = 0.237$), although the correlation was statistically significant owing to the large number of viral sequences being compared (all samples, Pearson's $r = 0.280$, $p < 2.2e\text{-}16$; individual samples, $r = 0.174$–0.487, $p < 1.94e\text{-}11$). Indeed, relative transcript abundances could be high for some of the low-abundant viral genomes as shown in Fig. 2b–d. We consider that possible factors affecting the low correlations between virion genome abundance and viral transcriptional activity include a wide range of viral burst sizes as well as viral latent periods where transcriptional activity precedes viral burst events.

In viral ecology, the Bank model [24, 26] states that only a handful of abundant viruses are "active" in a community of viruses and the rest are rare and inactive (i.e., "bank"), which may move from the bank to the active group upon environmental changes or host community shifts (Fig. 2a). Photoautotrophic cyanobacteria as well as heterotrophic bacteria are known to display population-specific diel cycles in their gene expression [12]. Previous laboratory experiments also showed that cyanophage production depends on host photosynthesis and exhibits a diel pattern [27, 28]. If the lytic cycles of marine viruses are synchronized with the host diel cycles in natural populations, we may observe coordinated viral gene expression followed by an increase of progeny virions. Indeed, our data show apparent signals of diel infection cycles for some of the viral populations (Supplementary Figure S3). For instance, the transcript abundance of cyanophage contigs reached its highest level in the afternoon (i.e., 15:00–18:00), whereas remaining at lower levels during the night. By contrast, the transcript abundance of SAR11 phages was high from 3:00 to 9:00 and at its lowest at 15:00. A recent metatranscriptome study has also shown diel cycling of transcription for 17 cyanophage contigs, one SAR11 phage contig, and eight unassigned phage contigs out of 170 contigs tested [8]. If viral production succeeds their transcriptional activity, viruses showing low genomic but high transcriptomic abundances (Fig. 2) might become abundant in the virome compartment at a later time point. We investigated the existence of such viral diel infection dynamics.

Our sampling site is close to a shore boarding the Kii Channel and water masses at this site fluctuate between low salinity-coastal and high salinity-oceanic waters according to the level of the tide. Therefore, both microbial and viral communities in the time-series samples could be influenced by both diel cycles of communities and water mass

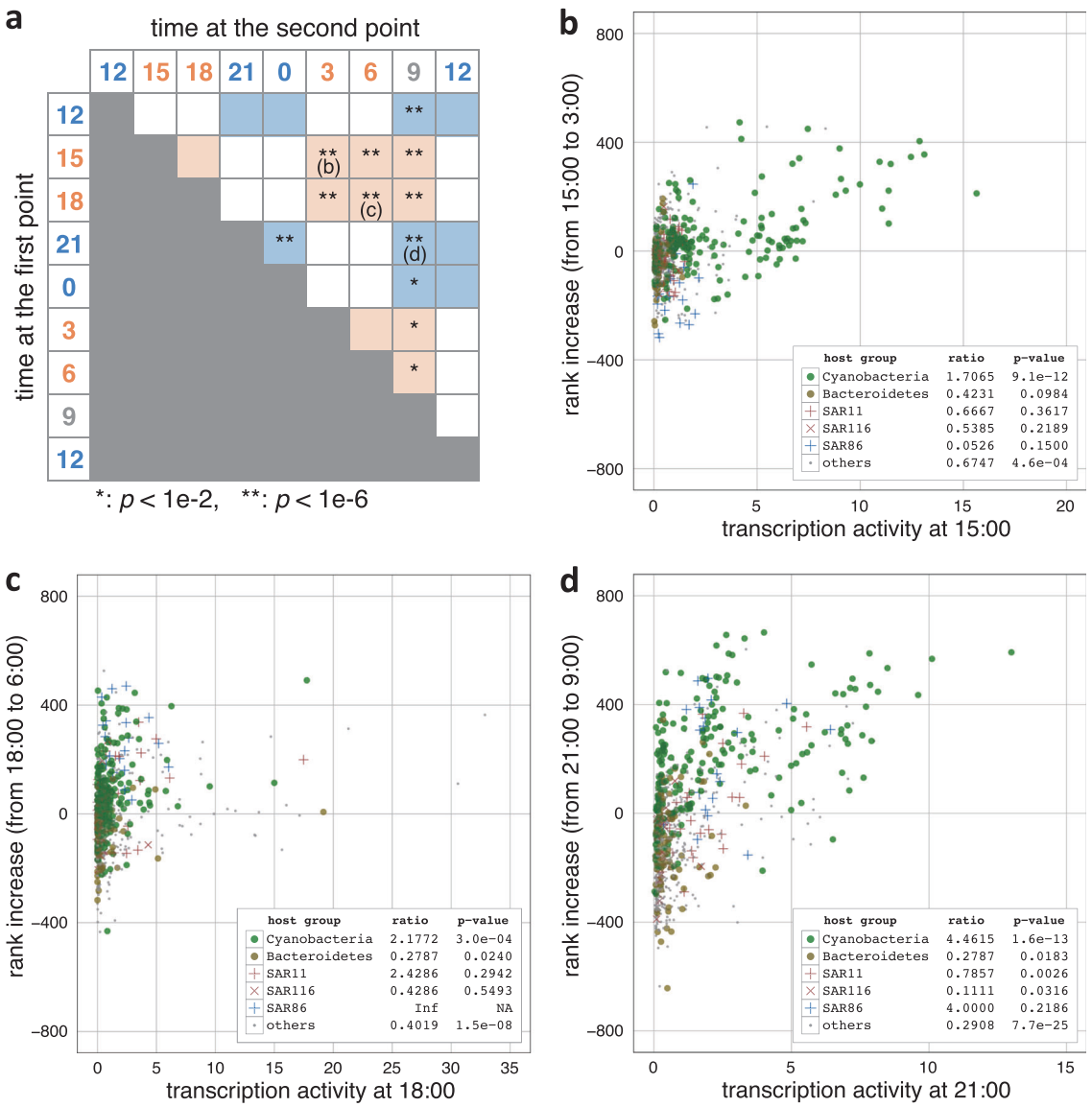

Fig. 3 Infection dependent rank increases in virome abundance. **a** For each pair of different time points, it was investigated whether rank increased viruses show higher transcriptional activity compared with rank decreased viruses. Significance was detected by the Mann–Whitney *U*-tests (asterisks). Colors of sampling times represent sample groups (blue: oceanic, orange: coastal, gray: mixed). The statistical tests were performed within oceanic/mixed samples (areas in blue), and within coastal/mixed samples (areas in orange). **b–d** The transcriptional activities (i.e., metatranscriptome FPKM divided by virome FPKM) at the first time point were plotted against the virome abundance rank increases from the first time point to the second time point. Predicted host groups are color coded as in the panels. Ratios of rank increased viruses to rank decreased viruses, and *p*-values of the tests for each predicted host group are shown in the panels. For **a–d**, statistical analyses and plots were performed on OBV viral genomes whose FPKM values in the virome at the first time point were more than 10. Inf: infinite. NA: not available

movement (i.e., coastal vs. oceanic). We thus evaluated water mass movement by physicochemical parameters and microbial community compositions prior to investigating viral diel infection dynamics.

Salinity, temperature, and nutrients oscillated twice during the day of sampling in response to the rise and fall of the tide (Supplementary Figure S4a,b). Oligotrophic oceanic water thus intruded into the bay as the tide rose and eutrophic coastal water moved outwards as the tide fell. Major microbial community components as well as viral genotype communities also oscillated in response to the mass movement of water (Supplementary Figure S4c,d). These observations suggest that our time-series samples correspond to an oscillation between two distinct host–viral communities (i.e., oceanic and coastal communities).

To minimize the effect of water mass movement, the nine samples were divided into three groups ("oceanic group": 12:00, 21:00, 0:00, and 12:00; “coastal group”: 15:00,

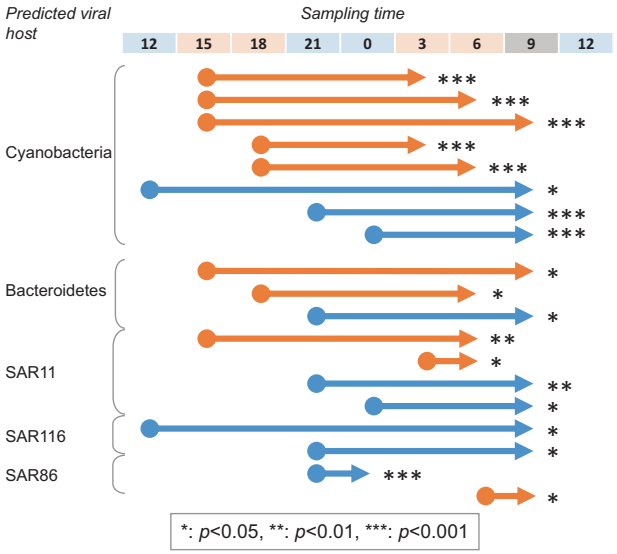

Fig. 4 Summary of infection dependent increases. Different groups of viruses show different timings in their infection dependent rank increases. The tails (indicated by circles) and heads (indicated by triangles) of arrows indicate the two time points in comparisons. Colors of sampling times represent sample groups (blue: oceanic, orange: coastal, gray: mixed). Orange and blue arrows correspond to coastal/mixed and oceanic/mixed waters. For each pair of different time points, it was investigated whether rank increased viruses show higher transcriptional activity compared with rank decreased viruses. Significance was detected by the Mann–Whitney *U*-tests for each predicted host group (asterisks). The statistical tests were performed on OBV viral genomes whose FPKM values in the virome at the first time point were >10

18:00, 3:00, and 6:00; "mixed group": 9:00) and inter-sample comparisons were made within the oceanic/mixed or coastal/mixed groups. We measured increases of virions between two time points by their relative genome abundance rank increases and compared the virion increases with viral transcriptional activities (Fig. 3, Fig. 4, Supplementary Figure S5). We observed a general tendency for viruses with relatively high transcriptional levels at a given time point to increase their genome abundance ranks at a later time point (4 out of 9 for the oceanic/mixed group; 8 out of 10 for the coastal/mixed group) (Fig. 3a). Hereafter, we refer to such time-delayed virion increases that correlate with viral transcriptional activities as "infection dependent virion increases (IDVIs)". Remarkable IDVIs of cyanophages were observed at 3:00–9:00 after their metatranscriptome peak at around 15:00–18:00 in the coastal communities (Fig. 3b, c). Similar cyanophage IDVIs were observed for the oceanic communities from 21:00/0:00 to 9:00 (Fig. 3d, Fig. 4). This dynamic linkage between viral transcription and virion production suggests cell lysis of cyanobacteria during the time period from night to dawn. IDVIs of SAR116 viruses were observed between 12:00 and 9:00 and between 21:00 and 9:00 in oceanic/mixed samples, whereas IDVIs of SAR11 viruses were observed in both oceanic and coastal water masses towards dawn (Fig. 4). Viruses putatively infecting SAR86 showed IDVIs twice during the sampling period (from 21:00 to 0:00 in oceanic water, and from 6:00 to 9:00 in coastal/mixed water). IDVIs were also observed for other viruses including viruses putatively infecting Bacteroidetes (Flavobacteriaceae) (Fig. 4).

Only a few studies have addressed the cyanophage infection cycle in association with the light cycle in natural populations [7]. Kimura et al. [7] investigated the infection dynamics of freshwater *Microcystis* cyanophage populations by monitoring the phage tail sheath *g91* gene and its transcript. They revealed peaks of *g91* transcripts during daylight hours and peaks of *g91* DNA copy numbers in host cell fractions in the afternoon, followed by an increase of *g91* DNA copies in phage particle fractions at night. Another study revealed increases of *Synechococcus* cyanophages in seawater samples at midnight using a culture-based method [29]. Interestingly, a recent study revealed that not only the growth but also the mortality of *Prochlorococcus* showed a remarkable light–dark cycle with a high cell mortality rate at night [30]. The mortality of cyanobacteria has been suggested to be caused by viral infection or grazing. By connecting cyanophage transcriptional activities with increases of free virions (i.e., host lysis), our data suggest that cyanophages represent a major driver of the cyclic mortality of marine cyanobacteria and the release of cyanobacteria-derived organic matter into the environment.

Less-coherent IDVI patterns were observed for viruses putatively infecting heterotrophic bacteria (Fig. 4). A previous study reported that transcriptional activities of heterotrophic bacteria displayed population-specific diel cycling, which was hypothesized to be initiated by the diel release cycle of dissolved organic matter from *Prochlorococcus* [12]. Successions of heterotrophic bacteria were also observed during a longer time span after a eukaryotic algal bloom in a coastal area [31, 32]. It has been suggested that these successions were caused by species-specific utilization of various substrates. Therefore, the rather complicated IDVI patterns for viruses infecting heterotrophic bacteria might originate in viral adaptation to the population-specific life cycles of heterotrophic bacteria with different substrate preferences.

Conclusion

Our cross-omics analysis revealed that the majority of viral genomes detected in viromes were derived from viral particles that were generated through local viral–host interactions. This observation has a significant consequence in the way that we interpret viral particle metagenomes (i.e.,

"viromes"). Namely, our results suggest that viral community structure and dynamics revealed by viromes reflect active virus–host interactions that occur in the same ecosystem. Our results also strongly support the hypothesis that cyanophages drive the diel mortality of marine cyanobacteria. The various IDVI patterns in viruses of heterotrophic bacteria as well as cyanophages will be key for understanding sequential cascades of population-specific host growth and viral infection following the input of organic matter from primary producers.

Acknowledgements We thank Dr. Takuji Yamada for laboratory space and support provided to YN. This work was supported by The Canon Foundation (No. 203143100025), JSPS/KAKENHI (Nos. 16KT0020, 17H03850, 26430184), Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (Nos. 16H06429, 16K21723, 16H06437), and the Collaborative Research Program of the Institute for Chemical Research, Kyoto University (No. 2016-28, 2017-25). PH was supported by the OCEANOMICS "Investissements d'Avenir" program of the French government (No. ANR-11-BTBR-0008). Computational work was completed at the SuperComputer System, Institute for Chemical Research, Kyoto University. We thank the *Tara* Oceans consortium, people, and sponsors who supported the *Tara* Oceans expedition (http://www.embl.de/tara-oceans/) for making the data accessible. This is contribution number 66 of the *Tara* Oceans Expedition 2009–2012. We thank Kate Fox, DPhil, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bergh O, Borsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature. 1989;340:467–8.
2. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. Nature. 2016;532:465–70.
3. Suttle CA. Viruses in the sea. Nature. 2005;437:356–61.
4. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. Nature. 1999;399:541–8.
5. Thurber RV. Current insights into phage biodiversity and biogeography. Curr Opin Microbiol. 2009;12:582–7.
6. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, et al. Environmental viral genomes shed new light on virus-host interactions in the ocean. mSphere. 2017;2:e00359–16.
7. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537:689–93.
8. Kimura S, Yoshida T, Hosoda N, Honda T, Kuno S, Kamiji R, et al. Diurnal infection patterns and impact of Microcystis cyanophages in a Japanese pond. Appl Environ Microbiol. 2012;78:5805–11.
9. Aylward FO, Boeuf D, Mende DR, Wood-Charlson EM, Vislova A, Eppley JM, et al. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. Proc Natl Acad Sci USA. 2017;114:11446–51.
10. Moniruzzaman M, Wurch LL, Alexander H, Dyhrman ST, Gobler CJ, Wilhelm SW. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. Nat Commun. 2017;8:16054.
11. Ozaki K, Uye SI, Kusumoto T, Hagino T. 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. 2004;13:65–79.
12. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R 3rd, Schuster SC, et al. Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science. 2014;345:207–12.
13. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. PLoS ONE. 2014;9:e105592.
14. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27:2957–63.
15. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.
16. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–596.
17. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38:e132.
18. Söding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005;21:951–60.
19. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.
20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
21. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B Stat Methodol. 2001;63:411–23.
22. Brum JR, Sullivan MB. Rising to the challenge: accelerated pace of discovery transforms marine virology. Nat Rev Microbiol. 2015;13:147–59.
23. Iranzo J, Krupovic M, Koonin EV. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. mBio. 2016;7:e00978–16.
24. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, et al. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science. 2015;348:1261498.
25. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. Nature. 2013;494:357–60.
26. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 2005;13:278–84.

27. Lin X, Ding H, Zeng Q. Transcriptomic response during phage infection of a marine cyanobacterium under phosphorus-limited conditions. Environ Microbiol. 2016;18:450–60.
28. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature. 2007;449:83–86.
29. Clokie MRJ, Millard AD, Mehta JY, Mann NH. Virus isolation studies suggest short-term variations in abundance in natural cyanophage populations of the Indian Ocean. J Mar Biol Assoc UK. 2006;86:499–505.
30. Ribalet F, Swalwell J, Clayton S, Jimenez V, Sudek S, Lin Y, et al. Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. Proc Natl Acad Sci USA. 2015;112:8008–12.
31. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol. 2016;1:16005.
32. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608–11.