J. Microbiol. Biotechnol. (2014), 24(5), 592–596
http://dx.doi.org/10.4014/jmb.1312.12062

jmb

Depth-Specific Distribution of the SAR116 Phages Revealed by Virome Binning

Ilnam Kang and Jang-Cheon Cho*

Department of Biological Sciences, Inha University, Incheon 402-751, Republic of Korea

Received: December 24, 2013
Revised: February 10, 2014
Accepted: February 11, 2014

First published online
February 19, 2014

**Corresponding author*
Phone: +82-32-860-7711;
Fax: +82-32-232-0541;
E-mail: chojc@inha.ac.kr

pISSN 1017-7825, eISSN 1738-8872

HMO-2011, a recently isolated lytic phage that infects the SAR116 bacterial clade, represents one of the most abundant phage types in the oceans. In this study, the HMO-2011 genome sequence was compared with virome sequences obtained from various depths of the Pacific Ocean regions using metagenome binning. HMO-2011 was confirmed to be one of the most highly assigned viruses, with a maximum of 7.6% of total reads assigned. The HMO-2011-type phages demonstrated a depth-specific distribution, showing more abundance in the euphotic zone of coastal, transition, and open ocean regions as compared with the dark ocean.

Keywords: SAR116, "*Candidatus* Puniceispirillum marinum" strain IMCC1322, marine virus, HMO-2011, viral metagenome, metagenome binning

The SAR116 clade ("*Candidatus* Puniceispirillum" group) is one of the major bacterial lineages found in the marine euphotic zone [4]. Genome information of two SAR116 isolates showed that they possess diverse metabolic potentials, including methylotrophy and photoheterotrophy [5, 11]. A recent metatranscriptomic study revealed that genes affiliated with IMCC1322, the first isolate of the SAR116 clade, were highly expressed in coastal water microbial communities [3]. All these studies have shown the ecological importance of the SAR116 clade; however, its population dynamics remains to be explored.

Considering that viruses are responsible for about 10–30% of marine prokaryotic mortality [2], studies of viruses infecting major bacterial groups, including the SAR116 clade, are important for understanding the dynamics of microbial communities in marine habitats. Recently, HMO-2011, a phage infecting strain IMCC1322 of the SAR116 clade, was isolated from the East Sea of Korea [7]. HMO-2011, the first lytic phage of the SAR116 clade, was revealed to be a podovirus and had an approximately 55 kb dsDNA genome that encoded many novel proteins, including an atypical methanesulfonate monooxygenase [7]. The HMO-2011 genome recruited a large fraction of the marine virome sequences obtained from the surface water samples of the Indian and Pacific Oceans, suggesting that this SAR116 phage type is one of the most successful viruses in the euphotic zone [7]. In that study, however, the distribution of this phage type along environmental gradients such as depth or proximity to land was not analyzed because of the shortage of available viromes. In the present study, we further investigated the vertical and spatial distribution of virome sequences assigned to the SAR116 phage, using mainly a recently released large-scale virome dataset (POV; Pacific Ocean Virome), which includes samples collected at various depths of the Pacific regions [6]. The virome binning analyses showed that the SAR116 phage type is prevalent in the euphotic zones of both coastal and open ocean regions where members of the SAR116 clade are known to be abundant, whereas it is much less abundant in the aphotic zones where the SAR116 bacterial population is known to be sparse.

All 32 virome sequences analyzed in this study (Table 1) were downloaded from the Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) database [16]. Four viromes from the SPOT (San Pedro Ocean Time series) station were analyzed after trimming and dereplication, as described previously [7]. Twenty-eight recently reported viromes (project CAM_P_0000915) were used without sequence processing [6], and five viromes of the above study originated from the Scripps

Table 1. Binning of viral metagenome reads to reference viral genomes

Sample						Binning					
Accn[a]	Station[b]	Date (2009)	Depth[c] (m)	Temp. (°C)	Total reads	All viruses[d]	Highly assigned viruses[e] 1st	2nd	3rd	4th	5th
002238	LineP:P26	04-Feb	*10*	6.0	192,685	16.2	HMO (3.1)	010P (1.5)	008M (1.5)	GAP32 (0.7)	RaK2 (0.5)
002239			500	3.9	167,616	6.2	HMO (0.3)	008M (0.3)	010P (0.3)	Ehv86 (0.2)	JL001 (0.2)
002240			1,000	2.9	147,537	5.9	008M (0.3)	HMO (0.3)	JL001 (0.3)	YuA (0.2)	010P (0.2)
002241			2,000	1.9	125,896	5.6	008M (0.3)	HMO (0.3)	010P (0.2)	YuA (0.2)	JL001 (0.2)
002242		14-Jun	*10*	9.1	75,036	12.6	008M (1.3)	010P (1.0)	S-SM2 (0.9)	S-CRM01 (0.7)	GAP32 (0.7)
002243			1,000	2.9	101,179	6.3	008M (0.5)	HMO (0.3)	010P (0.3)	Ehv86 (0.2)	S-SM2 (0.2)
002244			2,000	2.0	55,332	10.7	008M (1.1)	HMO (1.1)	GAP32 (0.6)	010P (0.5)	S-SM2 (0.4)
002234		27-Aug	*10*	12.5	165,256	14.9	010P (2.6)	HMO (2.6)	GAP32 (1.1)	008M (1.0)	RaK2 (0.8)
002235			500	3.9	42,118	6.8	HMO (0.4)	Ehv86 (0.3)	008M (0.3)	S-MbCM6 (0.3)	S-SM2 (0.2)
002236			1,000	2.9	70,596	4.7	Gfi (0.4)	HMO (0.2)	Ehv86 (0.2)	008M (0.2)	010P (0.1)
002237			2,000	2.0	68,516	7.1	HMO (0.9)	008M (0.4)	010P (0.4)	Gfi (0.3)	Ehv86 (0.2)
002230	LineP:P12	09-Jun	*10*	11.0	92,415	11.4	008M (1.2)	HMO (0.9)	010P (0.8)	S-SM2 (0.5)	S-SSM7 (0.3)
002231			500	4.3	58,108	9.9	HMO (0.7)	008M (0.6)	S-SM2 (0.4)	010P (0.3)	P-SSM2 (0.3)
002232			1,000	3.3	122,565	7.7	008M (0.5)	HMO (0.4)	Ehv86 (0.2)	010P (0.2)	RM378 (0.2)
002233			2,000	2.0	49,914	11.5	HMO (1.2)	S-SM2 (1.0)	010P (0.8)	P-SSM2 (0.6)	008M (0.5)
002245	LineP:P4	08-Jun	*10*	11.6	107,244	13.8	HMO (1.6)	008M (1.5)	GAP32 (0.9)	RaK2 (0.7)	010P (0.6)
002246			500	5.1	136,876	9.1	008M (0.9)	HMO (0.5)	RM378 (0.3)	HSIC (0.3)	JL001 (0.2)
002247			1,000	3.8	97,126	5.8	HMO (0.4)	008M (0.4)	S-SM2 (0.3)	010P (0.2)	P-SSM2 (0.2)
002248			1,300	3.1	98,478	5.0	HSIC (0.3)	Pr (0.3)	008M (0.2)	HMO (0.2)	JL001 (0.1)
002249	MBARI:67-155	05-Oct	*10*	19.9	203,238	18.5	HMO (7.6)	010P (2.4)	008M (0.8)	YuA (0.4)	S-SM2 (0.4)
002250			*105*	14.2	156,509	24.7	S-SM2 (3.7)	P-SSM2 (2.5)	010P (2.3)	HMO (2.2)	S-SSM7 (1.1)
002251			1,000	3.8	225,833	6.0	HMO (0.6)	008M (0.4)	010P (0.3)	JL001 (0.2)	YuA (0.2)
002252			4,300	1.5	144,588	5.7	HMO (0.5)	008M (0.3)	010P (0.3)	YuA (0.2)	S-SM2 (0.2)
002253	MBARI:67-70	02-Oct	*10*	16.4	321,754	19.7	HMO (5.6)	010P (1.9)	008M (1.5)	S-SM2 (0.9)	019P (0.7)
002254			*42*	11.7	31,528	11.3	HMO (1.9)	010P (0.7)	S-SM2 (0.6)	008M (0.6)	S-TIM5 (0.4)
002255	MBARI:H3	01-Oct	*10*	12.6	303,519	16.2	HMO (3.1)	008M (1.1)	P12053L (0.7)	010P (0.6)	SIO1 (0.4)
000990	SPOT	11-Mar,	*5-30*	12.0-19.2	165,223	22.4	HMO (3.0)	008M (2.4)	S-SM2 (1.3)	010P (1.2)	019P (1.2)
000961		13-May,	150	9.5-9.6	179,975	8.3	HMO (1.0)	010P (0.8)	008M (0.6)	YuA (0.3)	JL001 (0.3)
000962		and	500	6.5-7.0	116,687	8.8	HMO (1.1)	010P (0.7)	008M (0.6)	JL001 (0.4)	YuA (0.4)
001014		19-Aug	885	5.1-5.2	151,740	7.6	HMO (0.7)	010P (0.6)	008M (0.4)	YuA (0.2)	JL001 (0.2)
002228	GBR:F1	10-Oct	*9*	26.2	82,739	20.7	HMO (5.7)	S-SM2 (1.3)	P-SSP7 (1.1)	010P (0.9)	P-SSP2 (0.7)
002229	GBR:TT-3	13-Oct	*8*	26.6	116,855	22.7	HMO (4.5)	S-SM2 (1.7)	P-SSP7 (1.5)	P-SSP2 (1.0)	P-SSM2 (0.9)

[a]Accession numbers as used at CAMERA website (CAM_SMPL_). Note that the two viromes already analyzed in the previous study [7] are also included in this study: 000990 and 002238 (001011 in the previous study). These samples were included as components of depth profiling datasets.

[b]Stations' latitude and longitude: LineP:P26, 50.00N 145.00W; LineP:P12, 48.97N 130.67W; LineP:P4, 48.65N 126.66W; MBARI:67-155, 33.29N 129.43W; MBARI:67-70, 36.13N 123.49W; MBARI:H3, 36.80N 121.85W; SPOT, 33.55N 118.4W; GBR:F1, 16.92S 146.00E; GBR:TT-3, 17.93S 146.14E. See Hurwitz and Sullivan [6] for the station maps.

[c]Depths corresponding to the euphotic zone are indicated in italic. A sample from 150 m depth at the SPOT station is specified in metadata available at the CAMERA database as collected from the sub-euphotic zone.

[d]Proportion of reads assigned to viral genomes, %.

[e]The top five most highly assigned viruses in each sample. The proportion of reads assigned to each virus is indicated within parentheses as (%). Abbreviations for virus names: HMO, HMO-2011; 010P, HTVC010P; 008M, HTVC008M; 019P, HTVC019P; GAP32, *Cronobacter* phage vB_CsaM_GAP32; RaK2, *Enterobacteria* phage vB_KleM-RaK2; Ehv86, *Emiliania huxleyi* virus 86; JL001, Alphaproteobacteria phage phiJL001; YuA, *Pseudomonas* phage YuA; Gfi, *Glypta fumiferanae* ichnovirus; RM378, *Rhodothermus* phage RM378; HSIC, *Listonella* phage phiHSIC; Pr, *Brucella* phage Pr; P12053L, *Celeribacter* phage P12053L; SIO1, *Roseobacter* phage SIO1. All remaining phages (starting with "P-" or "S-") are cyanophages.

pier samples were excluded because they have been analyzed in our previous report [7]. Each of the virome reads was assigned by BLASTx to a best-matching viral protein in the search database comprising the RefSeq viral database (release 57), HMO-2011, and four pelagiphages [19], which infect "*Candidatus* Pelagibacter" belonging to the most abundant SAR11 clade [10], if the alignment satisfied the criteria of bitscore (≥40) and length (≥20 amino acids). All parameters were default except for "-seg no". All reads assigned to viral proteins were checked against the RefSeq Microbial Proteins database of CAMERA and discarded if the best match to a nonviral protein showed a higher bitscore with a length of ≥20. BLASTx at CAMERA was performed with the default parameters, except for the low complexity filter (false), gap open score (11), and gap extend score (1).

Virome binning results showed a depth-specific distribution of SAR116 phages: proportions of reads assigned to the HMO-2011 genome were higher in the euphotic than in the aphotic zone (Fig. 1A). Among the seven depth-profile virome sets that included samples from both euphotic and aphotic zones (Table 1)—with the exception of two sample sets collected at P12 and P26 stations in June—proportions of the HMO-2011-assigned reads in the euphotic zone samples (1.6–7.6%; depth ≤105 m) were significantly higher than those in the aphotic zone samples (0.2–1.1%; depth ≥150 m) ($p < 0.01$, Mann–Whitney U Test) (Table 1 and Fig. 1A). Decrease of the HMO-2011-assigned reads below the euphotic zones corresponds to the distribution pattern of the SAR116 bacteria, which is known to be abundant in the euphotic zone but rarely found in aphotic zones [4, 13, 17, 18]. The similarity between depth-specific distribution profiles of the SAR116 clade and HMO-2011-type phages suggests that SAR116 bacteria may be exposed to infection by the co-occurring phages.

Two other highly assigned viruses, pelagiphages HTVC010P and HTVC008M [19], also showed a depth-specific distribution pattern similar to that of HMO-2011. Proportions of reads assigned to pelagiphages were higher in the euphotic than in the aphotic zone samples (Table 1), confirming the findings of Zhao *et al*. [19].

The depth-specific distribution of these highly assigned viruses seems to have contributed to the overall depth profile of virus-assigned reads. The proportions of reads assigned to viruses were significantly higher in the euphotic (13.8–24.7%) than in the aphotic (4.7–9.1%) zones ($p < 0.01$, Mann–Whitney U Test) (Table 1 and Fig. 1B), as shown in the previous study by a taxonomic assignment of virome reads based on the NCBI taxonomy hierarchy and

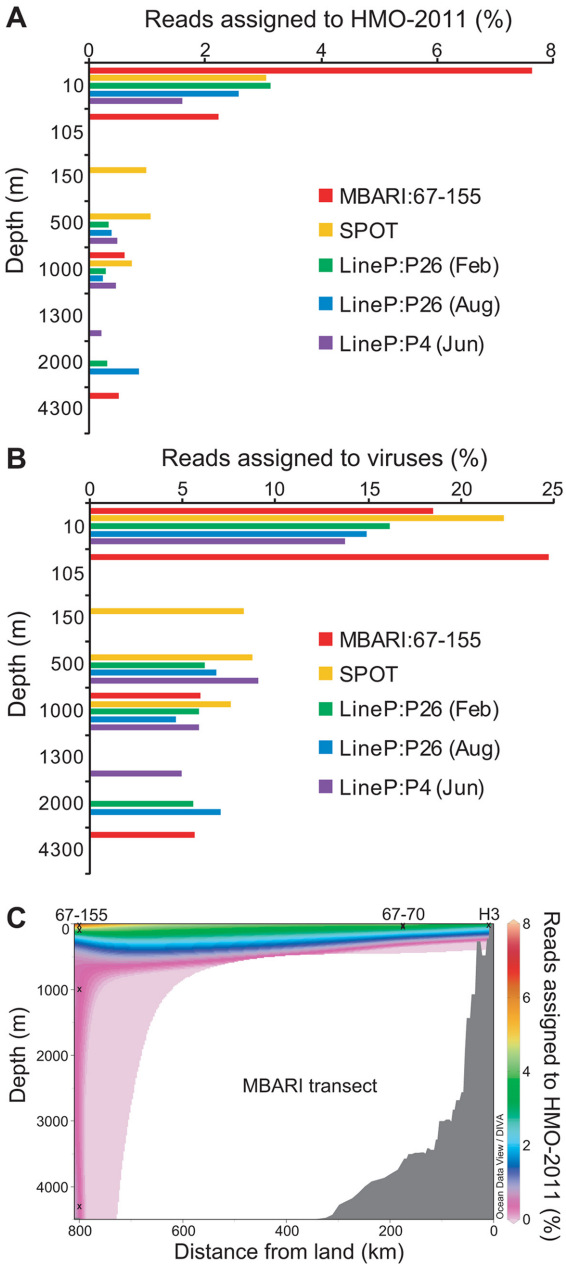

Fig. 1. Distribution of virome reads assigned to the HMO-2011 genome or viral genomes.

(**A**) Proportion of virome reads assigned to the HMO-2011 genome for five depth-profile virome sets. (**B**) Proportion of virome reads assigned to all viral genomes for the same depth-profile virome sets. Note that the data from CAM_SMPL_001014 (885 m at SPOT station) are indicated to be from 1,000 m for presentation purposes. (**C**) Proportion of virome reads assigned to the HMO-2011 genome in the MBARI transect samples. Color shading was performed using Data-Interpolating Variational Analysis (DIVA) gridding implemented in the Ocean Data View software, with default parameters except for X and Y scale-length (100 per mille for both). Symbol "x" denotes stations and depths where viromes listed in Table 1 were obtained.

the SIMAP database [6]. This depth-specific profile may reflect a relative insufficiency of reference viruses from the dark ocean in genome databases, which is caused by a significant lack of cultured isolates affiliated with prokaryotic clades predominantly found in the aphotic zone, such as *Thaumarchaeota*, SAR202, and SAR324.

This study also showed that the virome reads from the large number of Pacific Ocean regions were highly assigned to the HMO-2011 genome, confirming our previous findings obtained from several regions of the Indian and Pacific Oceans [7]. Except for the P26 sample collected in June, the HMO-2011 genome was either the most (n = 9), the second (n = 2), or the fourth (n = 1) most frequently assigned viral genome in the samples collected at the depth of ≤105 m (n = 12). The highest proportion of the HMO-2011-assigned reads (7.6%) was recorded in a surface water sample obtained at the station 67-155 in the MBARI transect (Table 1). Considering that water mass characteristics of this station have been reported to be similar to those of the North Pacific Subtropical Gyre (NPSG) [9, 12], SAR116 phages may be prevalent in NPSG, "the largest circulation feature on our planet" [8]. High proportions of SAR116 phages were also observed in surface samples from two other MBARI transect stations, 67-70 (coastal transition zone) and H3 (coastal upwelling zone)— at 5.6% and 3.1%, respectively (Table 1 and Fig. 1C) [1, 12]. These results suggest the prevalence of SAR116 phages in the euphotic zone along a coastal to open ocean transect, which may reflect the abundance of the SAR116 clade bacteria in the surface waters of various ocean regions, including coastal [14, 15, 17] and open ocean areas [13, 18].

In summary, HMO-2011-type phages were found to be abundant in the euphotic zone of the Pacific Ocean, including coastal and open ocean regions. This distribution pattern is similar to that of the host SAR116 clade bacteria reported from previous studies, which suggests that the dynamics of the SAR116 bacterial population is affected by the infection with the co-occurring phages.

Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2063660: to I.K.) and also by the Mid-Career Research Program through NRF funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A2A2A01068004).

References

1. Collins CA, Pennington JT, Castro CG, Rago TA, Chavez FP. 2003. The California Current system off Monterey, California: physical and biological coupling. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **50:** 2389-2404.
2. Danovaro R, Corinaldesi C, Dell'Anno A, Fuhrman JA, Middelburg JJ, Noble RT, *et al.* 2011. Marine viruses and global climate change. *FEMS Microbiol. Rev.* **35:** 993-1034.
3. Gifford SM, Sharma S, Booth M, Moran MA. 2013. Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J.* **7:** 281-298.
4. Giovannoni SJ, Vergin KL. 2012. Seasonality in ocean microbial communities. *Science* **335:** 671-676.
5. Grote J, Bayindirli C, Bergauer K, Moraes PC, Chen H, D'Ambrosio L, *et al.* 2011. Complete genome sequence of strain HIMB100, a cultured representative of the SAR116 clade of marine *Alphaproteobacteria*. *Stand. Genomic Sci.* **5:** 269-278.
6. Hurwitz BL, Sullivan MB. 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* **8:** e57355.
7. Kang I, Oh H-M, Kang D, Cho J-C. 2013. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc. Natl. Acad. Sci. USA* **110:** 12343-12348.
8. Karl DM. 1999. A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems* **2:** 181-214.
9. Lin Y-C, Campbell T, Chung C-C, Gong G-C, Chiang K-P, Worden AZ. 2012. Distribution patterns and phylogeny of marine stramenopiles in the North Pacific Ocean. *Appl. Environ. Microbiol.* **78:** 3387-3399.
10. Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, *et al.* 2002. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806-810.
11. Oh H-M, Kwon KK, Kang I, Kang SG, Lee J-H, Kim S-J, *et al.* 2010. Complete genome sequence of "*Candidatus* Puniceispirillum marinum" IMCC1322, a representative of the SAR116 clade in the *Alphaproteobacteria*. *J. Bacteriol.* **192:** 3240-3241.
12. Paerl RW, Johnson KS, Welsh RM, Worden AZ, Chavez FP, Zehr JP. 2011. Differential distributions of *Synechococcus* subgroups across the California Current system. *Front. Microbiol.* **2:** 59.
13. Pham VD, Konstantinidis KT, Palden T, DeLong EF. 2008. Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ. Microbiol.* **10:** 2313-2330.
14. Rappé MS, Kemp PF, Giovannoni SJ. 1997. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North

Carolina. *Limnol. Oceanogr.* **42:** 811-826.
15. Rappé MS, Vergin K, Giovannoni SJ. 2000. Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol. Ecol.* **33:** 219-232.
16. Sun S, Chen J, Li W, Altintas I, Lin A, Peltier S, *et al.* 2011. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res.* **39:** D546-D551.
17. Suzuki MT, Preston CM, Béjà O, de la Torre JR, Steward GF, DeLong EF. 2004. Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb. Ecol.* **48:** 473-488.
18. Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA, *et al.* 2009. Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J.* **3:** 1148-1163.
19. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, *et al.* 2013. Abundant SAR11 viruses in the ocean. *Nature* **494:** 357-360.