



# Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean

Francisco Martínez-Hernández<sup>1</sup> · Óscar Fornas<sup>2</sup> · Monica Lluesma Gomez<sup>1</sup> · Inmaculada Garcia-Heredia<sup>1</sup> · Lucia Maestre-Carballa<sup>1</sup> · Mario López-Pérez<sup>3</sup> · Jose M. Haro-Moreno<sup>3</sup> · Francisco Rodriguez-Valera<sup>3</sup> · Manuel Martínez-García<sup>1</sup>

Received: 24 May 2018 / Revised: 31 July 2018 / Accepted: 26 August 2018 / Published online: 18 September 2018  
© International Society for Microbial Ecology 2018

## Abstract

The identification of relevant virus–host pairs that globally account for a large pool of carbon and nutrients in the ocean is paramount to build accurate ecological models. A previous work using single-virus genomics led to the discovery of the uncultured single-virus vSAG 37-F6, originally sorted from the Mediterranean Sea (Blanes Bay Microbial Observatory), that represents one of the most abundant dsDNA viral population in the marine surface virosphere. Here, from same sampling site, we report that a Pelagibacter single-cell contained a viral member of vSAG 37-F6 population, by means of PCR screening of sorted, genome-amplified single cells with vSAG 37-F6-specific primers and whole-genome sequencing. Furthermore, viruses from this population were also found in three other Pelagibacter single cells from the South Pacific and Atlantic oceans. These new uncultured pelagiphages were genetically different from the previously characterized pelagiphage isolates. Data showed that the uncultured vSAG 37-F6 population represents the Pelagibacter phages that inhabit the sunlit ocean better, and contains a vast unrecognized microdiversity.

Our understanding of the biogeochemical impact of marine viruses is limited by our ability of carbon and nutrient transfer measurement in the “viral shunt” and identification of ecologically relevant viral–host pairs in marine foodwebs [1–3]. The former remains virtually unquantified due to technical challenges in natural populations [2, 3]. Concerning the latter, thanks to metagenomics and single-cell genomics, the number of viral and bacterial genomes has

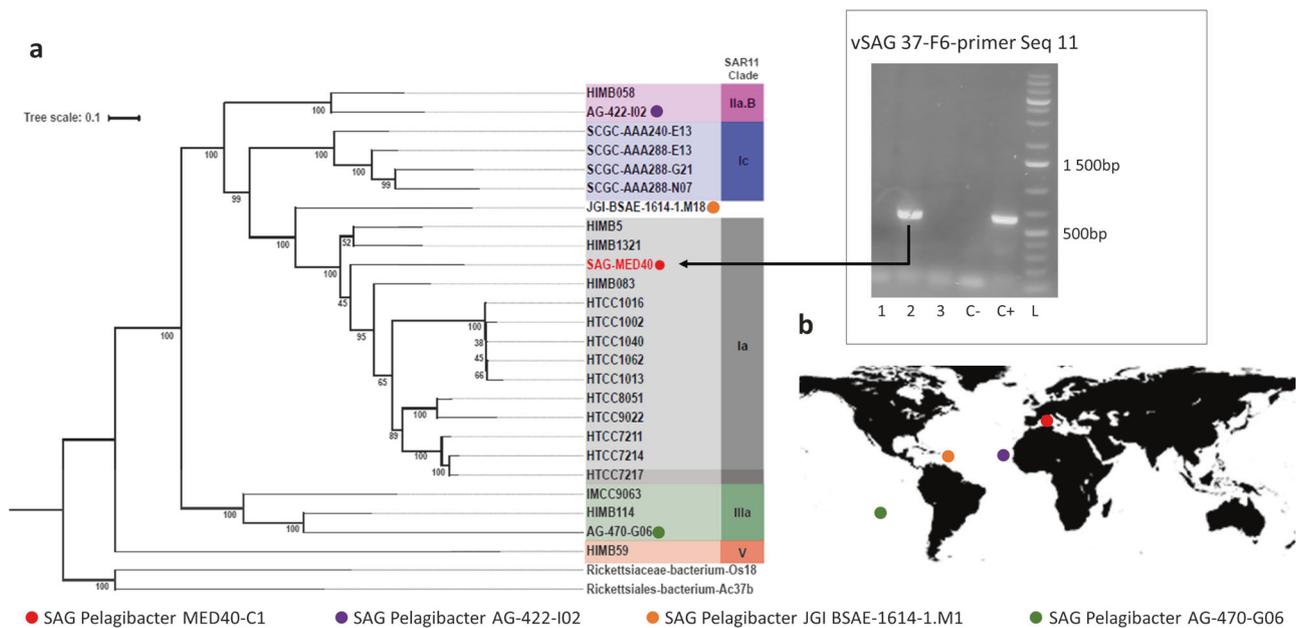
increased exponentially in the last 5 years [4–12], although the experimental identification of virus–host remains elusive and complex to address in natural uncultured populations [13, 14].

Recently, the new emergent single-virus genomic (SVGs) approach [15] in viral ecology has tackled the natural viral diversity by sequencing one virus at a time from different ecosystems [16–18]. In the marine surface, SVGs has enabled the discovery of some of the likely most abundant and ecologically relevant uncultured viral populations, such as the dsDNA single-virus 37-F6 [16]. This virus in particular, sorted from the Mediterranean Sea (Blanes Bay Microbial Observatory (BBMO)), was more abundant in the global marine surface virosphere than any other dsDNA virus previously described. Furthermore, its capsid protein showed the highest rate of proteomic recruitment in the *Tara* viral proteomes [16, 19]. Despite its overwhelming abundance and ubiquity, the viral single-amplified genome (vSAG) 37-F6 remained undiscovered until the advent of SVGs. We now know that the vast natural microdiversity found in the 37-F6 viral population has prevented metagenome assembling [16]. Viral metagenomics yet struggles to assemble genomes from natural

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41396-018-0278-7>) contains supplementary material, which is available to authorized users.

✉ Manuel Martínez-García  
m.martinez@ua.es

- <sup>1</sup> Department of Physiology, Genetics, and Microbiology, University of Alicante, Alicante, Spain
- <sup>2</sup> Flow Cytometry Unit: Pompeu Fabra University (UPF) and Centre for Genomic Regulation (CRG), The Barcelona Institute for Sciences and Technology (BIST), Barcelona, Spain
- <sup>3</sup> Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan, Alicante 03550, Spain



**Fig. 1** *Pelagibacter* single-amplified genomes (SAGs) analyzed in this study. PCR results and screening of SAGs with vSAG 37-F6-specific primers Seq11 are shown. A PCR band with the expected size was obtained for the *Pelagibacter* SAG MED40 (lane 2). Lanes 1 and 3 correspond to other SAGs that did not yield positive amplification. Lanes 4 and 5 are negative and positive (DNA template of 37-F6

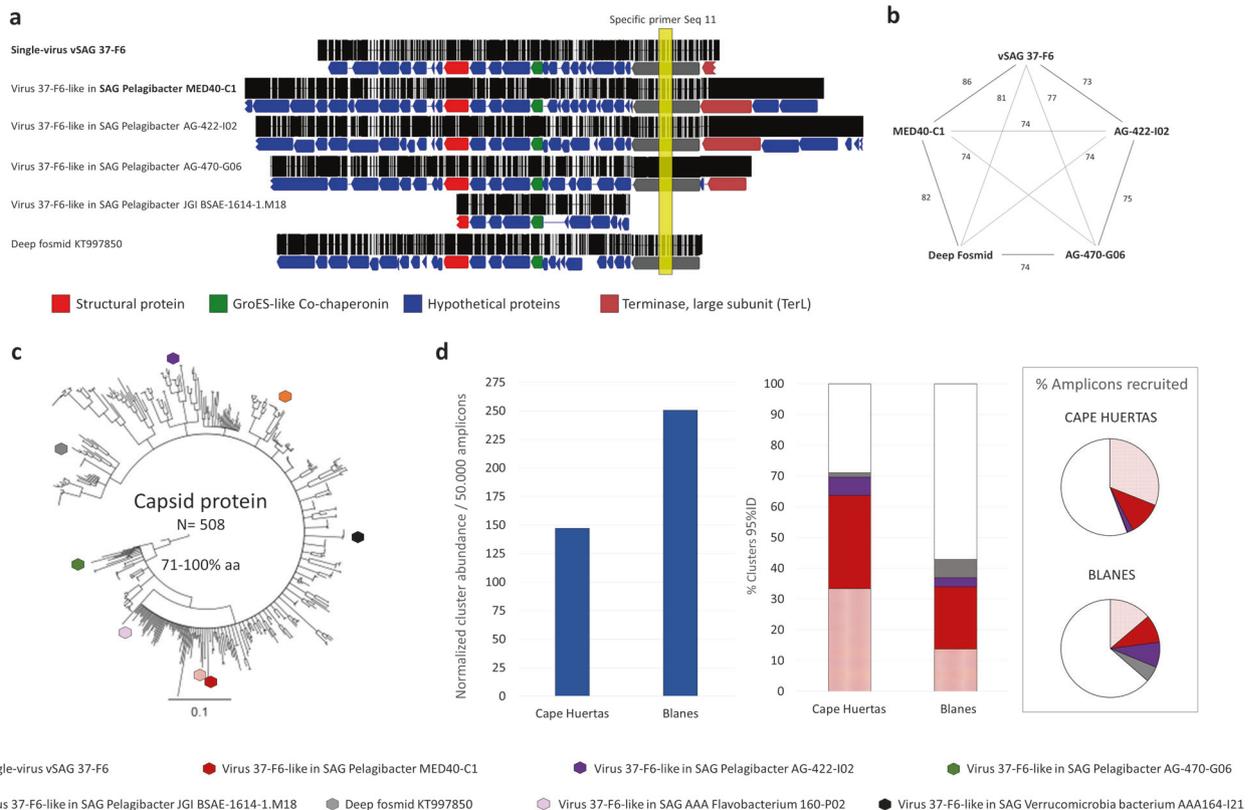
single-virus MDA product) controls, respectively. Maximum-likelihood phylogeny using a concatenation of 25 conserved proteins from SAGs of this study and from SAR11 representative genomes (see also Supplementary Figure S2). Bootstrap values (%) are indicated at each node (a). Circles indicate geographical location of the different SAGs analyzed in this study (b)

viral populations with  $<2\text{--}5\times$  coverage, high microdiversity, and/or uneven abundance [16, 20].

In this study, we use single-cell genomics [13, 21] to unveil one of the mysteries of the relevant virus vSAG 37-F6 in the marine “viral shunt”: the identity of its host. For that, we designed a specific primer set (named as Seq11, Fw: AACACCGATTGCTTCGTACAT, Rv: AGAGGGG TGCGAGTAAGAGA) for a gene of vSAG 37-F6 encoding a hypothetical protein (HP X). We ensured that primers did not match any of the 331,723 viral marine metagenomic contigs and viral genomes available in public datasets (Supplementary Information). Then, from a fresh cell sample taken from the same sampling site (BBMO) at which virus vSAG 37-F6 was originally found, we sorted 1992 single cells (Supplementary Figure 1) that were whole-genome amplified by multiple-displacement amplification, yielding a total of  $\approx 1300$  single-amplified genomes (SAGs). PCR screening and sequencing for the 16S rRNA gene identified a total of 241 SAGs (Supplementary Figure 2), that overall belonged to common marine bacterial groups, such as SAR11 clade ( $n = 79$  SAGs; 3.9% of total sorted cells), SAR116 ( $n = 21$ ) or SAR86 ( $n = 16$ ) (Supplementary Figure 2). PCR screening of SAGs library with vSAG 37-F6-specific primers (primer set Seq11) showed that SAG MED40 (Supplementary Table 1) yielded an expected PCR amplicon (Fig. 1a) that was further confirmed by Sanger sequencing. Subsequently, this SAG was

whole-genome sequenced by Illumina technology and assembled with SPAdes program, yielding a 455.7 kb assembly size (Supplementary Table 2). Genome annotation, 16S rRNA gene and phylogenomic analyses confirmed that SAG MED40 was affiliated to the Pelagibacterales (Fig. 1, Supplementary Figure 3 and Supplementary Tables 3–6). The phylogenomic tree showed that this SAG was close to members of the SAR11 clade Ia (Fig. 1 and Supplementary Figure 3). In fact, the closest genome was a publicly available metagenome-assembled genome (MAG) MED-G39 (GCA\_002457565.1), which could possibly represent the most abundant and endemic *Pelagibacter* in the Mediterranean Sea [22]. The average nucleotide identity (ANI) with the nearest genome (ANI of ca. 75–80%; Supplementary Table 5) indicated that SAG MED40 could belong to a new clade within the Pelagibacterales family.

Genome annotation and analyses of SAG MED40 confirmed the presence of a 20.2 kb viral contig in the assembly (contig name MED40-C1) belonging to the vSAG 37-F6 viral population (Fig. 2). Both, single-virus vSAG 37-F6 and the viral contigs present in the *Pelagibacter* single cell SAG MED40 showed gene synteny and a very high average amino acid similarity (86%) (Fig. 2a, b). The high number of reads of viral origin ( $\approx 86\%$  of total sequenced reads from SAG MED40) compared to bacterial reads from *Pelagibacter* genome suggests that multiple viral genome copies were present in this single cell, which favors the MDA



**Fig. 2** Uncultured pelagiphage population vSAG 37-F6. Genome comparison is shown for the viral members belonging to the uncultured pelagiphage population vSAG 37-F6 recovered by single-cell and single-virus genomics. Black lines in whole-genome alignment denotes homologous genomic regions shared among all viral members. Genomic region (encoding a hypothetical protein) targeted by PCR with the specific viral primer set Seq11 used for SAGs screening is highlighted in yellow (**a**). Average amino acid similarity was calculated by considering 12 orthologous genes shared and present in all genomes. Higher values indicate a closer relationship between the compared viral pairs (**b**). Consensus phylogenetic tree based on neighbor-joining (bootstrapping = 1000) of the signature capsid protein found in the uncultured vSAG 37-F6-like viral population. This viral capsid protein has been shown to be the most abundant in viral marine proteomes from *Tara* expedition. Other homologous proteins ( $n = 502$ ) have been detected in the viral database released by IMG-VR and have been included in the analyses. Protein names are omitted in branches for convenience, except for those capsid proteins from vSAG 37-F6 viral population that are indicated as colored hexagons. vSAG 37-F6 and the genome variant found in the Pelagibacter SAG

MED40 contained nearly identical capsid protein sequences. None of the previously reported pelagiphage isolates have that above mentioned gene encoding a structural capsid protein. Branches with bootstrap values <50% are collapsed. Branches non-collapsed displayed >50% bootstrap value (numbers are omitted for convenience) (**c**). Illumina sequencing results of PCR amplicons obtained with the specific 37-F6-viral primers Seq11 from two environmental viral samples from the Mediterranean Sea. Data showed a vast microdiversity mostly dominated by the viral members of vSAG 37-F6 and that found in the single cell MED40. Total amplicons and amplicon clusters (cut-off 95% of nucleotide identity for putative species demarcation) were compared to uncultured pelagiphages and assigned to the most similar virus according to the higher bit-score. Bars diagram indicates the percentage of amplicon clusters assigned to each of the uncultured pelagiphages. Circle diagrams represent percentage of the total number of amplicons assigned to each one of the uncultured pelagiphages. White regions denoted unassigned amplicons or amplicons clusters. Viral DNA was obtained from Blanes Bay and Cape Huertas (**d**)

amplification and sequencing of the viral genome. As previously discussed [13], this feature along with a low bacterial genome coverage (46%, based on number of recovered orthologous genes by program CheckM) (Supplementary Table 2), as here with Pelagibacter SAG MED40, indicated that this SAG undergo mid-late lytic infection.

Furthermore, data mining of recently released Pelagibacter SAGs at the Joint Genome Institute (JGI-IMG) obtained by independent cell sorting experiments from different

laboratories [23] showed that viral contigs belonging to vSAG 37-F6 population were also present in three Pelagibacter SAGs from the Southern Ocean (SAG AG-470-G06; scaffold IMG-ID name Ga0172187\_101; 54 kb), Sargasso Sea (SAG JGI BSAE-1614-1.M18; scaffold IMG-ID name Ga0171398\_127; 4.9 kb) and North Atlantic (SAG AG-422-I02; scaffold IMG-ID name Ga0172161\_107; 59 kb) (Fig. 1b, Supplementary Table 2). According to their genome sizes, viruses found in the Pelagibacter SAGs AG-470-G06 and AG-422-I02 were likely complete.

Comparative analyses of the recovered partial and complete genomes from SVGs and SCGs belonging to the vSAG 37-F6 viral population showed that all viruses conserved gene synteny and shared a total of 12 orthologous genes with an overall amino acid similarity between each other ranging from 65 to 91% (Fig. 2 and Supplementary Table 7). Genomic data and phylogenetic analyses of the capsid protein from this cluster indicated that vSAG 37-F6 and the variant in the *Pelagibacter* SAG—recovered from the same sample—showed more similarity between each other (97% amino acid similarity) compared to those from other locations (Fig. 2c). The fact that one of the vSAG 37-F6 related viral genomes was cloned in a Mediterranean fosmid (ID KT997850) from 3000 m depth [24], suggests that viral members of this population also infect SAR11 in the deep sea. Among the shared genes of this viral population 37-F6, three had known functions (i.e. “gene signatures”) encoding the above mentioned capsid protein (Fig. 2c), previously characterized to be the most abundant in marine viral proteomes [19], a terminase (TerL), and a chaperonin Gro-Es, recently described to be widespread in marine virosphere and always placed upstream from capsid genes [25], as found in this case. Furthermore, the uncultured pelagiphages in SAG AG-470-G06 and AG-422-I02 had a gene annotated as “genome maintenance exonuclease 1” at JGI-IMG, that was also present in the bacterial genome of 18 different *Pelagibacter* SAGs (Supplementary Table 8), including SAG MED40; all of them from oligotrophic environments. Although the exact metabolic function of that gene is unknown, data suggest that it could be an auxiliary metabolic gene (AMG) involved in the recycling of dNTPs.

Based on previous results from vSAG 37-F6 [16], we speculate that all viral members of this population belong to Podoviridae and apparently they lack integrases and excisionases, suggesting thus a strictly lytic life style, in contrast to pelagiphage isolates HTVC011P and HTVC019P [26]. As previously demonstrated by gene viral-network analyses [16] and genome ANI, vSAG 37-F6 was not genetically related to pelagiphage isolates and indeed belonged to different marine viral clusters [16]. It is interesting to remark that our data showed that pelagiphage isolates [26] lacked those above described shared “gene signatures” and indeed, the TerL and capsid proteins were so divergent that no similarity was detected between the uncultured pelagiphages described here. In fact, the evolutionary genome distance was so high, that the computed ANI was  $\approx 0$ , albeit, at the amino acid level, between the podovirus HTVC010P and the uncultured pelagiphages in SAGs AG-470-G06 (89 ORFs) and AG-422-I02 (88 ORFs), we could find 4 and 7 homologous proteins, that accounted for <8% of total ORFs, with a mean amino acid identity of 49% and 58%, respectively; well below the values obtained

for shared homologs among the uncultured pelagiphage population 37-F6 (65–91%; Supplementary Table 9).

In our previous report on the single-virus vSAG 37-F6, a vast genetic microdiversity was apparent in the viral marine metagenomes [16]. Here, we empirically corroborated and expanded these data by sequencing the PCR amplicons obtained with specific viral Seq11 primers from two environmental viral samples from the Mediterranean Sea: BBMO, where the single-virus vSAG 37-F6 and its putative host were obtained, and the Cape Huertas (Alicante coast, Spain). Amplicons were clustered at 95% nucleotide identity (Supplementary Information), a proposed cut-off for viral species demarcation [27]. The sequencing data of a specific gene (HP X) from the vSAG 37-F6 viral population showed an unexpected high number of variants in Cape Huertas ( $n = 135$ ) and BBMO ( $n = 2029$ ). Results also indicated that precisely vSAG 37-F6 and the virus in SAG MED40-C1 were likely on of the most abundant members within the uncultured pelagiphage population in the Mediterranean Sea (Fig. 2d). Overall, results point to high gene variability that, along with previous metavirome data on genomic heterogeneity of vSAG 37-F6 [16], suggests a vast microdiversity of co-existing similar genotypes within this novel uncultured pelagiphage population. Therefore, although technically feasible, but likely high cost-effectiveness of whole genome sequencing on a large-scale, it would be difficult to find identical viral genomes of vSAG 37-F6 population in the free viral fraction and in infected cells.

In conclusion, in this study, we have unveiled the host identity of vSAG 37-F6 population, and that overall, SVG and SCG data indicate that we have revealed a major uncultured population of highly microdiverse pelagiphages represented by the viral species vSAG 37-F6, as the putative most abundant viral population of dsDNA viruses in the sunlit ocean [16]. Viral taxonomy of uncultured viruses and species and genera demarcation are currently under debate at the ICTV [27]. Thus, whether the uncultured pelagiphages of this viral population 37-F6 belong to the same viral species or genus remains unresolved and is postponed for future investigations, although we anticipate that the latter is most likely. Finally, the fact that two other viruses similar to vSAG 37-F6 were previously found in single cells of Verrucomicrobia and Bacteroidetes [13] rises an interesting question about host range specificity of this ubiquitous and abundant cluster among distantly related bacterial taxa that needs further investigation.

**Acknowledgements** This work has been supported by Spanish Ministry of Economy and Competitiveness (refs. CGL2013-40564-R and SAF2013-49267-EXP), Generalitat Valenciana (refs. ACOM/2015/133 and ACIF/2015/332), and Gordon and Betty Moore Foundation (grant 5334). We thank Ms. Maria Africa McMullen for English editing and review. We thank Sallie W. Crisholm and Stephen Giovannoni for generating a valuable public marine SAGs dataset available through the IMG-JGI platform that has been used in this study.

## Compliance with ethical standards

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

## References

1. Suttle CA. Marine viruses-major players in the global ecosystem. *Nat Rev Microbiol.* 2007;5:801–12.
2. Weinbauer M, Chen F, Wilhelm S. Virus-mediated redistribution and partitioning of carbon in the global oceans. *Microbial Carbon Pump in the Oceans.* Jiao, N., et al. (eds). 2011;54–6.
3. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, et al. Ocean plankton. Determinants of community structure in the global plankton interactome. *Science.* 2015;348:1262073.
4. 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.
5. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. *PLoS Genet.* 2013;9:e1003987.
6. Paez-Espino D, Eloë-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature.* 2016;536:425–30.
7. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun.* 2017;8:15955.
8. Swan BK, Martínez-García M, Preston CM, Sczyrba A, Woyke T, Lamy D, et al. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark. *Ocean Sci.* 2011;333:1296–300.
9. López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodríguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLOS Genet.* 2017;13:e1007018.
10. Martínez-García M, Brazel DM, Swan BK, Arnosti C, Chain PSG, Reitenga KG, et al. Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia. *PLoS ONE.* 2012;7:e35314.
11. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature.* 2013;499:431–7.
12. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol.* 2017;2:1533–42.
13. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* 2015;9:2386–99.
14. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, et al. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat Commun.* 2017;8:84.
15. Allen LZ, Ishoey T, Novotny Ma, McLean JS, Lasken RS, Williamson SJ. Single virus genomics: a new tool for virus discovery. *PLoS ONE.* 2011;6:e17722.
16. Martínez-Hernández F, Fornas O, Lluésma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun.* 2017;8:15892.
17. de la Cruz Peña M, Martínez-Hernández F, García-Heredia I, Lluésma Gomez M, Fornas O, Martínez-García M. Deciphering the human virome with single-virus genomics and metagenomics. *Viruses.* 2018;10:113.
18. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, et al. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat Commun.* 2017;8:84.
19. Brum JR, Ignacio-Espinoza JC, Kim E-H, Trubl G, Jones RM, Roux S et al. Illuminating structural proteins in viral “dark matter” with metaproteomics. *Proc Natl Acad Sci USA.* 2016;113:2436–41.
20. Roux S, Emerson JB, Eloë-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ.* 2017;5:e3817.
21. Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife.* 2014;3:e03125.
22. Haro-Moreno JM, López-Pérez M, de la Torre JR, Picazo A, Camacho A, Rodríguez-Valera F. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome.* 2018;6:128.
23. Berube PM, Biller SJ, Hackl T, Hogle SL, Satinsky BS, Becker JW, et al. Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Sci Data.* 2018;5:180154.
24. Mizuno CM, Ghai R, Saghafi A, López-García P, Rodríguez-Valera F. Genomes of abundant and widespread viruses from the deep ocean. *mBio.* 2016;7:e00805–16.
25. Marine RL, Nasko DJ, Wray J, Polson SW, Wommack KE. Novel chaperonins are prevalent in the viroplankton and demonstrate links to viral biology and ecology. *ISME J.* 2017;11:2479–91.
26. 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.
27. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, et al. Consensus statement: virus taxonomy in the age of metagenomics. *Nat Rev Microbiol.* 2017;15:161–8.