Efficient Dilution-to-Extinction isolation of novel virus-host model systems for fastidious heterotrophic bacteria.

Holger H. Buchholz[1], Michelle L. Michelsen[1], Michael J. Allen[1,2], Ben Temperton[1*]

[1]University of Exeter, School of Biosciences, Exeter, UK;
[2]Plymouth Marine Laboratory, Plymouth, UK

*Correspondence

Ben Temperton

b.temperton@exeter.ac.uk

Running Title: High-throughput cultivation of viruses of fastidious bacteria

Keywords: Dilution-to-Extinction, bacteriophages, high-throughput culturing, marine microbiology, virus-host systems, Western English Channel

Abstract

Microbes and their associated viruses are key drivers of biogeochemical processes in marine and soil biomes. While viruses of phototrophic cyanobacteria are well-represented in model systems, challenges of isolating marine microbial heterotrophs and their viruses have hampered attempts to study the Earth's most abundant virus-host systems. A recent resurgence in cultivation efforts has improved the cultivability of fastidious bacteria for hypothesis testing and synthetic community ecology. Yet, these improvements have not been matched by similar efforts to cultivate their associated bacteriophages. Here, we describe a high-throughput method for isolating ecologically important virus-host systems for fastidious heterotrophic bacteria, that couples advances in Dilution-to-Extinction culturing of hosts with sequential enrichment and isolation of associated phages, bringing viral cultivation efforts in line with those of their cellular hosts. Applied to five monthly samples from the Western English Channel, we successfully isolated 117 new phages including the first known siphovirus infecting SAR11 and the first known phages infecting members of the methylotrophic bacterial clade OM43. Genomic analyses of 13 novel viruses revealed representatives of three new viral genera. Similar to the abundant human-associated phage ϕCrAss001, infection dynamics within the majority of isolates suggested either prevalent lysogeny, irrespective of detectable associated genes, and/or host phenotypic bistability, with lysis putatively maintained within a susceptible subpopulation. Broader representation of virus-host systems in culture collections and genomic databases will improve both our understanding of virus-host interactions, and accuracy of computational approaches to evaluate ecological patterns from metagenomic data.

Introduction

It is estimated that viral predation kills ~15% of bacterial cells in marine surface water each day [1] and it has been shown to be a major contributor to nutrient recycling via the viral shunt, where marine viruses make cell-bound nutrients available to the neighbouring microbial community through viral lysis of host cells [2, 3]. Viruses are key players in the modulation of carbon fluxes across the oceans (150 Gt/yr), increasing particle aggregation and sinking to depth [2, 4], and accounting for 89% of the variance in carbon export from surface water to the deep ocean [5]. Viruses alter host metabolism through auxiliary metabolic genes (AMGs), increasing and altering the cellular carbon intake of infected cells [6]. Virus-host interactions also increase co-evolutionary rates of both predator and prey, via Red Queen dynamics [7, 8]. While recent metagenomic advances have provided major insight into global viral diversity and abundance [9–12], mechanistic understanding of virus-host interactions in ecologically important taxa is still reliant on experimental co-culturing of model systems. In cyanobacteria, such systems have shown that viruses increase duration of photosynthetic function [13] and can inhibit CO_2 fixation, providing direct evidence that viruses of abundant phototrophs have a direct and major role in nutrient regulation and global carbon cycling [14]. Furthermore, isolation of novel viral model systems provides complete or near-complete viral genomes with known hosts, enabling the development, ground-truthing and application of computational methods to identify viral genomes in metagenomic data (e.g. VirSorter [15]; VirFinder [16]; MARVEL [17]); quantify boundaries for viral populations [12, 18] and genera (VConTACT2 [19]); understand the importance of AMGs in altering nutrient flux in natural communities [20] and to predict host ranges of the vast diversity of viral genomes that remain to be cultured (e.g. WIsH [21, 22]).

However, while cyanophage model systems are numerous and have been well-studied in the laboratory, virus-host model systems for similarly important and abundant marine heterotrophic bacteria are rare. Current model virus-host systems are heavily biased towards those with a fast-growing copiotrophic host that grows readily on solid agar, enabling the use of plaque assays to isolate associated viruses. In contrast, nutrient-limited environments such as many aquatic and soil biomes are dominated by slow-growing, oligotrophic taxa with few regulatory mechanisms and complex auxotrophies that do not lend themselves to growth on solid or rich medium [23–26]. Advances in Dilution-to-Extinction culturing of important hosts have enabled the cultivation of many fastidious bacterial taxa that are not amenable to growth on solid media from soil, marine [27, 28] and freshwater environments [29]. However, isolation of associated viruses is challenging due to long host doubling times; complex nutrient requirements; challenges of identifying positive infection using plaque assays or epifluorescent microscopy to detect the release of viral progeny, and low efficiency of successful viral isolation and purification. These challenges were overcome with the recent isolation of phages associated with the ubiquitous marine chemoheterotrophic SAR11 clade [30, 31] and revealed them to be among the most abundant viruses on Earth [30], with novel genomes that had previously been missed in metagenomic studies due to challenges of assembling abundant microdiverse populations [12, 18, 32]. Isolation of model virus-host systems for ecologically important, but fastidious taxa both within and beyond the SAR11 clade require a more efficient approach. Otherwise our experimentally-derived understanding of virus-host interactions and our computational tools used to evaluate viral ecology in natural communities will remain biased by limitations of culturing model systems for ecologically representative taxa.

Here, we coupled recent advances in Dilution-to-Extinction culturing of hosts [28] and previous methods to isolate viruses from liquid media [33] with two adaptations to improve efficiency of recovering novel virus-host systems for fastidious taxa: (i) Sequential enrichment of viruses from natural communities on target hosts to improve rates of viral isolation [34]; (ii) replacement of time-intensive epifluorescent microscopy with comparative flow cytometry between infected and uninfected hosts to screen for putative viral infection, followed by confirmation using Transmission Electron Microscopy (TEM). We tested the efficacy of our protocol for isolating new virus-host systems for fastidious taxa, focusing on two model systems: (i) New representatives of phages infecting the ubiquitous SAR11 clade from the Western English Channel; (ii) the first phage isolates infecting the important and similarly fastidious methylotrophic OM43 clade. Novel hosts were isolated using Dilution-to-Extinction methods and subsequently used as bait to isolate associated viruses using our workflow. Initial concentration of viruses in inocula by tangential flow filtration, followed by one to three rounds of sequential enrichment on target hosts in 96-well plates yielded 117 viral isolates from 218 inoculated cultures from six water samples spanning a five-month period. Isolation of novel SAR11 viruses had an efficiency of ~90% (94 out of 105 inoculated cultures), including *Kolga* the first known siphovirus infecting SAR11. In comparison, efficiency in OM43 was ~20% (23 out of 113 inoculated cultures).

Results and Discussion

Isolation of a novel SAR11 strain and three new OM43 strains from the Western English Channel to use as bait for phage isolation.

As a proof of concept, we selected the ecologically important SAR11 clade as example fastidious taxa, for two reasons: First, viruses have previously been isolated on this model system [30, 31, 35] and growth conditions are well-established [36, 37]; Second, metagenomic analyses of marine viromes have shown that viruses infecting the SAR11 clade are highly abundant and genetically diverse [12, 38]. OM43 was selected as a second example as a streamlined and globally-distributed heterotroph with challenging culturing requirements; known ecological importance in coastal environments; and, importantly, lacking any current isolated virus-host model systems [39, 40]. Members of OM43 typically have much lower abundance compared to SAR11 in marine systems, but have complex population dynamics associated with phytoplankton blooms [40–42]. Dilution-to-Extinction culturing for host taxa using natural seawater-based medium on three 96-well plate samples in September 2017 yielded a new SAR11 strain (H2P3α) and three new OM43 strains (C6P1, D12P1, H5P1) from the Western English Channel. Phylogeny of full-length 16S rRNA genes for each isolate revealed H2P3α to be most closely related to the warm water SAR11 ecotype *Pelagibacter bermudensis* HTCC7211 (subclade 1A) (Fig. S1A) with a nucleotide identity of 100%, and was considered a local HTCC7211 variant. All three novel OM43 isolates (named CP61, D12P1, H5P1) were most closely related to *Methylophilales bacterium* HTCC2181 (C6P1 95.70%, D12P1 96.20%, H5P1 95.7% average nucleotide identity (ANI) across the full 16S rRNA gene), a streamlined member of the OM43 clade with a 1.3 Mbp genome, isolated from surface water of the North Eastern Pacific [43] (Fig. S1B). The ANI of the 16S rRNA gene of these isolates to each other was ~97.7%, suggesting they are representatives of the same genus [44].

117 new viral isolates for SAR11 and OM43 taxa. Using the four new hosts and established SAR11 isolates *Pelagibacter ubique* HTCC1062 (cold water ecotype) and HTCC7211 (warm water ecotype), we developed an optimised viral isolation pipeline (Fig. 1) and applied it to five monthly water samples from the Western English Channel, taken between October 2018 to April 2019. Additional samples were taken in October 2018, to attempt to isolate phages on HTCC1062 as a proof of concept, and July 2019 to attempt viral isolation on OM43 strains D12P1 and H5P1 to attempt isolation of additional methylophages. In total, out of 218 cultures inoculated with concentrated natural viral communities, 117 viruses were isolated and still infective after at least three passages, demonstrating an overall efficiency of 53% and an average yield of 18 isolates per sample (Fig. 2). All viral isolations required between one and three rounds of host enrichment (see methods) before changes in host growth curves between infected and uninfected cultures could be observed by flow cytometry, suggesting a minimal viral density is required to identify infected wells using this approach. Efficiencies varied between different hosts: For SAR11 the success rates were: 80% for HTCC7211, 93% for HTCC1062 and 96.6% for H2P3α. The increased efficiency in H2P3α can possibly be attributed to matching geography for both host and virus isolation, with the host being more susceptible to infection from viruses within the Western English Channel. This suggests an advantage of isolating relevant strains from the same locale to improve isolation efficiency. A linear model of isolation efficiency of viruses for ecotypes of SAR11 across sampling months showed no influence of temperature, nor any significant interaction terms between ecotype and water temperature, which

152 ranged from 9.3 °C - 17.3 °C. Thus, across this temperature range, thermal ecotype selection did not
153 significantly contribute to viral isolation efficiencies across warm and cold-water host ecotypes.
154 Different viruses were isolated on cold (HTCC1062) compared to warm (H2P3α) ecotypes,
155 suggesting that continued persistence of ecotype-specific viruses throughout the sampling period
156 better explained isolation success across both host ecotypes than broad viral host-range.
157
158 Efficiencies of viral isolation on OM43 strains ranged from 0% (no viruses isolated for C6P1) to 45%
159 for H5P1. To the best of our knowledge these are the first reported isolates of viruses infecting
160 members of the OM43 clade. One explanation for the lower efficiency of isolation of OM43 viruses
161 is simply one of lower host abundance concomitant with lower phage abundance in the viral
162 community, reducing the likelihood of infective viruses coming into contact with susceptible and
163 permissive cells. OM43 are closely associated with metabolism of extracellular substrates from
164 phytoplankton blooms [45], yet are of low abundance outside of phytoplankton spring blooms [41].
165 Our water samples were not associated with high *in-situ* fluorescence, and just missed the April 2019
166 Spring bloom (Fig. S7). Therefore, OM43 and their viruses were likely of low abundance at sampling
167 times, decreasing the success rate of viral isolation from these samples. Following three rounds of
168 purification through Dilution-to-Extinction, we subselected 16 of the 44 viral isolates available in
169 November 2018 across four different hosts (HTCC1062, HTCC7211, H2P3 and H5P1) for
170 sequencing at 30x coverage to maximise viral diversity. Three out of 16 samples (two from host
171 HTCC7211, one from host H5P1) failed to assemble into single viral contigs. Eleven of the remaining
172 13 (twelve from SAR11 hosts and one from OM43) individual assemblies were identified as complete
173 viral genomes by VirSorter [15], which returned a single viral genome (categories 1 or 2, >15kbp)
174 per sample, indicating that our purification process was effective in recovering pure viral isolates.
175
176 Our protocol required ~7 hours of handling time per sample for a complete 96-well plate. In addition,
177 each sample required ~8 hours run time on a 96-well plate flow cytometer. Due to the slow growth
178 rate of these fastidious taxa (doubling times of ~36 hours and 30 hours for SAR11 strains and OM43
179 strains, respectively), incubation times in between initial isolation from environmental water samples
180 and three rounds of purification were ~10 weeks. The most time-consuming step of our pipeline
181 occurred post-isolation in the generation of sufficiently high viral titres to generate enough DNA for
182 sequencing (approximately two weeks incubation time, and roughly four hours handling time over
183 two days). We estimate the cost of isolating a single virus is ~£20 for cultivation, flow cytometry and
184 DNA extraction consumables, and ~£50-100 for genome sequencing at 30-fold coverage required for
185 successful assembly, giving a total cost of £70-120 per sequenced viral isolate. Thus, our protocol
186 facilitates a high-throughput and scalable approach to viral isolation, with minimal personnel time
187 and high rates of efficiency.
188
189 ***Phage isolates belong to novel viral populations.*** The 12 sequenced pelagiphage genomes from this
190 study were clustered into viral populations with 85 other SAR11 viruses (isolates as well as selected
191 metagenomically-derived viral contigs) [31, 38, 46] at 95% ANI over 85% length [18] using
192 ClusterGenomes v5.1 (https://github.com/simroux/ClusterGenomes). Viral sequences obtained from
193 our isolates clustered into six viral populations, ranging from singletons to a population with four
194 members. Phages within the same population were all isolated from the same source water sample on
195 the same host, with the largest cluster (*Bylgja*, four isolates) isolated from a water sample taken in
196 October 2018 using H2P3α as host. Intrapopulation ANI ranged from 99.58-100% (calculated using

197 the OrthoANI algorithm [47]), indicating that some viruses had been isolated more than once, and
198 other populations contained microdiverse members (Table S9). No viral populations isolated from
199 the Western English Channel contained representatives from either known isolates [30, 31] or fosmid-
200 derived [38] genomes from other studies, despite high recruitment of metagenomic reads to their
201 genomes from a previous virome study in this location [12] (Fig. 3D). Curiously, we did not isolate
202 viruses that shared a population assignment with the most ubiquitous and abundant pelagiphages
203 HTVC010P [30] or vSAG-37-F6 [48], both of which were highly abundant in the Western English
204 Channel virome. For vSAG-37-F6, a possible explanation could be that its preferred host was not one
205 of the strains we used for viral isolation. However, HTVC010P was the most abundant pelagiphage
206 in the Western English Channel and was previously isolated on host HTCC1062 also used in this
207 study. Another possibility is that HTVC010P was isolated, but then subselection of representatives
208 for sequencing missed it by chance.
209
210 Representatives of each of our isolated viral populations as well as known isolates and fosmid-derived
211 phage sequences were clustered into ICTV-recognised genera using vConTACT2 [19] and by
212 hypergeometric testing of shared-gene content (Fig. S8). The six viral populations from this study
213 clustered into four main clusters and one singleton based on shared-protein content (Figure 3). The
214 largest cluster of which (Cluster A) contained 23 members (nine from fosmid-derived contigs [38];
215 eleven previously isolated pelagiphages [31, 49] and the viruses *Ran*, *Bylgja* and *Eistla* from this
216 study). Cluster B contained two previously isolated pelagiphages, one fosmid-derived contig and
217 *Eyrgjafa* from this study. According to VConTACT2, all viruses in Clusters A and B belonged to a
218 single viral genus-level taxon that also contained 12 previously isolated pelagiphages [30, 31]. No
219 known isolates, either from this or previous studies were assigned to Cluster C, which only contained
220 fosmid-derived contigs from the Mediterranean [38]. This may indicate that members of this cluster
221 require a different host than has previously been used in isolation experiments, or was missed by our
222 sequencing by chance. Cluster D contained eight fosmid-derived contigs, the pelagiphage
223 HTVC010P, and one new representative (*Greip*) from the Western English Channel (Figure 3C).
224 *Greip* and *Kolga* were the only members in their respective VConTACT2 assignments, hence are
225 likely the first representatives of novel viral genera.
226
227 Competitive mapping of metagenomic reads from a Western English Channel virome [12] against
228 existing pelagiphage genomes and those from this study using FastViromeExplorer [50] identified
229 HTVC010P as the most abundant virus within the known pelagiphages (536,531 Reads Per Million
230 [RPM - see methods]). Other isolates from this and previous studies recruited 1-3 orders of magnitude
231 fewer reads than HTVC010P (684 - 53,591 RPM)(Fig. 3D), indicating that viral community
232 structures for pelagiphages in the Western English Channel support the Bank Model, which proposes
233 that viral communities consist of a small number of viral species that are highly abundant, and long
234 tail of low abundance members [51].
235
236 ***Kolga - the first siphovirus infecting SAR11***. Most marine bacteriophages fall into one of four
237 different groups: non-tailed, podoviruses, myoviruses and siphoviruses. Members of these groups are
238 shown to be highly diverse and abundant in all marine environments [52, 53]. Yet, out of 25
239 previously known viral isolates infecting SAR11, 24 were podoviruses and one myovirus [30, 31,
240 35]. Previous cultivation efforts for viruses of SAR11 has not isolated any siphoviruses, nor are any
241 known from viral metagenomic studies. We isolated and sequenced the two viruses *Kolga* and *Aegir*

using H2P3α as bait, which appear to have a long tail on transmission electron microscopy (TEM) images, indicative of a siphovirus (Fig. 3B). *Aegir* and *Kolga* were members of the same population using a boundary cutoff of 95% ANI over 85% contig length, however, *Aegir* had a length of 18,297 bp compared to 48,659 bp in *Kolga*, therefore we considered *Aegir* to be a partial genome of the same viral population. *Kolga* did not share a significant number of genes with known SAR11 podoviruses (Fig. 3B), and did not cluster with other pelagiphages using a hypergeometric analysis based on gene content. VConTACT2 also grouped *Aegir* and *Kolga* into one cluster without any other known viruses, suggesting it represents a novel viral genus. A number of genes found in *Kolga* were shared with other known siphoviruses, and phylogenetic analysis of concatenated genes shows *Kolga* is most closely related to LK3, a known siphovirus infecting the Betaproteobacterial genus Bordetella (Fig. 4C). This suggests that *Kolga* is more closely related to siphoviruses of different hosts compared to podoviruses infecting the same host. 67% of genes encoded by Kolga could not be functionally annotated, and out of all hypothetical genes identified on *Kolga*, only three hypothetical genes were shared with SAR11 podoviruses. In contrast, for on average ~90% of genes without known function identified within our novel SAR11 podoviruses are shared between different pelagiphages. *Kolga* was the 10th most abundant viral isolate in our Western English Channel virome (Fig, 3D). *Kolga* possesses a tail tip J protein (Fig. 4A), often found in phages with long non-contractile tails such as λ, where it plays a role in DNA injection during cell entry and tail assembly [54]. *Kolga* also encodes a small S21 subunit of the 30S ribosomal gene structurally similar to the ones found in HTVC008M, and hosts HTCC7211 and H2P3α, a feature found in numerous myoviruses and sipoviruses [55]. S21 is involved in translation initiation and needed for the binding of mRNA [56]. Virally-encoded S21 genes may provide a competitive advantage for the phage as it could replace cellular S21 and assist in the translation of viral transcripts. *Kolga* may need its S21 gene for shifting the translational frame, as it has been shown that for some members of the Caudovirales the production of tail components is dependent on programmed translational frameshifting [57].

Venkman - the first known virus infecting OM43. Application of our method in the Western English Channel resulted in the first isolation of a bacteriophage infecting the methylotroph OM43. Named *Venkman*, in homage to the actor Bill Murray, this virus has a genome 38,624 bp long (31.9% GC content), encoding genes with similar synteny and function to the siphovirus P19250A (38,562 bp) that infects freshwater *Methylophilales* LD28, which are often considered a freshwater variant of OM43 [58, 59]. TEM images indicated a short tail (Fig. 5) similar to podoviruses, though it is possible that tail structures had been lost during grid preparation.

Phylogenetic analysis of concatenated TerL and exonuclease genes further indicated that these genes in *Venkman* are most closely related to other siphoviruses infecting different Proteobacteria (Fig. 5). We identified a number of common phage proteins such as a capsid protein, terminase, nucleases and tail structural proteins, the remaining 54% of genes were hypothetical. *Venkman* and P19250 appear to have similar genomic synteny and function but with an average ANI of >50% over 80% of length for all genes and an average 31% amino acid identity over 23% query length. VConTACT2 assigned *Venkman* and P19250 to the same genus-level cluster. Based on this *Venkman* may be a marine variant of the freshwater phage P19250. In the Western English Channel viral metagenomes, *Venkman*-like viruses were the second most abundant viruses (148,681 RPM) after HTVC010P (536,531 RPM) identifying them as ecologically important viruses in this coastal microbial community.

287 ***Unusual host-virus dynamics are prevalent in isolated phages*** Viral isolates *Bylgja*, *Eyrgjafa*, *Ran*
288 and *Eistla* (Cluster A and B) are presumed to be temperate phages, with all encoding endonucleases
289 and exonucleases (Fig. S2) and clustering by shared protein content with other pelagiphages shown
290 previously to integrate into host genomes (Fig. 3)[31]. *Eyrgjafa* encodes a tRNA-Leu, that has 85%
291 nucleotide identity over the first 34 bases of the tRNA-Leu of its host HTCC1062, suggesting a
292 putative integration site into the host genome [60]. To date, 16 of the 29 viruses isolated on SAR11
293 strains have either been shown to be capable of lysogeny, and/or encode genes associated with a
294 temperate infection cycle. In contrast, viruses such as *Greip,* and the abundant HTVC010P in Cluster
295 D (Fig. 2C) do not possess any genes associated with lysogeny, and would therefore presumed to be
296 exclusively lytic. Viruses in Clusters A and B, presumed temperate, were of much lower abundance
297 in the environment compared to *Greip* or HTVC010P in Cluster D which were among the most
298 abundant pelagiphages in the Western English Channel (Fig 3D), suggesting a possible ecological
299 difference between two groups of viruses with different infectivity strategies.
300
301 Interestingly, growth curves of hosts infected with *Greip* and other isolated phages deviated from the
302 expected decay in cell abundance associated with viral lysis and previously observed in isolated
303 pelagiphages [30]. In the initial isolation of the first pelagiphages HTVC010P and HTVC008M were
304 isolated from the warm waters of the Sargasso Sea, whereas HTVC011P and HTVC019P were
305 isolated from the colder waters of the Oregon Coast. All four strains were propagated on the cold
306 water SAR11 ecotype *P. ubique* HTCC1062. In all cases, host density was reduced from ~8×10^6
307 cells per mL at T_0 to <10^6 cells per mL over a 60-72h period, although viruses from warm waters took
308 17% longer than those from cold water to do so [30], suggesting that suboptimal hosts reduced the
309 rate of infection as shown in cyanophages [61], but not the nature of the interaction. In contrast,
310 infection dynamics of our isolates often resulted in host density of infected cultures growing to a
311 steady state, but at a lower cell density than uninfected cells. (e.g. Fig. 1Bii, Fig. S11), irrespective of
312 cluster or population assignment. Out of 117 viruses isolated in this study, only 16 infections reduced
313 host abundance below their inoculum density of 10^6 cells per mL. In 53 infections, densities of
314 infected cells increased to within an order of magnitude of uninfected cells (Fig. S12), but
315 demonstrated clear evidence of viral infection in cytograms, TEMs and subsequent recovery of viral
316 genomes in selected samples.
317
318 Similar patterns of infection were recently reported in the extremely abundant bacteriophage
319 ϕCrAss001 found in the human gut, where infection of exponentially growing cells of *B. intestinalis*
320 919/174 did not result in complete culture lysis, but caused a delay in stationary phase onset time and
321 final density, despite lacking genes associated with lysogeny. As with our study, the authors observed
322 that this only occurred in liquid culture, and isolation of the virus required numerous rounds of
323 enrichment They postulated that the virus may cause a successful infection in only a subset of host
324 cells, with the remainder exhibiting alternative interactions such as pseudolysogeny or dormancy
325 [34]. The prevalence of similar infection dynamics in the phages isolated in this study offer two
326 intriguing possibilities: (i) Many of the viruses isolated in this study are either not fully lytic, but fall
327 somewhere on the continuum of persistence [62], controlled by genes currently lacking a known
328 function, with lysogeny (and associated superinfection immunity) favoured at high cell density, in
329 support of the Piggyback-the-Winner hypothesis [63, 64]; (ii) the steady-state of host and virus
330 densities observed here are a function of host phenotypic bistability in these model streamlined
331 heterotrophic taxa, with viral propagation occurring in a subset of cells, potentially explaining the

requirement of multiple rounds of enrichment before sufficient viral load is reached to observe the effects of lytic infection on the host population. Either strategy, or a combination of both, would provide an ecological advantage of long-term stable coexistence between viruses and hosts and offer an explanation of the paradox of stable high abundances of both predator and prey across global oceans [12, 30, 48, 65]; the low lytic activity observed in pelagiphages *in situ* at high host densities [66]; and the small dynamic range and decoupled nature of virioplankton:SAR11 abundances in the Sargasso Sea [67]. Limited lysis of subpopulations of hosts such as SAR11 and OM43 that specialise in the uptake of labile carbon enriched through viral predation [68, 69] would facilitate efficient intra-population carbon recycling and explain the limited influence of SAR11 and associated viral abundances on carbon export to the deep ocean [5]. We propose the moniker the 'Soylent Green Hypothesis' for this mechanism, after the 1973 cult film in which the dead are recycled into food for the living. Further investigation leveraging our new virus-host model will provide greater insight into viral influence on ocean carbon biogeochemistry.

In conclusion, our method coupled Dilution-to-Extinction cultivation of hosts and associated viruses, resulting in the isolation of three new strains of OM43; a Western English Channel variant of a warm water ecotype of SAR11; the first known methylophage for OM43; the first siphovirus infecting SAR11, as well as eleven other viruses infecting this important marine heterotrophic clade and >100 more isolates to be sequenced and explored. The described method represents an efficient and cost-effective strategy to isolate novel virus-host systems for experimental evaluation of co-evolutionary dynamics of important fastidious taxa from marine and other biomes. Coupling these methods to existing advances in host cultivation requires minimal additional effort and will provide valuable representative genomes to improve success rates of assigning putative hosts to metagenomically-derived viral contigs. Broader representation of model systems beyond cyanophages and viruses of copiotrophic, *r*-strategist hosts will reduce bias in developing methods to delineate viral population boundaries [70, 71], increasing the accuracy with which we derive ecological meaning from viral metagenomic data. We therefore hope that this method will enable viruses to be included in the current resurgence of cultivation efforts to better understand the biology and ecology of phages, and the influence of the world's smallest predators on global biogeochemistry.

Materials and Methods

Water sampling and viral concentration: A total of 20 L of seawater was collected in rosette-mounted Niskin bottles at a depth of 5m from the Western Channel Observatory (WCO; http://www.westernchannelobservatory. org.uk/) coastal station 'L4' (50°15.00N; 4°13.00W) on the following dates: 17.10.2018, 05.11.2018, 11.02.2019, 11.03.2019, 01.04.2019, 22.07.2019 (for details see table S1). Seawater was transferred immediately to a clean 2 L acid-washed polycarbonate bottle and placed in a coolbox at ambient temperature. Upon return to shore, water was transported to the University of Exeter for immediate processing.

Isolation of SAR11 strain H2P3a and OM43 strains: 1 L seawater was collected as described above in September 2017, filtered through a 142 mm Merck Millipore pore 0.2 µm PC filter and autoclaved in 2 L Duran bottles with the lid tightly shut to maintain carbon chemistry [28]. Upon cooling, nutrients were added to make natural seawater medium (NSW): 1 mM NH_4Cl, 10 µM K_2HPO_4, 1 µM $FeCl_3$, 25 µM glycine, 25 µM methionine, 100 µM pyruvate [36]. 1 mL of medium was placed into each well in three sterile 96-well acid washed Teflon plates (Radleys, UK). Surface water from station

377 L4 was filtered through a 142 mm Merck Millipore 0.2 µm PC filter to remove larger plankton and
378 bacteria and enrich for the low-nucleic acid community fraction. A 200 µL aliquot was stained with
379 SybrGreen and then quantified on a C6 Accuri flow cytometer. The remaining retained fraction was
380 diluted in NSW to a density of ~ 1 cell per µL. 1 µL of diluted inoculum was added to each well, with
381 8 wells left blank as negative controls and eight wells with undiluted low-nucleic acid community
382 fraction as positive controls. Plates were covered and incubated in the dark at 15 °C for three months
383 and checked monthly for positive growth by flow cytometry as described above. Positive wells (>10^6
384 cells per mL) were transferred into 50 mL cultures in NSW medium and passaged three times once
385 they achieved >10^6 cells per mL). To verify culture purity and to taxonomically classify new isolates,
386 the 16S rRNA gene was amplified with PCR (25 cycles, 30 s @ 98 °C; 45 s @ 59 °C; 45 s @ 72 °C;
387 final extension of 90 s @ 72 °C) using commercially available 27F and 1492R primers by Eurofins
388 Genomics. Amplicon DNA was purified for Sanger sequencing (Eurofins) using a QIAquick PCR
389 purification kit (Qiagen) following the manufacturers standard protocol. 16S rRNA sequences were
390 compared to known bacterial sequences by ANI, calculated using the OrthoANI algorithm [47].
391
392 *Host cultivation:* SAR11 strains HTCC1062 and HTCC7211 were kindly provided by the Giovannoni
393 lab and were kept in continuous culture in 50 mL polycarbonate flasks at 15 °C on artificial seawater
394 medium ASM1 [37] throughout the duration of the study. Following isolation, cultures of OM43 and
395 H2P3α were maintained on ASM1; medium for OM43 was further amended with 100 mM MeOH
396 and 32 µL of commercially available MEM amino acid solution per 1.6 L of medium (Sigma,
397 M5550), to improve cellular yields and reduce occurrences of chaining cells (observed through EM,
398 Fig. 5, and a phenotype of alanine starvation in SAR11 [36]), respectively.
399
400 *Viral concentration and inoculation:* 1 mL of exponentially growing bacterial host in ASM1 medium
401 (amended with nutrients for OM43 strains as described for host cultivation above) was placed into
402 each well of a 96-well, acid washed Teflon plate. Eight wells were used as blank medium controls
403 (no cell controls). Monthly samples of 2 L of surface water were collected from station L4 as
404 described above and sequentially filtered through a 142 mm Whatman GF/D filter (2.7 µm pore size)
405 and a 142 mm 0.2 µm pore polycarbonate filter (Merck Millipore) using a peristaltic pump, to remove
406 larger sized bacterioplankton. The filtrate was concentrated to 50 mL (40-fold concentration) using a
407 Sartorius 50R VivaFlow tangential flow filtration unit. The concentrate was filtered through a 0.1 µm
408 pore PVDF membrane syringe-filter to remove any residual small cells and used as inoculum for viral
409 isolation. 100 µL of viral inoculum was added to each well in the 96-well plate, with 8 wells amended
410 with 10% (v/v) medium instead of viral inoculum to serve as no-virus controls. The plates were
411 incubated for ~2 weeks until no-virus controls would reach maximum cell density (~2 weeks for
412 strains HTCC1062, HTCCC7211 and H2P3α; about 1 week for OM43 strains H5P1, C6P1, D12P1).
413 Growth was monitored at the end of the incubation period by flow cytometry. Cytograms of virus
414 amended wells and no-virus controls were compared to spot infections as described below. A detailed
415 protocol of the viral isolation process is available on protocols.io
416 (dx.doi.org/10.17504/protocols.io.bb73irqn). Once the no-virus controls reached maximum cell
417 density, inoculated wells that did not meet the above criteria were pooled, filtered through a 0.1 µm
418 pore PVDF membrane syringe (Durapore) filter to remove the cells and then used as fresh inoculum
419 in a new 96-well plate of target host. This was repeated up to three times to amplify host-specific
420 viral particles to a density whereby signs of viral infectivity could be observed on the flow cytometer.
421

422 *Identification of positive infection via flow cytometry:* We observed positive infections in cells via
423 flow cytometry by comparing cytograms of wells inoculated with viruses against no-virus controls.
424 Cytograms of wells containing infected cultures were identifiable by a population shift of up to a 10-
425 fold increase in green fluorescence compared to uninfected cultures, an increase in the 'noise' fraction
426 (presumably from cellular debris following lysis as well as viral particles) over time (Supplementary
427 material S11) and a reduced maximum cell density (ranging from 10 to 1000-fold depending on initial
428 cell density and viral inoculum concentration) in infected cultures compared to no-virus controls. The
429 presence of viral-like particles in wells where a shift in green fluorescence, increased noise and
430 reduced maximum cell density were observed were verified by TEM and reinfection of new cultures
431 in fresh medium.
432
433 *Viral purification:* Following identification of a positive infection within a well, the contents of the
434 well were filtered through a 0.1 µm pore-size PVDF syringe filter (Durapore) and used as inoculum
435 in three rounds of viral purification. Briefly, a 96-well Teflon plate was inoculated with bacterial
436 hosts as described previously. A 10-fold serial dilution series (from 10^0-10^{-10}) of the inoculum in
437 ASM1 was added to a single row of the plate at 10 % v/v, with one well per row as no-virus controls
438 (total of eight no-virus controls/plate). Plates were incubated at 18 °C for ~2 weeks until the no-virus
439 controls reached maximum cell density and then screened for signs of viral infection using flow
440 cytometry as described above. For each dilution series, the well amended with the lowest number of
441 viruses in the inoculum that showed positive signs of infection was identified, and used as the
442 inoculum in another round of viral purification [33]. Using this format, we were able to purify eight
443 to twelve viral isolates simultaneously per plate.
444
445 *Transmission Electron Microscopy of viral isolates.* For ultrastructural analysis, virus particles were
446 transferred onto pioloform-coated EM copper grids (Agar Scientific, Standsted, UK) by floating the
447 grids on droplets of virus-containing suspension for 3 min. Following a series of four washes on
448 droplets of deionized water, the bound virus particles were contrasted with 2 % (w/v) uranyl acetate
449 in 2 % (w/v) methyl cellulose (mixed 1:9) on ice for 8 min and the grids then air-dried on a wire loop
450 after carefully removing excess staining solution with a filter. Dried grids were inspected with a JEOL
451 JEM 1400 transmission electron microscope operated at 120 kV and images taken with a digital
452 camera (ES 1000W CCD, Gatan, Abingdon, UK).
453
454 *DNA preparation and sequencing of viral isolates.* 50 mL ASM1 (amended with 100 mM MeOH and
455 amino acid solution for OM43 as described above) in 250 mL acid-washed, polycarbonate flasks were
456 inoculated with host cultures to 10^6 cells/mL. The cultures were infected with 10% v/v viruses in
457 ASM1 medium and incubated until after host cell lysis was detected using flow cytometry. The
458 cultures were transferred to 50 mL Falcon tubes and the cellular fraction was removed by
459 centrifugation at 8,500 rpm for 120 minutes. The supernatant was subsequently filtered through pore-
460 size 0.1 µm PVDF syringe filter membranes to remove any remaining smaller cellular debris. Phages
461 in the filtrate were precipitated using a modified version of an established PEG800/NaCl DNA
462 isolation method [72]. Briefly, the filtered phage lysate was transferred into 50 mL Falcon tubes with
463 pre-weighted 5 g PEG8000 and 3.3 g NaCl, shaken until both dissolved and incubated on ice
464 overnight. The phage particles were then pelleted by centrifugation at 8,500 rpm for 90 min at 4 °C.
465 Supernatant was discarded and phage particles resuspended by rinsing Falcon tubes twice with 1 mL
466 SM buffer (100 mM NaCl, 8 mM $MgSO_4$ -$7H_2O$, 50 mM Tris-Cl). DNA was extracted using the

467 Wizard DNA Clean-Up system (Promega) following manufacturer's instructions, but instead of TE
468 we used pre-heated 60 °C nuclease free water to elute DNA. DNA was sequenced by MicrobesNG
469 (Birmingham, UK) with Illumina paired-end HiSeq 2500. All reads were trimmed, quality-controlled
470 and error-corrected using bbmap and tadpole [73]. Contigs were assembled using SPAdes v3.13 [74]
471 and evaluated with QUAST [75]. The trimmed reads were mapped back against the contigs for
472 scaffolding using bowtie2 [76], BamM (alignment 0.9, identity 0.95) (available at
473 https://github.com/minillinim/BamM) and samtools [77]. Viral contigs were confirmed with
474 VirSorter (categories 1 or 2, >15kbp) and only accepted if the representative contig showed mean
475 depth coverage an order of magnitude greater than the next highest recruiting contig (to filter out
476 cellular DNA carryover). Gene calls returned by VirSorter were imported into DNA Master for
477 manual curation [78]. Additional gene calls were made using GenMark [79], GenMarkS [80],
478 GenMarkS2 [81], GenMark.hmm [82], GenMark.heuristic [83], Glimmer v.3.02 [84], and prodigal
479 v.2.6.3 [85]. All gene calls were listed and compared using a scoring system which evaluates gene
480 length, gene overlap and coding potential of ORFs [78]. ORFs were then annotated using BLASTp
481 against the NCBI's non-redundant protein sequences [86], Phmmer [87] against UniProts UniProtKB
482 and uniprotrefprot [88], Swissprot [89] as well as InterProScan [90] and Pfam [91]. Virally-encoded
483 tRNAs were identified with tRNAScan-SE v2.0 using settings `tRNAscan-SE -qQ --detail -o# -
484 m# -f# -l# -c tRNAscan-SE.conf -s# -B` [92] to identify putative tRNA genes. Matching tRNA
485 genes (e.g. viral and host tRNA-Leu) were then aligned using BLASTn with default settings [93]. All
486 viral genomes were organised into viral clusters (ICTV-recognised genera) based on shared gene
487 content with VConTACT2 [19], and into populations using a boundary cutoff of 95% ANI over 85%
488 [18].
489

490 *Viral phylogenetic trees.* For the podoviruses, manually curated gene calls for all isolate genomes
491 from this study and all other known pelagiphage isolates were imported into anvi'o (v.6.1.) [94] and
492 clustered using the internal pangenome functions (MCL inflation 1.5, minbit 0.5). Additionally, we
493 picked random contigs from each cluster of phage contigs identified in metagenomes that the authors
494 speculated to be infective towards 1A SAR11s and included them in the phylogenetic analysis [38].
495 The gene clusters were exported into a correlation matrix of shared genes between all viral isolates
496 and a number. Viral major capsid protein phylogenetic trees were created using the Phylogeny.fr
497 online service [95]. The amino acid sequences of proteins of interest were aligned using T-coffee
498 alignment [96] and curated with Gblocks [97], both using their respective standard settings. For the
499 construction of the phylogenetic tree, Bayesian inference trees were created using MrBayes [98] with
500 standard settings, creating 100,000 generations, tree samples every 10 generations and burn-in of the
501 first 5000 trees sampled.

502 For the siphoviruses *Kolga* and *Venkman*, all 14,883 viral contigs publically available in January 2020
503 on http://millardlab.org/ were gene called using Prodigal v2.6.3. [85]. All called genes were aligned
504 to vFams [99] using hmmsearch [100]. Genes that were aligned to the same vFams as *Kolga* and
505 *Venkman* genes were aligned individually using MUSCLE [101] and concatenated afterwards to be
506 used for phylogenetic analysis. Alignments were curated using Gblocks [97] with stringent setting
507 and maximum likelihood trees were created using PhyML with 500 bootstraps [102].

508 *Viral metagenomic recruitment.* For assessing relative abundances of phage contigs, a virome from
509 the Western English Channel was used to recruit against [12]. Metagenomic reads were downloaded

from the European Nucleotide Archive ERR2625613 and subsampled to 10 million reads using seqtk (https://github.com/lh3/seqtk). A kallisto index was created from 15 known pelagiphage genomes from isolates [30, 31]; 13 putative pelagiphage genomes from a fosmid-derived viral metagenome [38]; and six viral population representative genomes isolated on SAR11 hosts from this study, and *Venkman* representing the novel OM43 methylophage. FastViromeExplorer v1.3 [50] was used to quantify relative abundance of each pelagiphage using default parameters by calculating Reads Per Million (RPM) as follows:

$$RPM_i = \frac{X_i}{\tilde{l}_i} \cdot \left(\frac{1}{\sum_j \frac{X_j}{\tilde{l}_j}} \right) \cdot 10^6$$

RPM_i for a given genome, i, is calculated by dividing the rate (X_i/l_i) at which the number of reads, X_i, align to a genome of effective length, l_i, by the sum of rates of all other genomes, scaled by 10^6 [103].

Accession Numbers

All reads can be found in the SRA database under BioProject number PRJNA625644 as BioSamples SAMN14604128-SAMN14604140. Annotated phage genomes are deposited as GenBank submissions under accession numbers MT375519- MT375531.

Acknowledgements

We would like to thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing the TE microscopy and imaging. We would also like to thank the crew of the *R/V Plymouth Quest* and our collaborators at PML for collecting water samples, and the driver Magic for delivering water samples from Plymouth to Exeter. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk) which is supported by the BBSRC (grant number BB/L024209/1).

Funding Information

The efforts of Holger Buchholz in this work were funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training program. Michelle Michelsen and Ben Temperton were funded by NERC (NE/R010935/1) and by the Simons Foundation BIOS-SCOPE program.

Author contributions

HHB led cultivation experiments, nucleic acid extractions, and analyses; MLM led sample collections and assisted in cultivation maintenance. BT designed the study and assisted with bioinformatics, HHB and BT led manuscript writing; and all authors contributed to and reviewed the text.

References

1. Eric Wommack K, Colwell RR. Virioplankton: Viruses in Aquatic Ecosystems. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS* 2000; **64**: 69–114.

549 2. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, et al. A multitrophic model
550 to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J*
551 2015; **9**: 1352–1364.
552 3. Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical
553 cycles. *F1000 Biol Rep* 2012; **4**: 17.
554 4. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles:
555 implications for marine biogeochemical cycles. *Nat Rev Microbiol* 2014; **12**: 519–528.
556 5. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving
557 carbon export in the oligotrophic ocean. *Nature* 2016; **532**: 465–470.
558 6. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how
559 phages command the microbial high seas. *Virol J* 2019; **16**: 15.
560 7. Weitz JS, Hartman H, Levin SA. Coevolutionary arms races between bacteria and bacteriophage. *Proc*
561 *Natl Acad Sci U S A* 2005; **102**: 9535–9540.
562 8. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. Long-term stability and Red Queen-like strain
563 dynamics in marine viruses. *Nat Microbiol* 2020; **5**: 265–271.
564 9. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant
565 temperate viruses affect Southern Ocean microbial dynamics. *ISME J* 2016; **10**: 437–449.
566 10. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential
567 biogeochemical impacts of globally abundant ocean viruses. *Nature* 2016; **537**: 689–693.
568 11. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA
569 Viral Macro- and Microdiversity from Pole to Pole. *Cell* 2019; **177**: 1109–1123.e14.
570 12. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral
571 metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic
572 islands. *PeerJ* 2019; **7**: e6800.
573 13. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. *Nature*
574 2003; **424**: 741.
575 14. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Viruses Inhibit CO2 Fixation in the Most Abundant
576 Phototrophs on Earth. *Curr Biol* 2016; **26**: 1585–1589.
577 15. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic
578 data. *PeerJ* 2015; **3**: e985.
579 16. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying
580 viral sequences from assembled metagenomic data. *Microbiome* 2017; **5**: 69.
581 17. Amgarten D, Braga LPP, da Silva AM, Setubal JC. MARVEL, a Tool for Prediction of Bacteriophage
582 Sequences in Metagenomic Bins. *Front Genet* 2018; **9**: 304.
583 18. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation
584 of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 2017; **5**: e3817.
585 19. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment
586 of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* 2019.
587 20. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in
588 the 'core' and 'flexible' Pacific Ocean Virome. *ISME J* 2015; **9**: 472–484.
589 21. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts
590 from metagenomic phage contigs. *Bioinformatics* 2017; **33**: 3113–3114.
591 22. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d_2^* oligonucleotide frequency
592 dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences.
593 *Nucleic Acids Res* 2017; **45**: 39–53.
594 23. Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, et al. One carbon metabolism in
595 SAR11 pelagic marine bacteria. *PLoS One* 2011; **6**: e23973.
596 24. Becker JW, Hogle SL, Rosendo K, Chisholm SW. Co-culture and biogeography of Prochlorococcus and
597 SAR11. *ISME J* 2019.
598 25. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA. Diversity and biogeography of
599 SAR11 bacteria from the Arctic Ocean. *ISME J* 2020; **14**: 79–90.
600 26. Robert M. Morris*, Michael S. Rappe´ *, Stephanie A. Connon*, Kevin L. Vergin*, William A.
601 Siebold*, Craig A. Carlson† & Stephen J. Giovannoni*. SAR11 clade dominates ocean surface
602 bacterioplankton communities. *Nature* 2002; **420**.
603 27. Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-
604 nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* 2002; **68**: 3878–3885.
605 28. Henson MW, Pitre DM, Weckhorst JL, Celeste Lanclos V, Webber AT, Cameron Thrash J. Artificial

606 Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico.
607 *American Society for Microbiology* 2016; **1**: 1–9.
608 29. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. Cultivation and genomics of the first freshwater
609 SAR11 (LD12) isolate. *ISME J* 2018; **12**: 1846–1860.
610 30. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11
611 viruses in the ocean. *Nature* 2013; **494**: 357–360.
612 31. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family
613 integrate into host genomes. *Environ Microbiol* 2018.
614 32. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et
615 al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 2017; **8**:
616 15892.
617 33. Nagasaki K, Bratbak G. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists.
618 *Manual of aquatic viral ecology ASLO* 2010; 92–101.
619 34. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001
620 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis.
621 *Nat Commun* 2018; **9**: 4781.
622 35. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Novel pelagiphages prevail in the ocean.
623 *bioRxiv* . 2019. , 2019.12.14.876532
624 36. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme
625 oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J* 2013; **7**: 592–
626 602.
627 37. Carini P, Campbell EO, Morré J, Sañudo-Wilhelmy SA, Thrash JC, Bennett SE, et al. Discovery of a
628 SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea.
629 *ISME J* 2014; **8**: 1727–1738.
630 38. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using
631 metagenomics. *PLoS Genet* 2013; **9**: e1003987.
632 39. Taubert M, Grob C, Howat AM, Burns OJ, Dixon JL, Chen Y, et al. XoxF encoding an alternative
633 methanol dehydrogenase is widespread in coastal marine environments. *Environ Microbiol* 2015; **17**:
634 3937–3948.
635 40. Ramachandran A, Walsh DA. Investigation of XoxF methanol dehydrogenases reveals new
636 methylotrophic bacteria in pelagic marine and freshwater ecosystems. *FEMS Microbiol Ecol* 2015; **91**.
637 41. Morris RM, Longnecker K, Giovannoni SJ. Pirellula and OM43 are among the dominant lineages
638 identified in an Oregon coast diatom bloom. *Environ Microbiol* 2006; **8**: 1361–1370.
639 42. Halsey KH, Carter AE, Giovannoni SJ. Synergistic metabolism of a broad range of C1 compounds in
640 the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* 2012; **14**: 630–640.
641 43. Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, et al. The small genome of an
642 abundant coastal ocean methylotroph. *Environ Microbiol* 2008; **10**: 1771–1782.
643 44. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification
644 of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*
645 2014; **12**: 635–645.
646 45. Neufeld JD, Boden R, Moussard H, Schäfer H, Murrell JC. Substrate-specific clades of active marine
647 methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Appl Environ*
648 *Microbiol* 2008; **74**: 7321–7328.
649 46. Våge S, Storesund JE, Thingstad TF. SAR11 viruses and defensive host strains. *Nature* . 2013. , **499**:
650 E3–4
651 47. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average
652 nucleotide identity. *Antonie Van Leeuwenhoek* 2017; **110**: 1281–1286.
653 48. Martinez-Hernandez F, Fornas Ò, Lluesma Gomez M, Garcia-Heredia I, Maestre-Carballa L, López-
654 Pérez M, et al. Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant
655 uncultured 37-F6 viral population in the ocean. *ISME J* 2018.
656 49. Giovannoni S, Temperton B, Zhao Y. Giovannoni et al. reply. *Nature* . 2013. , **499**: E4–5
657 50. Tithi SS, Aylward FO, Jensen RV, Zhang L. FastViromeExplorer: a pipeline for virus and phage
658 identification and abundance profiling in metagenomics data. *PeerJ* 2018; **6**: e4227.
659 51. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* 2005;
660 **13**: 278–284.
661 52. Ackermann H-W. 5500 Phages examined in the electron microscope. *Arch Virol* 2007; **152**: 227–243.
662 53. Brum JR, Schenck RO, Sullivan MB. Global morphological analysis of marine viruses shows minimal

663 regional variation and dominance of non-tailed viruses. *ISME J* 2013; **7**: 1738–1751.
664 54. Tam W, Pell LG, Bona D, Tsai A, Dai XX, Edwards AM, et al. Tail tip proteins related to
665 bacteriophage λ gpL coordinate an iron-sulfur cluster. *J Mol Biol* 2013; **425**: 2450–2462.
666 55. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous
667 cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun* 2019; **10**: 752.
668 56. Van Duin J, Wijnands R. The function of ribosomal protein S21 in protein synthesis. *Eur J Biochem*
669 1981; **118**: 615–619.
670 57. Xu J, Hendrix RW, Duda RL. Conserved translational frameshift in dsDNA bacteriophage tail assembly
671 genes. *Mol Cell* 2004; **16**: 11–21.
672 58. Moon K, Kang I, Kim S, Kim S-J, Cho J-C. Genome characteristics and environmental distribution of
673 the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. *Environ*
674 *Microbiol* 2017; **19**: 4714–4727.
675 59. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. The ecology of pelagic freshwater
676 methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J* 2015; **9**: 2442–
677 2453.
678 60. Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages.
679 *Genome Res* 2007; **17**: 1486–1495.
680 61. Enav H, Kirzner S, Lindell D, Mandel-Gutfreund Y, Béjà O. Adaptation to sub-optimal hosts is a driver
681 of viral diversification in the ocean. *Nat Commun* 2018; **9**: 4698.
682 62. Weitz JS, Li G, Gulbudak H, Cortez MH, Whitaker RJ. Viral invasion fitness across a continuum from
683 lysis to latency. *Virus Evol* 2019; **5**: vez006.
684 63. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, et al. Lytic to temperate
685 switching of viral communities. *Nature* 2016; **531**: 466–470.
686 64. Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. *NPJ*
687 *Biofilms Microbiomes* 2016; **2**: 16010.
688 65. Kang I, Oh HM, Kang D, Cho JC. Genome of a SAR116 bacteriophage shows the prevalence of this
689 phage type in the oceans. *Proc Natl Acad Sci U S A* 2013; **110**: 12343–12348.
690 66. Alonso-Sáez L, Morán XAG, Clokie MR. Low activity of lytic pelagiphages in coastal marine waters.
691 *ISME J* 2018; **12**: 2100–2102.
692 67. Parsons RJ, Breitbart M, Lomas MW, Carlson CA. Ocean time-series reveals recurring seasonal
693 patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J* 2012; **6**: 273–284.
694 68. Middelboe M, Lyck PG. Regeneration of dissolved organic matter by viral lysis in marine microbial
695 communities. *Aquat Microb Ecol* 2002; **27**: 187–194.
696 69. Middelboe M, Jorgensen NOG. Viral lysis of bacteria: an important source of dissolved amino acids
697 and cell wall compounds. *J Mar Biol Assoc U K* 2006; **86**: 605–612.
698 70. Brum JR, Sullivan MB. Rising to the challenge: accelerated pace of discovery transforms marine
699 virology. *Nat Rev Microbiol* 2015; **13**: 147–159.
700 71. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, et al. Genomic
701 differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics*
702 2016; **17**: 930.
703 72. Solonenko N. Isolation of DNA from phage lysate. *protocols.io* 2016.
704 73. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS*
705 *ONE* . 2017. , **12**: e0185056
706 74. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome
707 assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**: 455–477.
708 75. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies.
709 *Bioinformatics* 2013; **29**: 1072–1075.
710 76. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357–359.
711 77. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
712 format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
713 78. Salisbury A, Tsourkas PK. A Method for Improving the Accuracy and Efficiency of Bacteriophage
714 Genome Annotation. *Int J Mol Sci* 2019; **20**.
715 79. Borodovsky M, McIninch J. GenMark: Parallel gene recognition for both DNA strands. *Computers*
716 *Chemistry* 1993; **17**: 123–133.
717 80. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene
718 starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic*
719 *Acids Res* 2001; **29**: 2607–2618.

720 81. Lomsadze A, Gemayel K, Tang S, Borodovsky M. Modeling leaderless transcription and atypical genes
721 results in more accurate gene prediction in prokaryotes. *Genome Res* 2018; **28**: 1079–1089.
722 82. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic*
723 *Acids Res* 2010; **38**: e132.
724 83. Besemer J, Borodovsky M. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res*
725 1999; **27**: 3911–3920.
726 84. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA
727 with Glimmer. *Bioinformatics* 2007; **23**: 673–679.
728 85. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene
729 recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.
730 86. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant
731 sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007; **35**: D61–5.
732 87. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic*
733 *Acids Res* 2018; **46**: W200–W204.
734 88. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019; **47**:
735 D506–D515.
736 89. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in
737 2000. *Nucleic Acids Res* 2000; **28**: 45–48.
738 90. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale
739 protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.
740 91. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families
741 database. *Nucleic Acids Res* 2014; **42**: D222–30.
742 92. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA
743 genes. *Nucleic Acids Res* 2016; **44**: W54–7.
744 93. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*
745 1990; **215**: 403–410.
746 94. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis
747 and visualization platform for 'omics data. *PeerJ* 2015; **3**: e1319.
748 95. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust
749 phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 2008; **36**: W465–9.
750 96. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple
751 sequence alignment. *J Mol Biol* 2000; **302**: 205–217.
752 97. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously
753 aligned blocks from protein sequence alignments. *Syst Biol* 2007; **56**: 564–577.
754 98. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*
755 2001; **17**: 754–755.
756 99. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection of
757 viruses within metagenomic sequence data. *PLoS One* 2014; **9**: e105067.
758 100. Eddy S. HMMER: profile HMMs for protein sequence analysis. 1998.
759 101. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*
760 *Acids Res* 2004; **32**: 1792–1797.
761 102. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--a web server for fast maximum
762 likelihood-based phylogenetic inference. *Nucleic Acids Res* 2005; **33**: W557–9.
763 103. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with
764 read mapping uncertainty. *Bioinformatics* 2010; **26**: 493–500.

765

766

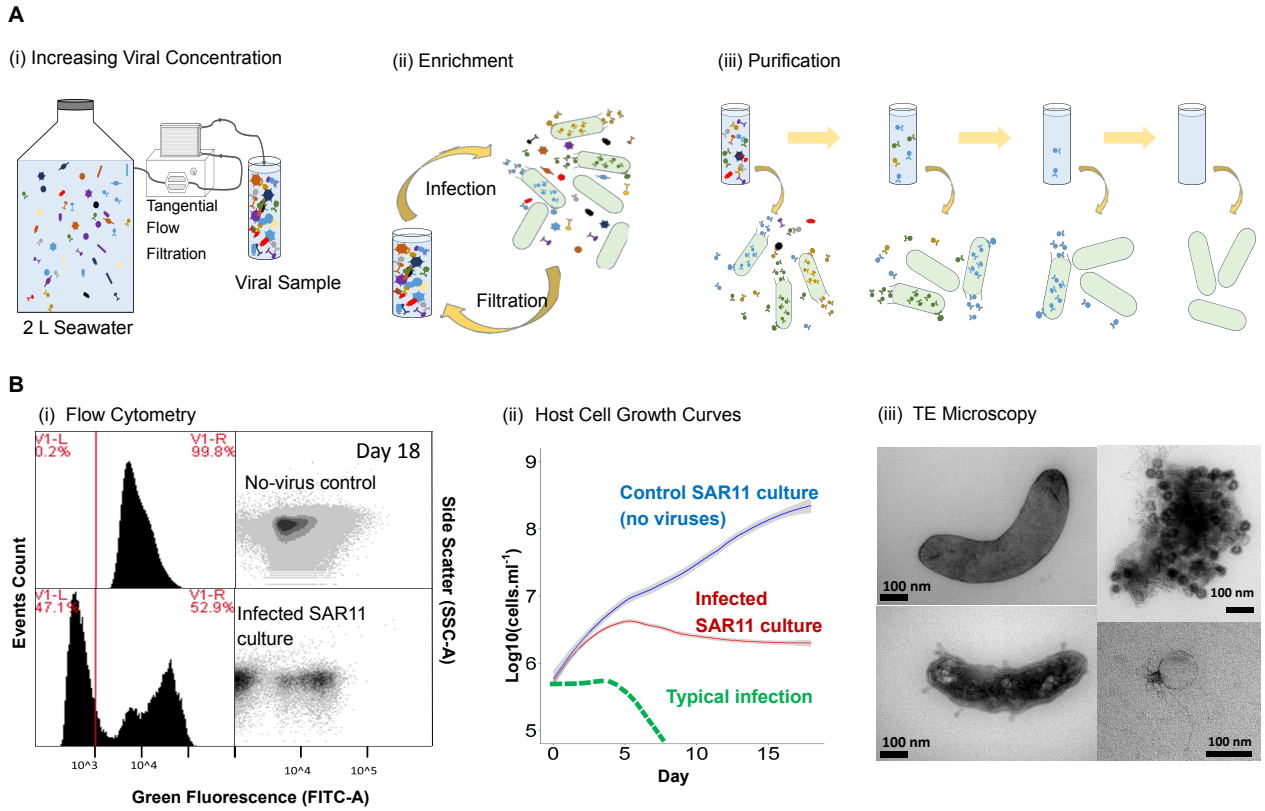


Fig. 1 Workflow for high-throughput isolation **(A)** i. Increasing concentration of viruses in water samples by TFF; ii. Initial infection of host cultures to enrich the sample for specific viruses; iii. Purification of viral isolates through three rounds of Dilution-to-Extinction; **(B)** Initial screen of viral infections using i. Flow cytometry, by comparing populations of no-virus controls and infected cultures; ii. Comparing growth curves of no-virus control culture (HTCC1062) against infected SAR11 cultures; iii. Confirming the presence of viruses in infected SAR11 cultures using TE microscopy: Top left: HTCC1062 no-virus control, bottom left: infected HTCC1062; top right: aggregated cellular debris and viruses, bottom right: virus found in infected HTCC1062 culture.

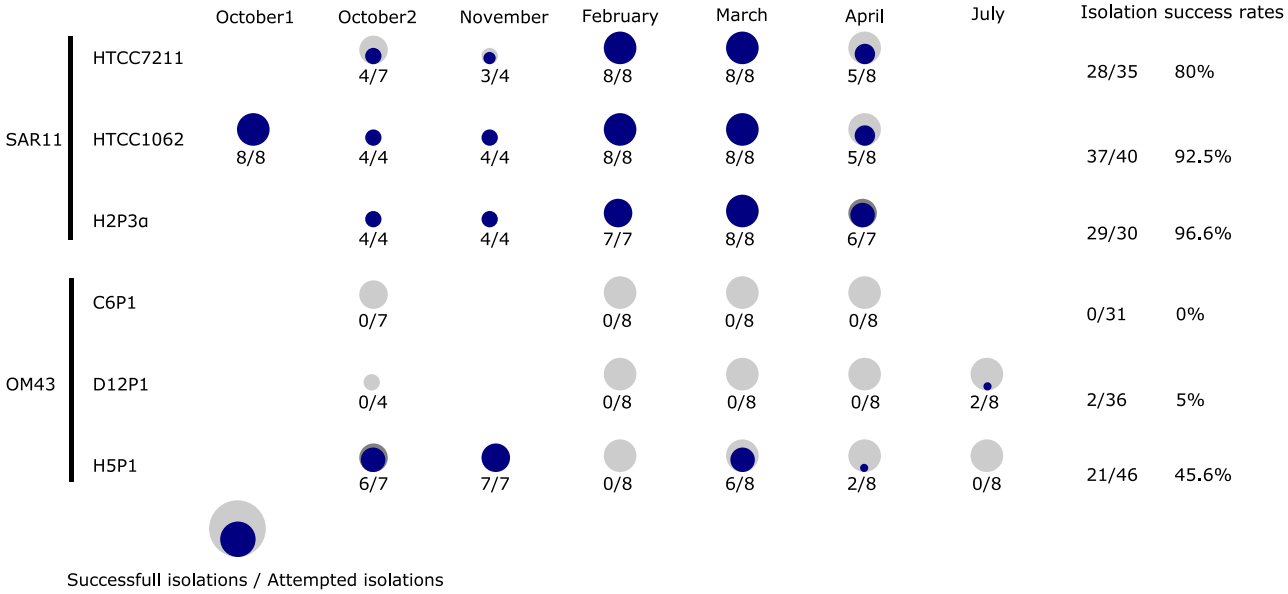
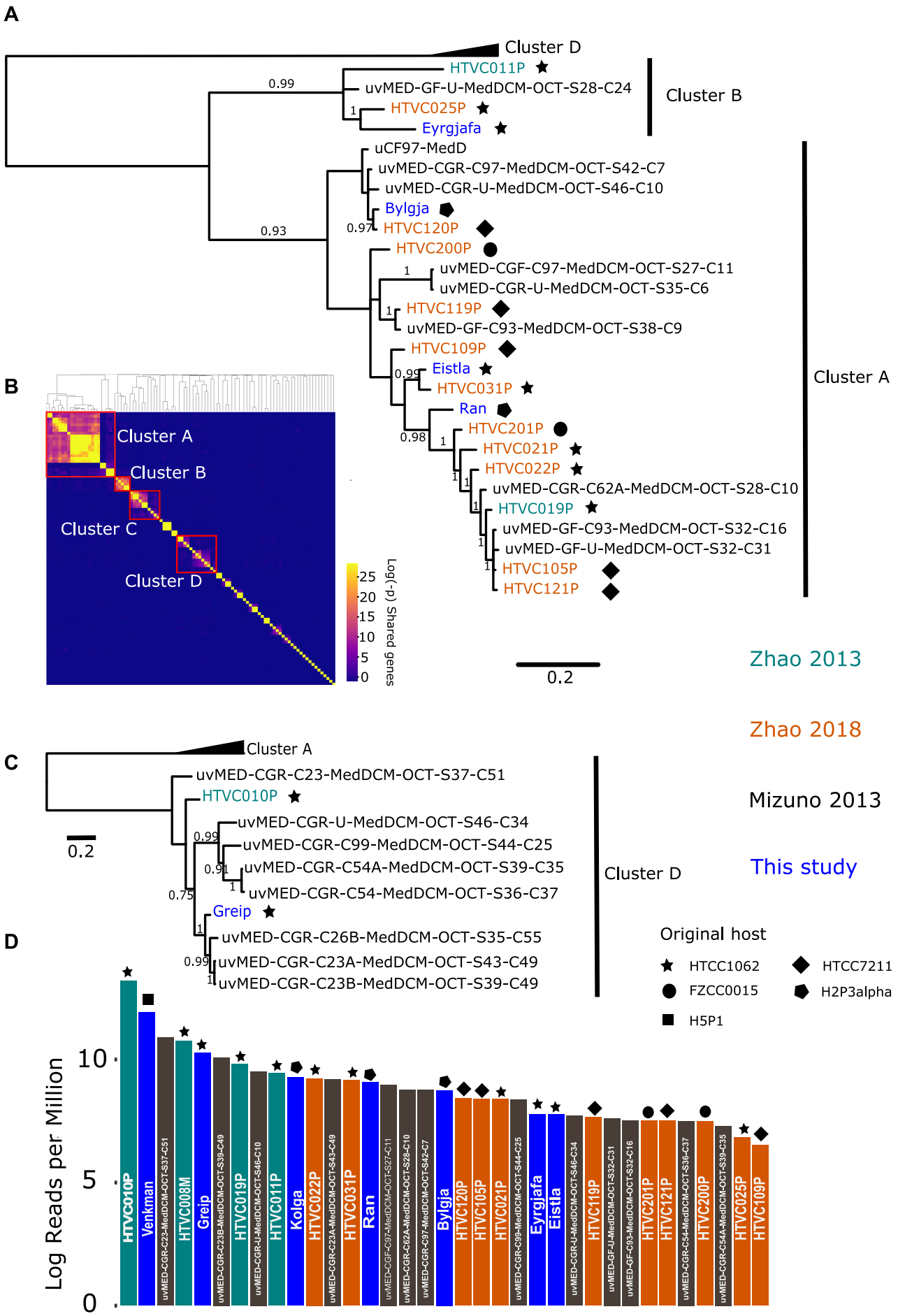


Fig. 2 Summary of success rates for each bacterial host (three SAR11 and three OM43) and water sample combination used for infections; the numerator indicates successful isolation of a virus from that water sample (i.e. successful passage through three rounds of Dilution-to-Extinction), denominator shows the isolation attempts made for a host-water sample combination (i.e. how many cultures were treated with viral samples after multiple rounds of enrichment).


A
Cluster D
HTVC011P
uvMED-GF-U-MedDCM-OCT-S28-C24
HTVC025P
Eyrgjafa
Cluster B
uCF97-MedD
uvMED-CGR-C97-MedDCM-OCT-S42-C7
uvMED-CGR-U-MedDCM-OCT-S46-C10
Bylgja
HTVC120P
HTVC200P
uvMED-CGF-C97-MedDCM-OCT-S27-C11
uvMED-CGR-U-MedDCM-OCT-S35-C6
HTVC119P
uvMED-GF-C93-MedDCM-OCT-S38-C9
HTVC109P
Eistla
HTVC031P
Ran
HTVC201P
HTVC021P
HTVC022P
uvMED-CGR-C62A-MedDCM-OCT-S28-C10
HTVC019P
uvMED-GF-C93-MedDCM-OCT-S32-C16
uvMED-GF-U-MedDCM-OCT-S32-C31
HTVC105P
HTVC121P
Cluster A
0.99
0.93
0.97
0.98
0.2
B
Cluster A
Cluster B
Cluster C
Cluster D
Log(-p) Shared genes
25
20
15
10
5
0
Zhao 2013
Zhao 2018
Mizuno 2013
This study
C
Cluster A
uvMED-CGR-C23-MedDCM-OCT-S37-C51
HTVC010P
uvMED-CGR-U-MedDCM-OCT-S46-C34
uvMED-CGR-C99-MedDCM-OCT-S44-C25
uvMED-CGR-C54A-MedDCM-OCT-S39-C35
uvMED-CGR-C54-MedDCM-OCT-S36-C37
Greip
uvMED-CGR-C26B-MedDCM-OCT-S35-C55
uvMED-CGR-C23A-MedDCM-OCT-S43-C49
uvMED-CGR-C23B-MedDCM-OCT-S39-C49
0.2
0.99
0.91
0.75
0.99
Cluster D
Original host
HTCC1062
HTCC7211
FZCC0015
H2P3alpha
H5P1
D
Log Reads per Million
10
5
0
HTVC010P
Venkman
uvMED-CGR-C23-MedDCM-OCT-S37-C51
HTVC008M
Greip
uvMED-CGR-C23B-MedDCM-OCT-S39-C49
HTVC019P
uvMED-CGR-U-MedDCM-OCT-S46-C10
HTVC011P
Kolga
HTVC022P
uvMED-CGR-C23A-MedDCM-OCT-S43-C49
HTVC031P
Ran
uvMED-CGF-C97-MedDCM-OCT-S27-C11
uvMED-CGR-C62A-MedDCM-OCT-S28-C10
uvMED-CGR-C97-MedDCM-OCT-S42-C7
Bylgja
HTVC120P
HTVC105P
HTVC021P
uvMED-CGR-C99-MedDCM-OCT-S44-C25
Eyrgjafa
Eistla
uvMED-CGR-U-MedDCM-OCT-S46-C34
HTVC119P
uvMED-GF-U-MedDCM-OCT-S32-C31
uvMED-GF-C93-MedDCM-OCT-S32-C16
HTVC201P
HTVC121P
uvMED-CGR-C54-MedDCM-OCT-S36-C37
HTVC200P
uvMED-CGR-C54A-MedDCM-OCT-S39-C35
HTVC025P
HTVC109P

800

801 **Fig. 3** Phylogeny and abundance of the first sequenced pelagiphages from the Western English
802 Channel **(A)** Clustering of viral genomes by shared protein content into ICTV-recognised genera
803 (VConTACT2) identified four viral clusters and numerous singletons; **(B)** Bayesian inference
804 phylogenetic tree of pelagiphage major capsid proteins from cluster A and B (most diverse
805 pelagiphage cluster, includes most known pelagiphages isolated previously (30, 31) as well as derived
806 from metagenomes (Mizuno et al. 2013)), outgroup Cluster D; **(C)** Bayesian inference phylogenetic
807 tree of pelagiphage major capsid proteins from Cluster D (containing the most abundant pelagiphages,
808 including HTVC010P) with Cluster A as outgroup; **(D)** Metagenomic recruitment (reads recruited
809 per million reads (RPM)) of isolate and metagenomically-derived viral genomes against a virome
810 from the Western English Channel (12).
811

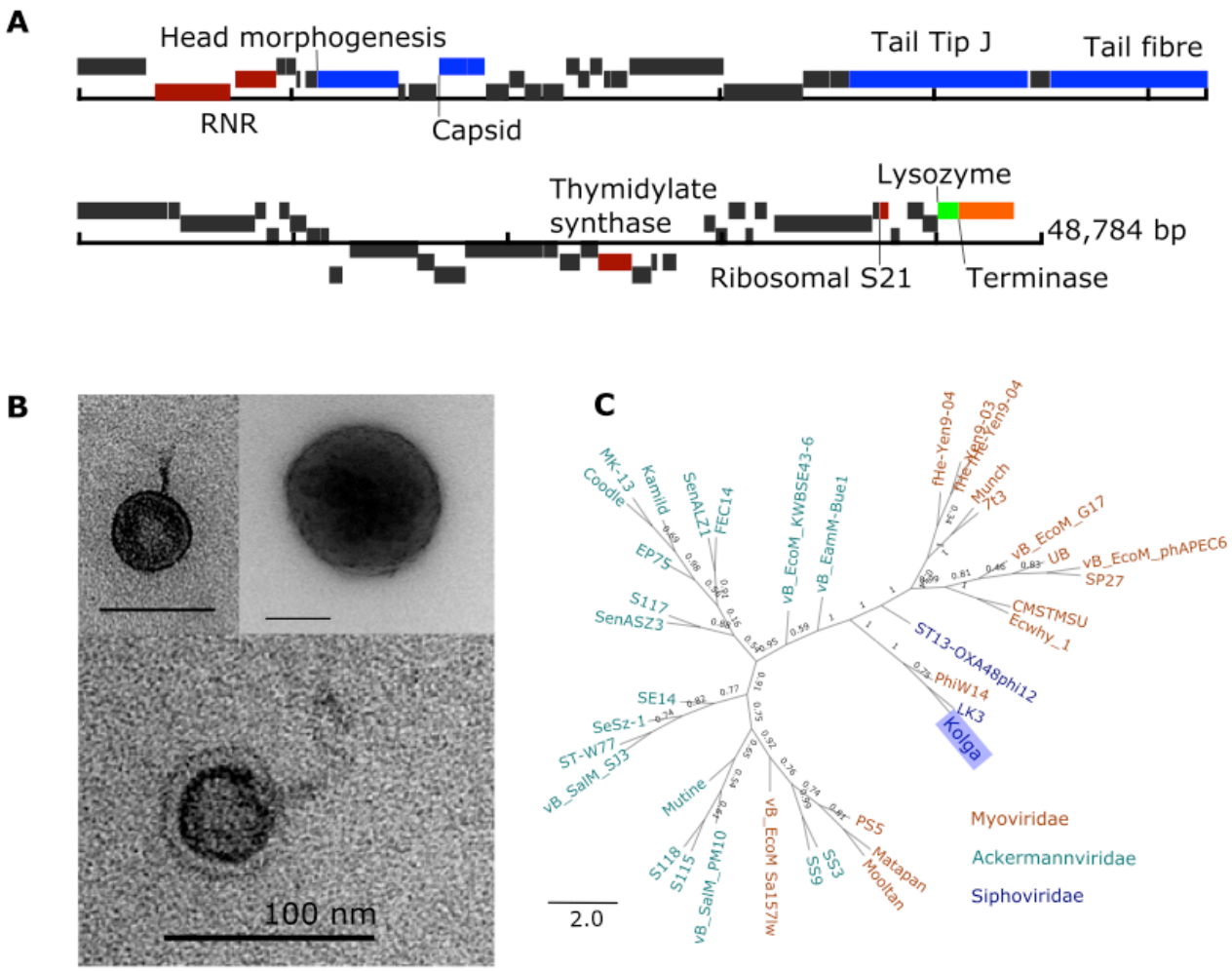


Fig. 4 The first reported siphovirus *Kolga* infecting SAR11 and isolated on novel host H2P3α; **(A)** gene map of the 48,784 bp genome, which contains 80% hypothetical genes without known function; **(C)** TEMs of *Kolga* and a H2P3α host cell infected with *Kolga*; **(B)** Bayesian inference phylogenetic tree of TerL genes of known siphoviruses and *Kolga*

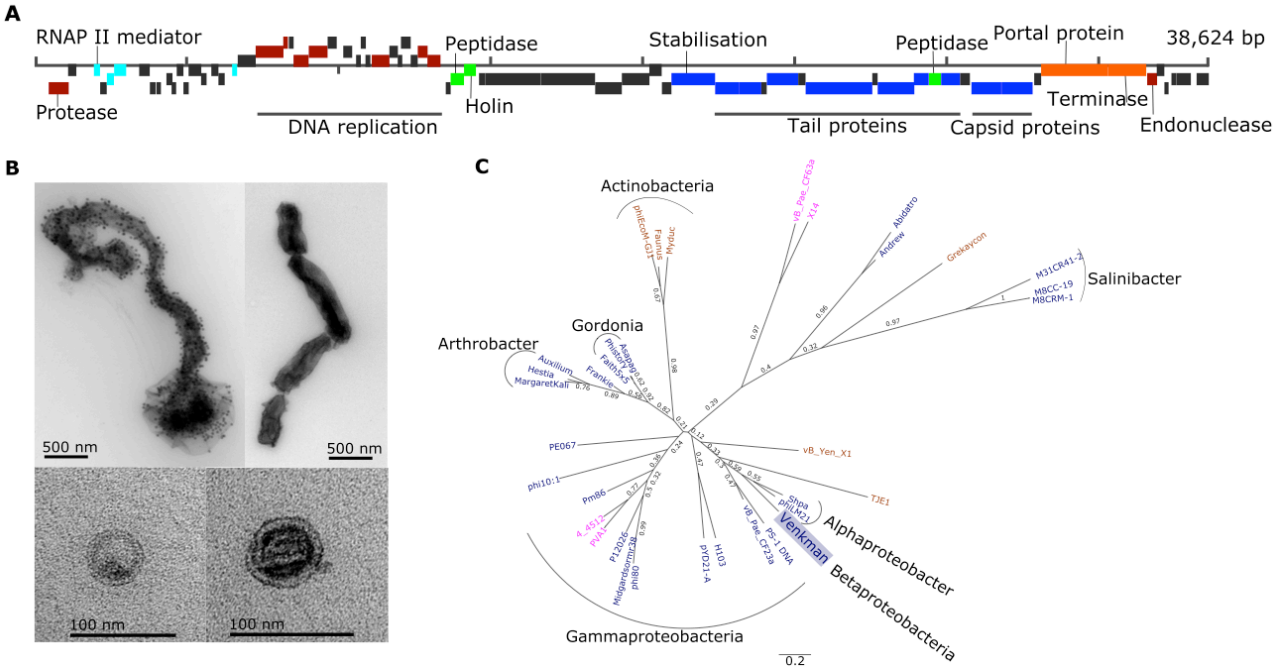
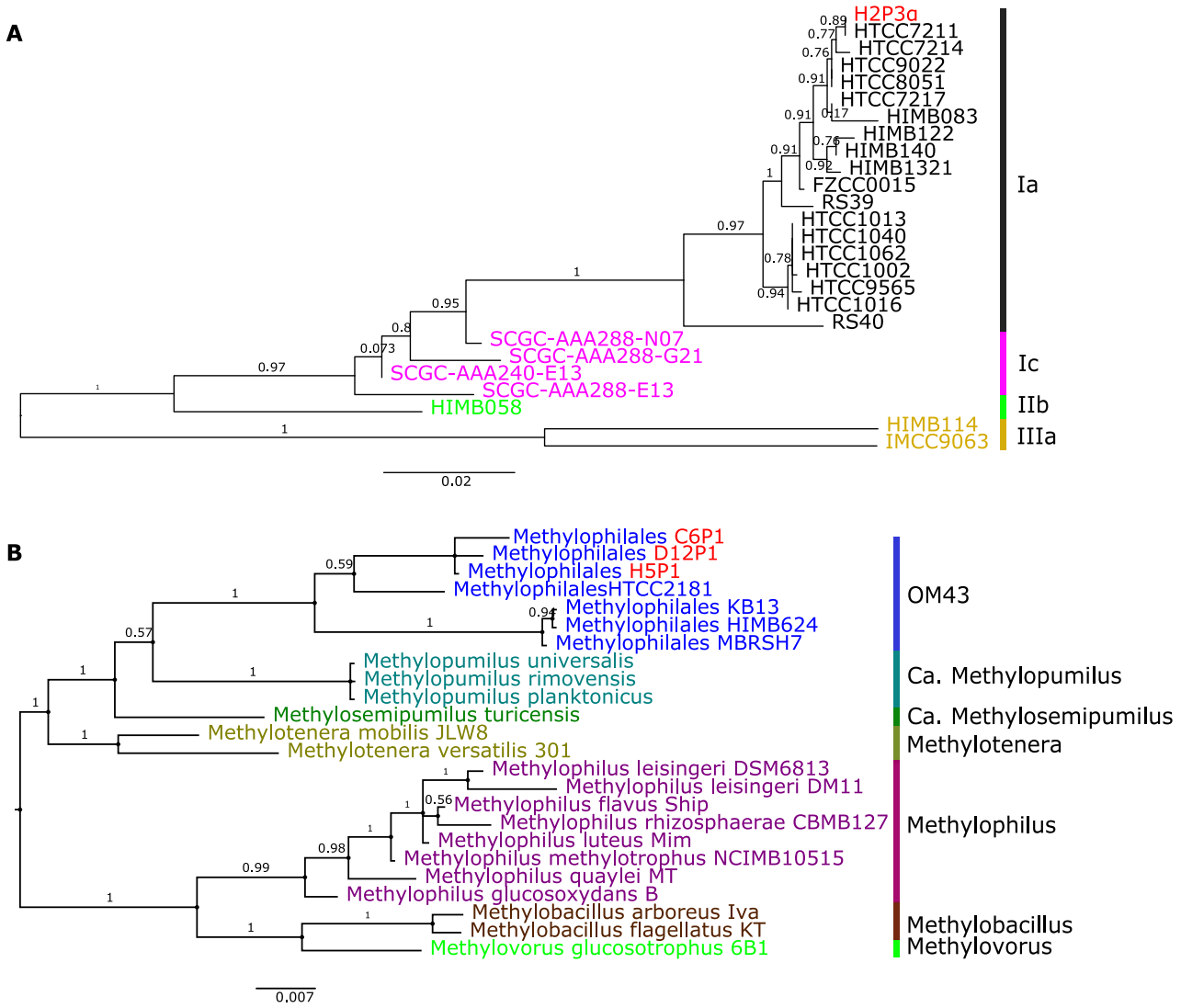


Fig. 5 Evidence for the first reported virus EXEVC282S *Venkman* infecting a member of the OM43 clade (H5P1); **(A)** Gene map displaying protein coding genes. Structural genes are shown in blue, DNA replication genes in red, lysis related genes in green, transcription genes in turquoise, packaging genes in orange, hypothetical genes are grey; **(B)** TEM images of: infected and chaining H5P1 cells (top left), uninfected H5P1 chaining cells (top right), *Venkman* viral particles (bottom left and right); **(C)** Maximum likelihood tree (500 bootstraps) of concatenated viral TerL and exonuclease genes.

826



827

828 **Fig. S1** Maximum likelihood trees (100,000 generations, sampled every 10 generations, first 5000
829 trees burn-in) of 16S genes of **(A)** A range of SAR11 strains and our isolate H2P3α; **(B)** OM43 and
830 other closely related *Methylophilales*

831

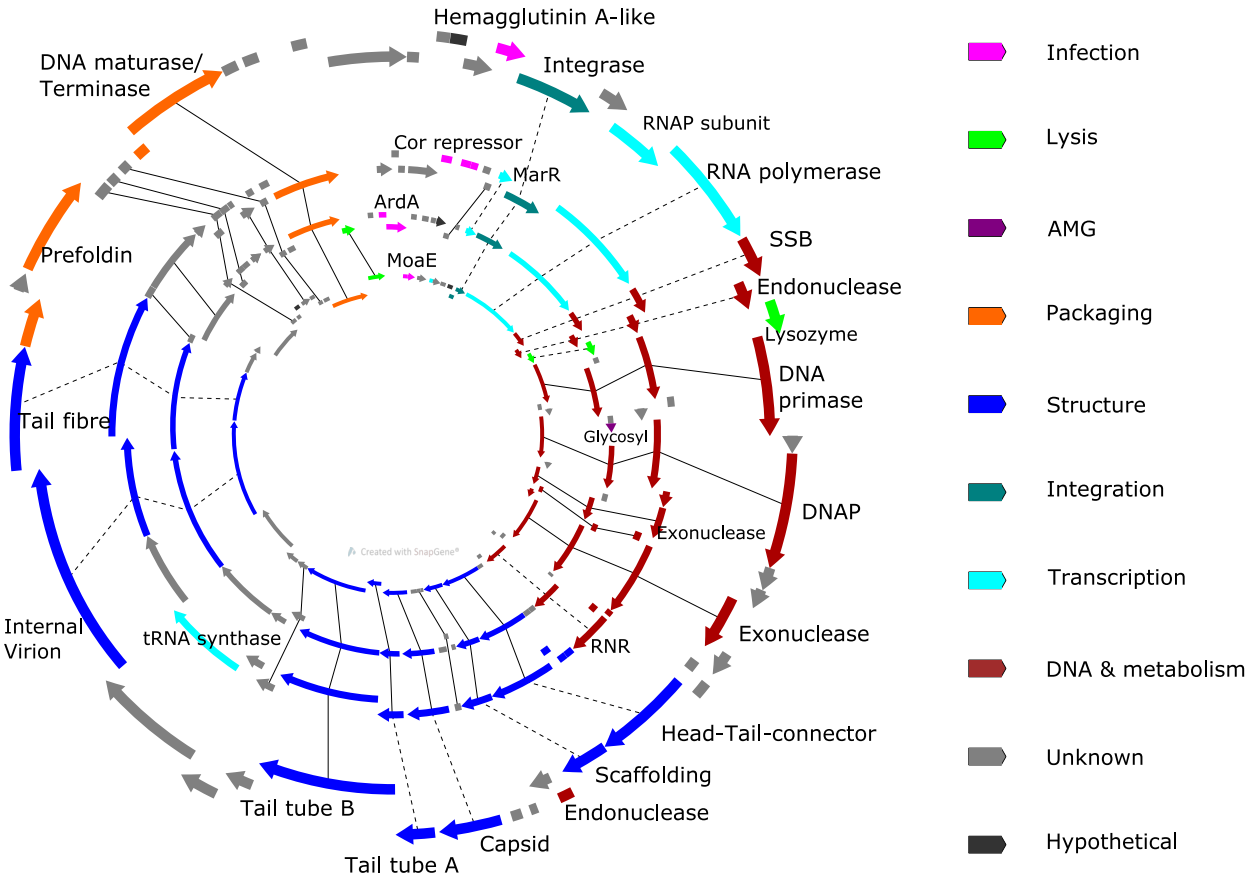


Fig. S2 Circular gene map of pelagiphages in Clusters A and B arranged inner to outer: *Eistla* (EXEVC025P), *Ran* (EXEVC014P), *Bylgja* (EXEVC010P) and *Eyrgjafa* (EXEVC018P). Solid lines indicate orthologous genes, dashed lines indicate genes of similar function.

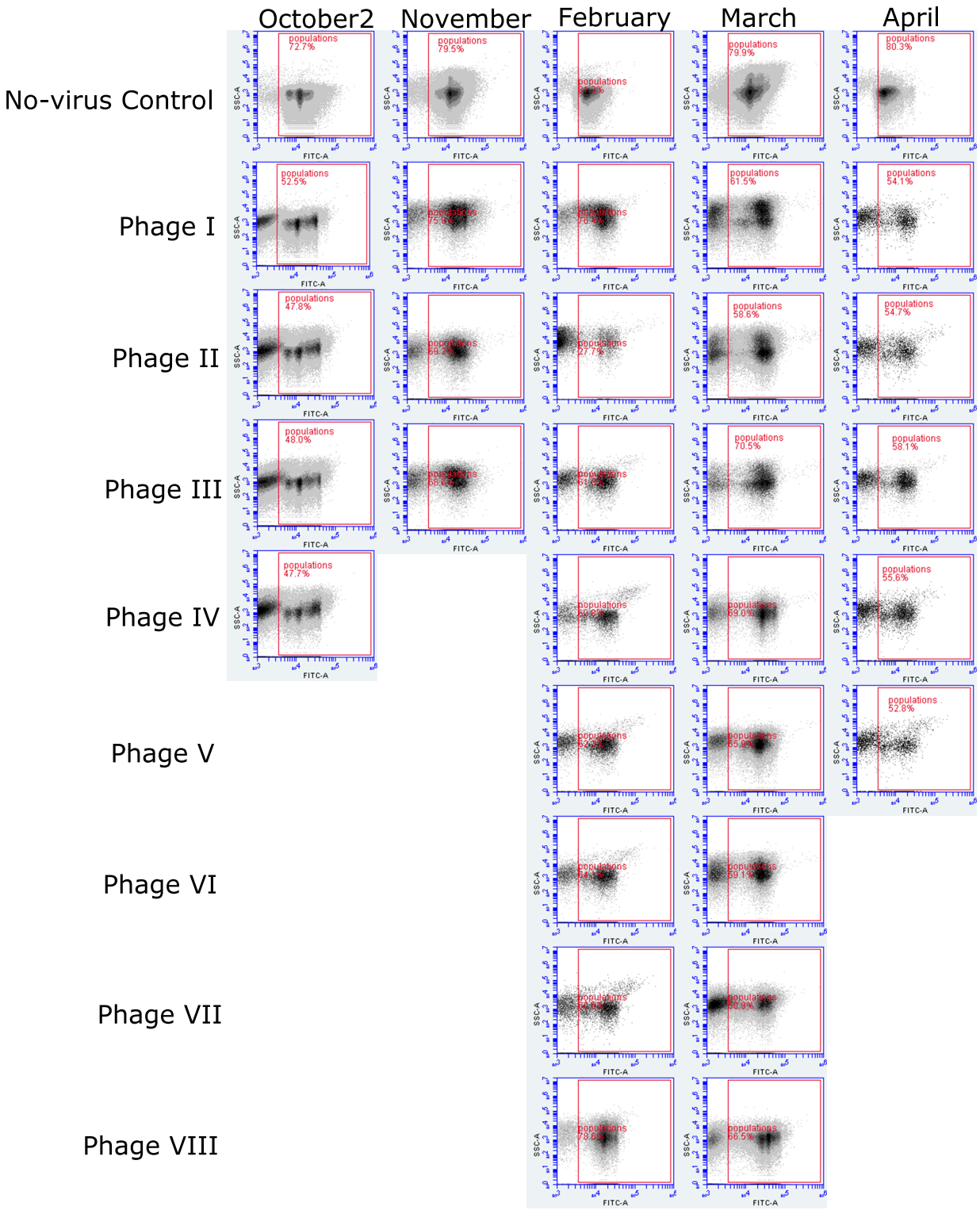
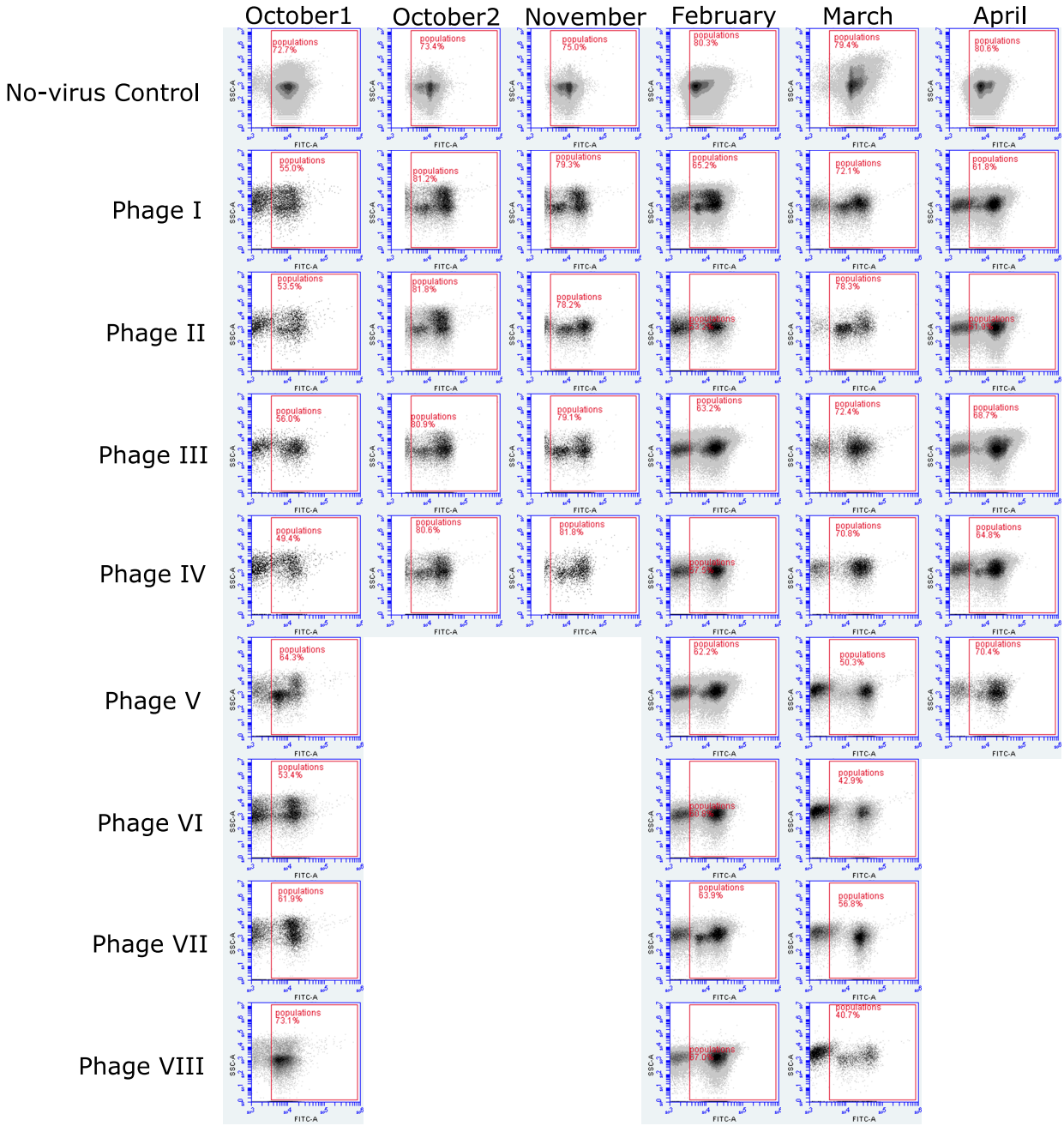


Fig. S3 Cytograms of infected vs. uninfected SAR11 warm-water ecotype HTCC7211 cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 72.7% - 84.3% of total events, whereas infected cultures range from 27.7% - 78.6%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to identify putative infections in cultures.



845
846 **Fig. S4** Cytograms of infected vs uninfected SAR11 cold-water ecotype HTCC1062 cultures to
847 isolate phages. Events within the acquisition range ("populations") of no-virus controls range from
848 72.7% - 80.5% of total events, whereas infected cultures range from 40.7% - 81.8%. Note the increase
849 in low fluorescence (FITC-A) events for phage cultures, which can be used to identify putative
850 infections in cultures.
851

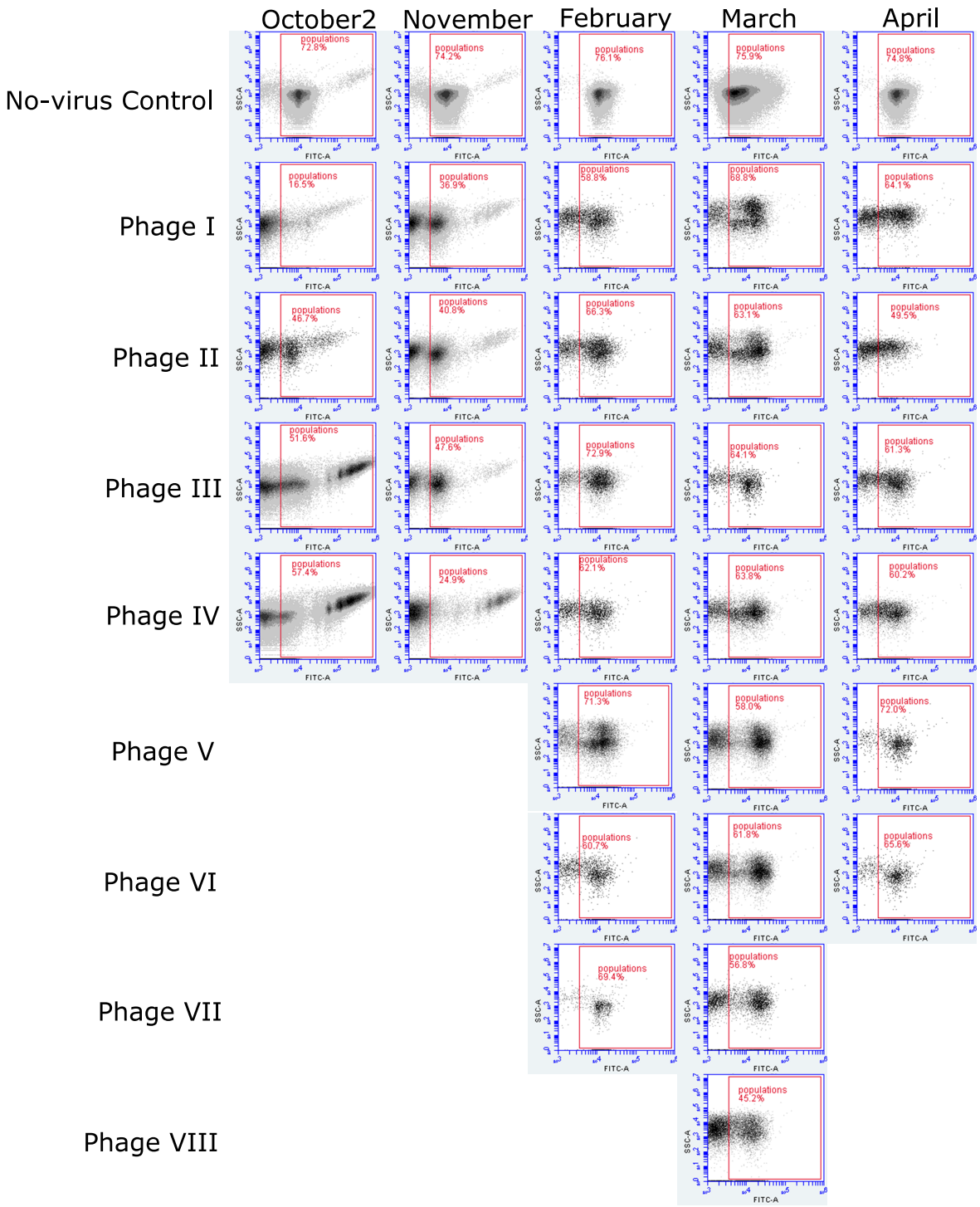


Fig. S5 Cytograms of infected vs uninfected new SAR11 isolate H2P3α cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 72.8% - 76.1% of total events, whereas infected cultures range from 16.5% - 72.0%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to identify putative infections in cultures.

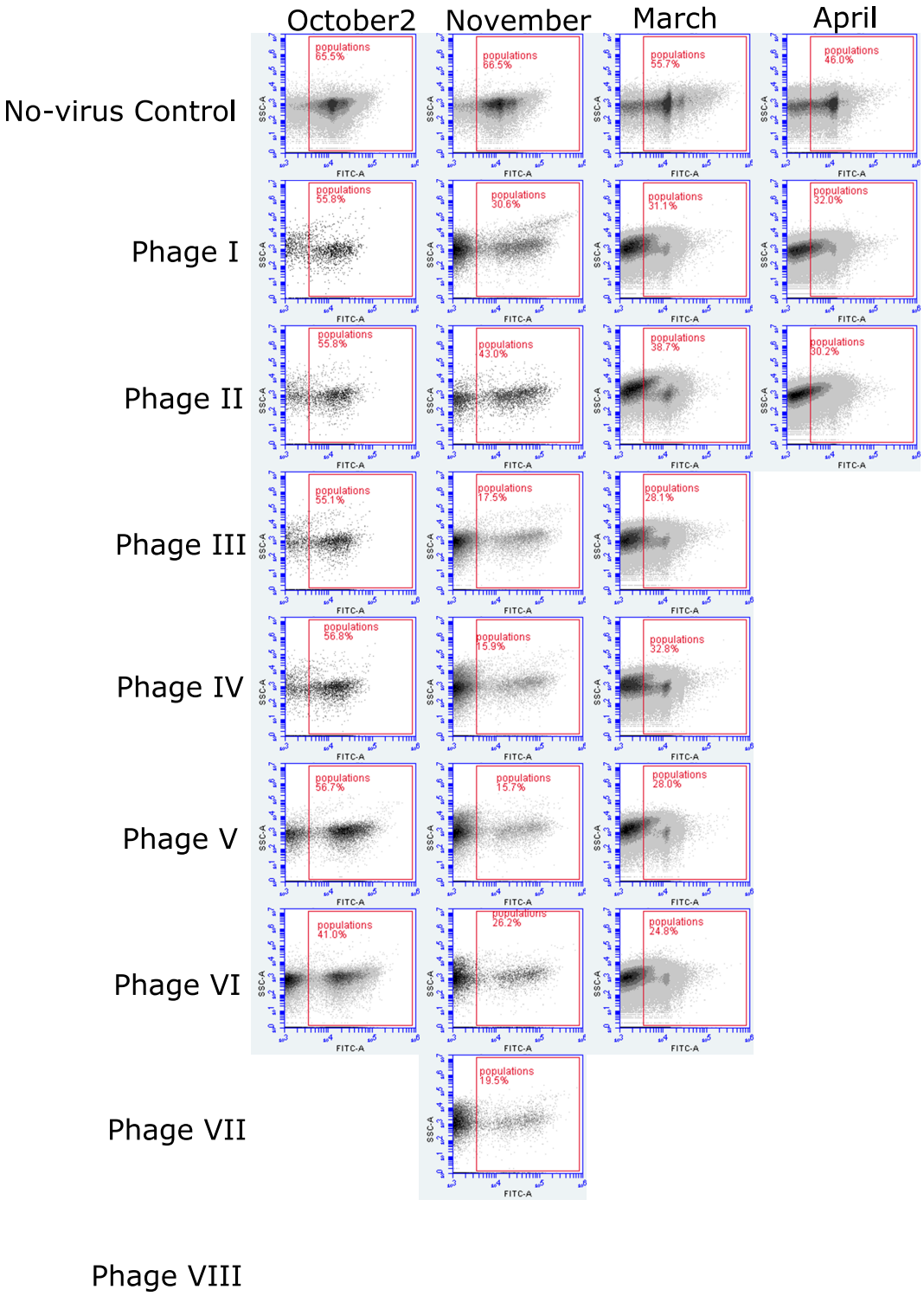


Fig. S6. Cytograms of infected vs uninfected new OM43 isolate H5P1 cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 46.0% - 66.5% of total events, whereas infected cultures range from 15.7% - 56.8%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to spot infections in cultures. The line between noise and the population peaks of the cytogram for March and April infected cultures is not, therefore the population percentage is likely biased by the acquisition range settings.

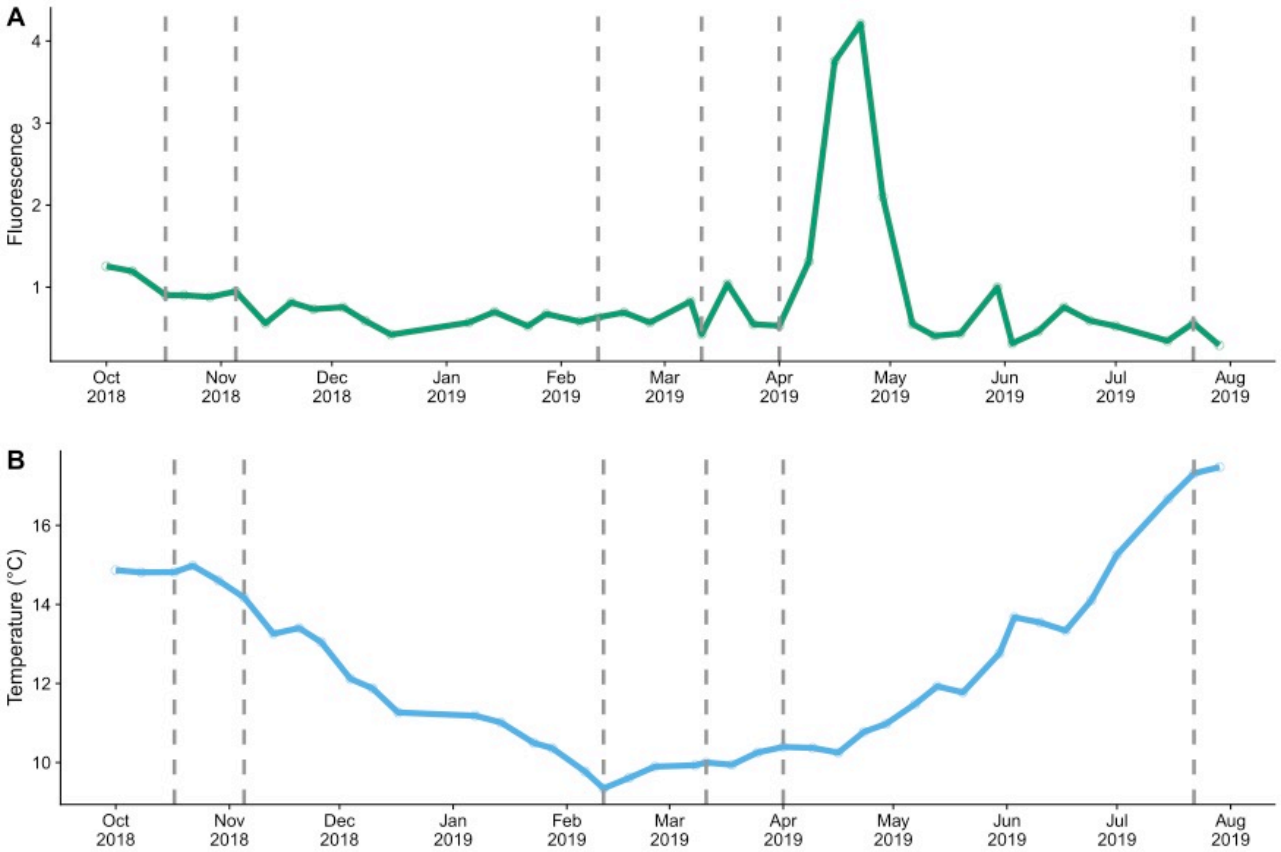


Fig. S7 (A) Fluorescence and **(B)** Temperature at Western English Channel sampling station L4 (50°15.00N; 4°13.00W) over the course of the study. Grey dashed lines represent samples for viral isolation.

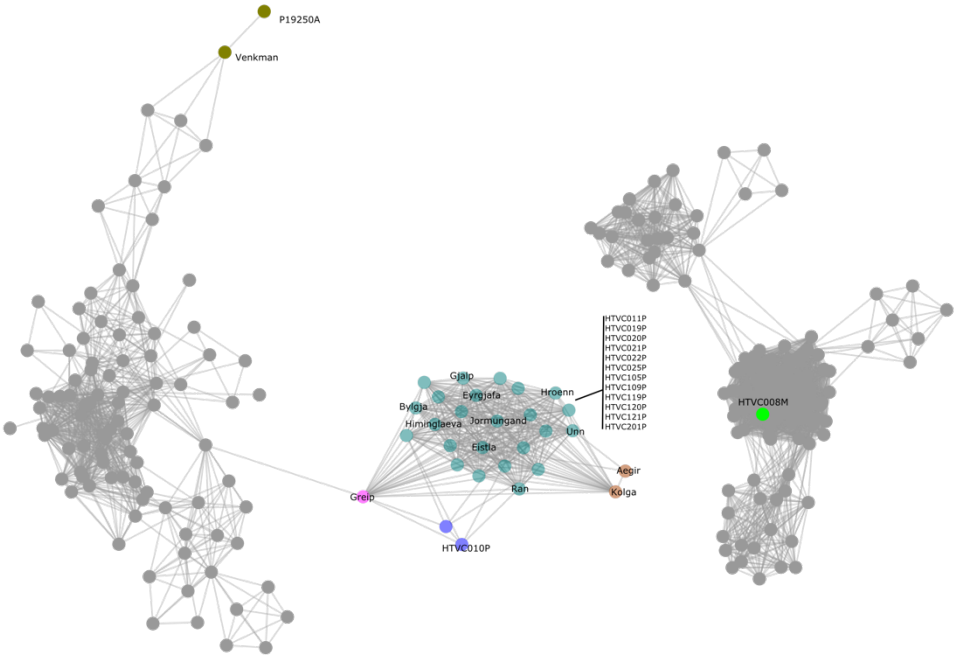


Fig. S8 Shared protein cluster network content of known phage isolates (30, 31) and viral isolates from this study and viruses within the NCBI Bacterial and Archaeal Viral RefSeq V85 with ICTV + NCBI taxonomy database (VConTACT2, default settings); different colours represent the separate viral clusters of ICTV-recognised genera

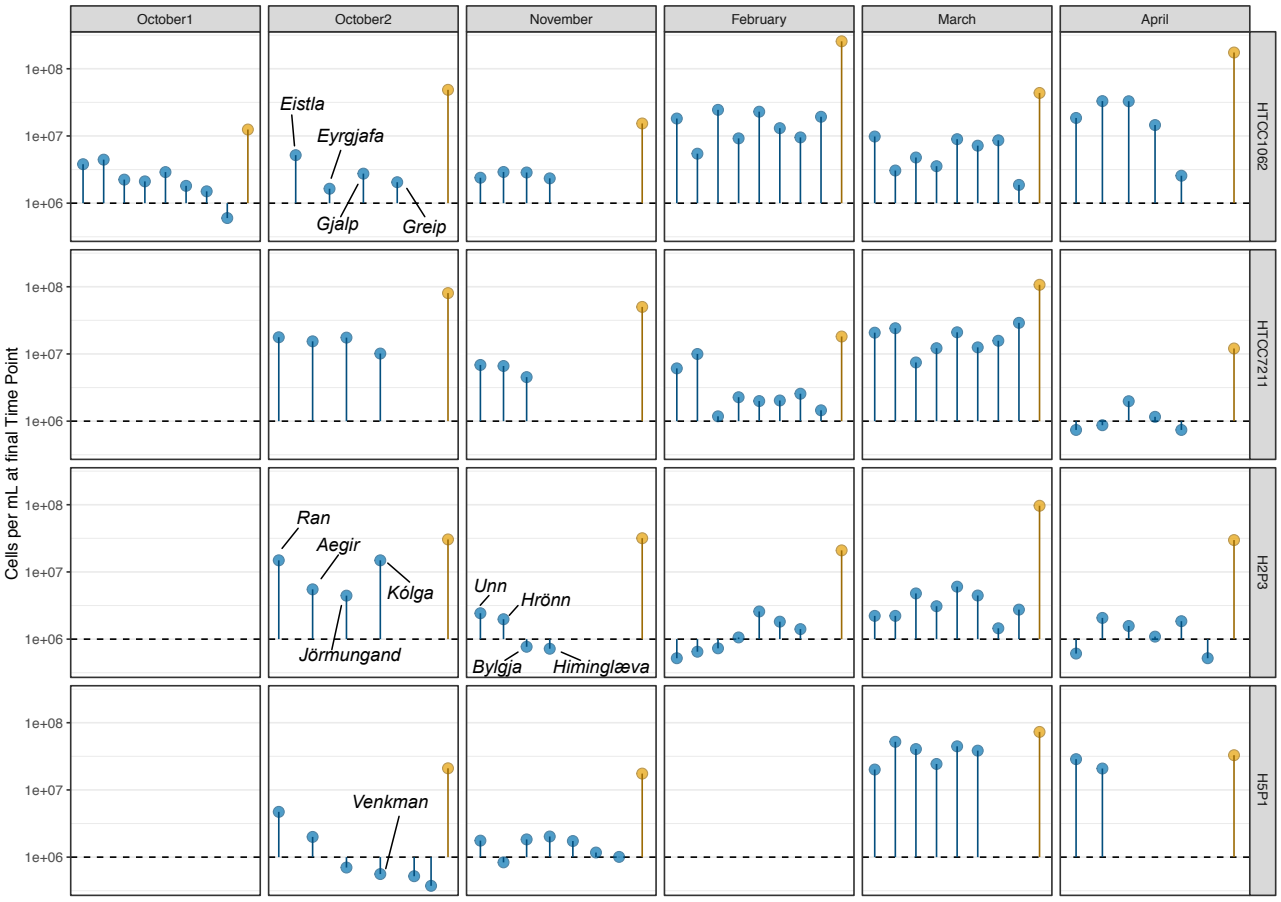


Fig. S9. Final cell densities of infected (blue) vs. uninfected (orange) hosts for 115 of the 117 viruses isolated in this study (two viruses isolated on OM43 D12P1 in July omitted). Dashed line at 10^6 cells per mL represents host density at T_0. Points below this line indicate viruses reduced host density below the inoculum density. Named points refer to genomes sequenced in this study.

Supplementary File S10 List of sample information and descriptions of viral isolates.

Supplementary File S11 List of average nucleotide identities and average amino acid identities within and between viral populations.

Supplementary File S12 Video of an infection of *Pelagibacter ubique* HTCC1062 showing how host cell density achieves a steady state in infected cultures that is lower than that of uninfected cultures. Associated cytograms are shown to demonstrate shifts in fluorescence associated with viral infection.