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

Holger H. Buchholz[1], Michelle Michelsen[1], Luis M. Bolaños[1], Emily Browne[1], Michael J. Allen[1,2], Ben Temperton[1*]

[1]University of Exeter, School of Biosciences, Exeter, UK;

[2]Plymouth Marine Laboratory, Plymouth, UK

*Corresponding Author

b.temperton@exeter.ac.uk

Running Title: Cultivation of viruses of fastidious bacteria

Keywords: 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 experimental approaches to quantify the importance of viruses in nutrient recycling. A resurgence in cultivation efforts has improved the availability of fastidious bacteria for hypothesis testing, but this has not been matched by similar efforts to cultivate their associated bacteriophages. Here, we describe a high-throughput method for isolating important virus-host systems for fastidious heterotrophic bacteria that couples advances in culturing of hosts with sequential enrichment and isolation of associated phages. Applied to six monthly samples from the Western English Channel, we first isolated one new member of the globally dominant bacterial SAR11 clade and three new members of the methylotrophic bacterial clade OM43. We used these as bait to isolate 117 new phages including the first known siphophage infecting SAR11, and the first isolated phage for OM43. Genomic analyses of 13 novel viruses revealed representatives of three new viral genera, and infection assays showed that the viruses infecting SAR11 have ecotype-specific host-ranges. Similar to the abundant human-associated phage ϕCrAss001, infection dynamics within the majority of isolates suggested either prevalent lysogeny or chronic infection, despite a lack of associated genes; or host phenotypic bistability with lysis putatively maintained within a susceptible subpopulation. Broader representation of important 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 is 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 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 play an important role in nutrient cycling and global carbon budgets [14]. Furthermore, isolation of new viruses provides complete or near-complete viral genomes, with concrete evidence of known hosts. Such systems are critical to the development, ground-truthing and application of computational methods to identify and classify 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 uncultured viruses in metagenomic data (e.g. WIsH [21, 22]).

63 Viruses of primary producers, such as cyanophages are both well represented with model systems
64 and well-studied in the laboratory. In contrast, virus-host model systems for similarly important and
65 abundant marine heterotrophic bacteria are rare. Isolated viruses infecting heterotrophs are heavily
66 biased towards those with fast-growing, copiotrophic hosts that grow readily on solid agar, enabling
67 the use of plaque assays for viral isolation. Such systems are not representative of the vast majority
68 of heterotrophs in nutrient-limited soil and aquatic environments, which are dominated by slow-
69 growing, oligotrophic taxa with few regulatory mechanisms and complex auxotrophies that limit
70 growth on solid media [23–26]. Advances in Dilution-to-Extinction culturing of ecologically
71 important hosts have enabled the cultivation of many fastidious bacterial taxa that are not amenable
72 to growth on solid media from soil [27], marine [28, 29] and freshwater environments [30]. Without
73 plaque assays to facilitate isolation and purification of viral isolates, cultivation of viruses infecting
74 fastidious taxa in liquid media is challenging, and further exacerbated by the slow growth rates and
75 complex nutrient requirements of their hosts. Paucity of such model systems introduces significant
76 bias in our understanding of viral influence on global carbon biogeochemical cycles. Therefore, it is
77 important that the efforts to isolate heterotrophic bacterial taxa for experimentation and synthetic
78 ecology is matched by efforts to isolate their associated viruses.
79
80 Here, we adapted recent advances in Dilution-to-Extinction culturing of hosts [29], and protocols to
81 isolate viruses from liquid media [31] to improve efficiency of cultivating novel virus-host systems
82 for fastidious taxa: First, we selected the ecologically significant SAR11 and OM43 heterotrophic
83 marine clades as models for viral isolation; Second, we used sequential enrichment of viruses from
84 natural communities on target hosts to improve rates of viral isolation [32]; Third, we replaced the
85 requirement for time-intensive epifluorescent microscopy with identification of putative viral
86 infection by comparing infected and uninfected hosts by flow cytometry, followed by confirmation
87 using Transmission Electron Microscopy (TEM). These clades are abundant and important to global

88 carbon biogeochemistry [23, 33–37], but little is known about their associated viruses. In the case of
89 viruses infecting SAR11, two challenges limit our ability to evaluate host-virus ecology in natural
90 communities: (1) Assembly of abundant and microdiverse genomes from viral metagenomes presents
91 a challenge to short-read assembly methods, resulting in underrepresentation in subsequent datasets
92 [12, 18]. This was demonstrated in the successful isolation of the first known pelagiphages by
93 culturing, including the globally dominant HTVC010P, which, prior to its isolation, was entirely
94 missed in marine viromes [38]; (2) poor representation of viral taxa in databases limits our capacity
95 to accurately train machine learning approaches for *in silico* host prediction [12, 39, 40]. This results
96 in either a lack of host information for abundant viral contigs, or worse, incorrect assignment of host
97 to viral contigs, confounding ecological interpretation of data. Some important taxa, such as the
98 OM43 clade, which plays an important role in oxidation of volatile carbon associated with
99 phytoplankton blooms [33, 35, 41], lack any isolated viruses with experimentally confirmed hosts. In
100 addition, both SAR11 and OM43 represent model organisms for genome streamlining as a result of
101 nutrient-limited selection [42]. The effect of genome minimalism on viral infection dynamics is
102 poorly understood, but critical to evaluating the impact of predator-prey dynamics on global marine
103 carbon budgets. In this study, novel SAR11 and OM43 representatives from the Western English
104 Channel were isolated and used as bait to isolate associated viruses. We increased the initial
105 concentration of viruses in natural seawater samples by tangential flow filtration, followed by
106 inoculation of cultures and one to three rounds of sequential enrichment on target hosts in 96-well
107 plates. This yielded 117 viral isolates from 218 inoculated cultures from seven monthly water samples
108 (September 2018 to July 2019). A subsample of putative viral isolates for both clades were sequenced,
109 providing 13 novel viral genomes, including the first known siphovirus to infect SAR11 and the first
110 known virus-host model for OM43.
111

112 **Results and Discussion**

113

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

116 Dilution-to-Extinction culturing for host taxa using natural seawater-based medium was performed
117 from a water sample collected in September 2017 from the Western English Channel, and yielded the
118 first SAR11 strain (named H2P3α) and the first three OM43 strains (named C6P1, D12P1, H5P1)
119 from this region. The full-length 16S rRNA gene of *Pelagibacter sp.* H2P3α was 100% identical to
120 that of the warm water SAR11 ecotype *Pelagibacter bermudensis* HTCC7211 (subclade 1a.3) and
121 was considered to be a local variant [43] (Supplementary Fig. 1A). All three novel OM43 isolates
122 were most closely related to *Methylophilales sp*. HTCC2181, a streamlined member of the OM43
123 clade with a 1.3 Mbp genome, isolated from surface water of the North Eastern Pacific [44] (C6P1
124 96.17%, D12P1 96.62%, H5P1 97.79% nucleotide identity across the full 16S rRNA gene)
125 (Supplementary Fig. 1B). The average nucleotide identity of the 16S rRNA gene of isolates CP61,
126 D12P1 and H5P1 to each other was ~98.46% (Supplementary Table 1), suggesting they are
127 representatives of the same genus [45].

128

129 **An efficient, low-cost method of isolating new viruses yielded 117 new viral isolates for SAR11**
130 **and OM43 taxa**

131 Using the four new hosts from above and established SAR11 isolates *Pelagibacter ubique* HTCC1062
132 (subclade 1a.1) and HTCC7211 (subclade 1a.3), we developed an optimised viral isolation pipeline
133 (Fig. 1) and applied it to six monthly water samples from the Western English Channel, taken between
134 September 2018 to April 2019 (Supplementary Table 2). Each month, we concentrated a natural viral
135 community with tangential flow filtration and used it to inoculate one to two 96-well Teflon plates
136 containing host cultures at ~10^6 cells·mL^{-1}. Plates were monitored by flow cytometry and growth of
137 putatively infected cultures was compared to those of unamended controls over the course of ~2

138 weeks to account for the slow growth rates of SAR11 and OM43 [35, 46]. An additional sample was
139 taken in July 2019 to attempt viral isolation on OM43 strains D12P1 and H5P1. Out of a total of 218
140 cultures amended with concentrated viral populations, 117 viruses were isolated, purified and still
141 infective after at least three passages, with repeatable differences observed in cytograms of infected
142 cultures and control cultures (Supplementary Fig. 2-5). This represents an overall isolation efficiency
143 of 53% and an average yield of 18 viruses per environmental sample (Fig. 2). For 90% of inoculated
144 SAR11 cultures (94 out of 105) we observed evidence of viral infection, and the putative viral isolate
145 could be propagated and purified, fulfilling Koch's postulates for confirming a pathogenic agent. For
146 OM43, 23 out of 113 (20%) inoculated cultures yielded positive infections that could be similarly
147 propagated. All viral isolations required between one and three rounds of virus enrichment
148 (Supplementary Table 3) before changes in host growth curves between infected and uninfected
149 cultures could be observed. This suggests putatively rare viruses can be enriched within three rounds
150 to a level at which infection can be observed on a flow cytometer.
151
152 **New viruses represent novel viral populations and support established ANI cut-offs for**
153 **ecologically discrete viral ecotypes**
154 Due to the rate-limiting step of culturing sufficient biomass for extraction of viral DNA, we
155 subselected 16 viral isolates based on availability in November 2018 across four different hosts
156 (HTCC1062, HTCC7211, H2P3α and H5P1) for Illumina sequencing to >30-fold coverage (Table
157 1). Three out of 16 sequenced samples (~19%, two from host HTCC7211, one from OM43 host
158 H5P1) failed to assemble into single viral contigs, in line with previously reported failure rates of 18-
159 39% for phages of *Escherichia coli* and *Salmonella spp.* [47]. For 11 of the remaining 13 samples (12
160 from SAR11 hosts and one from OM43), each individual sequence assembly was identified as a
161 complete viral genome by VirSorter [15] and 95-100% complete using CheckV [48]. All assemblies
162 yielded a single viral contig (categories 1 or 2, >15kbp) per sample, indicating that our purification

process was effective in recovering pure viral isolates. Interestingly, all viral isolate genomes were classified by VirSorter as Category 2 ("likely virus") rather than Category 1 ("most confident"), despite being complete - indicating either a lack of viral hallmark genes or a lack of enrichment of viral genes on the contigs. This finding matches our own observations for other isolated viruses infecting SAR11 (data not shown), and suggests VirSorter classification of pelagiphages is conservative. Viral isolates were named in accordance with current ICTV guidelines for prokaryotic virus taxonomy [49], with names selected from Norse mythology and folklore, and contemporary culture (Table 1).

Viral populations are defined as discrete ecological and evolutionary units that form non-overlapping clouds of sequence space, and previous work in cyanophage populations have shown that viral populations can be delineated into populations using an average nucleotide identity (ANI) cutoff of 95% [50]. Pairwise ANI was calculated between the thirteen successfully sequenced viral genomes from this study and 85 other known or putative pelagiphages [39, 51, 52]. Pairwise ANI ranged between 77.5-100%, with a discrete distribution between 96.4-100.0% (Supplementary Fig. 6, Supplementary Table 4). This is in agreement with previous work in cyanophages [50, 53] and supports the broad use of proposed boundary cutoffs to define viral populations within viral metagenomic assemblies [10, 11]. At the proposed ANI cutoffs of 95% over 85% length [18], our 13 new viruses clustered into six viral populations, ranging from singletons to a viral population with four members (Table 1). Phages within the same populations were all isolated from the same environmental sample and on the same host, in agreement with their classification as discrete ecological and evolutionary units. All viruses isolated in this study formed their own exclusive viral populations, with no representatives from either known isolates [38, 52] or fosmid-derived [39] genomes from other studies, indicating that a high degree of viral population diversity remains to be discovered in the Western English Channel and beyond.

188 **Current pelagiphage isolates can be organised into five distinct phylogenetic clades**

189 To evaluate isolate diversity at higher taxonomic organisation, we picked one representative from
190 each of our six viral populations and compared them to previous isolates and fosmid-derived phage
191 sequences using three approaches: First, phylogenetic analyses were performed based on conserved
192 genes in known pelagiphage isolates and closely-related taxa from viral metagenomic surveys [11,
193 12]; Second, raw hypergeometric probability of shared gene content was calculated (to capture
194 broader relationships and account for genomic mosaicism) [54]; Third, genomes were organised into
195 ICTV-recognized genera using vConTACT2. vConTACT2 initially derives viral clusters using a
196 hypergeometric approach, with subsequent refinement with Euclidean-distance-based hierarchical
197 clustering to split mismatched, 'lumped' viral clusters [19]. All three approaches were congruent -
198 clustering on probability of shared gene content organised pelagiphage genomes into four main
199 clusters and numerous singleton genomes (Fig. 3). This was broadly supported by phylogenetic
200 (Supplementary Fig. 7) and vConTACT2 (Supplementary Fig. 8) classification approaches. Cluster
201 A contained 23 members (nine from fosmid-derived contigs [39]; eleven previously isolated
202 pelagiphages [52, 55]) and Pelagibacter phages *Ran* (EXVC014P), *Bylgja* (EXVC010P) and *Eistla*
203 (EXVC025P) from this study. Cluster B contained two previously isolated pelagiphages, one fosmid-
204 derived contig and Pelagibacter phage *Eyrgjafa* (EXVC018P) from this study. All viruses in Clusters
205 A and B were assigned to a single viral genus by vConTACT2 that also contained 12 previously
206 isolated pelagiphages [38, 52]. Cluster C only contained fosmid-derived contigs from the
207 Mediterranean [39], with no isolated representatives, marking it an important target for future
208 isolation attempts. Cluster D contained eight fosmid-derived contigs, Pelagibacter phage HTVC010P,
209 and Pelagibacter phage *Greip* (EXVC021P) from this study. Pelagibacter phage *Kolga* (EXVC016S)
210 from this study and the only known Pelagimyophage HTVC008M from a previous study [38] fell
211 outside the four main clusters. VConTACT2 split Cluster D, leaving Pelagibacter phages *Greip* and

212 *Kolga* as members of two singleton clusters, suggesting they are the first cultured representatives of
213 novel viral genera and distinct from the globally ubiquitous Pelagibacter phage HTVC010P.
214
215 **Novel pelagiphages are ecotype-specific and persist in the community.**
216 Community composition analysis using 16S rRNA genes showed that during the sampling period
217 (September 2018 to July 2019) the SAR11 contribution relative to the total number of sequences
218 ranged from a minimum of 28.0% in July 2019 to a maximum of 42.2% in September 2018
219 (Supplementary Fig.9). At the clade level, SAR11 composition was relatively stable over time, except
220 for clade II roughly doubling its relative contribution from 15.1% to about 29.6% between February
221 and April. The SAR11 community was dominated by clade I overall throughout the sampling period.
222 Within clade I, the ratio between SAR11 subclade Ia.1 (cold-water ecotype) and Ia.3 (warm-water
223 ecotype) showed that the warm-water ecotype dominated from September to November as well as
224 July (Figure 2A). During the coldest months in the Western English Channel (February to April) the
225 cold-water ecotype became as abundant as the warm-water ecotype with roughly a 1:1 ratio. Overall,
226 temperature ranged from 9.3 °C - 14.1 °C from October to April, when isolations were attempted on
227 all strains (Supplementary Table 2). HTCC1062 and HTCC7211 have specific growth rates of ~0.22
228 and ~0.12 divisions per day at 10 °C, respectively [46]. Our new SAR11 isolate from the Western
229 English Channel, *Pelagibacter sp.* H2P3α, showed similar growth rates to *Pelagibacter bermudensis*
230 HTCC7211 (Supplementary Fig. 10), with specific growth rates of ~0.10, ~0.45 and ~0.84 divisions
231 per day at 10, 18 and 25 °C, respectively. Therefore, measured *in situ* temperatures during our
232 sampling period were sufficient to support slow growth of warm-water ecotypes even during winter
233 months, potentially providing sufficient prey to support a population of warm-water ecotype specific
234 phages. This result shows that slow growth of warm water ecotypes at *in situ* temperatures is possible,
235 supporting the finding of the 16S community analysis that this ecotype persists throughout the year.

236 An alternative explanation is that isolated viruses have a broad host range that encompasses both
237 warm- and cold-water ecotypes of the SAR11 subclade Ia. We tested the host range (Table 2) of six
238 pelagiphages (one of each viral population), isolated from samples in October 2018 and November
239 2018 (*in situ* water temperature of 14.8 °C and 14.2 °C respectively) across the three SAR11 strains
240 (Table 1). Pelagibacter phages *Eistla, Eyrgjafa* and *Greip* all infected cold-water ecotype HTCC1062
241 exclusively, while Pelagibacter phages *Ran* and *Kolga* only infected warm-water ecotypes
242 HTCC7211 and H2P3α. Pelagibacter phage *Bylgia* was the only virus that could infect both warm
243 and cold-water ecotypes. Therefore, our new pelagiphages appear to be broadly ecotype-specific,
244 confirming previous findings [52]. Our results suggest overall that pelagiphages persist in the water
245 column throughout the year in sufficient densities to be isolated by our enrichment method, despite
246 ecotypic specificity and fluctuations in warm- and cold-water ecotype community contributions of
247 SAR11 subclade 1a. If concentration and enrichment of viruses during isolation is sufficient to
248 successfully isolate even low abundance phages then a comprehensive library of representative phage
249 isolates could be generated with relatively modest sampling effort across a few locations.
250

251 **Pelagibacter phages Kolga and Aegir- the first siphovirus infecting SAR11**

252 The 25 previously known viral isolates infecting SAR11 comprise 24 podoviruses and one myovirus
253 [38, 52, 56]. Previous cultivation efforts for viruses of SAR11 have not isolated any siphoviruses, nor
254 are any known from viral metagenomic studies. We isolated and sequenced the two Pelagibacter
255 phages *Kolga* and *Aegir* using host H2P3α as bait (Fig. 4A-C). Transmission electron microscopy
256 (TEM), which showed evidence of a long tail (Fig. 4B), suggested classification as the first reported
257 siphoviruses infecting members of the SAR11 clade.
258 Pelagibacter phages *Aegir* and *Kolga* were classified as members of the same population using a
259 boundary cutoff of 95% ANI over 85% contig length, however, *Aegir* had a length of 18,297 bp
260 compared to 48,659 bp in *Kolga*, therefore we considered *Aegir* to be a partial genome of the same

261 viral population. *Kolga* did not share a significant number of genes with known SAR11 podoviruses
262 (Fig. 3), and did not cluster with other pelagiphages using hypergeometric analysis based on shared
263 gene content. VConTACT2 also grouped *Aegir* and *Kolga* into one cluster without any other known
264 viruses, suggesting it represents a novel viral genus. A number of genes found in *Kolga* were shared
265 with other known siphoviruses (associated with different hosts) such as the Bordetella phage LK3.
266 Screening of contigs from the Global Ocean Virome dataset [11] identified six contigs from various
267 ecological zones which shared a viral cluster with *Kolga*, but which belonged to different viral
268 populations based on network analysis using vConTACT2. Phylogenetic analysis of all Pelagibacter
269 phage TerL genes indicates that the closest known relatives to *Kolga* and *Kolga*-like contigs are
270 members of the Myoviridae family, supporting a classification as a distinct and novel viral group
271 (Fig. 4C, Supplementary Fig. 7). In *Kolga* 67% of encoded genes could not be functionally annotated,
272 and out of all hypothetical genes identified on *Kolga*, only three hypothetical genes were shared with
273 SAR11 podoviruses. In contrast, on average ~90% of genes without known function identified within
274 our novel SAR11 podoviruses are shared between different SAR11 podoviruses. *Kolga* possesses a
275 tail tip J protein (Fig. 4A), often found in phages with long non-contractile tails such as E. coli phage
276 λ, where it plays a role in DNA injection during cell entry and tail assembly [57]. *Kolga* also encodes
277 a small S21 subunit of the 30S ribosomal gene structurally similar to the ones found in Pelagibacter
278 phage HTVC008M, Pelagimyophage-like contigs [58], and hosts HTCC7211 and H2P3α. Encoding
279 ribosomal genes is a feature found in numerous myoviruses and siphoviruses [59]. The S21 gene is
280 involved in translation initiation and needed for the binding of mRNA [60]. Virally-encoded S21
281 genes may provide a competitive advantage for the phage as it could replace cellular S21 and assist
282 in the translation of viral transcripts. *Kolga* may need its S21 gene for shifting the translational frame,
283 as it has been shown that for some members of the Caudovirales the production of tail components is
284 dependent on programmed translational frameshifting [61]. Given the constitutive nature of gene

285 expression in genomically streamlined bacteria [62], genes such as S21 may also provide the virus
286 with a mechanism to manipulate host metabolism in the absence of typical promoters and repressors.
287
288 **First methylophages for marine OM43 isolated**
289 Isolation of novel phages for the OM43 clade yielded 23 positive infections, with efficiencies ranging
290 from 0% (no viruses isolated on host C6P1) to 45% on H5P1 (Fig. 2). To the best of our knowledge
291 these are the first reported viruses infecting members of the OM43 clade. One explanation for the
292 lower efficiency of isolation of phages infecting OM43 is simply one of lower host abundance
293 concomitant with lower phage abundance in the viral community, reducing the likelihood of infective
294 viruses coming into contact with susceptible and permissive cells. OM43 are closely associated with
295 metabolism of extracellular substrates from phytoplankton blooms [63], but have low abundance
296 outside of phytoplankton spring blooms [33]. Our water samples were not associated with high *in-*
297 *situ* fluorescence (used as proxy measurement for phytoplankton), and missed the April 2019 spring
298 bloom by about two weeks (Supplementary Fig. 11). Based on 16S community analysis, the OM43
299 contribution to the bacterial community (0.7%) was lowest during sampling in April 2019 and highest
300 in October 2018 (1.9%). Relative to all other OM43, H5P1 was the most abundant OM43 in the
301 Western English Channel, contributing more than half of all OM43 in November (Fig. 2B). This
302 could explain the higher success rate of isolating phages on host H5P1 compared to C6P1 and D12P1.
303 Sequencing of the first OM43 phages isolate, Methylophilales phage *Venkman* (EXVC282S),
304 returned a single genome 38,624 bp long (31.9% GC content), which was linear but complete
305 (Supplementary Table 5). *Venkman* encodes genes (Fig. 5A) with similar synteny and function to the
306 siphovirus P19250A (38,562 bp) that infects freshwater *Methylophilales* LD28, which are often
307 considered a freshwater variant of OM43 [64, 65]. Unlike the siphovirus P19250A, TEM images
308 indicated that OM43 phage *Venkman* had a short tail (Fig. 5B) similar to podoviruses, though it is
309 possible that tail structures were lost during grid preparation. *Venkman* shared a viral population with

310 a 23kbp contig from the Western English Channel virome assembled from short-read data
311 (WEC_HYBRID_01170), suggesting this viral type does not suffer from the issues of high abundance
312 and microdiversity that challenge assembly of some pelagiphages. Phylogenetic analysis of
313 concatenated TerL and exonuclease genes indicated that *Venkman* is most closely related to P19250A
314 and other siphoviruses infecting different Proteobacteria (Fig. 5C). However, branch support values
315 were low, despite numerous attempts to refine the tree with different approaches (see Supplementary
316 Methods). We identified a number of common phage proteins such as a capsid protein, terminase,
317 nucleases and tail structural proteins, and the remaining 54% of genes were hypothetical.
318 VConTACT2 assigned OM43 phage *Venkman* and LD28 phage P19250A to the same genus-level
319 cluster, therefore the OM43 phage *Venkman* may be a marine variant of the freshwater LD28
320 siphovirus P19250A.
321
322 **Global abundance of novel isolates highlights niche-specificity and low representation in**
323 **existing datasets**
324 To evaluate relative global abundance of our new phage isolates, existing virome datasets from the
325 Global Ocean Virome survey [11] and the Western English Channel [12] were randomly subsampled
326 to 5 million reads and competitively recruited against genomes of viral population representatives
327 from this study as well as previously isolated pelagiphages (see Supplementary methods). Overall,
328 with the exception of Pelagibacter phage HTVC010P, pelagiphages were poorly represented in
329 samples from temperate-tropical mesopelagic (TT-MES) and Antarctic (ANT) ecological zones, as
330 defined in [11] (Fig. 6). Phages isolated in this study were neither globally ubiquitous, nor abundant
331 in the single Western English Channel virome (with the exception of phage *Ran*). Phages *Bylgja,*
332 *Himinglaeva, Eistla* and *Eyrgjafa* did not achieve the minimum cutoff of 40% genome coverage to
333 be classified as present [18] in any of the viromes tested. *Ran* was the only phage isolated in this study
334 with representation in at least two samples from temperate-tropical epipelagic (TT-EPI) zones,

335 concordant with its host specificity of warm-water SAR11 ecotypes (*Pelagibacter bermudensis*
336 HTCC7211 and H2P3α from this study (Table 2). Similarly, the Western English Channel virome
337 was taken in September 2016, when waters are usually highly stratified after summer heating [66]
338 and warm water ecotypes of SAR11 dominate the microbial community (Fig. 2A). Other viral
339 populations from this study also isolated on warm water SAR11 ecotypes (phages *Kolga, Bylgja* and
340 *Himinglaeva*) were either absent or below limits of detection in this sample. This coupled with the
341 low global abundance of these viruses suggests that pelagiphage communities comprise few highly
342 abundant taxa and a long rare tail. Pelagibacter phage *Greip* was detected in seven samples, six of
343 which were Arctic (ARC) samples from a discrete region of the Arctic characterized by low nutrient
344 ratios [11]. In three of those samples (191_SRF, 193_SRF, 196_SRF), HTVC010P was not detected
345 and in the other three (206_SRF, 208_SRF and 209_SRF) *Greip* was 1.5 to 6.6-fold more abundant
346 than HTVC010P, identifying *Greip* as an abundant arctic pelagiphage ecotype. The fact that *Greip*
347 was isolated on the cold-water ecotype of SAR11 (Pelagibacter ubique HTCC1062) and does not
348 infect either warm-water ecotype HTCC7211 or H2P3 supports the hypothesis that host niche
349 specificity shapes the phylogeography of associated viral taxa.
350
351 In the Western English Channel viral metagenomes, OM43 phage *Venkman* was the third most
352 abundant (2,699 RPKM) after Pelagibacter phages HTVC027P (5,023 RPKM) and HTVC010P
353 (5,026 RPKM) identifying it as an ecologically important virus in this coastal microbial community
354 (Fig. 6). In contrast, neither *Venkman* nor LD28 phage P19250A were identified in the majority of
355 samples from GOV2, except in one sample (67_SRF) in the South Atlantic Ocean (257 RPKM against
356 *Venkman*). Sample 67_SRF was also classified as a coastal biome by the authors [11]. These results
357 coincide with previous reports that OM43, and presumably their associated viruses as a consequence
358 of host abundance, are important in some coastal regions, but largely absent in open-ocean systems
359 [33, 41].

360

361 **Unusual host-virus dynamics are prevalent in isolated phages**

362 Pelagibacter phages *Bylgja*, *Eyrgjafa*, *Ran* and *Eistla* all encode endonucleases and exonucleases
363 (Supplementary Fig. 12) and cluster by shared protein content with other pelagiphages (Cluster A and
364 B), such as e.g. HTVC011P and HTVC025P, shown previously to integrate into host genomes (Fig.
365 3) [52]. We therefore predict that all our phages within Clusters A and B are temperate phages.
366 *Eyrgjafa* also encodes a tRNA-Leu, that has 85% nucleotide identity over the first 34 bases of the
367 tRNA-Leu of its host HTCC1062, suggesting a putative integration site into the host genome [67].
368 To date, 16 of the 29 viruses previously isolated on SAR11 strains have either been shown to be
369 capable of lysogeny, and/or encode genes associated with a temperate infection cycle. In contrast,
370 viruses such as the Pelagibacter phages *Greip,* and the abundant HTVC010P in Cluster D do not
371 possess any genes associated with lysogeny, and would therefore be classified as exclusively lytic.
372 Viruses in Clusters A and B (putatively temperate phage), were of much lower abundance in the
373 environment compared HTVC010P which was among the most abundant pelagiphages in the Western
374 English Channel (Fig. 6), suggesting a possible ecological difference between two groups of viruses
375 with different infectivity strategies.

376 Interestingly, growth curves of hosts infected with Pelagibacter phage *Greip* and other isolated phages
377 deviated from the expected decay in cell abundance associated with viral lysis and previously
378 observed in isolated pelagiphages [38] (Fig.1 Bii). The first pelagiphages HTVC010P and
379 HTVC008M were isolated from the warm waters of the Sargasso Sea, and HTVC011P and
380 HTVC019P were isolated from the colder waters of the Oregon Coast. All four strains were
381 propagated on the cold-water SAR11 ecotype *P. ubique* HTCC1062. In all cases, host density was
382 reduced from ~8×10^6 cells·mL^{-1} at T_0 to <10^6 cells·mL^{-1} over a 60-72h period. Viruses isolated
383 from warm waters took 17% longer than those from cold water to do so [38], suggesting that
384 suboptimal hosts reduced the rate of infection as shown in cyanophages [68]. In contrast, infection

385 dynamics of our isolates often resulted in host density of infected cultures growing to a steady state,
386 but at a lower cell density than uninfected cells. (Supplementary Video 1), irrespective of cluster,
387 population assignment or evidence of genes associated with temperate lifestyles. Out of 117 viruses
388 isolated in this study, only 16 infections reduced host abundance below their inoculum density of 10^6
389 cells·mL^{-1}. In 53 infections, densities of infected cells increased to within an order of magnitude of
390 uninfected cells (Supplementary Fig. 13), but demonstrated clear evidence of viral infection in
391 cytograms, TEMs and subsequent recovery of viral genomes in selected samples.
392 Similar patterns of infection were recently reported in the extremely abundant bacteriophage
393 ϕCrAss001 found in the human gut, where infection of exponentially growing cells of *B. intestinalis*
394 919/174 did not result in complete culture lysis, but caused a delay in stationary phase onset time and
395 final density, despite lacking genes associated with lysogeny. As with our study, the authors observed
396 that this only occurred in liquid culture, and isolation of the virus required numerous rounds of
397 enrichment. They postulated that the virus may cause a successful infection in only a subset of host
398 cells, with the remainder exhibiting alternative interactions such as pseudolysogeny or dormancy
399 [32]. The prevalence of similar infection dynamics in the phages isolated in this study offer two
400 intriguing possibilities: (i) Many of the viruses isolated in this study are either not fully lytic, but fall
401 somewhere on the continuum of persistence [69]. This could be controlled by genes currently lacking
402 a known function, with lysogeny (and associated superinfection immunity) favoured at high cell
403 density, which would support the Piggyback-the-Winner hypothesis [70, 71]; (ii) the steady-state of
404 host and virus densities observed here are an indicator of host phenotypic bistability in these
405 streamlined heterotrophic taxa. Viral propagation occurring in only a subset of cells could explain the
406 requirement of multiple rounds of enrichment before sufficient viral load is reached to be able to
407 observe lytic infection on the host population. Either strategy, or a combination of both, would
408 provide an ecological advantage of long-term stable coexistence between viruses and hosts and offer
409 an explanation of the paradox of stable high abundances of both predator and prey across global

410 oceans [12, 38, 72, 73]. Infection in a subset of the population could also explain the low lytic activity
411 observed in pelagiphages *in situ*, despite high host densities [74]; and the small dynamic range and
412 decoupled abundances of SAR11 and virioplankton in the Sargasso Sea [75]. Limited lysis of
413 subpopulations of hosts such as SAR11 and OM43 that specialize in the uptake of labile carbon
414 enriched through viral predation [76, 77] could facilitate efficient intra-population carbon recycling
415 and explain the limited influence of SAR11 and associated viral abundances on carbon export to the
416 deep ocean [5]. We propose the moniker the 'Soylent Green Hypothesis' for this mechanism, after
417 the 1973 cult film in which the dead are recycled into food for the living. Further investigation
418 leveraging our new virus-host model will provide greater insight into viral influence on ocean carbon
419 biogeochemistry.
420
421 In conclusion, our method coupled Dilution-to-Extinction cultivation of hosts and associated viruses,
422 resulting in the isolation of three new strains of OM43; a Western English Channel variant of a warm
423 water ecotype of SAR11; the first known methylophages for OM43; the first siphovirus infecting
424 SAR11, as well as eleven other viruses infecting this important marine heterotrophic clade and >100
425 more isolates to be sequenced and explored. The described method represents an efficient and cost-
426 effective strategy to isolate novel virus-host systems for experimental evaluation of co-evolutionary
427 dynamics of important fastidious taxa from marine and other biomes. Coupling these methods to
428 existing advances in host cultivation requires minimal additional effort and will provide valuable
429 representative genomes to improve success rates of assigning putative hosts to metagenomically-
430 derived viral contigs. Broader representation of model systems beyond cyanophages and viruses of
431 copiotrophic, *r*-strategist hosts will reduce bias in developing methods to delineate viral population
432 boundaries [78, 79], increasing the accuracy with which we derive ecological meaning from viral
433 metagenomic data. We therefore hope that this method will enable viruses to be included in the

434 current resurgence of cultivation efforts to better understand the biology and ecology of phages, and
435 the influence of the world's smallest predators on global biogeochemistry.
436

437 **Methods Summary**

438 A complete description of the materials and methods is provided in the Supplementary Information.
439 Four bacterial strains (*Methylophilales sp*. C6P1, D12P1 and H5P1; *Pelagibacter sp*. H2P3α) were
440 isolated from Western English Channel station L4 seawater samples using Dilution-to-Extinction
441 methods [29]. All four bacteria and two additional SAR11 strains *Pelagibacter bermudensis*
442 HTCC7211 and *Pelagibacter ubique* HTCC1062 were used as bait to isolate phages from six monthly
443 Western English Channel L4 seawater samples (50°15.00N; 4°13.00W). Briefly, water samples were
444 concentrated for viruses using tangential flow filtration (100k Da Hydrosart membrane) and used as
445 viral inoculum (10% v/v) in exponentially growing cultures of host bacteria in artificial seawater
446 medium [80] in 96-well Teflon plates (Radleys, UK). Cells of the resulting lysate were filtered out
447 (0.1 µm PVDF syringe filters) and the filtrate was used as viral inoculum in another round of isolation.
448 This process was repeated until viral infection could be detected by flow cytometry - comparing
449 cytograms and maximum density of infected cultures against uninfected cultures. Phages were
450 purified by dilution-to-extinction methods (detailed protocol available here:
451 dx.doi.org/10.17504/protocols.io.c36yrd). Phage genomes were sequenced using Illumina 2x150 PE
452 sequencing, assembled and manually annotated as described in [81]. Phylogenetic classification of
453 phages was performed on concatenated shared genes using a combination of Bayesian inference trees,
454 maximum likelihood trees and shared-gene likelihood analyses, depending on the availability of
455 appropriate taxa. ICTV-recognised genera based on shared gene content were assigned with
456 VConTACT2 [19]. The relative abundance of novel phages in the Western English Channel [12] and
457 Global Ocean viromes [11] was estimated (RPKM) by competitive read recruitment of five million

458 randomly subsampled reads against pelagiphage and methylophage genomes from this study and
459 others [38, 52, 64, 82].

460

461 **Data availability** All reads can be found in the SRA database under BioProject number
462 PRJNA625644 as BioSamples SAMN14604128 - SAMN14604140. Annotated phage genomes are
463 deposited as GenBank submissions under accession numbers MT375519 - MT375531. Sequences
464 used for phylogenetic analysis are deposited under https://github.com/ViralPirates/viral-dte.

465

466 **Acknowledgements** We would like to thank Christian Hacker and the Bioimaging Centre of the
467 University of Exeter for performing the TE microscopy and imaging. We would also like to thank the
468 crew of the *R/V Plymouth Quest* and our collaborators at PML for collecting water samples, and the
469 driver Magic for delivering water samples from Plymouth to Exeter. Genome sequencing was
470 provided by MicrobesNG (http://www.microbesng.uk) which is supported by the BBSRC (grant
471 number BB/L024209/1). This project used equipment funded by the Wellcome Trust Institutional
472 Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award
473 (WT101650MA) and BBSRC LOLA award (BB/K003240/1). Bioinformatic analyses were
474 conducted using the high-performance computing, ISCA, provided by the University of Exeter.

475

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

480

481 **Conflict of interests** The authors declare no conflict of interest.

482

483 ## References

484

485 1. Eric Wommack K, Colwell RR. Virioplankton: Viruses in Aquatic Ecosystems. *MICROBIOLOGY AND*
486 *MOLECULAR BIOLOGY REVIEWS* 2000; **64**: 69–114.
487 2. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, et al. A multitrophic model
488 to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J*
489 2015; **9**: 1352–1364.
490 3. Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical
491 cycles. *F1000 Biol Rep* 2012; **4**: 17.
492 4. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles:
493 implications for marine biogeochemical cycles. *Nat Rev Microbiol* 2014; **12**: 519–528.
494 5. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving
495 carbon export in the oligotrophic ocean. *Nature* 2016; **532**: 465–470.
496 6. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how
497 phages command the microbial high seas. *Virol J* 2019; **16**: 15.
498 7. Weitz JS, Hartman H, Levin SA. Coevolutionary arms races between bacteria and bacteriophage. *Proc*
499 *Natl Acad Sci U S A* 2005; **102**: 9535–9540.
500 8. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. Long-term stability and Red Queen-like strain
501 dynamics in marine viruses. *Nat Microbiol* 2020; **5**: 265–271.
502 9. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant
503 temperate viruses affect Southern Ocean microbial dynamics. *ISME J* 2016; **10**: 437–449.
504 10. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential
505 biogeochemical impacts of globally abundant ocean viruses. *Nature* 2016; **537**: 689–693.
506 11. Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Temperton B, et al. Marine DNA Viral

507 Macro- and Microdiversity from Pole to Pole. *Cell* 2019; **177**: 1109–1123.
508 12. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral
509 metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic
510 islands. *PeerJ* 2019; **7**: e6800.
511 13. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. *Nature*
512 2003; **424**: 741.
513 14. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Viruses Inhibit CO2 Fixation in the Most Abundant
514 Phototrophs on Earth. *Curr Biol* 2016; **26**: 1585–1589.
515 15. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic
516 data. *PeerJ* 2015; **3**: e985.
517 16. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying
518 viral sequences from assembled metagenomic data. *Microbiome* 2017; **5**: 69.
519 17. Amgarten D, Braga LPP, da Silva AM, Setubal JC. MARVEL, a Tool for Prediction of Bacteriophage
520 Sequences in Metagenomic Bins. *Front Genet* 2018; **9**: 304.
521 18. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation
522 of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 2017; **5**: e3817.
523 19. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment
524 of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* 2019.
525 20. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in
526 the 'core' and 'flexible' Pacific Ocean Virome. *ISME J* 2015; **9**: 472–484.
527 21. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts
528 from metagenomic phage contigs. *Bioinformatics* 2017; **33**: 3113–3114.
529 22. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d_2^* oligonucleotide frequency
530 dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences.
531 *Nucleic Acids Res* 2017; **45**: 39–53.
532 23. Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, et al. One carbon metabolism in
533 SAR11 pelagic marine bacteria. *PLoS One* 2011; **6**: e23973.

24. Becker JW, Hogle SL, Rosendo K, Chisholm SW. Co-culture and biogeography of Prochlorococcus and SAR11. *ISME J* 2019.

25. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA. Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *ISME J* 2020; **14**: 79–90.

26. Robert M. Morris*, Michael S. Rappe´ *, Stephanie A. Connon*, Kevin L. Vergin*, William A. Siebold*, Craig A. Carlson† & Stephen J. Giovannoni*. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 2002; **420**.

27. Bartelme RP, Custer JM, Dupont CL, Espinoza JL, Torralba M, Khalili B, et al. Influence of Substrate Concentration on the Culturability of Heterotrophic Soil Microbes Isolated by High-Throughput Dilution-to-Extinction Cultivation. *mSphere* 2020; **5**.

28. Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* 2002; **68**: 3878–3885.

29. Henson MW, Pitre DM, Weckhorst JL, Celeste Lanclos V, Webber AT, Cameron Thrash J. Artificial Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico. *American Society for Microbiology* 2016; **1**: 1–9.

30. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J* 2018; **12**: 1846–1860.

31. Nagasaki K, Bratbak G. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. *Manual of aquatic viral ecology ASLO* 2010; 92–101.

32. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. *Nat Commun* 2018; **9**: 4781.

33. Morris RM, Longnecker K, Giovannoni SJ. Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* 2006; **8**: 1361–1370.

34. Ramachandran A, Walsh DA. Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems. *FEMS Microbiol Ecol* 2015; **91**.

35. Halsey KH, Carter AE, Giovannoni SJ. Synergistic metabolism of a broad range of C1 compounds in

561 the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* 2012; **14**: 630–640.
562 36. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic
563 Ocean: a biogeographical study of bacterial substrate utilisation. *ISME J* 2019; **13**: 1119–1132.
564 37. Carini P, White AE, Campbell EO, Giovannoni SJ. Methane production by phosphate-starved SAR11
565 chemoheterotrophic marine bacteria. *Nat Commun* 2014; **5**: 4346.
566 38. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11
567 viruses in the ocean. *Nature* 2013; **494**: 357–360.
568 39. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using
569 metagenomics. *PLoS Genet* 2013; **9**: e1003987.
570 40. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et
571 al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 2017; **8**:
572 15892.
573 41. Taubert M, Grob C, Howat AM, Burns OJ, Dixon JL, Chen Y, et al. XoxF encoding an alternative
574 methanol dehydrogenase is widespread in coastal marine environments. *Environ Microbiol* 2015; **17**:
575 3937–3948.
576 42. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial
577 ecology. *ISME J* 2014; **8**: 1553–1565.
578 43. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ. The presence of the glycolysis
579 operon in SAR11 genomes is positively correlated with ocean productivity. *Environ Microbiol* 2010;
580 **12**: 490–500.
581 44. Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, et al. The small genome of an
582 abundant coastal ocean methylotroph. *Environ Microbiol* 2008; **10**: 1771–1782.
583 45. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification
584 of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*
585 2014; **12**: 635–645.
586 46. Polyphasic taxonomy of Marine bacteria. *studylib.net*. https://studylib.net/doc/5577926/polyphasic-
587 taxonomy-of-marine-bacteria. Accessed 11 May 2020.

588 47. Olsen NS, Hendriksen NB, Hansen LH, Kot W. A New High-Throughput Screening Method for
589 Phages: Enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential.
590 *PHAGE* 2020; **1**: 137–148.
591 48. Nayfach S, Camargo AP, Eloe-Fadrosh E, Roux S, Kyrpides N. CheckV: assessing the quality of
592 metagenome-assembled viral genomes. *bioRxiv* . 2020. , 2020.05.06.081778
593 49. Adriaenssens EM, Sullivan MB, Knezevic P, van Zyl LJ, Sarkar BL, Dutilh BE, et al. Taxonomy of
594 prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee.
595 *Arch Virol* 2020; **165**: 1253–1260.
596 50. Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, et al. Viral tagging
597 reveals discrete populations in Synechococcus viral genome sequence space. *Nature* 2014; **513**: 242–
598 245.
599 51. Våge S, Storesund JE, Thingstad TF. SAR11 viruses and defensive host strains. *Nature* . 2013. , **499**:
600 E3–4
601 52. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family
602 integrate into host genomes. *Environ Microbiol* 2018.
603 53. Marston MF, Martiny JBH. Genomic diversification of marine cyanophages into stable ecotypes.
604 *Environ Microbiol* 2016; **18**: 4240–4253.
605 54. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Reticulate representation of evolutionary and
606 functional relationships between phage genomes. *Mol Biol Evol* 2008; **25**: 762–777.
607 55. Giovannoni S, Temperton B, Zhao Y. Giovannoni et al. reply. *Nature* . 2013. , **499**: E4–5
608 56. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Novel pelagiphages prevail in the ocean.
609 *bioRxiv* . 2019. , 2019.12.14.876532
610 57. Tam W, Pell LG, Bona D, Tsai A, Dai XX, Edwards AM, et al. Tail tip proteins related to
611 bacteriophage λ gpL coordinate an iron-sulfur cluster. *J Mol Biol* 2013; **425**: 2450–2462.
612 58. Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M. Metagenome Mining Reveals Hidden Genomic
613 Diversity of Pelagimyophages in Aquatic Environments. *mSystems* 2020; **5**.
614 59. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous

615 cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun* 2019; **10**: 752.
616 60. Van Duin J, Wijnands R. The function of ribosomal protein S21 in protein synthesis. *Eur J Biochem*
617 1981; **118**: 615–619.
618 61. Xu J, Hendrix RW, Duda RL. Conserved translational frameshift in dsDNA bacteriophage tail assembly
619 genes. *Mol Cell* 2004; **16**: 11–21.
620 62. Giovannoni SJ, Thrash JC, Temperton B. Implications of streamlining theory for microbial ecology.
621 *ISME J* 2014; **8**: 1553–1565.
622 63. Neufeld JD, Boden R, Moussard H, Schäfer H, Murrell JC. Substrate-specific clades of active marine
623 methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Appl Environ*
624 *Microbiol* 2008; **74**: 7321–7328.
625 64. Moon K, Kang I, Kim S, Kim S-J, Cho J-C. Genome characteristics and environmental distribution of
626 the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. *Environ*
627 *Microbiol* 2017; **19**: 4714–4727.
628 65. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. The ecology of pelagic freshwater
629 methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J* 2015; **9**: 2442–
630 2453.
631 66. Sargeant SL, Murrell JC, Nightingale PD, Dixon JL. Seasonal variability in microbial methanol
632 utilisation in coastal waters of the western English Channel. *Mar Ecol Prog Ser* 2016; **550**: 53–64.
633 67. Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages.
634 *Genome Res* 2007; **17**: 1486–1495.
635 68. Enav H, Kirzner S, Lindell D, Mandel-Gutfreund Y, Béjà O. Adaptation to sub-optimal hosts is a driver
636 of viral diversification in the ocean. *Nat Commun* 2018; **9**: 4698.
637 69. Weitz JS, Li G, Gulbudak H, Cortez MH, Whitaker RJ. Viral invasion fitness across a continuum from
638 lysis to latency. *Virus Evol* 2019; **5**: vez006.
639 70. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, et al. Lytic to temperate
640 switching of viral communities. *Nature* 2016; **531**: 466–470.
641 71. Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. *NPJ*

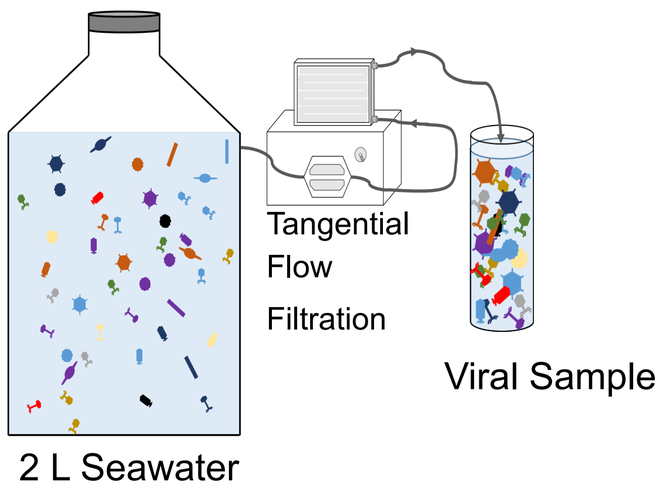
642 *Biofilms Microbiomes* 2016; **2**: 16010.
643 72. Kang I, Oh HM, Kang D, Cho JC. Genome of a SAR116 bacteriophage shows the prevalence of this
644 phage type in the oceans. *Proc Natl Acad Sci U S A* 2013; **110**: 12343–12348.
645 73. Martinez-Hernandez F, Fornas Ò, Lluesma Gomez M, Garcia-Heredia I, Maestre-Carballa L, López-
646 Pérez M, et al. Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant
647 uncultured 37-F6 viral population in the ocean. *ISME J* 2018.
648 74. Alonso-Sáez L, Morán XAG, Clokie MR. Low activity of lytic pelagiphages in coastal marine waters.
649 *ISME J* 2018; **12**: 2100–2102.
650 75. Parsons RJ, Breitbart M, Lomas MW, Carlson CA. Ocean time-series reveals recurring seasonal
651 patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J* 2012; **6**: 273–284.
652 76. Middelboe M, Lyck PG. Regeneration of dissolved organic matter by viral lysis in marine microbial
653 communities. *Aquat Microb Ecol* 2002; **27**: 187–194.
654 77. Middelboe M, Jorgensen NOG. Viral lysis of bacteria: an important source of dissolved amino acids
655 and cell wall compounds. *J Mar Biol Assoc U K* 2006; **86**: 605–612.
656 78. Brum JR, Sullivan MB. Rising to the challenge: accelerated pace of discovery transforms marine
657 virology. *Nat Rev Microbiol* 2015; **13**: 147–159.
658 79. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, et al. Genomic
659 differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics*
660 2016; **17**: 930.
661 80. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme
662 oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J* 2013; **7**: 592–
663 602.
664 81. Salisbury A, Tsourkas PK. A Method for Improving the Accuracy and Efficiency of Bacteriophage
665 Genome Annotation. *Int J Mol Sci* 2019; **20**.
666 82. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Culturing novel and abundant pelagiphages in
667 the ocean. *Environ Microbiol* 2020.

ıeter Viral Collection Numb	Virus	Host	Viral Population Representative	Hypergeometric cluster	Genome	Length (bp)	G + C %	Taxon	Accession number	Simplified phonetic spelling	Meaning and origin of the names
EXVC025P	Eistla	HTCC1062	Eistla	Cluster A	Circular	39638	32.7	Podoviridae	MT375521	aisːtlaː	"stormy one", Giantess in the poetic Edda
EXVC018P	Eyrgjafa	HTCC1062	Eyrgjafa	Cluster B	Circular	38005	32.6	Podoviridae	MT375523	eːɪrgjaːfaː	"scar donor", Giantess in the poetic Edda
EXVC020P	Gjalp	HTCC1062	Eyrgjafa	Cluster B	Circular	37857	32.5	Podoviridae	MT375524	gjaːlp	"Roaring one", Giantess in the poetic Edda
EXVC021P	Greip	HTCC1062	Greip	Cluster D	Circular	34916	31.5	Podoviridae	MT375525	graip	"Grasp", Giantess in the poetic Edda
EXVC014P	Ran	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375530	ran	"plundering", Goddess of the Sea
EXVC013S	Aegir	H2P3α	Kolga	Singleton	Fragment	18297	31.1	Siphoviridae	MT375519	æːgir	"Sea", God of the Sea
EXVC012P	Jörmungand	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375528	jɔrmun̩gand	"huge monster", Giant sea serpent
EXVC016S	Kólga	H2P3α	Kolga	Singleton	Circular	48659	30.5	Siphoviridae	MT375529	kɔlgaː	"Cool-wave", Daughter of Ran and Aegir
EXVC019P	Unn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375531	ʊːn	"Wave", Daughter of Ran and Aegir
EXVC015P	Hroenn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375527	ɦrøn	"Wave", Daughter of Ran and Aegir
EXVC010P	Bylgja	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375520	bɪlgjaː	"Billow"-wave, Daughter of Ran and Aegir
EXVC011P	Himinglæva	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375526	hiːmɪŋɡlæfa	"transparent-on-top"-wave, Daughter of Ran and Aegir
EXVC282S	Venkman	H5P1	Venkman	Singleton	Linear	38624	34.4	Siphoviridae	MT375522	vɛŋkmæn	Bill Murray's Character in the Ghostbusters movie

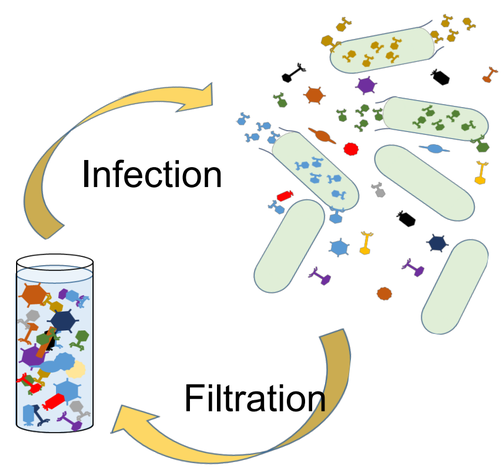
Phage	H2P3	HTVC1062	HTVC7211
Eistla		+	
Eyrgjafa		+	
Greip		+	
Ran	+		+
Kolga	+		+
Bylgja	+	+	+

A

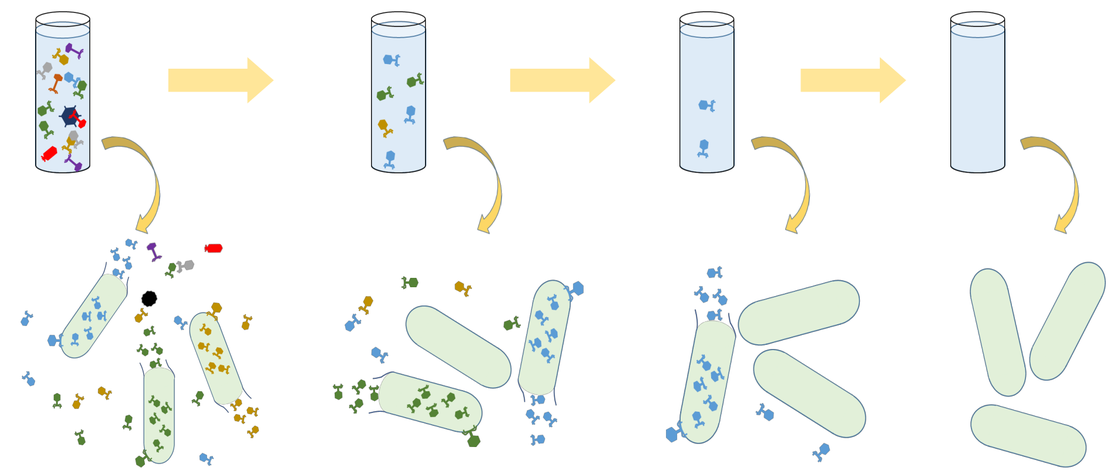
(i) Increasing Viral Concentration



(ii) Enrichment

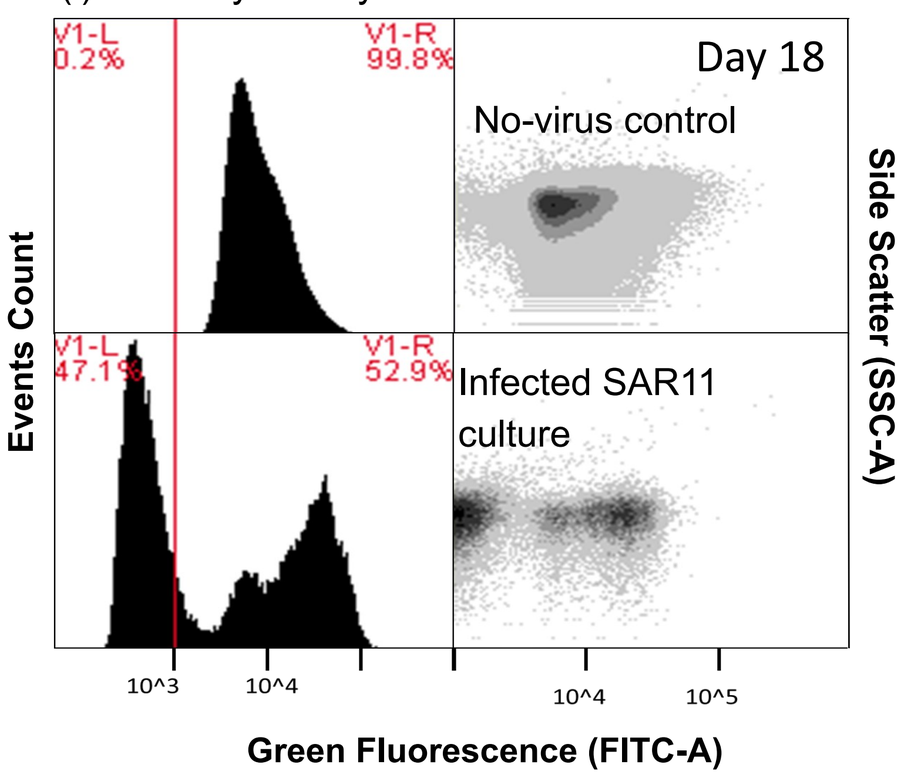


(iii) Purification

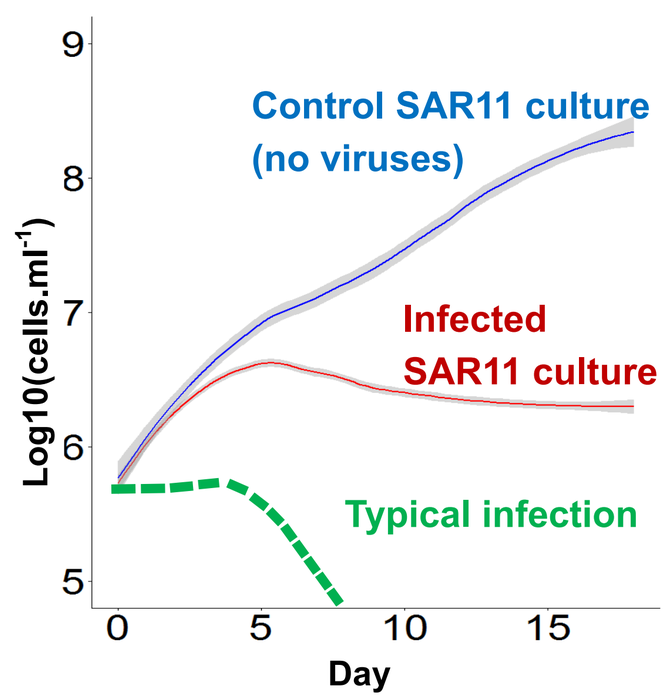


B

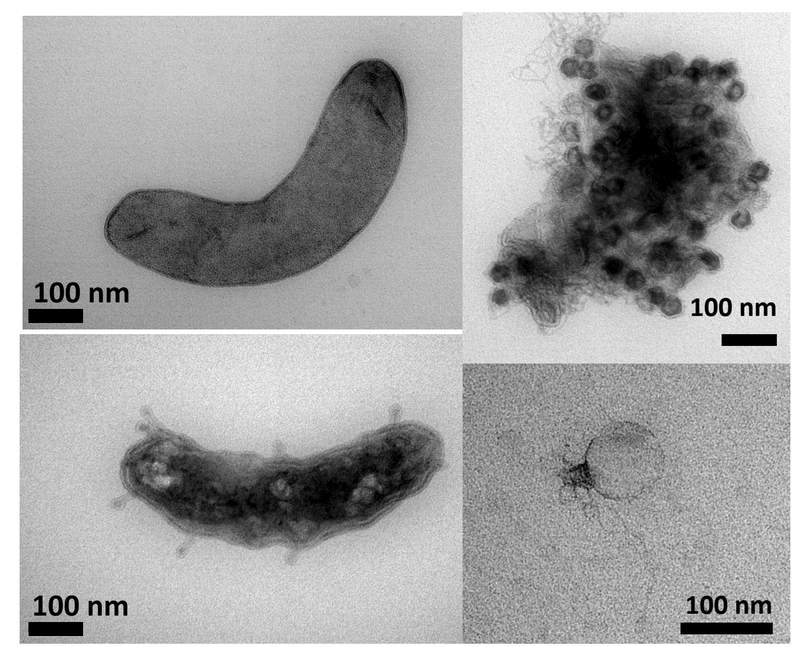
(i) Flow Cytometry

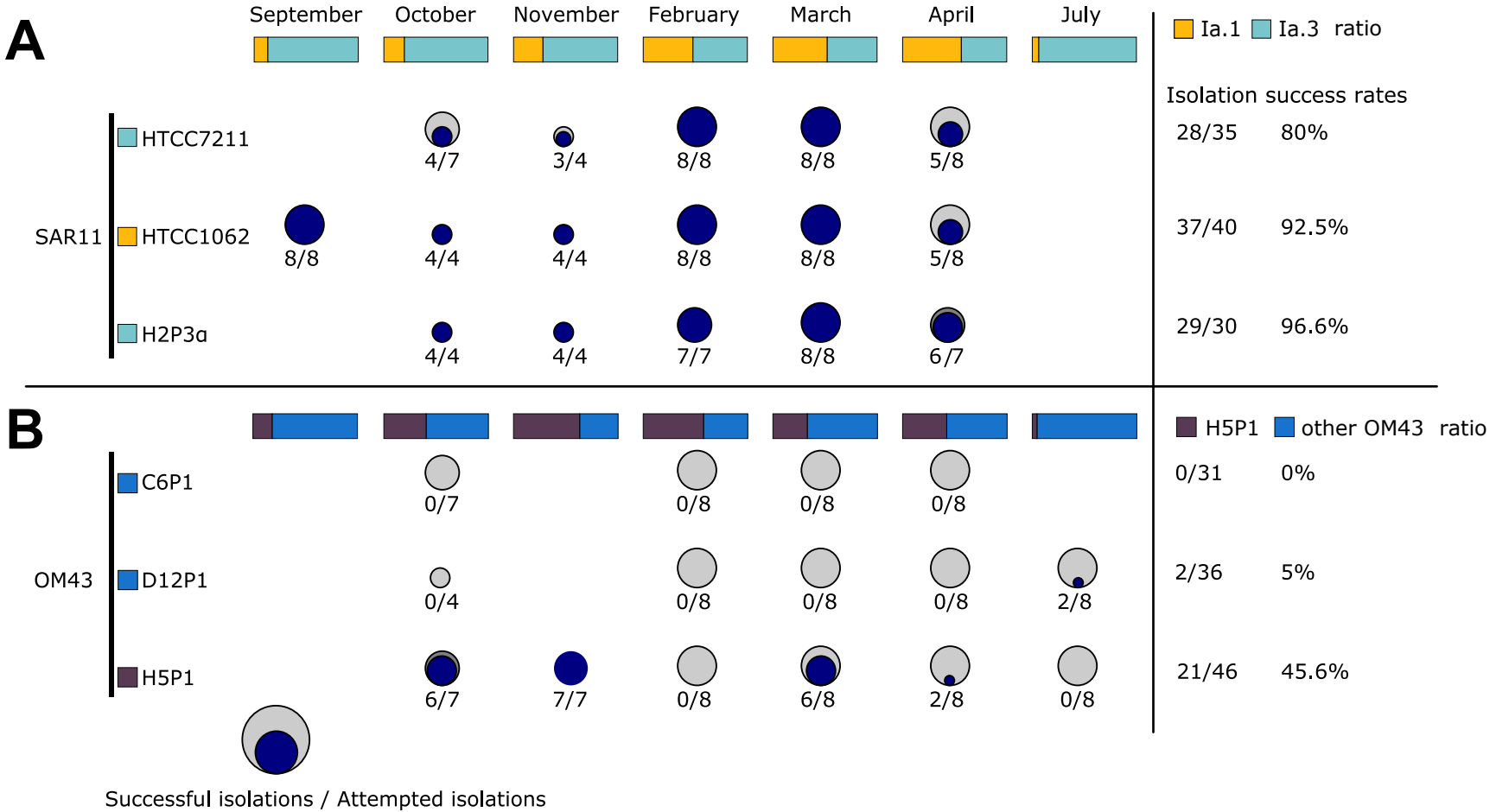


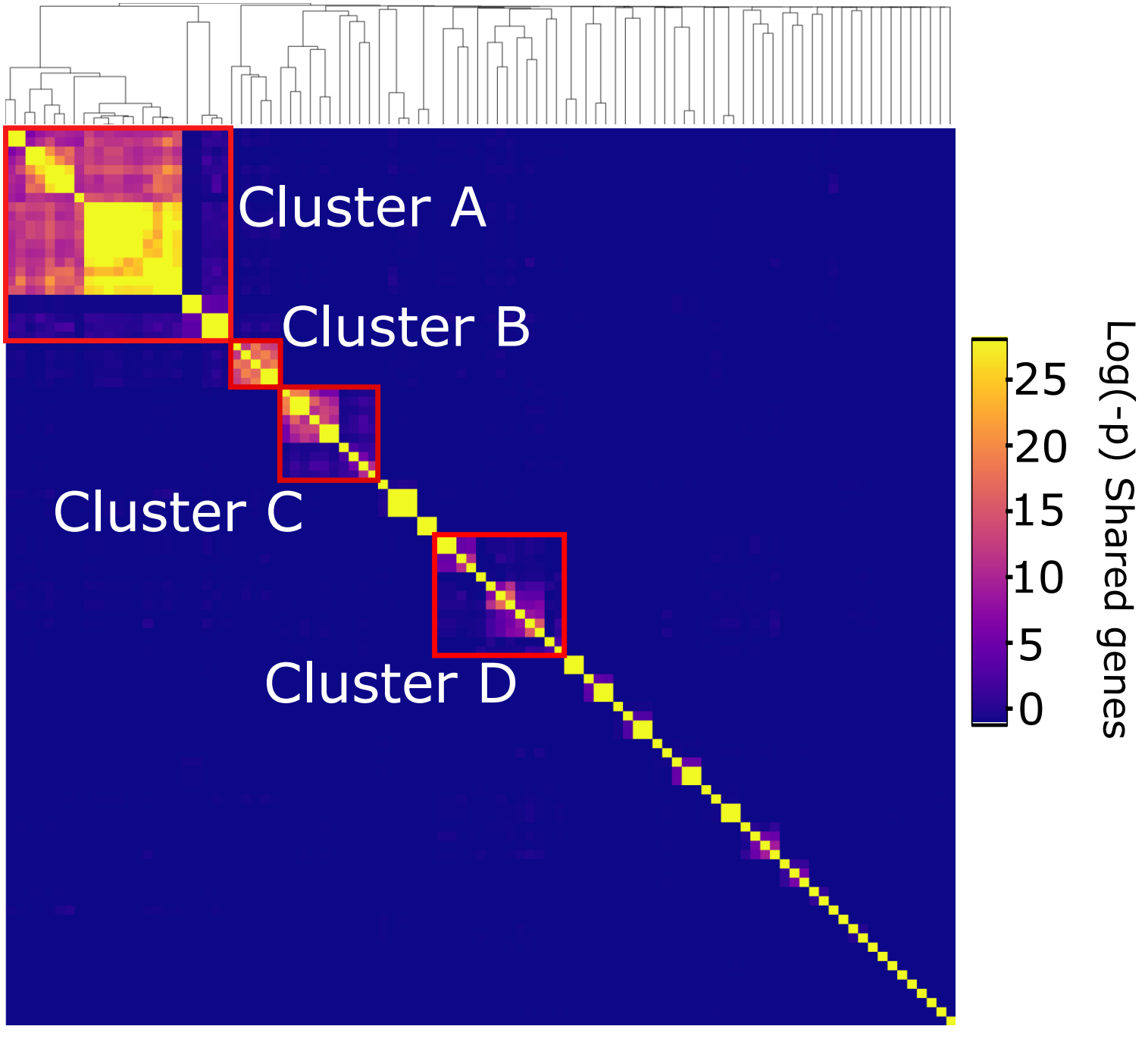
(ii) Host Cell Growth Curves

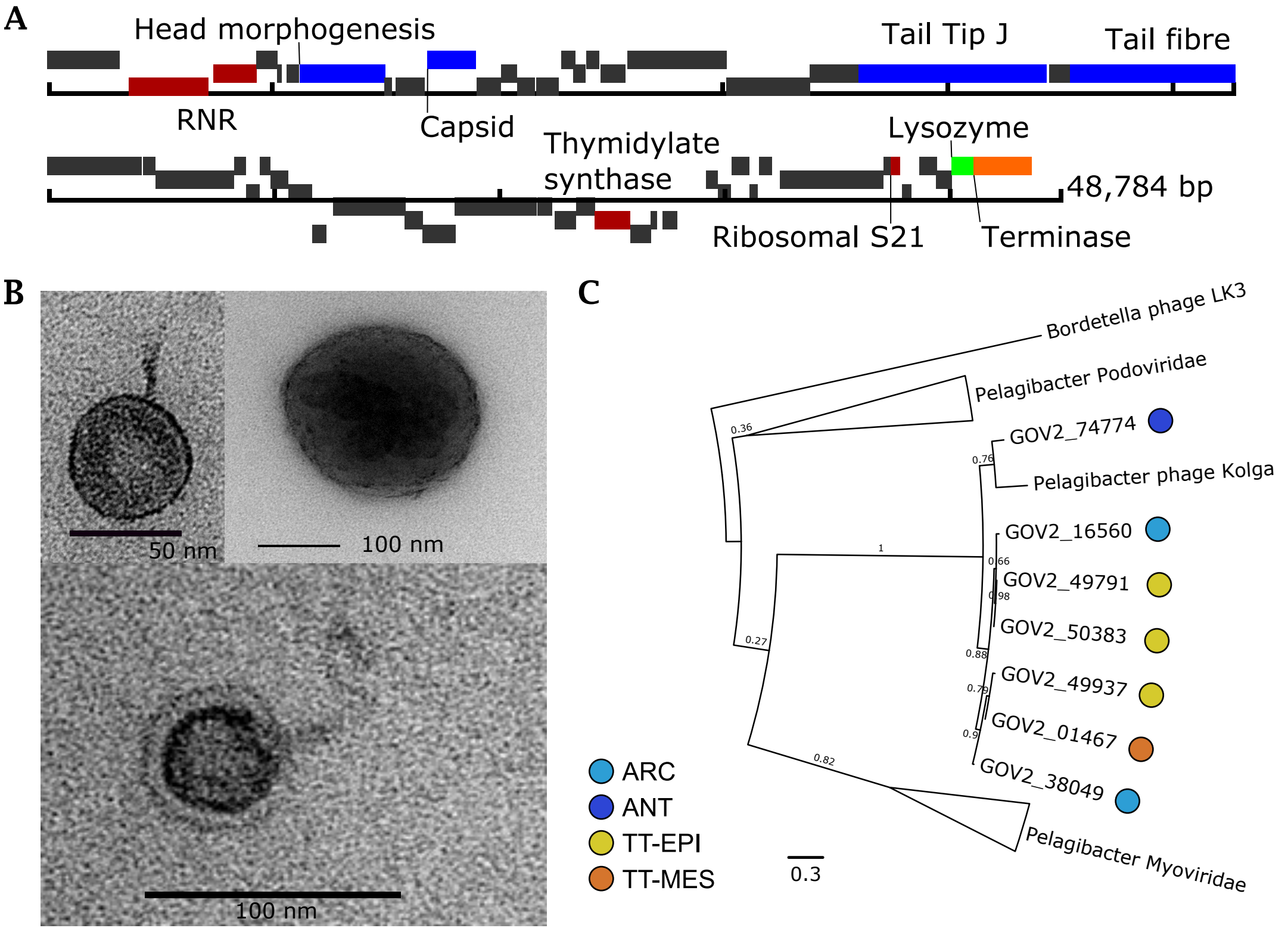


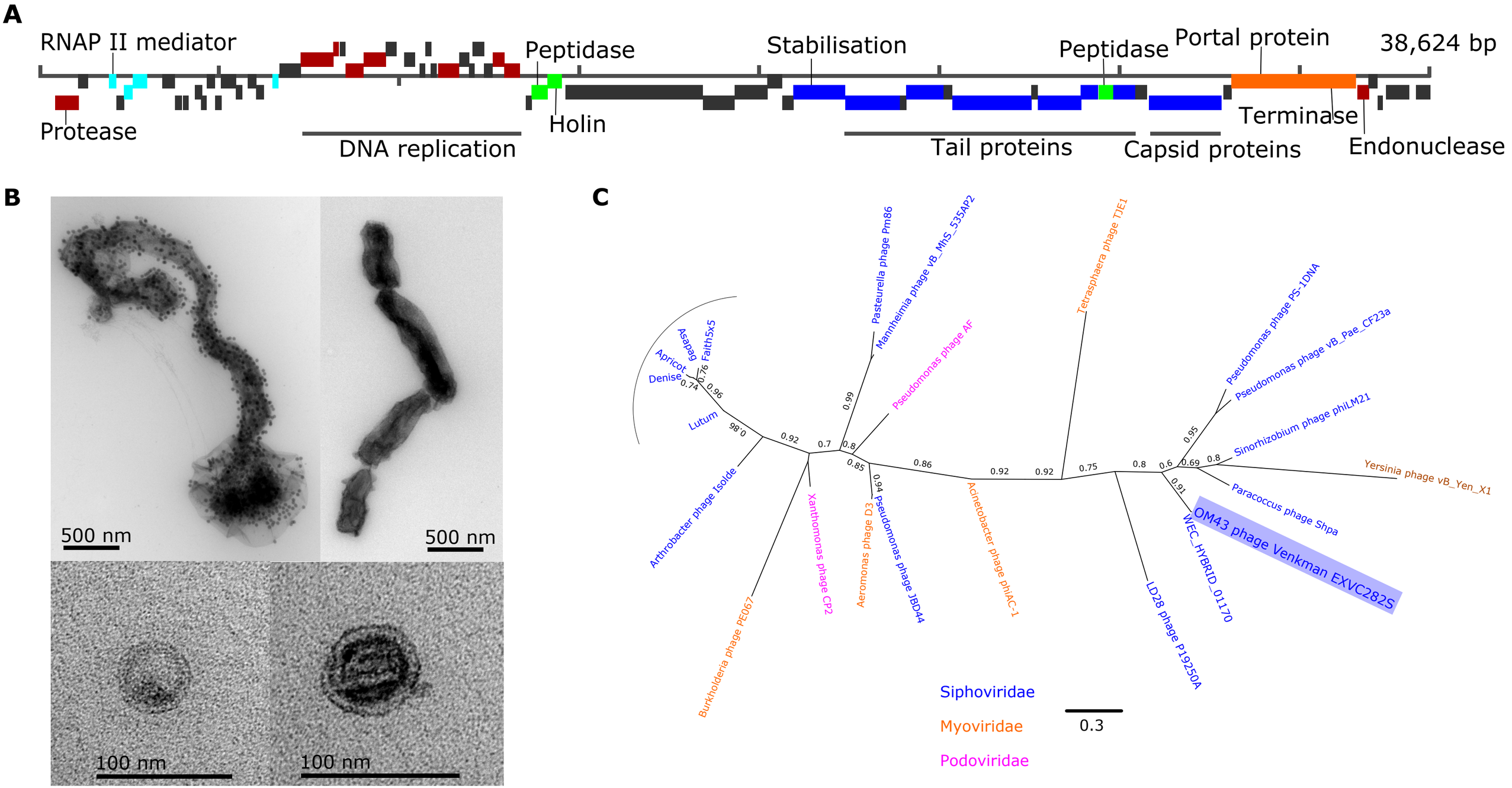
(iii) TE Microscopy




A
September
October
November
February
March
April
July
Ia.1
Ia.3
ratio
Isolation success rates
SAR11
HTCC7211
4/7
3/4
8/8
8/8
5/8
28/35
80%
HTCC1062
8/8
4/4
4/4
8/8
8/8
5/8
37/40
92.5%
H2P3α
4/4
4/4
7/7
8/8
6/7
29/30
96.6%
B
H5P1
other OM43
ratio
OM43
C6P1
0/7
0/8
0/8
0/8
0/31
0%
D12P1
0/4
0/8
0/8
0/8
2/8
2/36
5%
H5P1
6/7
7/7
0/8
6/8
2/8
0/8
21/46
45.6%
Successful isolations / Attempted isolations


Cluster A
Cluster B
Cluster C
Cluster D
25
20
15
10
5
0
Log(-p) Shared genes


A
Head morphogenesis
Tail Tip J
Tail fibre
RNR
Capsid
Thymidylate synthase
Lysozyme
48,784 bp
Ribosomal S21
Terminase
B
50 nm
100 nm
100 nm
C
Bordetella phage LK3
Pelagibacter Podoviridae
GOV2_74774
Pelagibacter phage Kolga
GOV2_16560
GOV2_49791
GOV2_50383
GOV2_49937
GOV2_01467
GOV2_38049
Pelagibacter Myoviridae
0.36
0.76
1
0.66
0.98
0.27
0.88
0.79
0.9
0.82
0.3
ARC
ANT
TT-EPI
TT-MES


A
RNAP II mediator
Peptidase
Stabilisation
Peptidase
Portal protein
38,624 bp
Protease
Holin
Terminase
DNA replication
Tail proteins
Capsid proteins
Endonuclease
B
500 nm
500 nm
100 nm
100 nm
C
Pasteurella phage Pm86
Mannheimia phage vB_MhS_535AP2
Tetraspaera phage TJE1
Pseudomonas phage PS-1DNA
Pseudomonas phage vB_Pae_CF23a
Sinorhizobium phage phiLM21
Yersinia phage vB_Yen_X1
Paracoccus phage Shpa
OM43 phage Venkman EXVC282S
WEC_HYBRID_01170
LD28 phage P19250A
Pseudomonas phage AF
Acinetobacter phage phiAC-1
Pseudomonas phage JBD44
Aeromonas phage D3
Xanthomonas phage CP2
Burkholderia phage PE067
Arthrobacter phage Isolde
Lutum
Denise
Apricot
Asapag
Faith5x5
Siphoviridae
Myoviridae
Podoviridae
0.3

ARC
TT-EPI
TT-MES
ANT
WEC
Other cultured pelagiphages
HTVC010P
HTVC106P
HTVC026P
HTVC027P
HTVC115P
HTVC104P
HTVC103P
HTVC202P
HTVC023P
HTVC111P
HTVC022P
HTVC021P
HTVC120P
HTVC112P
HTVC019P
HTVC109P
HTVC105P
HTVC008M
HTVC121P
HTVC011P
HTVC119P
HTVC031P
HTVC201P
HTVC025P
HTVC200P
Pelagiphages from this study
Greip
Ran
Kolga
Bylgja
Himinglaeva
Eistla
Eyrgjafa
Other methylophage
P19250A
Methylophage from this study
Venkman
Temperature
Depth
Latitude
RPKM
30
20
10
0
Temperature (°C)
30
20
10
0
−10
Depth (m)
0
500
1000
0
500
1000
1500
2000
total RPKM
WEC virome 2016

Fig. 1 Workflow for high-throughput isolation **(A)** i. Increasing concentration of viruses in water samples by tangential flow filtration (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 transmission electron 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.

692 Fig. 2 Summary of success rates for each bacterial host as well as relative contribution to the clade
693 community during sampling based on 16S rRNA analysis: (**A**) Three SAR11 strains HTCC1062,
694 HTCC7211 and H2P3α and (**B**) three OM43 strains C6P1, D12P1 and H5P1. Circles and associated
695 numbers represent number of successful isolations (i.e. successful passage through three rounds of
696 Dilution-to-Extinction), over the total number of attempts made. Bars at the top of each panel
697 represent the relative abundance (SAR11 subclades 1a.1 and 1a.3, and OM43 strains for A and B,
698 respectively) of strains in the microbial community sampled at the time of collection (based on 16S
699 rRNA community profiling). Percentages on the right represent total successes for all attempts per
700 host strain.
701
702 Fig. 3 Hypergeometric probability of shared gene content between known pelagiphage genomes
703 identified four main viral phylogenetic clusters (outlined with red boxes).
704
705 Fig. 4 The first reported siphovirus infecting SAR11 (Pelagibacter phage *Kolga* EXVC016S) isolated
706 on novel host *Pelagibacter sp.* H2P3α. **(A)** Gene map of the 48,784 bp genome, which contains 80%
707 hypothetical genes without known function; **(B)** Transmission Electron Micrographs of *Kolga* (left
708 and bottom) and an H2P3α host cell infected with *Kolga* (top right). **(C)** Unrooted maximum
709 likelihood tree of *terL* genes found in pelagiphages. Branches containing members of the Podoviridae
710 and Myoviridae family infecting SAR11 are collapsed for clarity (full tree is available in
711 Supplementary Fig. 7). Closely related contigs prefixed with 'GOV' represent viral contigs from the
712 Global Ocean Virome dataset and the ecological zone from which they were assembled is marked:
713 Temperate-tropical epipelagic (TT-EPI); temperate-tropical mesopelagic (TT-MES); Antarctic
714 (ANT), Arctic Ocean (ARC).
715

716 Fig. 5 The first reported cultured virus known to infect a member of the OM43 clade (Methylophilales
717 phage *Venkman* EXVC282S) which was isolated on host strain H5P1 from this study; **(A)** Gene map
718 displaying protein coding genes. Gene function is colour coded as follows: Structural genes (blue);
719 DNA replication (red); lysis (green); transcription (turquoise); packaging (orange); hypothetical
720 genes (grey). **(B)** TEM images of infected and chaining H5P1 cells (top left), uninfected H5P1
721 chaining cells (top right), *Venkman* viral particles (bottom left and right); **(C)** Maximum likelihood
722 tree (500 bootstraps) of concatenated viral *terL* and exonuclease genes. Host families of the phages
723 are indicated on the figure.
724

725 Fig.6 Relative abundance (reads recruited per kilobase of contig per million reads (RPKM)) of known
726 pelagiphage isolates and methylophage genomes in viromes from the Western English Channel [12]
727 and the Global Ocean Virome [11]. Samples are organised by decreasing latitude per ecological zone:
728 ARC: Arctic; TT-EPI: temperate-tropical epipelagic; TT-MES: temperate-tropical mesopelagic;
729 ANT: Antarctic; WEC: Western English Channel.

730

731 Table 1. Summary of phages isolated and sequenced in this study.

732

733 Table 2. Host infectivity of viral populations isolated and sequenced in this study. *Pelagibacter sp.*
734 H2P3α and *Pelagibacter bermudensis* HTCC7211 are warm-water ecotypes of SAR11 subclade 1a;
735 HTCC1062 is a cold-water ecotype. *Pelagibacter* phage *Bylgja* was the only virus capable of
736 infecting both ecotypes.
737
738
739

Supplementary Information

Supporting Methods

Water sampling: 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: 2018-09-24, 2018-10-17, 2018-11-05, 2019-02-11, 2019-03-11, 2019-04-01, 2019-07-22 (for details see Supplementary Table 2). Seawater was transferred immediately to a clean 2 L acid-washed polycarbonate (PC) Nalgene bottles (Thermo Fisher Scientific, Waltham, USA) and placed in a cooler box (Igloo, Katy, USA) at ambient temperature. Upon return to shore, water was transported to the University of Exeter for immediate processing (Two hours maximum duration from collection to 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-size 0.2 µm PC filter and autoclaved in 2 L Duran bottles (DWK Life Sciences GmbH, Mainz, Germany) with the lid tightly shut to maintain carbon chemistry [1]. 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 [2]. 1 mL of medium was placed into each 2 mL well in three sterile 96-well acid washed Teflon plates (Radleys, UK). Surface water from station L4 was filtered through a 142 mm Merck Millipore 0.2 µm PC filter to remove larger plankton and bacteria and enrich for smaller bacteria. A 200 µL aliquot was stained with SybrGreen and then quantified on a C6 Accuri flow cytometer. The remaining retained fraction was diluted in NSW to a density of ~ 1 cell·μL^{-1}. 1 µL of diluted inoculum was added to each well, with eight wells left blank as negative controls and eight wells with undiluted low-nucleic acid community fraction as positive controls. Plates were covered and incubated in the dark at 15 °C for three months and checked monthly for positive growth by flow cytometry as described above. Positive wells (>10^6 cells·mL^{-1}) were transferred into 50 mL cultures in NSW medium and passaged three times once they achieved >10^6 cells·mL^{-1}). To verify culture purity and to taxonomically classify new isolates, the 16S rRNA gene was amplified from 10 µL of exponentially growing culture, heated to 95°C for 10 minutes, followed by PCR (25 cycles, 30 s @ 98 °C; 45 s @ 59 °C; 45 s @ 72 °C; final extension of 90 s @ 72 °C) using commercially available 27F and 1492R primers by Eurofins Genomics. Amplicon DNA was purified for Sanger sequencing (Eurofins) using a QIAquick PCR purification kit (Qiagen) following the manufacturer's standard protocol. 16S rRNA sequences were compared to known bacterial sequences in the SILVA SSU database with SINA [3]. Sequences were aligned using T-coffee [4], curated using Gblocks [5] and maximum likelihood trees for both alignments were created using PhyML with standard settings and 500 bootstraps [6]. The trees were visualised using FigTree (v1.4.4, available https://github.com/rambaut/figtree).

41 *Host cultivation:* SAR11 strains HTCC1062 and HTCC7211 were kindly provided by the
42 Giovannoni lab (Oregon State University, USA) and were kept in continuous culture in 50 mL
43 polycarbonate flasks at 15 °C on artificial seawater medium ASM1 [7] throughout the duration of
44 the study. Following isolation, cultures of OM43 and H2P3α were maintained on ASM1; medium
45 for OM43 was further amended with 100 mM MeOH and 32 µL of commercially available MEM
46 amino acid solution per 1.6 L of medium (Sigma, M5550), to improve cellular yields and reduce
47 occurrences of chaining cells (observed through EM, Fig. 5, and a phenotype of alanine starvation
48 in SAR11 [2]), respectively.
49
50 *Viral concentration and inoculation:* Monthly samples of 2 L of surface water were collected from
51 station L4 as described above and sequentially filtered through a 142 mm Whatman GF/D filter
52 (2.7 µm pore size) and a 142 mm 0.2 µm pore polycarbonate filter (Merck Millipore) using a
53 peristaltic pump, to remove larger sized bacterioplankton. The filtrate was concentrated to 50 mL
54 (40-fold concentration) using a 50R VivaFlow tangential flow filtration unit (Sartorius Lab
55 instruments, Goettingen, Germany). The concentrate was filtered through a 0.1 µm pore PVDF
56 membrane syringe-filter to remove any residual small cells and used as inoculum for viral
57 isolation. 1 mL of exponentially growing bacterial host in ASM1 medium (amended with nutrients
58 for OM43 strains as described for host cultivation above) was placed into each 2 mL well of a 96-
59 well, acid washed Teflon plate. Eight wells were used as blank medium controls (no cell controls).
60 100 µL of viral inoculum was added to each well in the 96-well plate, with eight wells amended
61 with 10% (v/v) medium instead of viral inoculum to serve as no-virus controls. The plates were
62 incubated for ~2 weeks until no-virus controls reached maximum cell density (~2 weeks for strains
63 HTCC1062, HTCCC7211 and H2P3α; about 1 week for OM43 strains H5P1, C6P1, D12P1).
64 Growth was monitored at the end of the incubation period by flow cytometry; incubation periods
65 were determined by the average growth times and data of the strains used for infections. Cytograms
66 of virus-amended wells and no-virus controls were compared to identify infections as described
67 below. A detailed protocol of the viral isolation process is available on www.protocols.io (DOI:
68 dx.doi.org/10.17504/protocols.io.c36yrd). Once the no-virus controls reached maximum cell density,
69 inoculated wells that did not show signs of infection were pooled (according to host type), filtered
70 through a 0.1 µm pore PVDF membrane syringe (Durapore) filter to remove the cells and then
71 used as fresh inoculum in a new 96-well plate of target host. This was repeated up to three times
72 to amplify host-specific viral particles to a density whereby signs of viral infectivity could be
73 observed on the flow cytometer.
74
75 *Identification of positive infection via flow cytometry:* We observed positive infections in cells via
76 flow cytometry by comparing cytograms of wells inoculated with viruses against no-virus controls.
77 Cytograms of wells containing infected cultures were identifiable by (1) a population shift of up
78 to a 10-fold increase in green fluorescence compared to uninfected cultures; (2) an increase in the
79 'noise' fraction (presumably from cellular debris following lysis as well as viral particles) over
80 time (Supplementary Video 1); and (3) a reduced maximum cell density (ranging from 10 to 1000-

81 fold depending on initial cell density and viral inoculum concentration) in infected cultures
82 compared to no-virus controls (Fig. 1B). The presence of viral-like particles in wells where a shift
83 in green fluorescence, increased noise, and reduced maximum cell density were observed, were
84 verified by TEM and reinfection of new cultures in fresh medium.
85
86 *Viral purification:* Following identification of a positive infection within a well, the contents of
87 the well were filtered through a 0.1 µm pore-size PVDF syringe filter (Durapore) and used as
88 inoculum in three rounds of viral purification. Briefly, a 96-well Teflon plate was inoculated with
89 bacterial hosts as described previously. A 10-fold serial dilution series (from 10^{0}-10^{-10}) of the viral
90 inoculum in ASM1 was added to a single row of the plate at 10 % v/v, with one well per row as
91 no-virus controls (total of eight no-virus controls/plate). Plates were incubated at 18 °C for ~2
92 weeks until the no-virus controls reached maximum cell density and then screened for signs of
93 viral infection using flow cytometry as described above. For each dilution series, the well amended
94 with the lowest number of viruses that showed positive signs of infection was identified, and used
95 as the inoculum in another round of viral purification [8]. Using this format, we were able to purify
96 eight to twelve viral isolates simultaneously per plate.
97
98 *Viral isolation costs and handling time*. For each sample (processed environmental water sample),
99 all steps of the initial viral isolation process (counting one to two plates for a total of three times)
100 required ~6 hours run time on a flow cytometer. For the subsequent three rounds of purification,
101 one plate was required for each host-sample combination (another ~6 hours of cytometer run time
102 in total for one host-sample combination). Initial isolation of viruses from environmental water
103 samples and three rounds of purification took ~10 weeks of incubation time in total. Between all
104 steps, our protocol required ~7 hours of handling time per sample. Following three rounds of viral
105 purification, generating sufficiently high viral titres to extract enough DNA for sequencing
106 (approximately two weeks incubation time, and roughly four hours handling time over two days)
107 was the rate-limiting step, and so required subselection of available viruses for sequencing. Future
108 advances in DNA extraction efficiency, reducing DNA input requirements for sequencing and/or
109 automation of viral DNA extraction will enable all isolated viruses to be sequenced. We estimate
110 the cost of isolating a single virus is ~£20 for cultivation, flow cytometry and DNA extraction
111 consumables, and ~£50-100 for genome sequencing at 30-fold coverage required for successful
112 assembly, giving a total cost of £70-120 per sequenced viral isolate, not including the costs for
113 person-hours. Thus, our protocol provides a high-throughput and scalable approach to viral
114 isolation.
115
116 *Transmission Electron Microscopy of viral isolates.* For ultrastructural analysis, virus particles
117 were transferred onto pioloform-coated electron microscopy (EM) copper grids (Agar Scientific,
118 Standsted, UK) by floating the grids on droplets of virus-containing suspension for 3 min.
119 Following a series of four washes on droplets of deionized water, the bound virus particles were
120 contrasted with 2 % (w/v) uranyl acetate in 2 % (w/v) methyl cellulose (mixed 1:9) on ice for 8

121 min and the grids then air-dried on a wire loop after carefully removing excess staining solution
122 with a filter. Dried grids were inspected with a JEOL JEM 1400 transmission electron microscope
123 operated at 120 kV and images taken with a digital camera (ES 1000W CCD, Gatan, Abingdon,
124 UK).
125
126 *Host ranges of viral isolates*. To test the infectivity of phage isolates against different hosts, 2 mL
127 96-well Teflon plates were prepared with 1 mL of exponentially growing bacterial host
128 (HTCC1062, HTCC7211, H2P3α) in ASM1 medium. As medium control, ASM1 without any
129 bacteria was added to eight wells. 100 µL of phage inoculum was added to eight wells per host
130 type. As no-virus control, for each host further eight wells were left without any viruses added to
131 them. The plate was incubated for ~2 weeks and analysed for signs of infection using flow
132 cytometry as described above. A phage was classified as unable to infect a certain host if none of
133 the eight amended replicates showed signs of infection.
134
135 *DNA preparation and sequencing of viral isolates.* 50 mL ASM1 (amended with 100 mM MeOH
136 and amino acid solution for OM43 as described above) in 250 mL acid-washed, polycarbonate
137 flasks were inoculated with host cultures to 10^6 cells·mL^{-1}. The cultures were infected with 10%
138 v/v viruses in ASM1 medium and incubated until after host cell lysis was detected using flow
139 cytometry. The cultures were transferred to 50 mL Falcon tubes and the cellular fraction was
140 removed by centrifugation (GSA rotor, Thermo Scientific 75007588) at 8,500 rpm/10,015 x g for
141 120 minutes. The supernatant was subsequently filtered through pore-size 0.1 µm PVDF syringe
142 filter membranes to remove any remaining smaller cellular debris. Phages in the filtrate were
143 precipitated using a modified version of an established PEG8000/NaCl DNA isolation method [9].
144 Briefly, the filtered phage lysate was transferred into 50 mL Falcon tubes with pre-weighted 5 g
145 PEG8000 and 3.3 g NaCl, shaken until both dissolved and incubated on ice overnight. The phage
146 particles were then pelleted by centrifugation at 8,500 rpm/10,015 x g for 90 min at 4 °C.
147 Supernatant was discarded and phage particles resuspended by rinsing Falcon tubes twice with 1
148 mL SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-Cl). DNA was extracted using
149 the Wizard DNA Clean-Up system (Promega) following manufacturer's instructions, but instead
150 of TE we used pre-heated 60 °C nuclease free water to elute DNA. DNA was sequenced by
151 MicrobesNG (Birmingham, UK) with Illumina paired-end HiSeq 2500. Default settings were used
152 for all programs listed below, unless specified otherwise. All reads were trimmed, quality-
153 controlled and error-corrected using bbmap and tadpole [10]. Contigs were assembled using
154 SPAdes v3.13 [11] and evaluated with QUAST [12]. The trimmed reads were mapped back against
155 the contigs for scaffolding using bowtie2 [13], BamM (alignment 0.9, identity 0.95) (available at
156 https://github.com/minillinim/BamM) and samtools [14]. Viral contigs were confirmed with
157 VirSorter (categories 1 or 2, >15kbp) and only accepted if the representative contig showed mean
158 depth coverage an order of magnitude greater than the next highest recruiting contig (to filter out
159 any cellular DNA carryover). Gene calls returned by VirSorter were imported into DNA Master
160 for manual curation [15]. Additional gene calls were made using GenMark [16], GenMarkS [17],

GenMarkS2 [18], GenMark.hmm [19], GenMark.heuristic [20], Glimmer v.3.02 [21], and Prodigal v.2.6.3 [22]. All gene calls were listed and compared using a scoring system which evaluates gene length, gene overlap and coding potential of ORFs [15]. ORFs were then annotated using BLASTp against the NCBI's non-redundant protein sequences [23], Phmmer [24] against UniProts UniProtKB and uniprotrefprot [25], Swissprot [26] as well as InterProScan [27] and Pfam [28]. Virally-encoded tRNAs were identified with tRNAScan-SE v2.0 using settings `tRNAscan-SE -qQ --detail -o# -m# -f# -l# -c tRNAscan-SE.conf -s# -B` [29] to identify putative tRNA genes. Matching tRNA genes (e.g. viral and host tRNA-Leu) were then aligned using BLASTn [30]. All viral genomes were organised into viral clusters (ICTV-recognised genera) based on shared gene content with VConTACT2 [31], and into populations using a boundary cutoff of 95% ANI over 85% contig length [32].

Evaluation of gene-sharing networks with hypergeometric testing. For the hypergeometric testing we gathered all annotated genomes from this study, other known pelagiphage isolates [33, 34] and we arbitrarily picked one representative contig from each cluster of complete phage genomes that were retrieved from a metagenomic Mediterranean deep chlorophyll maximum sample, that the authors speculated to be infective towards 1A SAR11s and included them in the phylogenetic analysis [35]. The contigs were imported into anvio v6.1 [36] where open reading frames were identified internally using Prodigal v2.6.3. The contigs were then used to run a pangenome analysis using the `--use-ncbi-blast` option. The binning function was then used to export the protein clusters, from which a matrix of the sum of shared proteins between all phages was created. The `hypergeom` function in the sciPy.stats python package was used to calculate the probability of sharing a protein between phages, based on which we created an average correlation cluster map.

Viral phylogenetic analysis. First we identified additional contigs to include in the phylogenetic analysis. The Global Ocean Virome (GOV2) [37] and a virome from the Western English Channel [38] were screened for contigs that belonged to the same population (95% ANI over 80% length using ClusterGenomes.pl - https://github.com/simroux/ClusterGenomes) or the same genus (clustered by VConTACT2 [39] into the same viral cluster, using default parameters) as isolate genomes. Additionally, all available genomes from Pelagibacter phage isolates were included, as well as selected fosmid derived contigs from Mediterranean metagenomes and 26 sequences from metagenomic mining that were identified as putative Pelagimyophages [33–35, 40, 41]. All sequences were subjected to gene calling using Prodigal v2.6.3. (default parameters) and screened for TerL genes by aligning all proteins against known TerL genes in annotated genomes (belonging to clusters A, B and D, as well as siphovirus *Kolga*) from this study using Protein BLAST (default parameter) [22, 30]. The TerL gene of Bordetella phage LK3 was added as an outgroup. The TerL genes were aligned within the Phylogeny.fr online service [42] opting for MUSCLE alignment [43] and Gblocks curation [5] with default settings. A maximum-likelihood tree was calculated with PhyML [44, 45], visualised using FigTree (v.1.4.4, available at

199 http://tree.bio.ed.ac.uk/software/figtree/) and edited in Inkscape (www.inkscape.org) for aesthetics
200 (font adjustments and addition of GOV2 metadata).

201 We only identified a single additional contig from metagenomes using the approach described
202 above that was similar to OM43 phage *Venkman*. Only one other known phage for a related host
203 (LD28 phage P19250A [46]) was available. These three contigs as well as all 14,883 viral contigs
204 publicly available in January 2020 on http://millardlab.org/ were gene called using Prodigal v2.6.3.
205 [22]. All called genes were compared to vFams [47] using hmmsearch [48] and sorted by E-value
206 for best hits. The best matching gene sequences that were identified as similar to shared vFams
207 found in *Venkman* and P19250A were important to the Phylogeny.fr online service [42]. Sequences
208 of were aligned individually using T-coffee [4], curated using Gblocks [5] and concatenated
209 manually. Maximum likelihood trees for both alignments were created using PhyML with standard
210 settings and 500 bootstraps [6, 44]. The trees were visualised with FigTree (v.1.4.4) and edited in
211 Inkscape (www.inkscape.org) for aesthetics. All sequences used for phylogenetic analysis are
212 available at https://github.com/ViralPirates/viral-dte.

213 *Calculation of Average Nucleotide Identity*. All-vs-all average nucleotide identity (ANI) for our
214 new isolates and existing pelagiphages using nucmer v.4.1 with the --nooptimize flag, followed by
215 show-coords to convert the delta file into tabular format. ANI was calculated by taking the mean
216 of all matching regions between two isolates.

217

218 *Viral metagenomic recruitment.* For assessing relative abundances of phage contigs in global
219 datasets, a virome from the Western English Channel and all samples of the Global Ocean
220 Virome dataset (GOV2) were used for recruitment [37, 38]. Metagenomic reads were
221 downloaded from the European Nucleotide Archive ERR2625613 and subsampled to 5 million
222 reads using bbmap's `reformat.sh` command. A bowtie2 [13] index of dereplicated contigs was
223 created from 25 known pelagiphage genomes from isolates [33, 34, 41], LD28 phage P19250A
224 [46] and from this study six viral population representative genomes isolated on SAR11 hosts
225 and one Methylophilales phage *Venkman* isolated on OM43 host. Reads were mapped against the
226 genomes for each metagenome sample using bowtie2 using these commands: `bowtie2 --seed`
227 `42 --non-deterministic`. To calculate coverage and Reads Per Kilobase of contig per Million
228 reads (RPKM) we used coverm [reference], with the following commands: `coverm contig --`
229 `bam-files *.bam --min-read-percent-identity 0.9 --methods rpkm --min-covered-`
230 `fraction 0.4`. To minimize false positive detection rates [32], contigs that did not meet the
231 40% minimum genome coverage in a given sample were given an RPKM value of 0.

232 *Evaluation of isolation efficiency as a function of matching host and virus geography.* In total, 105
233 attempts were made for isolating viruses for SAR11 across the three host strains (HTCC7211: 35,
234 HTCC1062: 40; H2P3α: 30). For each strain, the number of successful attempts were recorded in
235 a vector, v_s, of length n where n is the number of attempts made to isolate a virus, such that v_i=1 if

236 a successful attempt was made and v_i=0 if an unsuccessful attempt was made. To test whether
237 observed rates of success for H2P3 were significantly elevated compared to those for HTCC1062
238 and HTCC7211, the observed rates were evaluated against the null hypothesis that host strain did
239 not make a difference to isolation success. Briefly, for each pair (H2P3 vs HTCC1062; H2P3α vs
240 HTCC7211), the two vectors v_{s1} *and* v_{s2} were concatenated, and then randomly subsampled 999
241 times with replacement to a size of the smallest value of *n* between the vectors. The number of
242 successes in each subsample were counted and the distribution of values across the 999 bootstraps
243 was recorded. *P*-values were calculated by evaluating how many times within the 999 bootstraps
244 a value as extreme as the observed value had been recorded.

245 *16S rRNA high-throughput amplicon sequencing and analysis*. Genomic DNA was extracted using
246 Qiagen DNeasy PowerWater Kit (REF 14900-50-NF) from biomass retained on 0.2 µm PC filters
247 following the manufacturer's protocol with minor modifications to increase the yield. Step 7 was
248 modified from a 5 minute to 10 minute vortex bead beat. Step 21 was changed to have a 2 minute
249 incubation with EB warmed to 55C. Amplification of the hypervariable V4 region of the 16S rRNA
250 gene was performed using the 515fB (5'-GTGYCAGCMGCCGCGGTAA-3') and 806rB (5'-
251 GGACTACNVGGGTWTCTAAT-3') primers. Libraries for each reaction were done attaching
252 dual indices and Illumina adapters with the NexTera XT Index Kit (Illumina Inc.) Purified libraries
253 were pooled and sequenced in a MiSeq platform. Primer sequences from paired-end fastq files
254 were trimmed using CutAdapt [49]. Trimmed sequences were quality filtered, dereplicated and
255 merged with DADA2 R package (version 1.8 [50]). Taxonomic assignment was performed using
256 the "assignTaxonomy" command within dada2 pipeline and silva_nr_v132 database [51] as
257 training set. SAR11 ASVs were extracted for further characterization using Phyloassigner v6.166
258 [52] and oligotyping [53]. SAR11 phylogenetic database SAR11_Phy_DB used in this study is
259 available at https://www.github.com/lbolanos32/NAAMES_2020. Relative contribution barplots
260 were done in R using ggplot2 [54] and edited in inkscape (www.inkscape.org) for aesthetics.
261

262 **Supplementary References**

263

264 1. Henson MW, Pitre DM, Weckhorst JL, Celeste Lanclos V, Webber AT, Cameron Thrash J. Artificial
265 Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico.
266 *American Society for Microbiology* 2016; **1**: 1–9.

267 2. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme
268 oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J* 2013; **7**:
269 592–602.

270 3. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of
271 ribosomal RNA genes. *Bioinformatics* 2012; **28**: 1823–1829.
272 4. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple
273 sequence alignment. *J Mol Biol* 2000; **302**: 205–217.
274 5. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously
275 aligned blocks from protein sequence alignments. *Syst Biol* 2007; **56**: 564–577.
276 6. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--a web server for fast maximum
277 likelihood-based phylogenetic inference. *Nucleic Acids Res* 2005; **33**: W557–9.
278 7. Carini P, Campbell EO, Morré J, Sañudo-Wilhelmy SA, Thrash JC, Bennett SE, et al. Discovery of a
279 SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso
280 Sea. *ISME J* 2014; **8**: 1727–1738.
281 8. Nagasaki K, Bratbak G. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists.
282 *Manual of aquatic viral ecology ASLO* 2010; 92–101.
283 9. Solonenko N. Isolation of DNA from phage lysate. *protocols.io* 2016.
284 10. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. *PLOS*
285 *ONE* . 2017. , **12**: e0185056
286 11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new
287 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**:
288 455–477.
289 12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome
290 assemblies. *Bioinformatics* 2013; **29**: 1072–1075.
291 13. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357–
292 359.
293 14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
294 format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

295 15. Salisbury A, Tsourkas PK. A Method for Improving the Accuracy and Efficiency of Bacteriophage
296 Genome Annotation. *Int J Mol Sci* 2019; **20**.
297 16. Borodovsky M, McIninch J. GenMark: Parallel gene recognition for both DNA strands. *Computers*
298 *Chemistry* 1993; **17**: 123–133.
299 17. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene
300 starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic*
301 *Acids Res* 2001; **29**: 2607–2618.
302 18. Lomsadze A, Gemayel K, Tang S, Borodovsky M. Modeling leaderless transcription and atypical
303 genes results in more accurate gene prediction in prokaryotes. *Genome Res* 2018; **28**: 1079–1089.
304 19. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences.
305 *Nucleic Acids Res* 2010; **38**: e132.
306 20. Besemer J, Borodovsky M. Heuristic approach to deriving models for gene finding. *Nucleic Acids*
307 *Res* 1999; **27**: 3911–3920.
308 21. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont
309 DNA with Glimmer. *Bioinformatics* 2007; **23**: 673–679.
310 22. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene
311 recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 119.
312 23. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant
313 sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007; **35**: D61–5.
314 24. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update.
315 *Nucleic Acids Res* 2018; **46**: W200–W204.
316 25. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019; **47**:
317 D506–D515.
318 26. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL
319 in 2000. *Nucleic Acids Res* 2000; **28**: 45–48.

320 27. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale
321 protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.
322 28. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein
323 families database. *Nucleic Acids Res* 2014; **42**: D222–30.
324 29. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer
325 RNA genes. *Nucleic Acids Res* 2016; **44**: W54–7.
326 30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol*
327 *Biol* 1990; **215**: 403–410.
328 31. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic
329 assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat*
330 *Biotechnol* 2019; **37**: 632–639.
331 32. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico
332 evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*
333 2017; **5**: e3817.
334 33. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11
335 viruses in the ocean. *Nature* 2013; **494**: 357–360.
336 34. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae
337 family integrate into host genomes. *Environ Microbiol* 2018.
338 35. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using
339 metagenomics. *PLoS Genet* 2013; **9**: e1003987.
340 36. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced
341 analysis and visualization platform for 'omics data. *PeerJ* 2015; **3**: e1319.
342 37. Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Temperton B, et al. Marine DNA
343 Viral Macro- and Microdiversity from Pole to Pole. *Cell* 2019; **177**: 1109–1123.
344 38. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read

345 metagenomics reveals cryptic and abundant marine viruses. *bioRxiv* . 2018. , 345041

346 39. Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB. vConTACT: an iVirus tool to
347 classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* 2017; **5**: e3243.

348 40. Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M. Metagenome Mining Reveals Hidden
349 Genomic Diversity of Pelagimyophages in Aquatic Environments. *mSystems* 2020; **5**.

350 41. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Culturing novel and abundant pelagiphages
351 in the ocean. *Environ Microbiol* 2020.

352 42. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust
353 phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 2008; **36**: W465–9.

354 43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*
355 *Acids Res* 2004; **32**: 1792–1797.

356 44. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and
357 powerful alternative. *Syst Biol* 2006; **55**: 539–552.

358 45. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by
359 maximum likelihood. *Syst Biol* 2003; **52**: 696–704.

360 46. Moon K, Kang I, Kim S, Kim S-J, Cho J-C. Genome characteristics and environmental distribution
361 of the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. *Environ*
362 *Microbiol* 2017; **19**: 4714–4727.

363 47. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection
364 of viruses within metagenomic sequence data. *PLoS One* 2014; **9**: e105067.

365 48. Eddy S. HMMER: profile HMMs for protein sequence analysis. 1998.

366 49. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
367 *EMBnet.journal* 2011; **17**: 10–12.

368 50. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-
369 resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**: 581–583.