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

Holger H. Buchholz[1], Michelle L Michelsen[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 a new SAR11 and three new members of the methylotrophic bacterial clade OM43, and used these as bait to isolate 117 new phages including the first known siphophage infecting SAR11, and the first known phages for OM43. Genomic analyses of 13 novel viruses revealed representatives of three new viral genera, and infection assays showed that SAR11 viruses have ecotype-specific host-ranges. Similar to the abundant human-associated phage ϕCrAss001, infection dynamics within the majority of isolates suggested either prevalent lysogeny, irrespective of detectable associated genes, and/or host phenotypic bistability, with lysis putatively maintained within a susceptible subpopulation. Broader representation of virus-host systems in culture collections and genomic databases will improve both our understanding of virus-host interactions, and accuracy of computational approaches to evaluate ecological patterns from metagenomic data.

Introduction

40 It is estimated that viral predation kills ~15% of bacterial cells in marine surface water each day [1]
41 and is a major contributor to nutrient recycling via the viral shunt, where marine viruses make cell-
42 bound nutrients available to the neighbouring microbial community through viral lysis of host cells
43 [2, 3]. Viruses are key players in the modulation of carbon fluxes across the oceans (150 Gt/yr),
44 increasing particle aggregation and sinking to depth [2, 4], and accounting for 89% of the variance in
45 carbon export from surface water to the deep ocean [5]. Viruses alter host metabolism through
46 auxiliary metabolic genes (AMGs), increasing and altering the cellular carbon intake of infected cells
47 [6]. Virus-host interactions also increase co-evolutionary rates of both predator and prey via Red
48 Queen dynamics [7, 8]. While recent metagenomic advances have provided major insight into global
49 viral diversity and abundance [9–12], mechanistic understanding of virus-host interactions in
50 ecologically important taxa is reliant on experimental co-culturing of model systems. In
51 cyanobacteria, such systems have shown that viruses increase duration of photosynthetic function
52 [13] and can inhibit CO_2 fixation, providing direct evidence that viruses of abundant phototrophs
53 have a direct and major role in nutrient regulation and global carbon cycling [14]. Furthermore,
54 isolation of novel viral model systems provides complete or near-complete viral genomes with known
55 hosts, critical to the development, ground-truthing and application of computational methods to
56 identify viral genomes in metagenomic data (e.g. VirSorter [15]; VirFinder [16]; MARVEL [17]);
57 quantify boundaries for viral populations [12, 18] and genera (VConTACT2 [19]); understand the
58 importance of AMGs in altering nutrient flux in natural communities [20] and to predict host ranges
59 of uncultured viruses *in situ* (e.g. WIsH [21, 22]).
60
61 In comparison to viruses of primary producers, such as cyanophage, which are well represented with
62 model systems and well-studied in the laboratory, virus-host model systems for similarly important
63 and abundant marine heterotrophic bacteria are rare. Available models of viruses infecting
64 heterotrophs are heavily biased towards those with a fast-growing copiotrophic host that grow readily
65 on solid agar, enabling the use of plaque assays to isolate associated viruses. Such systems are not

representative of the vast majority of heterotrophs in nutrient-limited soil and aquatic environments, which are dominated by slow-growing, oligotrophic taxa with few regulatory mechanisms and complex auxotrophies that do not lend themselves to growth on solid or rich medium [23–26]. Advances in Dilution-to-Extinction culturing of ecologically important hosts have enabled the cultivation of many fastidious bacterial taxa that are not amenable to growth on solid media from soil, marine [27, 28] and freshwater environments [29]. Without plaque assays to identify infection and purification of new viral taxa, cultivation of associated viruses is challenging, and exacerbated by the slow growth rates and complex nutrient requirements of their hosts. Paucity of such model systems introduces significant bias in our understanding of viral influence on global carbon biogeochemical cycles. Therefore, it is important that the efforts to isolate heterotrophic bacterial taxa for experimentation and synthetic ecology is matched by efforts to isolate their associated viruses.

Here, we adapted recent advances in Dilution-to-Extinction culturing of hosts [28] and protocols to isolate viruses from liquid media [30] to improve efficiency of recovering novel virus-host systems for fastidious taxa: First, we used sequential enrichment of viruses from natural communities on target hosts to improve rates of viral isolation [31]; Second, we replaced the requirement for time-intensive epifluorescent microscopy with comparative flow cytometry between infected and uninfected hosts to screen for putative viral infection, followed by confirmation using Transmission Electron Microscopy (TEM). We selected the ecologically important, genomically streamlined SAR11 and OM43 heterotrophic marine clades as models for viral isolation. These clades are abundant and important to global carbon biogeochemistry [23, 32–36], but little is known of their associated viral diversity. In the case of viruses infecting SAR11, challenges in assembling abundant and microdiverse genomes from viral metagenomes limit our ability to determine associated host-virus ecology and representatives for *in silico* host prediction [12, 37, 38], particularly for novel species that lack any representation in the databases. This was demonstrated in the successful isolation of the first known pelagiphages by culturing, including the globally dominant HTVC010P, which, prior to

its isolation, was entirely missed in marine viromes [39]. To the best of our knowledge there are no known viruses for OM43, which plays an important role in remineralisation of volatile carbon associated with phytoplankton blooms [32, 34, 40]. As well as their importance in experimentally evaluating the impact of predator-prey dynamics on global marine carbon biogeochemistry, these systems are critical to understand the influence of genome minimalism on host-virus co-evolution. Novel SAR11 and OM43 representatives from the Western English Channel were isolated and used as bait to isolate associated viruses. Initial concentration of viruses in inocula by tangential flow filtration, followed by one to three rounds of sequential enrichment on target hosts in 96-well plates yielded 117 viral isolates from 218 inoculated cultures from seven monthly water samples from September 2018 to July 2019. 94 out of 105 (90%) inoculated SAR11 cultures yielded positive viral infections. 23 out of 113 (20%) inoculated OM43 cultures yielded positive infections. A subsample of positive infections for both clades were subsequently purified and sequenced, revealing the first known siphovirus infecting SAR11 and providing the first known host-virus model for OM43.

Results and Discussion

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

Dilution-to-Extinction culturing for host taxa using natural seawater-based medium on three 96-well plate samples in September 2017 yielded the first SAR11 strain to be isolated from the Western English Channel (named H2P3α) and the first three new OM43 strains (named C6P1, D12P1, H5P1) from the Western English Channel. The full-length 16S rRNA gene of H2P3α was 100% identical to that of the warm water SAR11 ecotype *Pelagibacter bermudensis* HTCC7211 (subclade 1A) and was considered to be a local variant [41] (Supplementary Fig. 1A). All three novel OM43 isolates (named CP61, D12P1, H5P1) were most closely related to *Methylophilales* sp. HTCC2181 (C6P1 96.17%,

118 D12P1 96.62%, H5P1 97.79% nucleotide identity across the full 16S rRNA gene), a streamlined
119 member of the OM43 clade with a 1.3 Mbp genome, isolated from surface water of the North Eastern
120 Pacific [42] (Supplementary Fig. 1B). The average identity of the 16S rRNA gene of isolates CP61,
121 D12P1 and H5P1 to each other was ~98.46% (Supplementary Table 1), suggesting they are
122 representatives of the same genus [43].
123
124 **An efficient, low-cost method of isolating new viruses yielded 117 new viral isolates for SAR11**
125 **and OM43 taxa**
126
127 Using the four new hosts and established SAR11 isolates *Pelagibacter ubique* HTCC1062 (cold-
128 water ecotype) and HTCC7211 (warm water ecotype), we developed an optimised viral isolation
129 pipeline (Fig. 1) and applied it to six monthly water samples from the Western English Channel, taken
130 between September 2018 to April 2019 (Supplementary Table 2). Briefly, each month, we inoculated
131 one to two 96-well Teflon plates containing our four new hosts, plus SAR11 isolates HTCC1062 and
132 HTCC7211 at ~10^6 cells per mL with a natural viral community concentrated by tangential flow
133 filtration. Plates were monitored by flow cytometry and growth of putatively infected cultures was
134 compared to those of uninfected controls over the course of ~2 weeks, due to the slow growth rates
135 of SAR11 and OM43 [34, 44]. Additional samples were taken July 2019 to attempt viral isolation on
136 OM43 strains D12P1 and H5P1. In total, out of 218 cultures inoculated with concentrated natural
137 viral communities for initial isolation, 117 viruses were purified and still infective after at least three
138 passages, evidenced by differences in cytograms between cultures inoculated with viruses and control
139 cultures (Supplementary Fig. 2-5), demonstrating an overall efficiency of 53% and an average yield
140 of 18 isolates per sample (Fig. 2). All viral isolations required between one and three rounds of host
141 enrichment (Supplementary Table 3) before changes in host growth curves between infected and
142 uninfected cultures could be observed by flow cytometry, suggesting a minimal viral density is

143 required to identify infected wells using this approach, but this threshold can be reached through
144 successive rounds of enrichment.
145
146 For each sample, all steps of the initial viral isolation process (counting one to two plates for a total
147 of three times) required ~6 hours run time on a flow cytometer. For the subsequent three rounds of
148 purification, one plate was required for each host-sample combination (another ~6 hours of cytometer
149 run time in total for one host-sample combination). Both initial isolation of viruses from
150 environmental water samples and three rounds of purification took ~10 weeks of incubation time in
151 total. Between all steps, our protocol required ~7 hours of handling time per sample. Following three
152 rounds of viral purification, generating sufficiently high viral titres to extract enough DNA for
153 sequencing (approximately two weeks incubation time, and roughly four hours handling time over
154 two days) was the rate-limiting step, and so required subselection of available viruses for sequencing.
155 Future advances in DNA extraction efficiency, reducing DNA input requirements for sequencing
156 and/or automation of viral DNA extraction will enable all isolated viruses to be sequenced. We
157 estimate the cost of isolating a single virus is ~£20 for cultivation, flow cytometry and DNA
158 extraction consumables, and ~£50-100 for genome sequencing at 30-fold coverage required for
159 successful assembly, giving a total cost of £70-120 per sequenced viral isolate, not including the costs
160 for person-hours. Thus, our protocol provides a high-throughput and scalable approach to viral
161 isolation.
162
163 **Novel pelagiphages are ecotype-specific and persist in the community**
164 We hypothesised that the density dependence of predator/prey dynamics would result in greater
165 isolation efficiency of viruses on the cold-water SAR11 ecotype, HTCC1062, during winter months,
166 when this host is more abundant. Contrary to our expectations, the isolation efficiencies (number of
167 successes / number of attempts) were: 80% for HTCC7211, 93% for HTCC1062 and 96.6% for
168 H2P3α, but these were not significantly different (see Supporting Methods: HTCC1062 vs. H2P3α

169 (*P*=0.5, 999 bootstraps); HTCC1062 vs. HTCC7211 (*P=0.18*, 999 bootstraps)). Temperature ranged
170 from 9.3 °C - 14.1 °C from October to April, when isolation was attempted on all strains
171 (Supplementary Table 2). Neither temperature, ecotype nor an ecotype*temperature interaction term
172 were significantly correlated with isolation efficiency based on a generalized linear model using a
173 quasi-binomial error distribution (see Supplementary Methods). One explanation for this result is that
174 host ecotype abundances are constant throughout the year and unaffected by temperature. HTCC1062
175 and HTCC7211 have specific growth rates of ~0.22 and ~0.12 divisions per day at 10 °C, respectively
176 [44], and our new SAR11 isolate from the Western English Channel, H2P3α, showed similar growth
177 rates to HTCC7211 (Supplementary Fig. 6), with specific growth rates of ~0.10, ~0.45 and ~0.84
178 divisions per day at 10, 18 and 25 °C, respectively. Measured lower specific growth rates of warm
179 water ecotypes at winter temperatures, and evidence of global ecotype niche partitioning [45] suggest
180 that uniform ecotype abundance across seasons in the Western English Channel is unlikely, but that
181 slow growth of warm water ecotypes at *in situ* temperatures is possible.
182
183 An alternative explanation is that isolated viruses have a broad host range that encompasses both
184 warm- and cold-water ecotypes. We tested the host range (Table 1) of six pelagiphage population
185 representatives out of the 12 pelagiphages isolated from samples in October 2018 and November
186 2018 (*in situ* water temperature of 14.8 °C and 14.2 °C respectively) across the three SAR11 strains
187 (Table 2). Viruses *Eistla, Eyrgjafa* and *Greip* all infected cold-water ecotype HTCC1062,
188 exclusively, while *Ran* and *Kolga* only infected warm-water ecotypes HTCC7211 and H2P3α.
189 Therefore, our new pelagiphages appear to be broadly ecotype-specific, confirming previous findings
190 [46], with important consequences for their global distribution. *Bylgia* was the only virus that could
191 infect both warm and cold-water ecotypes. Ecotypic specificity at first seems at odds with our ability
192 to isolate viruses on the warm water SAR11 ecotype at *in situ* temperature far below its optimal
193 growth temperature. A possible explanation is that pelagiphages are ecotype-specific, but persist in
194 the water column throughout the year in sufficient densities to be isolated by our enrichment method.

195 Measured *in situ* temperatures during our sampling period were sufficient to support slow growth of
196 warm-water ecotypes even during winter months, potentially providing sufficient prey to support a
197 population of warm-water ecotype specific phages. If concentration and enrichment of viruses during
198 isolation is sufficient to successfully isolate even low abundance phages then a comprehensive library
199 of representative phage isolates could be generated with relatively modest sampling effort across a
200 few locations.
201
202 **First methylophages for marine OM43 isolated.**
203 Viral isolation of novel methylophages for the OM43 clade yielded 23 positive infections, with
204 efficiencies ranging from 0% (no viruses isolated on host C6P1) to 45% on H5P1 (Fig. 2). To the best
205 of our knowledge these are the first reported viruses infecting members of the OM43 clade. One
206 explanation for the lower efficiency of isolation of OM43 viruses is simply one of lower host
207 abundance concomitant with lower phage abundance in the viral community, reducing the likelihood
208 of infective viruses coming into contact with susceptible and permissive cells. OM43 are closely
209 associated with metabolism of extracellular substrates from phytoplankton blooms [47], but have low
210 abundance outside of phytoplankton spring blooms [32]. Our water samples were not associated with
211 high *in-situ* fluorescence (used as proxy measurement for phytoplankton), and missed the April 2019
212 Spring bloom by about two weeks (Supplementary Fig. 7). Therefore, OM43 and their viruses were
213 likely of lower abundance at sampling times, decreasing the success rate of viral isolation from these
214 samples.
215
216 **New viruses represent novel viral populations and support established ANI cut-offs for**
217 **ecologically discrete viral ecotypes**
218 Following three rounds of purification through Dilution-to-Extinction, due to the rate-limiting step of
219 extracting viral DNA, we subselected 16 viral isolates based on availability in November 2018 across
220 four different hosts (HTCC1062, HTCC7211, H2P3α and H5P1) for sequencing at >30-fold coverage

221 (Table 2). Three out of 16 sequenced samples (two from host HTCC7211, one from OM43 host H5P1)
222 failed to assemble into single viral contigs. For 11 of the remaining 13 samples (12 from SAR11 hosts
223 and one from OM43) each individual sequence assembly was identified as a complete viral genome
224 by VirSorter [15] and 95-100% complete using CheckV [48]. OM43 phage *Venkman* possesses a
225 linear but complete genome (Supplementary Table 4). Sequence data yielded a single viral genome
226 (categories 1 or 2, >15kbp) per sample, indicating that our purification process was effective in
227 recovering pure viral isolates.
228
229 Viral populations are defined as discrete ecological and evolutionary units that form non-overlapping
230 clouds of sequence space, and average nucleotide identity (ANI) cutoffs of >95% have been proposed
231 from cyanophage studies to delineate viral populations [49]. To evaluate whether our new isolates
232 represented discrete ecological units with respect to other known pelagiphages, pairwise ANI was
233 calculated between the thirteen successfully sequenced viral genomes from this study and 85 other
234 known or putative pelagiphages [37, 46, 50]. Pairwise ANI ranged between 77.5-100%, with a
235 discrete distribution between 96.4-100.0% (Supplementary Fig. 8, Supplementary Table 5). This is
236 in agreement with previous work in cyanophages [49, 51] and thus supports their application to
237 define viral populations beyond cyanophages for viral genomes derived from metagenomic assembly
238 [10, 11]. At proposed ANI cutoffs of 95% over 85% length [18], the new viruses clustered into six
239 viral populations, ranging from singletons to a viral population with four members (Table 2). Phages
240 within the same populations were all isolated from the same source water sample and on the same
241 host, in agreement with their classification as discrete ecological and evolutionary units. Interestingly,
242 viral population membership was exclusive to viruses isolated in this study, with no associated
243 populations containing representatives from either known isolates [39, 46] or fosmid-derived [37]
244 genomes from other studies, including pelagiphage HTVC010P [39], which is ubiquitous and
245 dominant throughout global oceans, including the Western English Channel [12]. Such a distinct
246 separation between populations associated with viruses sampled in this study and all other

247 pelagiphages, either in culture or assembled from metagenomes, suggests a high degree of viral
248 population diversity remains to be discovered in the Western English Channel and beyond.
249
250 **Current pelagiphage isolates can be organised into five distinct phylogenetic clades**
251
252 To evaluate isolate diversity at higher taxonomic organisation, chosen representatives of each of the
253 six viral populations were compared to previous isolates and fosmid-derived phage sequences using
254 three approaches: First, where possible, phylogenetic analysis was performed based on conserved
255 genes; Second, to capture broader relationships and account for genomic mosaicism, raw
256 hypergeometric probability of shared gene content; Finally genomes were organised into ICTV-
257 recognized genera using vConTACT2 (Supplementary Fig. 9), which initially derives viral clusters
258 using a similar hypergeometric approach followed by refinement with Euclidean-distance-based
259 hierarchical clustering to split mismatched, 'lumped' viral clusters [19]. All three approaches were
260 congruent - clustering on probability of shared gene content organising pelagiphage genomes into
261 four main clusters and one singleton genome from this study (*Kolga*) (Fig. 3A-C), which was broadly
262 supported by phylogenetic and vConTACT2 classifications. Cluster A contained 23 members (nine
263 from fosmid-derived contigs [37]; eleven previously isolated pelagiphages [46, 52]), including *Ran*,
264 *Bylgja* and *Eistla* from this study. Cluster B contained two previously isolated pelagiphages, one
265 fosmid-derived contig and *Eyrgjafa* from this study. All viruses in Clusters A and B were assigned to
266 a single viral genus by vConTACT2 that also contained 12 previously isolated pelagiphages [39, 46].
267 Cluster C only contained fosmid-derived contigs from the Mediterranean [37], with no isolated
268 representatives, making it an important location for future isolation attempts of different hosts and
269 viruses using our method. Cluster D contained eight fosmid-derived contigs, pelagiphage
270 HTVC010P, and a new representative, *Greip*, from this study (Fig. 3B). VConTACT2 split Cluster
271 D, leaving *Greip* and *Kolga* as members of two singleton clusters, suggesting they are the first
272 cultured representatives of novel viral genera.

273

274 **Kolga - the first siphovirus infecting SAR11**

275

276 The 25 previously known viral isolates infecting SAR11 comprise 24 podoviruses and one myovirus
277 [39, 46, 53]. Previous cultivation efforts for viruses of SAR11 has not isolated any siphoviruses, nor
278 are any known from viral metagenomic studies. We isolated and sequenced the two viruses *Kolga*
279 and *Aegir* using H2P3α as bait, which to the best of our knowledge are the first known siphoviruses
280 infecting members of the SAR11 clade (Fig. 4A-C). The phages appeared to have a long tail on
281 transmission electron microscopy (TEM) images and are thus morphologically indicative of
282 belonging to the *Siphoviridae* (Fig. 4B). *Aegir* and *Kolga* were members of the same population using
283 a boundary cutoff of 95% ANI over 85% contig length, however, *Aegir* had a length of 18,297 bp
284 compared to 48,659 bp in *Kolga*, therefore we considered *Aegir* to be a partial genome of the same
285 viral population. *Kolga* did not share a significant number of genes with known SAR11 podoviruses
286 (Fig. 3B), and did not cluster with other pelagiphages using a hypergeometric analysis based on gene
287 content. VConTACT2 also grouped *Aegir* and *Kolga* into one cluster without any other known
288 viruses, suggesting it represents a novel viral genus. A number of genes found in *Kolga* were shared
289 with other known siphoviruses, and phylogenetic analysis of concatenated genes shows *Kolga* is most
290 closely related to LK3, a known siphovirus infecting the Betaproteobacterial genus *Bordetella* (Fig.
291 4C). This suggests that *Kolga* is more closely related to siphoviruses of different hosts compared to
292 podoviruses infecting the same host. 67% of genes encoded by Kolga could not be functionally
293 annotated, and out of all hypothetical genes identified on *Kolga*, only three hypothetical genes were
294 shared with SAR11 podoviruses. In contrast, for on average ~90% of genes without known function
295 identified within our novel SAR11 podoviruses are shared between different pelagiphages. *Kolga* was
296 the 10th most abundant viral isolate in our Western English Channel virome (Fig. 3D). *Kolga*
297 possesses a tail tip J protein (Fig. 4A), often found in phages with long non-contractile tails such as
298 λ, where it plays a role in DNA injection during cell entry and tail assembly [54]. *Kolga* also encodes

299 a small S21 subunit of the 30S ribosomal gene structurally similar to the ones found in HTVC008M,
300 and hosts HTCC7211 and H2P3α, a feature found in numerous myoviruses and sipoviruses [55]. S21
301 is involved in translation initiation and needed for the binding of mRNA [56]. Virally-encoded S21
302 genes may provide a competitive advantage for the phage as it could replace cellular S21 and assist
303 in the translation of viral transcripts. *Kolga* may need its S21 gene for shifting the translational frame,
304 as it has been shown that for some members of the Caudovirales the production of tail components is
305 dependent on programmed translational frameshifting [57]. Given the constitutive nature of gene
306 expression in genomically streamlined bacteria [58], genes such as S21 may provide the virus with a
307 mechanism to manipulate host metabolism in the absence of typical promoters and repressors.
308
309 **Kolga and Greip are the most abundant of the novel viruses in Western English Channel**
310
311 To evaluate relative abundance of our new isolates in the Western English Channel, we used
312 competitive mapping of metagenomic reads from a Western English Channel virome taken in
313 September 2016 (*in situ* surface water temperature of ~16.1 °C) [12] against existing pelagiphage
314 genomes and those from this study using FastViromeExplorer [59]. This identified HTVC010P as
315 the most abundant virus within the known pelagiphages (536,531 Reads Per Million [RPM - see
316 Supplementary Methods]). Other isolates from this and previous studies recruited 1-3 orders of
317 magnitude fewer reads than HTVC010P (684 - 53,591 RPM), indicating that viral community
318 structures for pelagiphages in the Western English Channel consist of a small number of viral species
319 that are highly abundant, and long tail of lower abundance members (Fig. 3D). *Greip* and *Kolga* were
320 the two most abundant of our new viral isolates in the Western English Channel. Their abundance
321 could be driven in part by the fact that either the *in situ* surface water temperature when the virome
322 sample was taken matches the optimum growth temperature of host HTCC1062 (~0.44 divisions per
323 day at 16 °C) [44], in line with proposed Kill-the-Winner predator/prey dynamics [60], and/or that
324 genomic novelty promotes recruitment of reads closely related taxa that have no other representation

325 in the reference dataset [39]. Future work will evaluate the global abundance of these and newly
326 sequenced isolates across spatial (Global Ocean Viromes) and temporal gradients (an ongoing time-
327 series project in the Western English Channel) to evaluate their biogeography.
328
329 **Venkman - the first known virus infecting OM43**
330
331 Our isolation efforts in the Western English Channel yielded the first cultured bacteriophage infecting
332 the methylotroph OM43 (Fig. 5A-C). Named *Venkman*, in homage to the actor Bill Murray in the
333 film Ghostbusters, this virus has a genome 38,624 bp long (31.9% GC content), encoding genes (Fig.
334 5A) with similar synteny and function to the siphovirus P19250A (38,562 bp) that infects freshwater
335 *Methylophilales* LD28, which are often considered a freshwater variant of OM43 [61, 62]. Unlike
336 P19250A, TEM images indicated that *Venkman* had a short tail (Fig. 5B) similar to podoviruses,
337 though it is possible that tail structures had been lost during grid preparation. Phylogenetic analysis
338 of concatenated TerL and exonuclease genes further indicated that these genes in *Venkman* are most
339 closely related to other siphoviruses infecting different Proteobacteria (Fig. 5C). However, branch
340 support values were low, despite numerous attempts to refine the tree with different approaches (see
341 Supplementary Methods). We identified a number of common phage proteins such as a capsid
342 protein, terminase, nucleases and tail structural proteins, the remaining 54% of genes were
343 hypothetical. VConTACT2 assigned *Venkman* and P19250 to the same genus-level cluster, therefore
344 *Venkman* may be a marine variant of the freshwater phage P19250. In the Western English Channel
345 viral metagenomes, *Venkman*-like viruses were the second most abundant viruses (148,681 RPM)
346 after HTVC010P (536,531 RPM) identifying them as ecologically important viruses in this coastal
347 microbial community (Fig. 3B).
348
349 **Unusual host-virus dynamics are prevalent in isolated phages**
350

351 Viral isolates *Bylgja*, *Eyrgjafa*, *Ran* and *Eistla* (Cluster A and B) are presumed to be temperate
352 phages, with all encoding endonucleases and exonucleases (Supplementary Fig. 10) and clustering
353 by shared protein content with other pelagiphages, such as e.g. HTVC011P and HTVC025P, shown
354 previously to integrate into host genomes (Fig. 3B) [46]. *Eyrgjafa* encodes a tRNA-Leu, that has 85%
355 nucleotide identity over the first 34 bases of the tRNA-Leu of its host HTCC1062, suggesting a
356 putative integration site into the host genome [63]. To date, 16 of the 29 viruses isolated on SAR11
357 strains have either been shown to be capable of lysogeny, and/or encode genes associated with a
358 temperate infection cycle. In contrast, viruses such as *Greip,* and the abundant HTVC010P in Cluster
359 D (Fig. 3B) do not possess any genes associated with lysogeny, and would therefore presumed to be
360 exclusively lytic. Viruses in Clusters A and B, presumed temperate, were of much lower abundance
361 in the environment compared to *Greip* or HTVC010P in Cluster D which were among the most
362 abundant pelagiphages in the Western English Channel (Fig. 3D), suggesting a possible ecological
363 difference between two groups of viruses with different infectivity strategies.
364
365 Interestingly, growth curves of hosts infected with *Greip* and other isolated phages deviated from the
366 expected decay in cell abundance associated with viral lysis and previously observed in isolated
367 pelagiphages [39]. The first pelagiphages HTVC010P and HTVC008M were isolated from the warm
368 waters of the Sargasso Sea, and HTVC011P and HTVC019P were isolated from the colder waters of
369 the Oregon Coast. All four strains were propagated on the cold-water SAR11 ecotype *P. ubique*
370 HTCC1062. In all cases, host density was reduced from ~8×10^6 cells per mL at T_0 to <10^6 cells per
371 mL over a 60-72h period, although viruses from warm waters took 17% longer than those from cold
372 water to do so [39], suggesting that suboptimal hosts reduced the rate of infection as shown in
373 cyanophages [64]. In contrast, infection dynamics of our isolates often resulted in host density of
374 infected cultures growing to a steady state, but at a lower cell density than uninfected cells.
375 (Supplementary Video 1), irrespective of cluster or population assignment. Out of 117 viruses isolated
376 in this study, only 16 infections reduced host abundance below their inoculum density of 10^6 cells per

377 mL. In 53 infections, densities of infected cells increased to within an order of magnitude of
378 uninfected cells (Supplementary Fig. 11), but demonstrated clear evidence of viral infection in
379 cytograms, TEMs and subsequent recovery of viral genomes in selected samples.
380
381 Similar patterns of infection were recently reported in the extremely abundant bacteriophage
382 ϕCrAss001 found in the human gut, where infection of exponentially growing cells of *B. intestinalis*
383 919/174 did not result in complete culture lysis, but caused a delay in stationary phase onset time and
384 final density, despite lacking genes associated with lysogeny. As with our study, the authors observed
385 that this only occurred in liquid culture, and isolation of the virus required numerous rounds of
386 enrichment. They postulated that the virus may cause a successful infection in only a subset of host
387 cells, with the remainder exhibiting alternative interactions such as pseudolysogeny or dormancy
388 [31]. The prevalence of similar infection dynamics in the phages isolated in this study offer two
389 intriguing possibilities: (i) Many of the viruses isolated in this study are either not fully lytic, but fall
390 somewhere on the continuum of persistence [65], controlled by genes currently lacking a known
391 function, with lysogeny (and associated superinfection immunity) favoured at high cell density, in
392 support of the Piggyback-the-Winner hypothesis [66, 67]; (ii) the steady-state of host and virus
393 densities observed here are an indicator of host phenotypic bistability in these streamlined
394 heterotrophic taxa. Viral propagation occurring in only a subset of cells could explain the requirement
395 of multiple rounds of enrichment before sufficient viral load is reached to be able to observe lytic
396 infection on the host population. Either strategy, or a combination of both, would provide an
397 ecological advantage of long-term stable coexistence between viruses and hosts and offer an
398 explanation of the paradox of stable high abundances of both predator and prey across global oceans
399 [12, 39, 68, 69]. Infection in a subset of the population could also explain the low lytic activity
400 observed in pelagiphages *in situ*, despite high host densities [70]; and the small dynamic range and
401 decoupled abundances of SAR11 and virioplankton in the Sargasso Sea [71]. Limited lysis of
402 subpopulations of hosts such as SAR11 and OM43 that specialise in the uptake of labile carbon

403 enriched through viral predation [72, 73] would facilitate efficient intra-population carbon recycling
404 and explain the limited influence of SAR11 and associated viral abundances on carbon export to the
405 deep ocean [5]. We propose the moniker the 'Soylent Green Hypothesis' for this mechanism, after
406 the 1973 cult film in which the dead are recycled into food for the living. Further investigation
407 leveraging our new virus-host model will provide greater insight into viral influence on ocean carbon
408 biogeochemistry.
409
410 In conclusion, our method coupled Dilution-to-Extinction cultivation of hosts and associated viruses,
411 resulting in the isolation of three new strains of OM43; a Western English Channel variant of a warm
412 water ecotype of SAR11; the first known methylophages for OM43; the first siphovirus infecting
413 SAR11, as well as eleven other viruses infecting this important marine heterotrophic clade and >100
414 more isolates to be sequenced and explored. The described method represents an efficient and cost-
415 effective strategy to isolate novel virus-host systems for experimental evaluation of co-evolutionary
416 dynamics of important fastidious taxa from marine and other biomes. Coupling these methods to
417 existing advances in host cultivation requires minimal additional effort and will provide valuable
418 representative genomes to improve success rates of assigning putative hosts to metagenomically-
419 derived viral contigs. Broader representation of model systems beyond cyanophages and viruses of
420 copiotrophic, *r*-strategist hosts will reduce bias in developing methods to delineate viral population
421 boundaries [74, 75], increasing the accuracy with which we derive ecological meaning from viral
422 metagenomic data. We therefore hope that this method will enable viruses to be included in the
423 current resurgence of cultivation efforts to better understand the biology and ecology of phages, and
424 the influence of the world's smallest predators on global biogeochemistry.
425
426 **Methods Summary**

427 A complete description of the materials and methods is provided in the Supplementary Information.
428 Four bacterial strains (*Methylophilales sp*. C6P1, D12P1 and H5P1; *Pelagibacter sp*. H2P3α) were
429 isolated from Western English Channel station L4 seawater samples using Dilution-to-Extinction
430 methods [28]. All four bacteria and two additional SAR11 strains *Pelagibacter bermudensis*
431 HTCC7211 and *Pelagibacter ubique* HTCC1062 were used as bait to isolate phages from six monthly
432 Western English Channel L4 seawater samples (50°15.00N; 4°13.00W). Briefly, water samples were
433 concentrated for viruses using tangential flow filtration and used as viral inoculum (10% v/v) in
434 exponentially growing cultures of host bacteria in artificial seawater medium [76] in 96-well Teflon
435 plates (Radleys, UK). Cells of the resulting lysate were filtered out and the filtrate was used as viral
436 inoculum in another round of isolation. This process was repeated until viral infection could be
437 detected by flow cytometry - comparing cytograms and maximum density of infected cultures against
438 uninfected cultures. Phages were purified by dilution-to-extinction methods (detailed protocol
439 available here: dx.doi.org/10.17504/protocols.io.c36yrd). Phage genomes were sequenced using
440 Illumina 2x150 PE sequencing, assembled and manually annotated as described in [77]. Phylogenetic
441 classification of phages was performed on concatenated shared genes using a combination of
442 Bayesian inference trees, maximum likelihood trees and shared-gene likelihood analyses, depending
443 on the availability of appropriate taxa. ICTV-recognised genera based on shared gene content were
444 assigned with VConTACT2 [19]. The relative abundance of novel phages in the Western English
445 Channel was calculated by competitive read recruitment using FastViromeExplorer v1.3 [59] against
446 an existing virome from the same location [12].

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

450

451 **Acknowledgements** We would like to thank Christian Hacker and the Bioimaging Centre of the
452 University of Exeter for performing the TE microscopy and imaging. We would also like to thank the

crew of the *R/V Plymouth Quest* and our collaborators at PML for collecting water samples, and the driver Magic for delivering water samples from Plymouth to Exeter. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk) which is supported by the BBSRC (grant number BB/L024209/1).

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

Conflict of interests The authors declare no conflict of interest.

References

1. Eric Wommack K, Colwell RR. Virioplankton: Viruses in Aquatic Ecosystems. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS* 2000; **64**: 69–114.
2. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, et al. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J* 2015; **9**: 1352–1364.
3. Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* 2012; **4**: 17.
4. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol* 2014; **12**: 519–528.
5. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 2016; **532**: 465–470.
6. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virol J* 2019; **16**: 15.

480 7. Weitz JS, Hartman H, Levin SA. Coevolutionary arms races between bacteria and bacteriophage. *Proc*
481 *Natl Acad Sci U S A* 2005; **102**: 9535–9540.
482 8. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. Long-term stability and Red Queen-like strain
483 dynamics in marine viruses. *Nat Microbiol* 2020; **5**: 265–271.
484 9. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant
485 temperate viruses affect Southern Ocean microbial dynamics. *ISME J* 2016; **10**: 437–449.
486 10. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential
487 biogeochemical impacts of globally abundant ocean viruses. *Nature* 2016; **537**: 689–693.
488 11. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA
489 Viral Macro- and Microdiversity from Pole to Pole. *Cell* 2019; **177**: 1109–1123.e14.
490 12. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral
491 metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic
492 islands. *PeerJ* 2019; **7**: e6800.
493 13. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. *Nature*
494 2003; **424**: 741.
495 14. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Viruses Inhibit CO2 Fixation in the Most Abundant
496 Phototrophs on Earth. *Curr Biol* 2016; **26**: 1585–1589.
497 15. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic
498 data. *PeerJ* 2015; **3**: e985.
499 16. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying
500 viral sequences from assembled metagenomic data. *Microbiome* 2017; **5**: 69.
501 17. Amgarten D, Braga LPP, da Silva AM, Setubal JC. MARVEL, a Tool for Prediction of Bacteriophage
502 Sequences in Metagenomic Bins. *Front Genet* 2018; **9**: 304.
503 18. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation
504 of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 2017; **5**: e3817.
505 19. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment
506 of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* 2019.
507 20. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in

508 the ‘core’ and ‘flexible’ Pacific Ocean Virome. *ISME J* 2015; **9**: 472–484.
509 21. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts
510 from metagenomic phage contigs. *Bioinformatics* 2017; **33**: 3113–3114.
511 22. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d_2^* oligonucleotide frequency
512 dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences.
513 *Nucleic Acids Res* 2017; **45**: 39–53.
514 23. Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, et al. One carbon metabolism in
515 SAR11 pelagic marine bacteria. *PLoS One* 2011; **6**: e23973.
516 24. Becker JW, Hogle SL, Rosendo K, Chisholm SW. Co-culture and biogeography of Prochlorococcus and
517 SAR11. *ISME J* 2019.
518 25. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA. Diversity and biogeography of
519 SAR11 bacteria from the Arctic Ocean. *ISME J* 2020; **14**: 79–90.
520 26. Robert M. Morris*, Michael S. Rappe´ *, Stephanie A. Connon*, Kevin L. Vergin*, William A.
521 Siebold*, Craig A. Carlson† & Stephen J. Giovannoni*. SAR11 clade dominates ocean surface
522 bacterioplankton communities. *Nature* 2002; **420**.
523 27. Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-
524 nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* 2002; **68**: 3878–3885.
525 28. Henson MW, Pitre DM, Weckhorst JL, Celeste Lanclos V, Webber AT, Cameron Thrash J. Artificial
526 Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico.
527 *American Society for Microbiology* 2016; **1**: 1–9.
528 29. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. Cultivation and genomics of the first freshwater
529 SAR11 (LD12) isolate. *ISME J* 2018; **12**: 1846–1860.
530 30. Nagasaki K, Bratbak G. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists.
531 *Manual of aquatic viral ecology ASLO* 2010; 92–101.
532 31. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001
533 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis.
534 *Nat Commun* 2018; **9**: 4781.
535 32. Morris RM, Longnecker K, Giovannoni SJ. Pirellula and OM43 are among the dominant lineages

536 identified in an Oregon coast diatom bloom. *Environ Microbiol* 2006; **8**: 1361–1370.
537 33. Ramachandran A, Walsh DA. Investigation of XoxF methanol dehydrogenases reveals new
538 methylotrophic bacteria in pelagic marine and freshwater ecosystems. *FEMS Microbiol Ecol* 2015; **91**.
539 34. Halsey KH, Carter AE, Giovannoni SJ. Synergistic metabolism of a broad range of C1 compounds in
540 the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* 2012; **14**: 630–640.
541 35. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic
542 Ocean: a biogeographical study of bacterial substrate utilisation. *ISME J* 2019; **13**: 1119–1132.
543 36. Carini P, White AE, Campbell EO, Giovannoni SJ. Methane production by phosphate-starved SAR11
544 chemoheterotrophic marine bacteria. *Nat Commun* 2014; **5**: 4346.
545 37. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using
546 metagenomics. *PLoS Genet* 2013; **9**: e1003987.
547 38. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et
548 al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 2017; **8**:
549 15892.
550 39. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11
551 viruses in the ocean. *Nature* 2013; **494**: 357–360.
552 40. Taubert M, Grob C, Howat AM, Burns OJ, Dixon JL, Chen Y, et al. XoxF encoding an alternative
553 methanol dehydrogenase is widespread in coastal marine environments. *Environ Microbiol* 2015; **17**:
554 3937–3948.
555 41. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ. The presence of the glycolysis
556 operon in SAR11 genomes is positively correlated with ocean productivity. *Environ Microbiol* 2010;
557 **12**: 490–500.
558 42. Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, et al. The small genome of an
559 abundant coastal ocean methylotroph. *Environ Microbiol* 2008; **10**: 1771–1782.
560 43. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification
561 of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*
562 2014; **12**: 635–645.
563 44. Polyphasic taxonomy of Marine bacteria. *studylib.net*. https://studylib.net/doc/5577926/polyphasic-

564 taxonomy-of-marine-bacteria. Accessed 11 May 2020.
565 45. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, et al. Single-amino acid variants reveal
566 evolutionary processes that shape the biogeography of a global SAR11 subclade. *Elife* 2019; **8**.
567 46. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family
568 integrate into host genomes. *Environ Microbiol* 2018.
569 47. Neufeld JD, Boden R, Moussard H, Schäfer H, Murrell JC. Substrate-specific clades of active marine
570 methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Appl Environ*
571 *Microbiol* 2008; **74**: 7321–7328.
572 48. Nayfach S, Camargo AP, Eloe-Fadrosh E, Roux S, Kyrpides N. CheckV: assessing the quality of
573 metagenome-assembled viral genomes. *bioRxiv* . 2020. , 2020.05.06.081778
574 49. Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, et al. Viral tagging
575 reveals discrete populations in Synechococcus viral genome sequence space. *Nature* 2014; **513**: 242–
576 245.
577 50. Våge S, Storesund JE, Thingstad TF. SAR11 viruses and defensive host strains. *Nature* . 2013. , **499**:
578 E3–4
579 51. Marston MF, Martiny JBH. Genomic diversification of marine cyanophages into stable ecotypes.
580 *Environ Microbiol* 2016; **18**: 4240–4253.
581 52. Giovannoni S, Temperton B, Zhao Y. Giovannoni et al. reply. *Nature* . 2013. , **499**: E4–5
582 53. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Novel pelagiphages prevail in the ocean.
583 *bioRxiv* . 2019. , 2019.12.14.876532
584 54. Tam W, Pell LG, Bona D, Tsai A, Dai XX, Edwards AM, et al. Tail tip proteins related to
585 bacteriophage λ gpL coordinate an iron-sulfur cluster. *J Mol Biol* 2013; **425**: 2450–2462.
586 55. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous
587 cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun* 2019; **10**: 752.
588 56. Van Duin J, Wijnands R. The function of ribosomal protein S21 in protein synthesis. *Eur J Biochem*
589 1981; **118**: 615–619.
590 57. Xu J, Hendrix RW, Duda RL. Conserved translational frameshift in dsDNA bacteriophage tail assembly
591 genes. *Mol Cell* 2004; **16**: 11–21.

592 58. Giovannoni SJ, Thrash JC, Temperton B. Implications of streamlining theory for microbial ecology.
593 *ISME J* 2014; **8**: 1553–1565.
594 59. Tithi SS, Aylward FO, Jensen RV, Zhang L. FastViromeExplorer: a pipeline for virus and phage
595 identification and abundance profiling in metagenomics data. *PeerJ* 2018; **6**: e4227.
596 60. Thingstad TF, Lignell R. Theoretical models for the control of bacterial growth rate, abundance,
597 diversity and carbon demand. *Aquat Microb Ecol* 1997; **13**: 19–27.
598 61. Moon K, Kang I, Kim S, Kim S-J, Cho J-C. Genome characteristics and environmental distribution of
599 the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. *Environ*
600 *Microbiol* 2017; **19**: 4714–4727.
601 62. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. The ecology of pelagic freshwater
602 methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J* 2015; **9**: 2442–
603 2453.
604 63. Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages.
605 *Genome Res* 2007; **17**: 1486–1495.
606 64. Enav H, Kirzner S, Lindell D, Mandel-Gutfreund Y, Béjà O. Adaptation to sub-optimal hosts is a driver
607 of viral diversification in the ocean. *Nat Commun* 2018; **9**: 4698.
608 65. Weitz JS, Li G, Gulbudak H, Cortez MH, Whitaker RJ. Viral invasion fitness across a continuum from
609 lysis to latency. *Virus Evol* 2019; **5**: vez006.
610 66. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, et al. Lytic to temperate
611 switching of viral communities. *Nature* 2016; **531**: 466–470.
612 67. Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. *NPJ*
613 *Biofilms Microbiomes* 2016; **2**: 16010.
614 68. Kang I, Oh HM, Kang D, Cho JC. Genome of a SAR116 bacteriophage shows the prevalence of this
615 phage type in the oceans. *Proc Natl Acad Sci U S A* 2013; **110**: 12343–12348.
616 69. Martinez-Hernandez F, Fornas Ò, Lluesma Gomez M, Garcia-Heredia I, Maestre-Carballa L, López-
617 Pérez M, et al. Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant
618 uncultured 37-F6 viral population in the ocean. *ISME J* 2018.
619 70. Alonso-Sáez L, Morán XAG, Clokie MR. Low activity of lytic pelagiphages in coastal marine waters.

620 *ISME J* 2018; **12**: 2100–2102.
621 71. Parsons RJ, Breitbart M, Lomas MW, Carlson CA. Ocean time-series reveals recurring seasonal
622 patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J* 2012; **6**: 273–284.
623 72. Middelboe M, Lyck PG. Regeneration of dissolved organic matter by viral lysis in marine microbial
624 communities. *Aquat Microb Ecol* 2002; **27**: 187–194.
625 73. Middelboe M, Jorgensen NOG. Viral lysis of bacteria: an important source of dissolved amino acids
626 and cell wall compounds. *J Mar Biol Assoc U K* 2006; **86**: 605–612.
627 74. Brum JR, Sullivan MB. Rising to the challenge: accelerated pace of discovery transforms marine
628 virology. *Nat Rev Microbiol* 2015; **13**: 147–159.
629 75. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, et al. Genomic
630 differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics*
631 2016; **17**: 930.
632 76. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme
633 oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J* 2013; **7**: 592–
634 602.
635 77. Salisbury A, Tsourkas PK. A Method for Improving the Accuracy and Efficiency of Bacteriophage
636 Genome Annotation. *Int J Mol Sci* 2019; **20**.

637
638
639
640
641
642
643
644
645
646

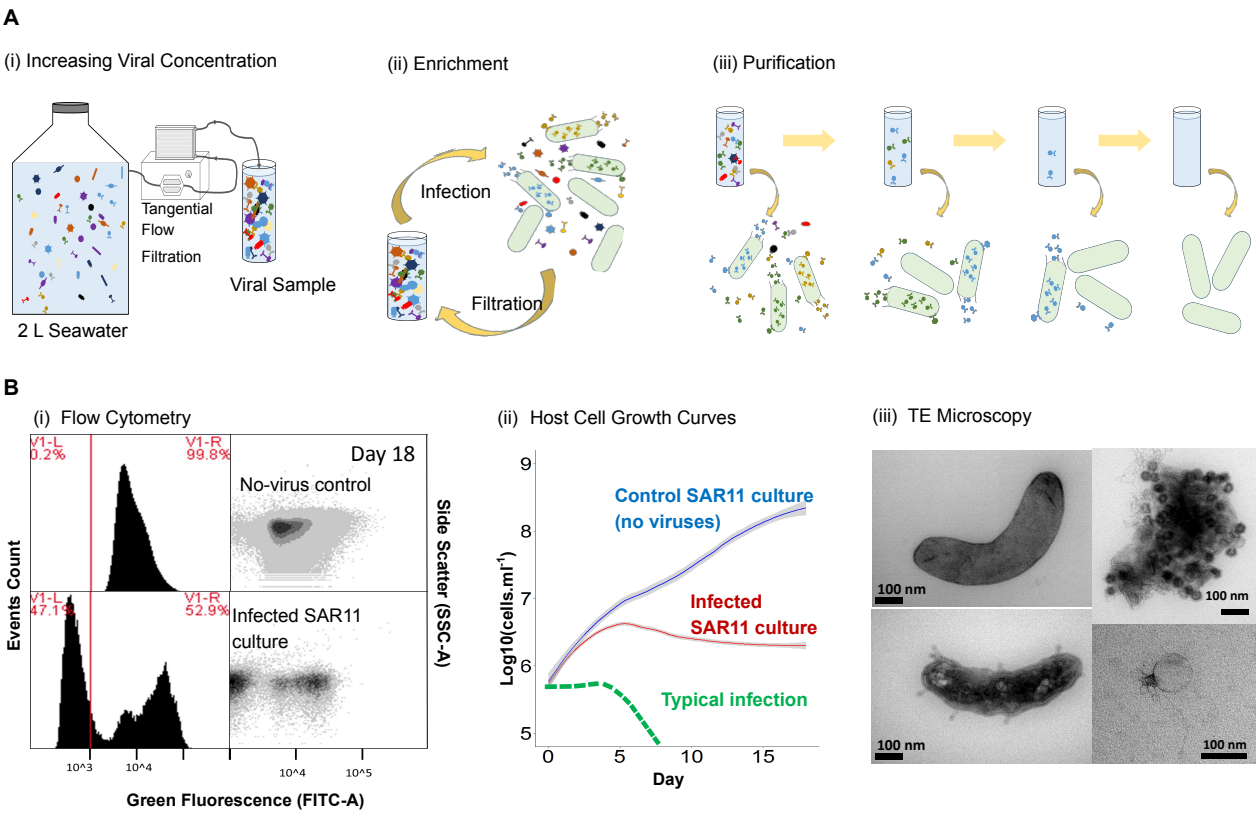


Fig. 1 Workflow for high-throughput isolation **(A)** i. Increasing concentration of viruses in water samples by TFF; ii. Initial infection of host cultures to enrich the sample for specific viruses; iii. Purification of viral isolates through 3 rounds of Dilution-to-Extinction; **(B)** Initial screen of viral infections using i. Flow cytometry, by comparing populations of no-virus controls and infected cultures; ii. Comparing growth curves of no-virus control culture (HTCC1062) against infected SAR11 cultures; iii. Confirming the presence of viruses in infected SAR11 cultures using TE microscopy: top left: HTCC1062 no-virus control, bottom left: infected HTCC1062, top right: aggregated cellular debris and viruses, bottom right: virus found in infected HTCC1062 culture.

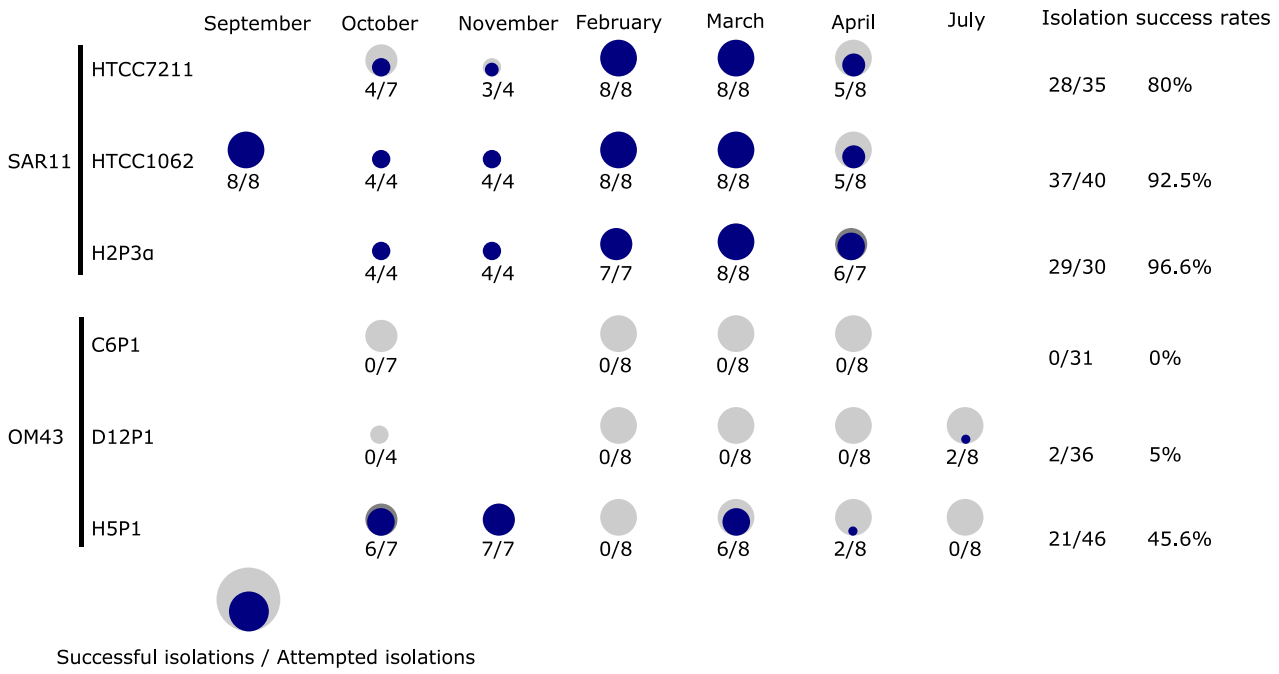
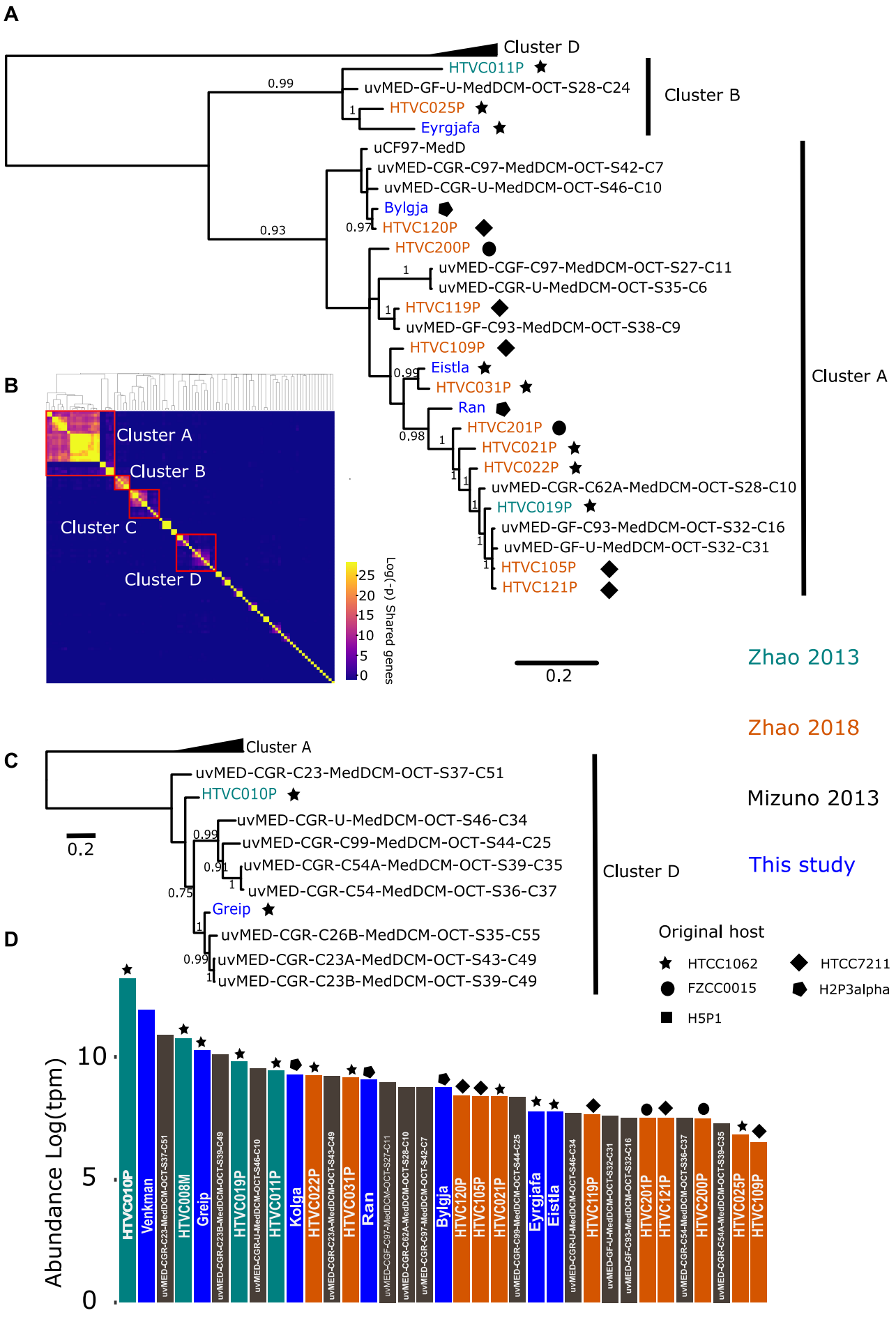


Fig. 2 Summary of success rates for each bacterial host (three SAR11 and three OM43) and water sample combination used for infections; the numerator indicates successful isolation of a virus from that water sample (i.e. successful passage through three rounds of Dilution-to-Extinction), denominator shows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


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

665

666 Fig. 3 Phylogeny and abundance of the first sequenced pelagiphages from the Western English
667 Channel **(A)** Hypergeometric testing on viral genomes by shared protein content identified four viral
668 clusters and numerous singletons; **(B)** Bayesian inference phylogenetic tree of pelagiphage major
669 capsid proteins from cluster A and B (most diverse pelagiphage cluster, includes most known
670 pelagiphages isolated previously [39, 46] as well as derived from metagenomes [37]), outgroup
671 Cluster D; **(C)** Bayesian inference phylogenetic tree of pelagiphage major capsid proteins from
672 Cluster D (containing the most abundant pelagiphages, including HTVC010P) with Cluster A as
673 outgroup; **(D)** Metagenomic recruitment (reads recruited per million reads (RPM)) of isolate and
674 metagenomically-derived viral genomes against a virome from the Western English Channel [12].
675

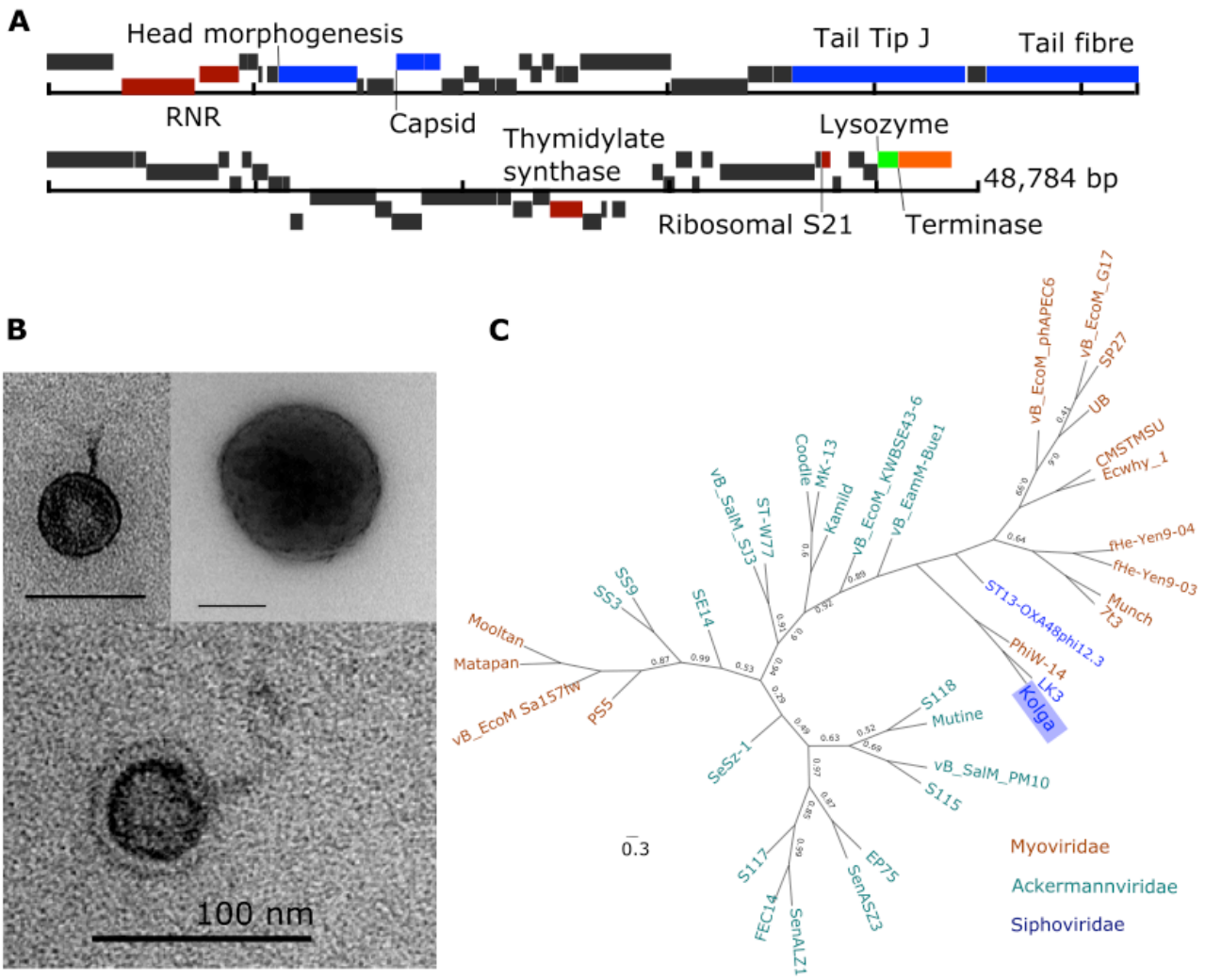


Fig. 4 The first reported siphovirus *Kolga* infecting SAR11 and isolated on novel host H2P3α ; **(A)** gene map of the 48,784 bp genome, which contains 80% hypothetical genes without known function; **(B)** TEMs of *Kolga* (top left and bottom) and a H2P3α host cell infected with *Kolga* (top right); **(C)** Bayesian inference phylogenetic tree of concatenated genes (Pyrophosphohydrolase, thymidylate synthase, two ribonucleotide reductase subunits), all branch support values of 1 have been omitted for clarity

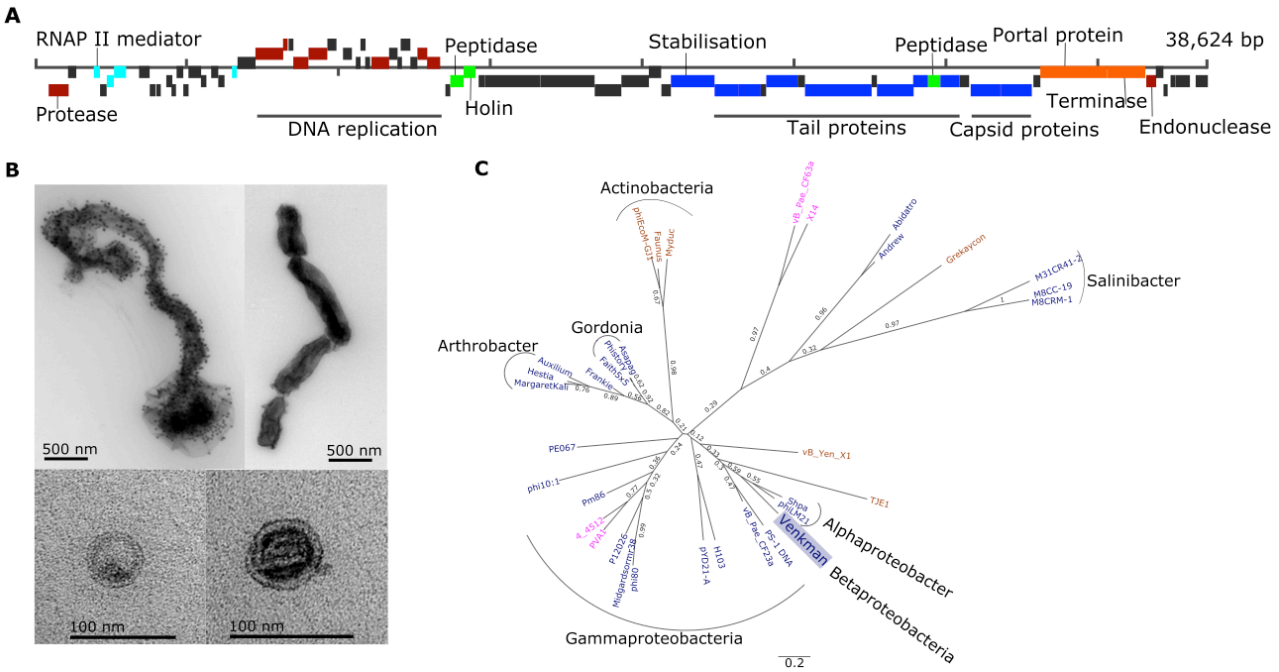


Fig. 5 Evidence for the first reported virus EXEVC282S *Venkman* infecting a member of the OM43 clade (H5P1); **(A)** Gene map displaying protein coding genes. Structural genes are shown in blue, DNA replication genes in red, lysis related genes in green, transcription genes in turquoise, packaging genes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

694

695 Table 1. Host infectivity of viral populations isolated and sequenced in this study. H2P3 and

696 HTCC7211 are warm-water ecotypes of *Pelagibacter spp*. subclade 1a; HTCC1062 is a cold-water

697 ecotype of *Pelagibacter spp.* Phage *Bylgja* is the only virus known to infect both ecotypes.

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

698

699

Table 2. Summary of phages isolated and sequenced in this study.

Virus	Host	Viral Population Representative	Hypergeometric cluster	Genome	Length (bp)	G + C %	Taxon	Accession number	Simplified phonetic spelling	Meaning and origin of the names
Eistla	HTCC1062	Eistla	Cluster A	Circular	39638	32.7	Podoviridae	MT375521	ais:tla:	"stormy one", Giantess in the poetic Edda
Eyrgjafa	HTCC1062	Eyrgjafa	Cluster B	Circular	38005	32.6	Podoviridae	MT375523	e:ɪrgja:fa:	"scar donor", Giantess in the poetic Edda
Gjalp	HTCC1062	Eyrgjafa	Cluster B	Circular	37857	32.5	Podoviridae	MT375524	gja:lp	"Roaring one", Giantess in the poetic Edda
Greip	HTCC1062	Greip	Cluster D	Circular	34916	31.5	Podoviridae	MT375525	graip	"Grasp", Giantess in the poetic Edda
Ran	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375530	ran	"plundering", Goddess of the Sea
Aegir	H2P3α	Kolga	Singleton	Fragment	18297	31.1	Siphoviridae	MT375519	æ:gir	"Sea", God of the Sea
Jörmungand	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375528	jɔrmunˌgand	"huge monster", Giant sea serpent
Kólga	H2P3α	Kolga	Singleton	Circular	48659	30.5	Siphoviridae	MT375529	kɔlga:	"Cool-wave", Daughter of Ran and Aegir
Unn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375531	ʊ:n	"Wave", Daughter of Ran and Aegir
Hroenn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375527	ɦrøn	"Wave", Daughter of Ran and Aegir
Bylgja	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375520	bɪ:lgja:	"Billow"-wave, Daughter of Ran and Aegir
Himinglæva	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375526	hi:mɪŋɡlæfa	"transparent-on-top"-wave, Daughter of Ran and Aegir
Venkman	H5P1	Venkman	Singleton	Linear	38624	34.4	Siphoviridae	MT375522	vɛŋkmæn	Bill Murray's Character in the Ghostbusters movie