



HAL
open science

“Quasigenus” among Phycodnaviridae: A diversity of chlorophyte-infecting viruses in response to a dense algal culture in a high-rate algal pond

Emily E Chase, Thomas M Pitot, Sonia Monteil-Bouchard, Christelle Desnues, Guillaume Blanc

► To cite this version:

Emily E Chase, Thomas M Pitot, Sonia Monteil-Bouchard, Christelle Desnues, Guillaume Blanc. “Quasigenus” among Phycodnaviridae: A diversity of chlorophyte-infecting viruses in response to a dense algal culture in a high-rate algal pond. 2023. hal-04281548

HAL Id: hal-04281548

<https://hal.science/hal-04281548>

Preprint submitted on 13 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **“Quasigenus” among *Phycodnaviridae*: A diversity of chlorophyte-infecting viruses in response to a**
2 **dense algal culture in a high-rate algal pond.**

3

4 Emily E. Chase,^{1,*} Thomas M. Pitot,² Sonia Monteil-Bouchard,³ Christelle Desnues,^{3,4} and Guillaume Blanc^{3,*}

5

6 ¹ *Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996, USA*

7 ² *Department of Biochemistry, Microbiology and Bioinformatics, Laval University, Quebec, Canada*

8 ³ *Microbiologie Environnementale Biotechnologie, Institut Méditerranéen d'Océanologie, Campus de Luminy,*
9 *163 Avenue de Luminy, 13009 Marseille, France*

10 ⁴ *Institut hospitalo-universitaire (IHU) Méditerranée infection, 19-21 Boulevard Jean Moulin, 13005 Marseille,*
11 *France*

12 *Corresponding authors

13

14 **Abstract**

15 This study approaches a high rate algal pond (HRAP) culture by metagenomic sequencing of the viral DNA
16 fraction, this includes the so-called giant virus fraction (phylum *Nucleocytoviricota*), with the goal of revealing
17 viruses coexisting within an intensified algal culture. A wealth of interesting novel viruses is revealed, including
18 members of *Nucleocytoviricota*, *Lavidaviridae*, and polinton-like viruses, which are taxa containing previously
19 characterized algal viruses. Our sequencing results are coupled with a virus targeted qPCR study and 18S rDNA
20 metabarcoding to elucidate potential virus-host interactions. Several species of green algae are identified
21 (Chlorophyta), likely representing the alternating dominant populations during the year of study. Finally, we
22 observe a bloom of viral diversity within the family *Phycodnaviridae* (*Nucleocytoviricota*), including highly
23 related but non-identical genotypes, appearing in the HRAP in September and October 2018. This bloom is most
24 likely the cause of a mass mortality event of the cultured algae that occurred during these same months. We
25 hypothesize that these related *Phycodnaviridae* lineages selectively infect different strains of the same algal
26 species of the Genus *Picochlorum* that have been identified in the HRAP by metabarcoding and coined this
27 phenomenon a "quasigenus" by analogy to the RNA virus quasispecies concept.

28

29 **Abbreviations:** High-rate algal pond (HRAP), Polinton-like viruses; PLVs, fluorescence activated cell sorting;
30 FACS, ultraviromes; UV, megaviromes; MV, open reading frame (ORF), major capsid protein; MCP, DNA
31 polymerase B; PolB, principal component analyses; PCA, similarity profile analysis; SIMPROF, auxiliary
32 metabolic genes; AMG, restriction modification; RM

33

34 **Keywords:** *Nucleocytoviricota*, polinton-like viruses, virophage, metagenomics, microalgae

35

36

37 Introduction

38

39 The importance of viruses in aquatic systems is well documented from nutrient cycling and algal bloom control
40 in the open ocean (Suttle and Wilhelm, 1999; Biggs et al., 2021), to modulating microbial communities in the
41 deepest ocean (Jian et al., 2021), and throughout the open ocean water column (Luo et al., 2020). In relation to
42 microalgae, specific viral groups of interest include members of Phylum *Nucleocytoviricota* or “giant viruses”,
43 virophages (*Lavidaviridae*), and polinton-like viruses. Family *Phycodnaviridae* and family *Mimiviridae* are two
44 taxa within *Nucleocytoviricota* known to infect microalgae (Wilson et al., 2009; Claverie and Abergel, 2018).
45 Family *Lavidaviridae* are a group of viruses infecting *Nucleocytoviricota* in a tripartite/co-infection system
46 between a giant virus and its host (Fischer, 2021). Previous work on family *Lavidaviridae* has given evidence
47 for algal *Nucleocytoviricota* hosts of these tripartite systems (Yau et al., 2011), and recently characterised this
48 interaction *in situ* with a *Chlorella* host (Sheng et al., 2022). Despite evidence of a virophage benefiting the
49 cellular host indirectly by reducing *Nucleocytoviricota* production in co-infection (La Scola et al., 2008), and
50 regulating host-virus dynamics (Yau et al., 2011), much is still unknown about family *Lavidaviridae*. Polintons
51 (or Mavericks), are relatively large (up to 40 kb) DNA transposons and have been found in a substantial diversity
52 of multicellular and unicellular species, including protists, fungi, beetles, fish, chicken, etc. (Kapitonov and
53 Jurka, 2006). In 2015, a 31 kb DNA virus (TsV-N1) similar in structure to polintons and with which it shares a
54 number of phylogenetically related genes, was identified as viral particles infecting the microalgae *Tetraselmis*
55 *striata* (Pagarete et al., 2015). This confirmed that some polinton-like viruses (PLVs) behaved as genuine viruses
56 (*i.e.*, having infectious virions) and not just transposable elements (Koonin and Krupovic, 2017). Notably, the
57 PLVs have some differences from polintons, including the absence of specific genes (see (Yutin et al., 2015)).
58 Recent metagenomic work concluded that there are at least eight major groups of PLVs and also uncovered
59 many associated with microalgae (Bellas and Sommaruga, 2021). Interestingly, members of family
60 *Lavidaviridae* can possess similarities to PLVs (Koonin and Krupovic, 2017) and PLVs also have potential as a
61 co-infection with extended *Mimiviridae* species (Yutin et al., 2015), therefore could contribute to viral-host
62 dynamics in a tripartite system as well.

63

64 Given the significance of viruses in major aquatic systems our study focuses on a specific system that is relatively
65 unexplored until now; the viral community of an industrial high-rate algal pond (HRAP). The said HRAP hosts
66 a non-specific polyculture of microalgae sourced from seawater of the Mediterranean Sea. Our objectives were
67 to understand the diversity of microalgal viruses in the system alongside their dynamics during a culturing
68 period. More specifically, the HRAP environment permitted a dense microalgal culture not unlike an algal bloom
69 of a natural aquatic environment, consequently providing an opportunity to investigate a bloom-like situation
70 and the viral community during this period. Each of the aforementioned virus groups (*i.e.*, *Nucleocytoviricota*;
71 family *Phycodnaviridae* and family *Mimiviridae*, family *Lavidaviridae*, and PLVs) were investigated within the

72 microalgae culturing HRAP of this study to gain an understanding of their role in a dense culture that we suggest
73 is analogous to a natural microalgal bloom.

74

75 **Materials and methods**

76

77 *Sampling, filtration, nucleic acid extraction, and next-generation sequencing of HRAP sampled water*

78

79 Water sampling and HRAP characteristics are as described in a previous publication (Chase et al., 2021).
80 Samples were processed (e.g., centrifuged, filtered, etc.) in the same manner to produce an “ultravirome” (UV)
81 of “small” DNA viruses (i.e., <0.2µm). As an attempt to isolate and subsequently sequence *Nucleocyotviricota*,
82 fluorescence activated cell sorting (FACS) was carried out in the size fraction 0.2 µm – 1.2 µm labelled using
83 SYBR-Green and analyzed according to the side scatter and FITC fluorescence parameters. The size fraction
84 was obtained by filtering samples through a 1.2 µm pore filter (Sartorius ref :17593), then concentrated 25× on
85 a tangential flow filtration column with a nominal pore size of 0.2 µm (Microkros ref: C02-P20U-05-N). Four
86 gated populations, “megaviromes” (MV), corresponding to areas of the SSC-FITC plot (where giant viruses are
87 generally observed (Khalil et al., 2016)) were sorted in each sample (**Figure S1**). Note that some bacterial
88 populations may emerge in the same area (i.e., gated population) and thus are considered “contaminants” in the
89 MVs. The number of isolated particles (events) per population ranged between 89 000 and 500 000. Lysis was
90 carried out as detailed in (Chase et al., 2021), and nucleic acid was extracted using EZ1 Advanced XL extraction
91 with a Virus Card (QIAGEN). Six of the 20 FACS populations did not generate sufficient DNA and were not
92 subsequently sequenced. Library preparation and paired-end 2 × 250 bp Illumina MiSeq sequencing of the UVs
93 were done as described in a previous publication (Chase et al., 2021). FACS population were amplified using
94 Ready-To-Go Genomiphi V3 kit (GE Healthcare), followed by purification and Illumina HiSeq sequencing
95 (paired-end 2 × 151 bp) performed by DOE Joint Genome Institute (JGI).

96

97 *Quality control and contig assembly of metagenomes and putative taxonomic assignments*

98

99 Quality control steps with raw reads were as described in previous work (Chase et al., 2021). Assemblies of
100 UVs and MVs were done with SPAdes version 3.15.0 using the metaviral algorithm and the single cell algorithm
101 respectively (Antipov et al., 2020). Only contigs over 2000 bp were kept for further analysis. Assembly statistics
102 were done using metaQUAST version 5.1 (Mikheenko et al., 2016) and are reported in **Table S1**. Taxonomic
103 assignments were given by an “open reading frames (ORFs) voting” procedure. To fulfill this, ORFs greater
104 than 100 codons were extracted from each contig in the UVs and MVs, aligned against TrEMBL (Boeckmann
105 et al., 2003) using MMSEQ (Steinegger and Söding, 2017) with e-value < 10⁻⁰⁵ and the taxonomic information
106 of the ORF best hits were recorded. Contig taxonomic classification was done using a majority rule criterion

107 over the ORF best hits. On average, 92.7% of the contigs were classified within the UVs and 97.1% within the
108 MV contigs. Overall, 3335 contigs were specifically assigned to viruses (prokaryotic or eukaryotic) from which
109 36,525 ORFs were extracted. Further ORFs functional annotation was performed by protein alignment against
110 the SWISS-PROT (Boutet et al., 2007) database using MMSEQS. We used the eggno-mapper (Cantalapiedra
111 et al., 2021) for assignment to COG (Tatusov et al., 2000) and KEGG (Kanehisa et al., 2016) functional
112 categories.

113 *Identifying potential microalgae hosts using 18S rDNA metabarcoding of HRAP water*

114

115 To characterize the diversity of eukaryotic microorganisms by metabarcoding, DNA was extracted from 0.2 μ m
116 filtered water and the V4 variable region of the 18S ribosomal RNA gene was amplified using eukaryote-specific
117 universal primers (forward CCAGCASCYGCGGTAATTCC and reverse ACTTTCGTTCTTGATYRA;
118 (Guillou et al., 2013)). The obtained amplicons were then subcontracted by the Genotoul facility (Toulouse;
119 get.genotoul.fr) which performed indexing and Illumina MiSeq sequencing. The sense and antisense sequences
120 obtained were assembled and then clustered as ASVs (amplicon sequence variants) and identified at different
121 taxonomic levels using the PR2 database version 4.12.0 (Guillou et al., 2013). These analyses were performed
122 under R software version 3.6.2 using the *DADA2* package (Callahan et al., 2016).

123

124 *Phylogenetic construction of Phylum Nucleocytoviricota, PLVs, and family Lavidaviridae relationships*

125

126 A major capsid protein (MCP), DNA polymerase B (PolB), and ATPase tree were produced using publicly
127 available data for *Nucleocytoviricota* using an alignment from MAFFT v.7 (Katoh and Standley, 2013), and
128 FastTree (Price et al., 2009) with default settings. Additionally, a PLV tree (MCP based using publicly available
129 data) was composed also using FastTree with default settings. MCP sequences were downloaded from National
130 Centre for Biotechnology Information (NCBI) for family *Lavidaviridae*. Corresponding sequences were also
131 aligned using MAFFT v.7 and the *Lavidaviridae* tree was produced using FastTree with default settings. All
132 trees were run with 1000 bootstrap replicates. In all cases putative sequences from relevant groups recovered
133 from the HRAP were included in the trees.

134

135 *Dynamics of putative viral targets by qPCR*

136

137 The selection of putative viruses for targeting, processing of raw water samples for qPCR, and the qPCR reaction
138 setup were conducted in the same way as described in previous work (Chase et al., 2021) without the need for
139 cDNA construction. Briefly, Primer3 (Untergasser et al., 2012) and PerlPrimer v.1.2.3 (Marshall, 2004) were
140 used to produce efficient primers *in silico* before *in vitro* testing. The number of reaction cycles (total of 45)
141 minus the Cq value were reported as the “inverted Cq”.

142

143 *Using correlation among viruses and potential hosts to uncover possible host-virus interactions*

144

145 Principle component analyses (PCA) were performed and visualised on a combined dataset of standardized
146 metabarcoding (18S) and qPCR results using R package *factoextra*, using time points where both metabarcoding
147 and qPCR results were available. Similarity profile analysis (SIMPROF) hierarchical clustering at $\alpha=0.1$ was
148 conducted using R package *clustsig*, and correlation networks were visualised using Cytoscape v3.9.1 (Smoot et
149 al., 2011). Publicly available sequenced alga of potential hosts (*e.g.*, *Picochlorum spp.*) were checked for recent
150 *Nucleocytoviricota* viral insertions using ViralRecall (Aylward and Moniruzzaman, 2021).

151

152 **Results and Discussion**

153

154 *Metagenomic sequencing data and biodiversity of putative DNA viruses in the HRAP*

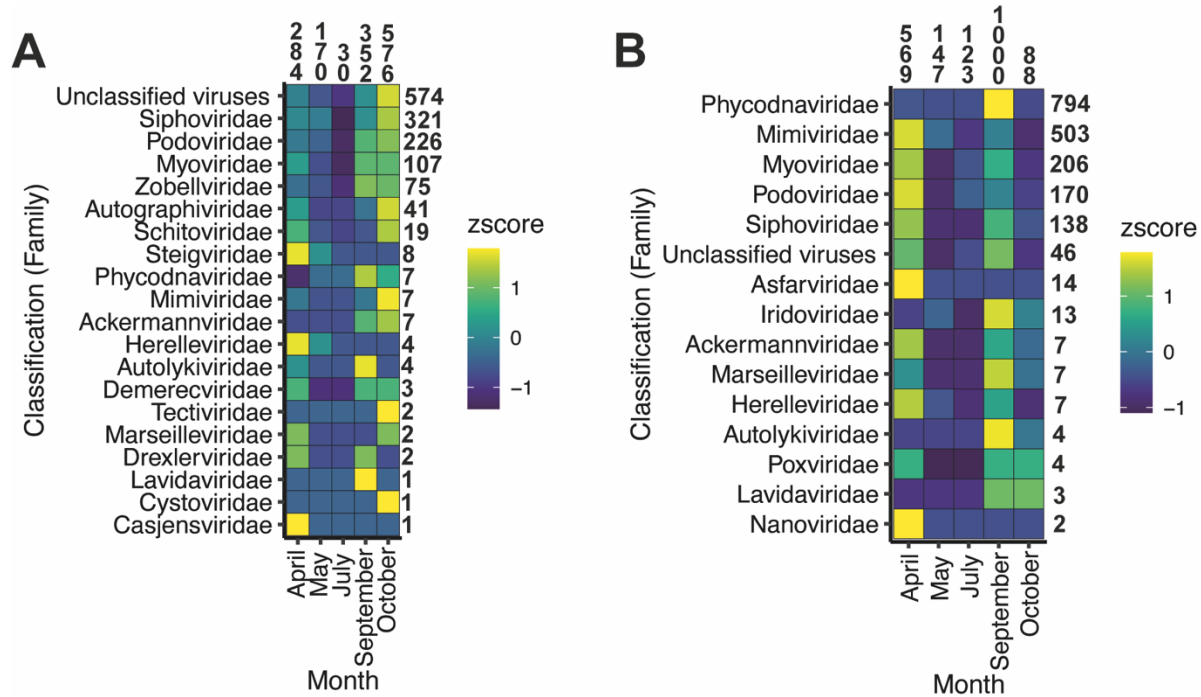
155

156 In total five ultraviromes (UV) from different months of 2018 (*i.e.*, April [04.17], May [05.17], July [07.05],
157 September [09.11], and October [10.23]) were sequenced, and 14 megavirome (MV) populations based on FACS
158 gated populations (see **Table S1**). Following genome assembly, the number of contigs greater than 2 kb ranged
159 from 74 to 1,968 for the UV samples, and 180 to 10,045 for the MVs, with the average number per sample being
160 1221 and 3995 for the UV and MV respectively. The GC content in the UV ranged from 47% to 54%, whereas
161 in the MV it ranged from 38% to 53%. Bacterial sequences were very abundant in all viromes (**Figure S2**).
162 Regardless of potential contamination issues that are common in viral metagenomics (Jurasz et al., 2021), this
163 result is not surprising for FAC-sorted MVs, as the size fraction of sorted particles includes that of small bacteria.
164 For UVs, this could be due to leaks in the filtration or inefficient digestion of free DNA (lysed bacteria) present
165 in the sample. Nevertheless, virus contigs still represented between 0.4% (05.17.P4) to 39% (07.05.UV) of the
166 cumulated contig length. Eukaryotic virus sequences were generally dominant over prokaryotic virus sequences
167 in MVs (the mean proportion was 2.5% for eukaryotic viruses versus 0.8% for prokaryotic viruses) while this
168 trend was opposite in the UVs (mean frequency was 18.1% versus 0.6%, respectively).

169

170 Using metagenomics, we employed two approaches to uncover DNA viruses of the HRAP, permitting the
171 observation of small DNA viruses within a UV, and large DNA viruses (*e.g.*, *Nucleocytoviricota*) through FACS
172 populations and further processing (MV). The UVs are primarily composed of phages (*e.g.*, *Siphoviridae*,
173 *Podoviridae*, and *Myoviridae*), and unclassified prokaryotic viruses (**Figure 1 A**). Polinton-like viruses were
174 also uncovered in the DNA UVs and MVs (appear in the “unclassified virus” category; see also phylogeny in
175 **Figure S3**) and are discussed more below. Of special importance to our study, is our effort to retrieve
176 *Nucleocytoviricota* by FACS methods, and we were able to recover a viral diversity community composition
177 different than the UV fraction (**Figure 1 B**). Recent work (Palermo et al., 2021) on metagenomic sample filtration
178 methods points out the almost definite loss of viral community taxonomic information in metagenomic studies

179 where larger ($> 0.45 \mu\text{m}$ pore size) water fractions are removed and left un-sequenced. No doubt, sequencing
 180 this large viral fraction permitted higher recovery of *Nucleocytoviricota* compared to the UV method. A putative
 181 member of family *Lavidaviridae* (a virophage) was detected and does not group closely with a *Lavidaviridae*
 182 known to infect a *Chlorella* species (**Figure S4**). As stated previously, virophages are found in association with
 183 algae infecting *Nucleocytoviricota*, and were recovered in our MV samples (**Figure 1 B**). This result will be
 184 explored in more depth.
 185



186
 187 **Figure 1.** Proportion (by z-score) of viral taxonomic assignments to assembled (A) DNA ultravirome contigs
 188 by each metagenomic sample in 2018, and (B) proportion of viral taxonomic assignments to assembled
 189 metagenomes produced from FACS population, grouped together by sampling month. Please note (B) is only
 190 to showcase the diversity and is not a quantitative assessment as some sample months produced more FACS
 191 populations of interest, and therefore were attributed more sequencing. Additionally, these classifications are
 192 resulting from our “ORF voting” procedure. Total numbers of contigs attributed to each family are shown on
 193 the y-axis, and total number of contigs attributed to each month are shown on the x-axis.

194
 195 *Genetic potential of viral communities*

196
 197 Of 21,896 ORFs extracted from eukaryotic virus contigs, 5407 (25%) had a significant match in sequence
 198 databases, and 811 could be assigned to COG (Cluster of Orthologous Groups of proteins) functional categories
 199 (**Table S2**). Also, 198 ORFs were identified as encoding an MCP. For prokaryotic viruses contigs, 6586 (45%)
 200 of the 14,629 extracted ORFs had a match in sequence databases, of which 154 were against a capsid protein

201 and 728 were assigned to a COG functional category. Contigs of prokaryotic viruses were found to encode a
202 higher proportion of proteins involved in DNA replication and repair, and in metabolisms of membranes and
203 cell walls (**Figure S5**). One possible reason is that these viruses were mainly found in the UV fractions and
204 therefore, likely had smaller genomes carrying genetic information more narrowly focused on essential viral
205 functions, *i.e.*, replication and entry into the cell. Conversely, eukaryotic viruses, which were more highly
206 represented in the MV fractions, encoded a greater proportion of proteins involved in accessory metabolic
207 functions, including nucleotide metabolism, protein translation and post-translation modification.

208 Genes of viral contigs that had significant matches in cellular organisms but not in viruses were further examined.
209 These genes are probably derived from horizontal transfers from cellular hosts. They represent candidate
210 auxiliary metabolic genes (AMGs) that have not yet been observed in viruses or whose sequence divergence is
211 such that the homology relationship with another virus is no longer detectable by protein alignment.

212 Eukaryotic virus contigs carried 99 of these potential AMGs, 62 of which had an unknown function. Among the
213 genes to which a function could be attributed, several were located on contigs generated from the 07.05.P3
214 population and encoded remarkable functions including a neutral sphingomyelinase which is a key enzyme in
215 sphingolipid metabolism and is implicated in stress-induced signaling pathways (Choezom and Gross, 2022).
216 Another gene encoding a myosin motor domain-containing protein suggests that the corresponding virus might
217 use actin-based movements in the infection process. Contigs from different populations of September (09.11)
218 also encoded a variety of interesting proteins never yet observed in eukaryotic viruses. They include homologues
219 of heme oxygenase and Phycocyanobilin:ferredoxin oxidoreductase which are both involved in synthesis of
220 bilins which play multiple role in chlorophyll metabolism and signal transduction (Zhang et al., 2018), a CP12-
221 like protein involved in regulation of the Calvin cycle responsible for CO₂ assimilation (Gontero and Maberly,
222 2012), a RNA polymerase subunit Rpb5, a phosphatidylserine decarboxylase involved in the synthesis of the
223 abundant phospholipid phosphatidylethanolamine of mitochondrial membranes (Nerlich et al., 2007), a
224 clavamate synthase-like proteins potentially involved in the biosynthesis of clavulanic acid and other 5S
225 clavams (Tahlan et al., 2004), which have antibacterial and antifungal activities, a non-histone chromosomal
226 MC1 family protein whose archeal homologue is used for DNA packaging (De Vuyst et al., 2005), and a
227 VIT1/CCC1 iron transporter family protein. We also found components of restriction modification (RM) system,
228 including several endonucleases and DNA methyltransferases yet unseen in viruses. The presence of RM
229 systems is unsurprising, given that they are found in members of *Nucleocytoviricota* (Filée, 2018), and more
230 specifically have been documented in *Phycodnaviridae* (chloroviruses) (Coy et al., 2020). As previously
231 summarized (Jeudy et al., 2020), nucleases within giant virus genomes could assist in the degradation of host
232 DNA to permit recycling into viral particles, or inhibit host gene expression (Agarkova et al., 2006). Prokaryotic
233 virus contigs encoded 242 proteins that match homologs in cellular organisms but not in viruses, of which most
234 (209) had an unknown function. Interesting functions were identified, including a ABC transporter efflux protein
235 encoded by a prokaryotic virus from the 09.11.P4 (MV) population, a N-acetylmuramoyl-L-alanine amidase
236 (09.11.UV) involved in cleavage of cell-wall glycopetides, a helix-turn-helix (HTH) transcription regulator

237 (05.17.UV), a non-viral-type sialidase (10.23.UV) involved in the degradation of extracellular mucin, a RNA
238 polymerase sigma factor (04.17.UV) and a trehalose utilization ThuA protein. Taken together, these results
239 suggest that HRAP viruses encode a specific subset of regulatory and/or enzymatic proteins that differs from the
240 proteomes of model viruses that serve as a template for comparative studies.

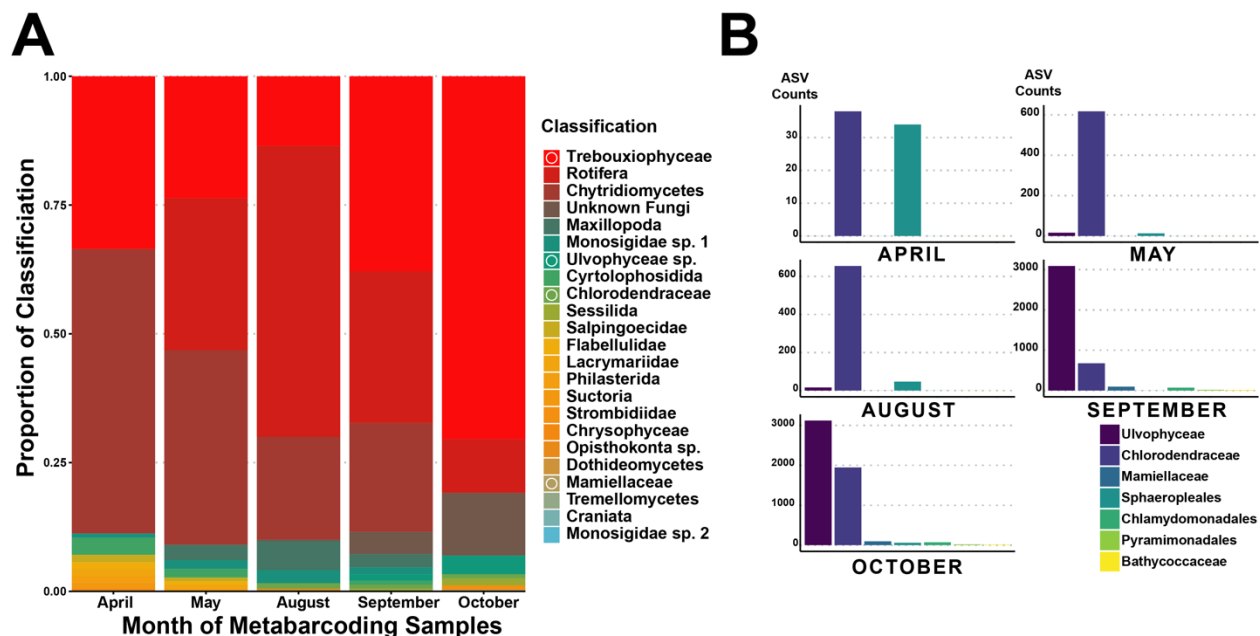
241

242 *Biodiversity of potential alga hosts in the HRAP*

243

244 When considering only the top ten classifications of V4-18S ASVs in each sample month (**Figure 2 A**) the most
245 abundant taxa are class *Trebouxiophyceae* (primarily family *Chlorellales*), and for the scope of this paper will
246 be the focus. It is worth noting, the lack of diatoms is most likely due to the HRAP culture lacking in biogenic
247 silica (*i.e.*, no silica was added). Of specific interest to our study, several groups of unicellular microalgae are
248 present (**Figure 2 A and B**). Including aforementioned family *Chlorellales*, alongside other members of phylum
249 Chlorophyta; *Chlorodendraceae*, *Chlamydomadales*, *Sphaeropleales*, *Pyramimonadales*, *Ulvophyceae*, and
250 *Mamiellaceae* in the HRAP during the 2018 culture samples (highlighted in **Figure 2 B**). Given previously
251 established associations with DNA viruses to Phylum Chlorophyta (Etten et al., 2020; Short et al., 2020), we
252 focused mostly on these microalgae and not additional unicellular plankton identified in the HRAP by
253 metabarcoding, as these species do not appear in substantial amounts (based on ASVs) in 2018 alongside our
254 metagenomic study or present evidence of relationships with the putative viruses tracked in our study.

255



256

257 **Figure 2.** Classification of HRAP eukaryotes by (A) proportion of amplicon sequence variants (ASVs)
258 assigned to taxonomy of 18S metabarcoding data for each month of 2018 that also has metagenomic sampling,
259 ASV from each month were summarised. Members of Phylum Chlorophyta are indicated on the legend with a

260 circle. Only the top ten ASV classifications for each month are represented in the schematic. Additionally, **(B)**
261 ASV counts of all Chlorophyta (excluding the abundant Class Trebouxiophyceae), this is not limited to the top
262 ten ASV classifications per month.

263

264 *Predicting potential microalgal hosts and their viruses.*

265

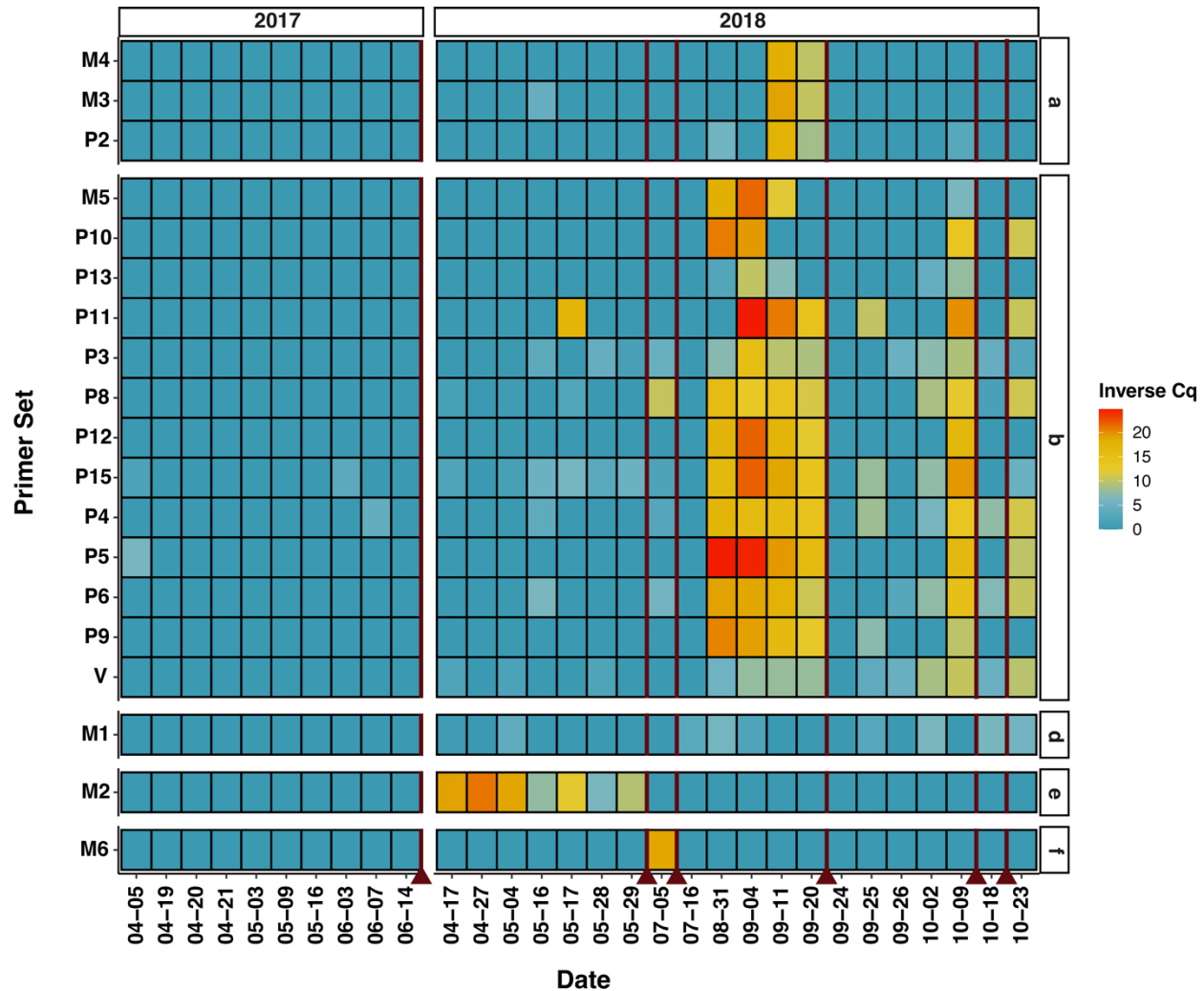
266 Using correlation of microalgal species and viral population presence we can try to resolve potential host and
267 virus relationships (Hingamp et al., 2013; Roux et al., 2017). As the approach is simply observing correlations
268 it cannot account for ecological nuances of host-virus relationships, namely it assumes that overlapping host and
269 virus presence and their relative abundances are an indicator of infection. With this understanding we interpreted
270 our results as a preliminary suggestion of possible host-virus relationships in the context of known viral ecology.
271 Correlation analyses were inconclusive for virus-host relationships outside of Chlorophyta alga, and
272 consequently we focused solely on this group.

273

274 Principal component analysis (PCA) with members of families *Mimiviridae*, *Phycodnaviridae*, and *Lavidaviridae*
275 (sequenced within MVs) showed a strong positive correlation (*i.e.*, higher covariance) among *Picochlorum* and
276 several members of these viruses (**Figure S6 A**), including a putative phycodnavirus that also had a strong
277 correlation with the detected member of family *Lavidaviridae* (*i.e.*, the putative virophage). A strong correlation
278 was shown between a member of family *Mimiviridae* (M1) and the alga group *Dictyosphaerium* (**Figure S6 A**).
279 These relationships are primarily associated with September and October (2018) sampling. Although much less
280 supported by these principal components, there was an additional association between a member of family
281 *Mimiviridae* (M2) and an unknown species from the order *Chlorellales*. The PCA for polinton-like viruses
282 (PLVs) does not show a close correlation for some of the putative viruses (**Figure S6 B**). Some PLVs are
283 associated with genera *Dictyosphaerium* and *Mantoniella*, and to some extent the genus *Picochlorum*. However,
284 these variables contribute much less to the ordination than the examples highlighted within the
285 *Nucleocytoviricota* and *Lavidaviridae* PCA (**Figure S5 A**). Hierarchical clustering indicates a significant
286 grouping among previously described MV viruses and hosts (**Figure S7**; b, d, and e). Additionally, PLVs and
287 their associated PCA suggested hosts also produce significant groupings, although with slightly weaker
288 correlations (although still significant) in most cases (**Figure S8**) in relation to *Nucleocytoviricota* and
289 *Lavidaviridae* groupings. Correlations among both the MVs and PLVs with genus *Picochlorum* reflect an
290 apparent bloom of *Picochlorum* in the area the HRAP is stationed during these months, however in comparison
291 to the PLVs the correlation among 11 *Phycodnaviridae*, one *Mimiviridae*, and the putative virophage are
292 specifically much stronger (**Figure S7 b**). In the context of viral ecology, it seems more likely that correlations
293 found among *Nucleocytoviricota* and *Lavidaviridae* could imply potential hosts more readily than PLV and alga
294 correlations. We attribute this to the likely rapid proliferation of the *Nucleocytoviricota* groups after infection
295 (*e.g.*, *Chlorella* virocells release PBCV-1 viral particles three hours post infection (Van Etten et al., 1983)). In

296 the case of PLVs, although relatively little is known, their dual lifestyle (Krupovic and Koonin, 2015; Yutin et
297 al., 2015) of insertion, and later proliferation may require specific cues for a virocell to develop and release viral
298 particles effectively uncoupling infection and release of viruses acutely, thereby confounding the interpretation
299 of a correlation analysis. Although *Lavidaviridae* could be uncoupled in the same way due to viral integration
300 capabilities, in this case a known trigger for viral particle production is present (*i.e.*, co-infection with
301 *Nucleocytoviricota*) and conclusively we expect all members of this tripartite infection to co-exist when
302 *Lavidaviridae* are detected in the HRAP. In this scenario, detection of the *Lavidaviridae* indicates viral particles
303 being produced and this would be in response to cellular hosts with a *Nucleocytoviricota* infection taking place
304 (*i.e.*, a virocell is formed). With ecology in mind, we expect to be able to produce some valid indication of both
305 *Nucleocytoviricota* and *Lavidaviridae* alga hosts through correlation analyses, however PLV hosts indication are
306 less straight forward. Additionally, some interactions could be occurring between PLVs and *Nucleocytoviricota*,
307 but these were not explored given the lack of ecological information on these putative PLVs.

308
309 Using our full qPCR dataset (**Figures 3 and 4**) and Chlorophyta metabarcoding results (**Figure 5**) we can further
310 hypothesize virus-host interactions by using our correlation and clustering work as an initial indication of
311 potential hosts. Overall viruses within hierarchically clustered groupings exhibited similar patterns of presence
312 and absence based on qPCR results in both *Nucleocytoviricota* and *Lavidaviridae* (**Figure 3**), and PLV analyses
313 (**Figure 4**). Notably, the *Nucleocytoviricota* and *Lavidaviridae* grouping (**Figure S6 b**) reflect *Picochlorum*
314 ASVs quite closely (**Figure 5 F**), strengthening our hypothesis that members from this grouping could be
315 infecting members of genus *Picochlorum*. Although overlap does occur between the single member of
316 *Mimiviridae* in group d (**Figure 3**) and correlated algae from genera *Dictyosphaerium*, and *Mantoniella* (**Figure**
317 **S7 c,d**), no obvious clarification is made by the full dataset and qPCR implied abundances are quite low. Finally,
318 the single *Mimiviridae* in group e (**Figure S7**) has further evidence for infecting the “unassigned” *Chlorella*
319 species (**Figure 5 F**), as previously indicated by our correlation analyses, given both of their nearly complete
320 disappearance after May 2018.



321

322

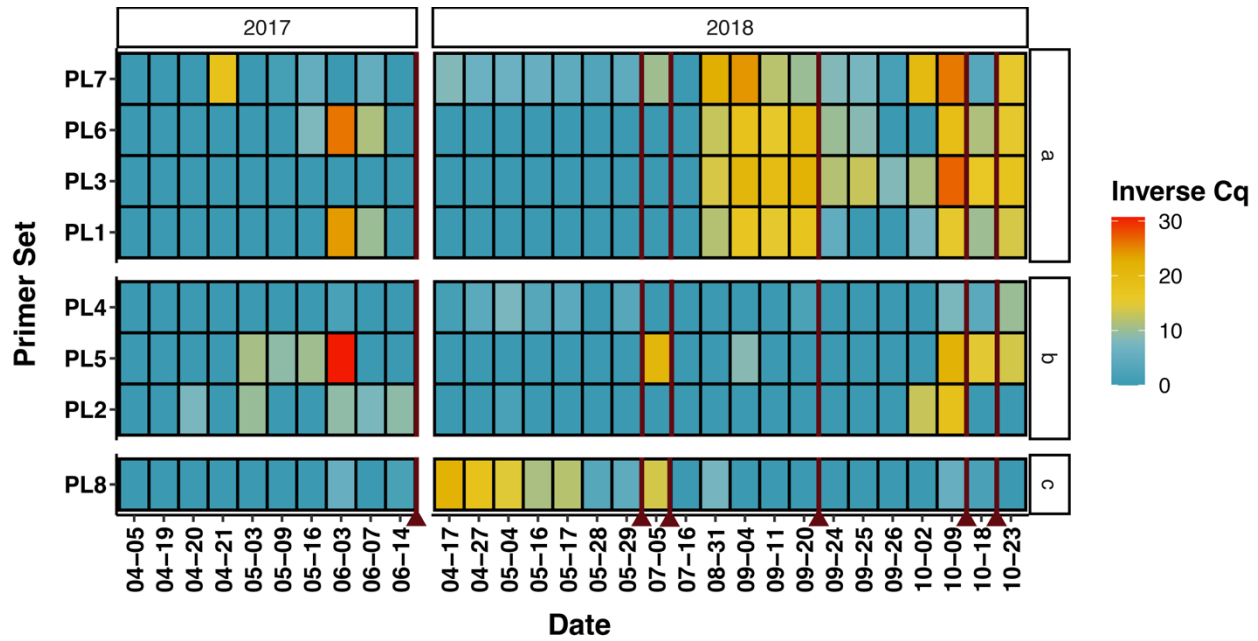
323 **Figure 3.** Putative Nucleocytoviricota and virophage tracking by qPCR throughout 2017 and 2018 samples.

324 Where V, P, and M indicate potential virophage (family *Lavidaviridae*), family *Phycodnaviridae*, and family

325 *Mimiviridae* respectively. An inverse Cq is calculated by number of cycles (total 45) minus the mean Cq value

326 across technical triplicate reactions for each instance (e.g., viral target and sample data). Redlines indicate HRAP

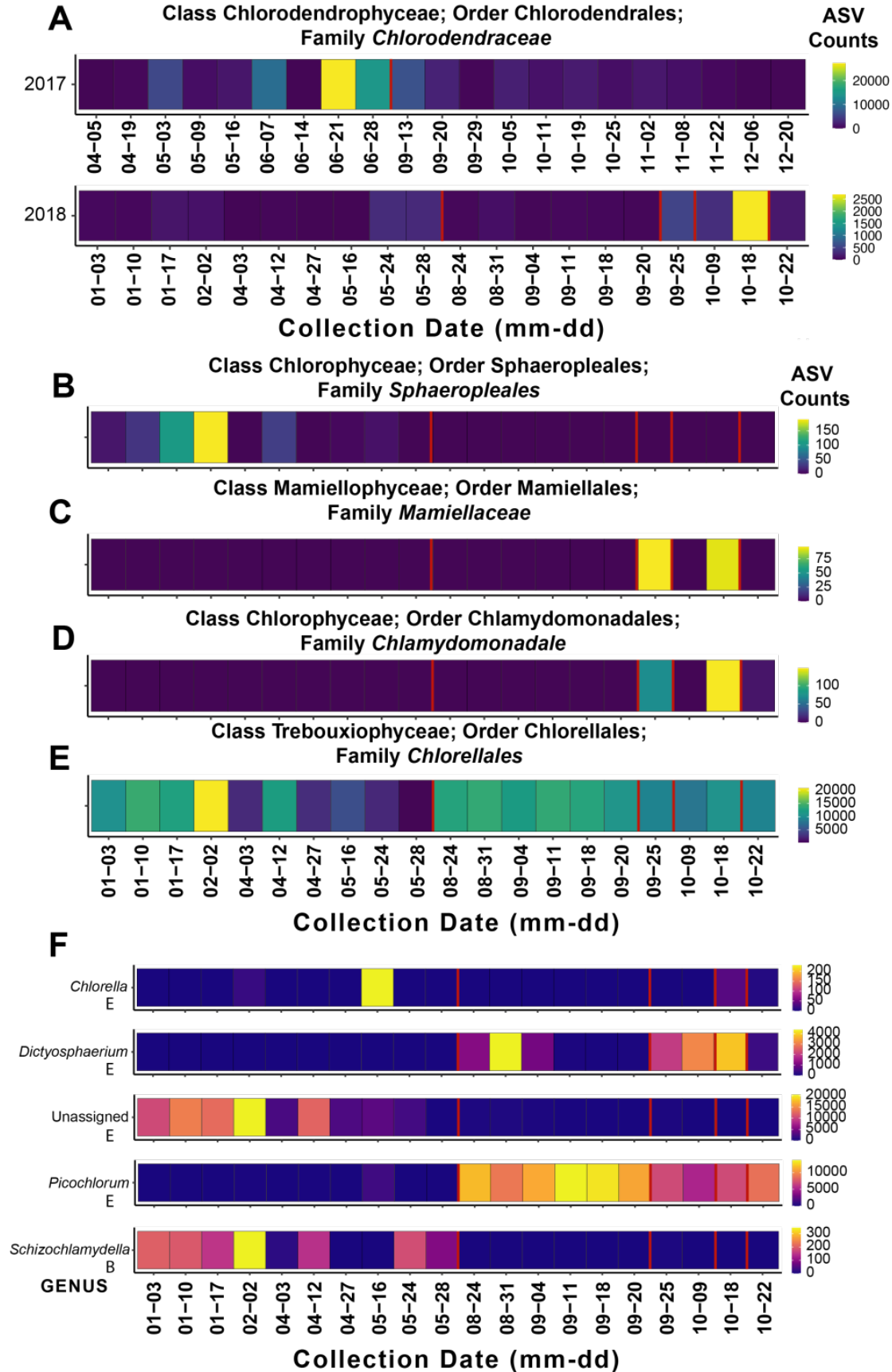
327 crash dates. Note, group c is not present given that no viruses were included during hierarchical clustering.



328

329 **Figure 4.** Putative polinton-like virus (PLV) tracking by qPCR throughout 2017 and 2018 samples. Where PLV
330 indicates potential polinton-like viruses. An inverse Cq is calculated by number of cycles (total 45) minus the
331 mean Cq value across technical triplicate reactions for each instance (*e.g.*, viral target and sample data). Redlines
332 indicate basin crash dates.

333



335 **Figure 5.** Presence and relative amounts (based on amplicon sequence variants (ASVs) of 18S data) of potential
336 microalgae hosts using 2018 data (with the exception of 2017 being included in **A**) collected by water samples
337 of the basin system. (**A-E**) The most abundant potential microalgal hosts at the Family level are shown, and (**F**)
338 Genus level for family *Chlorellales* and family *Sphaeropleales* (as indicated by E or B). ASV counts are
339 indicated by colour legends. Instances where a new culture run is made, after a die-off, are indicated by a red
340 line.

341

342 Interestingly, the PLVs most closely related to a known PLV infecting *Tetraselmis striata* (**Figure S3**; PLV 2,
343 5, and 7) do not correlate closely with *Tetraselmis* spp. in the HRAP based on our PCA analysis (**Figure S6 B**),
344 however two out of three significantly cluster together (**Figure S8 b**). We cannot confidently conclude that none
345 of the tracked PLVs infect *Tetraselmis* spp. not only because the theoretical decoupling of PLVs and hosts
346 discussed above, but also because an important peak of *Tetraselmis* spp. ASVs occurs during a period where
347 qPCR data is unavailable (mid-June to autumn 2017; **Table S3**). This peak would likely be an important part of
348 the data to assess the specific relationship (if any) between PLV 2, 5, and 7 and *Tetraselmis* spp. Phylogenetically
349 closely related PLVs could be infecting completely different organisms, therefore the relationship of PLVs 2, 5,
350 7, to TsV-N1 (and each other) does not completely confirm infection of the same species. Overall, correlating
351 PLVs and hosts by these methods are made difficult by their ecology (and our minimal understanding of PLVs
352 at present). No other classifications (by ASVs) outside of Chlorophyta correlate significantly (including
353 hierarchical clustering analysis) with these PLVs (data not included), despite the discovery of PLVs infecting
354 several groups outside of Chlorophyta (Bellas and Sommaruga, 2021). However, alternative hosts (*i.e.*, not
355 members of Chlorophyta) cannot be ruled out by these *in silico* analyses alone. An important point is that viruses
356 found in different taxonomic groups can also be infecting the same organism (Bellas and Sommaruga, 2021),
357 consequently it is difficult to infer a host of these tracked PLVs. Regardless of the true host, this HRAP appears
358 to be an interesting and significant reservoir of PLVs, some quite likely to be infecting microalgae.

359

360 Given the similarity of *Picochlorum* (**Figure 5 F**) and group c virus (**Figure 3**) dynamics (*i.e.*, presence and
361 abundance) with visual inspection and the results of our correlation analyses, the possibility of a previously
362 undocumented infection of *Nucleocytovirivota* and a tripartite infection with a *Lavidaviridae* to a *Picochlorum*
363 required further investigation. Analysis of eukaryotic genomes has revealed that endogenous *Nucleocytovirivota*
364 are common in some lineages, particularly in green algae, and thus play an important role in host genome
365 evolution (Blanc et al., 2015; Gallot-Lavallée and Blanc, 2017; Moniruzzaman et al., 2020). These viral
366 insertions provide evidence of past interaction between the original virus and the organism that captured its
367 DNA. Available *Picochlorum* genome assemblies were downloaded from NCBI GenBank to check for
368 *Nucleocytovirivota* integrated *Nucleocytovirivota* genes/viral elements (GenBank accessions GCA011316045,
369 GCA010909725, GCA00281821, GCA009650465, GCA_000876415). Several regions with *Nucleocytovirivota*
370 signatures were recovered among these genomes (median size of ~15 kb; **Table S4**) using the dedicated

371 ViralRecall program (Aylward and Moniruzzaman, 2021), interestingly a *Picochlorum* genome recovered from
372 the Mediterranean Sea (Costa Vermeille) (Krasovec et al., 2018) possessed a relatively large region (~94 kb)
373 with both superfamily II helicase and D5 primase-helicase markers detected. These results suggest that
374 *Nucleocytoviricota* do or have been infected with *Picochlorum* in the past (this particular sample was taken in
375 June 2011), and specifically in the geographical area where the HRAP was situated.

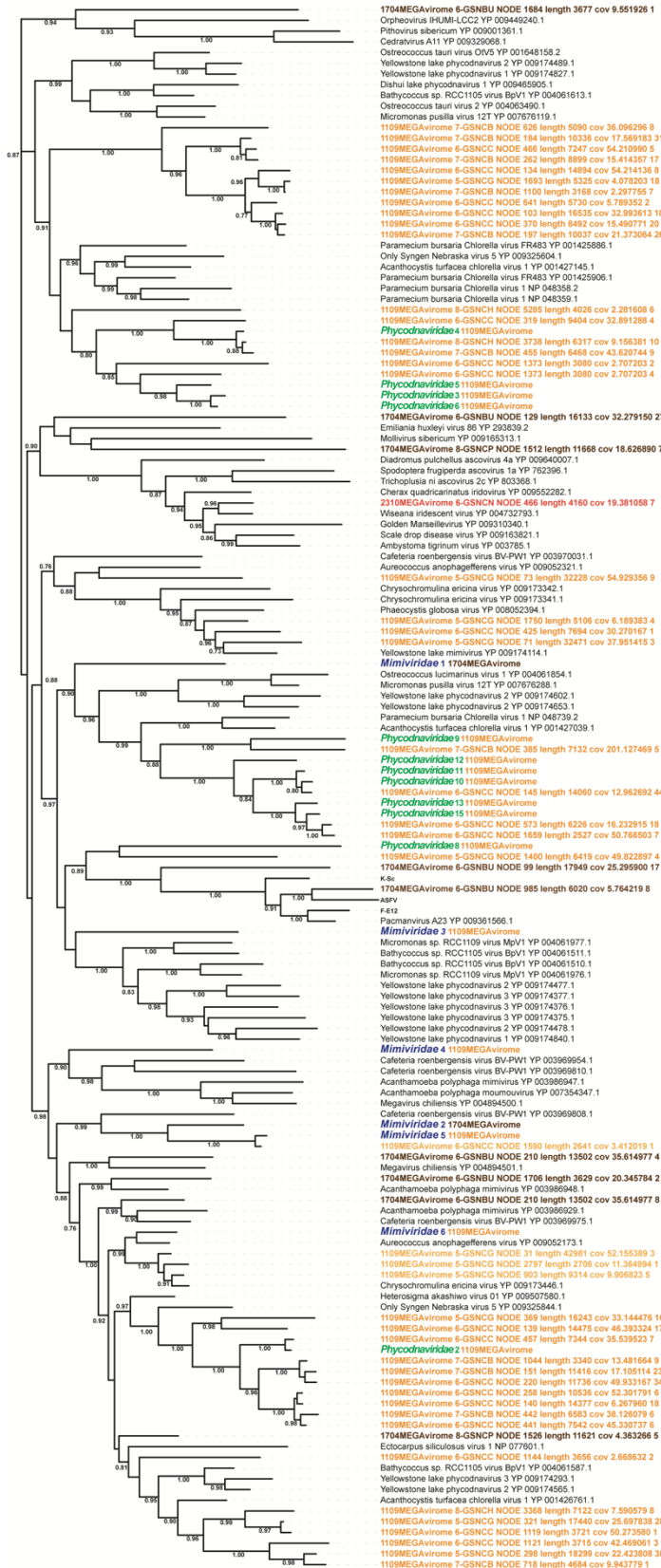
376

377 *Diversity of Phycodnaviridae may reflect diversity of the host*

378

379 Of substantial interest in this study is the scale of *Phycodnaviridae* (phycodnaviruses) uncovered in our MVs
380 (**Figure 6**). *Mimiviridae* representatives also appear, although in less abundance. It is known that
381 *Phycodnaviridae* genomes of genera *Prasinovirus* and *Chlorovirus* contain multiple copies of the MCP gene
382 (*i.e.*, paralogs) (Clerissi et al., 2014), therefore the sheer number of MCP hits in this phylogeny may be at least
383 partially explained by intragenome duplication. Nonetheless, as a core gene of *Nucleocytoviricota* (Yutin et al.,
384 2009), putative MCPs are a vital finding in the HRAP because they signal true viral hits. With respect to DNA
385 polymerase (PolB), a relatively robust single-copy gene for phylogenetic reconstruction of *Nucleocytoviricota*
386 (Chen and Suttle, 1996; Clerissi et al., 2014), there are still a considerable number of putative *Phycodnaviridae*
387 (19 total; **Figure S9**). Finally, a considerable number of ATPase sequences (another marker of
388 *Nucleocytoviricota*) with *Phycodnaviridae* hits are also recovered from the MVs (18 total; **Figure S10**). Our
389 results appear to suggest that in an environment such as an HRAP, where a dense culture of microalgae
390 simulates an algal bloom, a viral “bloom” reflecting the algal abundance can occur. In a similar fashion to the
391 RNA quasispecies (Andino and Domingo, 2015), the HRAP *Phycodnaviridae* form a repertoire of closely
392 related genomes. RNA quasispecies are said to involve a high copy number of genome variants arising from
393 the high mutation rates of RNA viruses (Duffy, 2018). This phenomenon produces so-called “mutant swarms”
394 that foster unique dynamics between viruses where variants are acting “within replicate complexes, within
395 cells, or outside cells” (Domingo et al., 2012; Domingo and Perales, 2019) in competition amongst themselves.
396 The results of this, to the host(s)’ detriment and the virus population’s benefit, is the production of a “genome
397 repertoire” (Domingo et al., 2020) in advance of a host’s response, or other factors that could affect the host-
398 virus dynamics of a system in the future. However, instead of multiple viral, quasi-identical strain-like species
399 of the RNA quasispecies, HRAP phycodnaviruses exhibited substantial genomic sequence divergence given
400 dotplot alignments of contigs containing a single-copy DNAP gene (**Figure S11**). RNA viruses are capable of
401 evolving as quasispecies because their error-prone (*i.e.*, low fidelity) RNA polymerase, resulting in a
402 mutational rate up to 1 million times that of their hosts (Duffy, 2018). *Nucleocytoviricota* do not feature the
403 same high rate of mutation, largely attributed to high fidelity PolB repair machinery (Fischer et al., 2014;
404 Redrejo-Rodríguez and Salas, 2014). Thus, it seems likely that the phycodnaviral lineages had begun to
405 diverge well before they entered in the culture basin and on a larger time scale than between members of a
406 typical RNA quasispecies. Assuming that the HRAP phycodnaviruses infect the same species they would have

407 to compete for access to the host. Under competitive conditions, the maintenance of a large diversity of viruses
408 is not the most likely scenario, since ecological theory rather predicts the hegemony of the most fit virus. A
409 possibly more likely scenario would be that the different viral lineages have evolved different strain specificity
410 within the same host species. Under the very favourable condition of a HRAP culture, multiple algal strains
411 could co-exist durably and support replication of a diversity of related viruses. Alternatively, this diversity
412 could reflect the diversity of closely related species present within one alga genus (*i.e.*, *Picochlorum*).
413



- Putative Phycodnaviridae
- Putative Mimiviridae
- April Sample
- September Sample
- October Sample

415 **Figure 6.** Phylogenetic relationship of *Nucleocyotoviricota* found within the study (denoted as “unclassified”),
416 and sequences obtained from NCBI GenBank (accession numbers as indicated) based on major capsid protein
417 (MCP) amino acid sequence. Groups are assigned based on clades of closely related (by sequence similarity)
418 novel *Nucleocyotoviricota* for further study. Bootstrap support is indicated and based on 1000 replicates.

419

420 Within RNA quasispecies, diverse genotypes are permissible in order to handle the diversity of strains occurring
421 within their host target species, such as the bloom species *Heterosigma akashiwo*, which has a wide distribution
422 (Lawrence and Suttle, 2004). An array of virus strains is kept as a repertoire of infectious agents to handle the
423 dynamics of their host strains. It may not be the case that new algal strains are developing, but rather a dance
424 between dominant strains, and strains that are simply maintained, which changes over time or space. Although
425 the exact mechanism is not the same between the HRAP phycodnaviruses and the RNA viruses, the pressure put
426 on the viruses to evolve (*i.e.*, the presence and maintenance of specific traits of a species pangenome) and the
427 evolutionary response, or “solution”, are similar. The said pressure being the dynamics and repertoire of
428 microalgae strains and closely related species (in our case members of family *Chlorellales* – potentially genus
429 *Picochlorum*), and the solution being having a complementary genome repertoire of related viruses at hand; in
430 other words a genome repertoire to handle host diversity similar to RNA quasispecies evolutionary “behaviour”.
431 In a study of *Prasinoviruses* found in *Ostreococcus* spp. (Yau et al., 2018), authors investigated a mechanism of
432 dynamics by which *Nucleocyotoviricota* and their hosts interactions centred around the host species (a microalgae
433 species or strain) being resistant or susceptible to virus infection. Simplified, it was shown that in different
434 environments, and in presence or absence of viruses, different microalgal life history strategies involving virus-
435 susceptible and virus-resistant microalgae would be favoured or disfavoured. We thereby suggest that these
436 phycodnaviruses forming this level of diversity in association with a dense bloom could be defined as a
437 “quasigenus”, akin to the RNA quasispecies. This quasigenus consist of a group of related, but individual
438 distinct, *Phycodnaviridae* that are capable of maintenance during a dense bloom of their host alga. Alga strains
439 within one alga species can be virus susceptible or resistant (*e.g.*, with the system of *Aureococcus*
440 *anophagefferens* and AaV (Gobler et al., 2007; Brown and Bidle, 2014; Gann et al., 2022)), thereby a dense
441 culture with different strains could support a group of closely related viruses (the “quasigenus”) that are capable
442 of infecting some (one or more) strains of an alga species and not others. It will be interesting to determine if
443 this “quasigenus” phenomenon also happens so prominently during algal blooms in natural environments or if
444 it meets only in the very favourable conditions of the HRAP. Although the concept of several different viruses
445 infecting one species is not new, it is specifically intriguing that the maintenance of several phycodnaviruses
446 could reflect the persistence of algal strains within possibly one host species.

447

448 *Viruses may be a factor of mass mortality and bloom control within the HRAP*

449 Several culture crashes occurred in the HRAP throughout 2017 and 2018, some were during periods where our
450 dataset is not at a high enough resolution to make implications about whether viruses were a significant factor

451 or not. For example, *Mimiviridae* 1 (**Figure 3 e**) and PLV 8 (**Figure 4 c**) are the only viruses tracked that
452 seemingly persists before the July 2018 crash (excluding 2017 data), although not many samples were collected
453 near the July 2018 crash. The resolution around the three crashes occurring in Autumn 2018 are quite well
454 defined overall, and consequently the abundance of viruses (**Figure 3 b, Figure 4 a**) appear to coincide well with
455 these crashes, thereby we hypothesize that the group of mostly phycodnaviruses are contributing to the crash of
456 primarily *Picochlorum*. This is not unusual as viruses are known to participate in bloom termination (Gastrich
457 et al., 2004; Lawrence and Suttle, 2004; Brussaard et al., 2005). In context of the co-occurring PLVs, it is possible
458 they are simply present and absent based on the presence and the removal of their potential hosts (*e.g.*, induced
459 by other predators, parasites or fluctuating seasonal conditions) and the overall crash and re-initiation of the
460 system, but more significantly we do not know the effect of these PLVs on their hosts and if PLVs can contribute
461 to or cause bloom termination in marine environments (*e.g.*, if these PLVs lyse their hosts versus behave like
462 *Lavidaviridae* with a tripartite infection system). Broadly speaking, the HRAP may not completely reflect the
463 trajectory of Mediterranean sourced water used to initiate the HRAP culture, given that some algae only appear
464 around times of culture restarts and then disappear after (family *Mamiellaceae* and *Chlamydomonadales*; **Figure**
465 **5**), however for the microalgae that persist and the viruses occurring alongside there appears to be an algal bloom
466 like maintenance and resulting virus bloom termination.

467

468 **Conclusion**

469 Overall, this study demonstrates the incredible diversity and dynamics of DNA viruses within an HRAP.
470 Ultimately, we suggest that the HRAP environment can mimic a marine microalgae bloom. This resulting
471 diversity of putative *Phycodnaviridae* contains a repertoire of viruses available to infect, at least one genera (and
472 potentially the strains of one or more species) of microalgae, giving way to the concept “quasispecies” of viruses.
473 Although the exact progression to reach this result are different than that of RNA quasispecies, we cannot deny
474 similarities in the end results of this potential virus-host dynamic. Ultimately, it appears that viruses (specifically
475 those involved in a “quasispecies” behaviour) could be meaningfully contributing to the termination of an algal
476 bloom in the HRAP system.

477

478 **References**

479 Agarkova, I. V., Dunigan, D. D., and Van Etten, J. L. (2006). Virion-Associated Restriction Endonucleases of
480 Chloroviruses. *Journal of Virology* 80, 8114–8123. doi: 10.1128/JVI.00486-06.

481 Andino, R., and Domingo, E. (2015). Viral quasispecies. *Virology* 479–480, 46–51. doi:
482 10.1016/j.virol.2015.03.022.

483 Antipov, D., Raiko, M., Lapidus, A., and Pevzner, P. A. (2020). MetaviralSPAdes: assembly of viruses from
484 metagenomic data. *Bioinformatics* 36, 4126–4129. doi: 10.1093/bioinformatics/btaa490.

485 Aylward, F. O., and Moniruzzaman, M. (2021). ViralRecall—A Flexible Command-Line Tool for the Detection of
486 Giant Virus Signatures in ‘Omic Data. *Viruses* 13, 150. doi: 10.3390/v13020150.

- 487 Bellas, C. M., and Sommaruga, R. (2021). Polinton-like viruses are abundant in aquatic ecosystems. *Microbiome* 9,
488 13. doi: 10.1186/s40168-020-00956-0.
- 489 Biggs, T. E. G., Huisman, J., and Brussaard, C. P. D. (2021). Viral lysis modifies seasonal phytoplankton dynamics
490 and carbon flow in the Southern Ocean. *ISME J*, 1–8. doi: 10.1038/s41396-021-01033-6.
- 491 Blanc, G., Gallot-Lavallée, L., and Maumus, F. (2015). Provirophages in the Bigelowiella genome bear testimony to
492 past encounters with giant viruses. *PNAS* 112, E5318–E5326. doi: 10.1073/pnas.1506469112.
- 493 Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-
494 PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31, 365–370. doi:
495 10.1093/nar/gkg095.
- 496 Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. (2007). UniProtKB/Swiss-Prot. *Methods*
497 *Mol Biol* 406, 89–112. doi: 10.1007/978-1-59745-535-0_4.
- 498 Brown, C. M., and Bidle, K. D. (2014). Attenuation of virus production at high multiplicities of infection in
499 *Aureococcus anophagefferens*. *Virology* 466–467, 71–81. doi: 10.1016/j.virol.2014.07.023.
- 500 Brussaard, C. P. D., Kuipers, B., and Veldhuis, M. J. W. (2005). A mesocosm study of *Phaeocystis globosa*
501 population dynamics: I. Regulatory role of viruses in bloom control. *Harmful Algae* 4, 859–874. doi:
502 10.1016/j.hal.2004.12.015.
- 503 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2:
504 High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13, 581–583. doi:
505 10.1038/nmeth.3869.
- 506 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2:
507 Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol*
508 *Biol Evol* 38, 5825–5829. doi: 10.1093/molbev/msab293.
- 509 Chase, E. E., Monteil-Bouchard, S., Gobet, A., Andrianjakarivony, F. H., Desnues, C., and Blanc, G. (2021). A High
510 Rate Algal Pond Hosting a Dynamic Community of RNA Viruses. *Viruses* 13, 2163. doi:
511 10.3390/v13112163.
- 512 Chen, F., and Suttle, C. A. (1996). Evolutionary Relationships among Large Double-Stranded DNA Viruses That
513 Infect Microalgae and Other Organisms as Inferred from DNA Polymerase Genes. *Virology* 219, 170–178.
514 doi: 10.1006/viro.1996.0234.
- 515 Choezom, D., and Gross, J. C. (2022). Neutral sphingomyelinase 1 regulates cellular fitness at the level of ER stress
516 and cell cycle. 2022.02.23.481585. doi: 10.1101/2022.02.23.481585.
- 517 Claverie, J.-M., and Abergel, C. (2018). Mimiviridae: An Expanding Family of Highly Diverse Large dsDNA
518 Viruses Infecting a Wide Phylogenetic Range of Aquatic Eukaryotes. *Viruses* 10, 506. doi:
519 10.3390/v10090506.
- 520 Clerissi, C., Grimsley, N., Ogata, H., Hingamp, P., Poulain, J., and Desdevises, Y. (2014). Unveiling of the
521 Diversity of Prasinoviruses (Phycodnaviridae) in Marine Samples by Using High-Throughput Sequencing
522 Analyses of PCR-Amplified DNA Polymerase and Major Capsid Protein Genes. *Appl. Environ. Microbiol.*
523 80, 3150–3160. doi: 10.1128/AEM.00123-14.
- 524 Coy, S. R., Gann, E. R., Papoulis, S. E., Holder, M. E., Ajami, N. J., Petrosino, J. F., et al. (2020). SMRT
525 Sequencing of *Paramecium Bursaria* Chlorella Virus-1 Reveals Diverse Methylation Stability in Adenines
526 Targeted by Restriction Modification Systems. *Frontiers in Microbiology* 11. Available at:
527 <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00887> [Accessed December 9, 2022].

- 528 De Vuyst, G., Aci, S., Genest, D., and Culard, F. (2005). Atypical Recognition of Particular DNA Sequences by the
529 Archaeal Chromosomal MC1 Protein. *Biochemistry* 44, 10369–10377. doi: 10.1021/bi0474416.
- 530 Domingo, E., and Perales, C. (2019). Viral quasispecies. *PLoS Genetics* 15, e1008271. doi:
531 10.1371/journal.pgen.1008271.
- 532 Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. *Microbiol Mol Biol Rev* 76, 159–
533 216. doi: 10.1128/MMBR.05023-11.
- 534 Domingo, E., Soria, M. E., Gallego, I., de Ávila, A. I., García-Crespo, C., Martínez-González, B., et al. (2020). A
535 new implication of quasispecies dynamics: Broad virus diversification in absence of external perturbations.
536 *Infection, Genetics and Evolution* 82, 104278. doi: 10.1016/j.meegid.2020.104278.
- 537 Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biol* 16, e3000003. doi:
538 10.1371/journal.pbio.3000003.
- 539 Etten, J., Dunigan, D., Nagasaki, K., Schroeder, D., Grimsley, N., Brussaard, C., et al. (2020). “Phycodnaviruses
540 (Phycodnaviridae),” in *Reference Module in Life Sciences* doi: 10.1016/B978-0-12-809633-8.21291-0.
- 541 Filée, J. (2018). Giant viruses and their mobile genetic elements: the molecular symbiosis hypothesis. *Current
542 Opinion in Virology* 33, 81–88. doi: 10.1016/j.coviro.2018.07.013.
- 543 Fischer, M. G. (2021). The Virophage Family Lavidaviridae. *Current Issues in Molecular Biology* 40, 1–24. doi:
544 10.21775/cimb.040.001.
- 545 Fischer, M. G., Kelly, I., Foster, L. J., and Suttle, C. A. (2014). The virion of Cafeteria roenbergensis virus (CroV)
546 contains a complex suite of proteins for transcription and DNA repair. *Virology* 466–467, 82–94. doi:
547 10.1016/j.virol.2014.05.029.
- 548 Gallot-Lavallée, L., and Blanc, G. (2017). A Glimpse of Nucleo-Cytoplasmic Large DNA Virus Biodiversity
549 through the Eukaryotic Genomics Window. *Viruses* 9, 17. doi: 10.3390/v9010017.
- 550 Gann, E. R., Truchon, A. R., Papoulis, S. E., Dyhrman, S. T., Gobler, C. J., and Wilhelm, S. W. (2022).
551 *Aureococcus anophagefferens* (Pelagophyceae) genomes improve evaluation of nutrient acquisition
552 strategies involved in brown tide dynamics. *Journal of Phycology* 58, 146–160. doi: 10.1111/jpy.13221.
- 553 Gastrich, M. D., Leigh-Bell, J. A., Gobler, C. J., Roger Anderson, O., Wilhelm, S. W., and Bryan, M. (2004).
554 Viruses as potential regulators of regional brown tide blooms caused by the alga, *Aureococcus*
555 *anophagefferens*. *Estuaries* 27, 112–119. doi: 10.1007/BF02803565.
- 556 Gobler, C., Anderson, O., Gastrich, M., and Wilhelm, S. (2007). Ecological aspects of viral infection and lysis in the
557 harmful brown tide alga *Aureococcus anophagefferens*. *Aquat. Microb. Ecol.* 47, 25–36. doi:
558 10.3354/ame047025.
- 559 Gontero, B., and Maberly, S. C. (2012). An intrinsically disordered protein, CP12: jack of all trades and master of
560 the Calvin cycle. *Biochem Soc Trans* 40, 995–999. doi: 10.1042/bst20120097.
- 561 Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2013). The Protist Ribosomal Reference
562 database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated
563 taxonomy. *Nucleic Acids Research* 41, D597–D604. doi: 10.1093/nar/gks1160.
- 564 Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., et al. (2013). Exploring nucleo-
565 cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J* 7, 1678–1695. doi:
566 10.1038/ismej.2013.59.

- 567 Jeudy, S., Rigou, S., Alempic, J.-M., Claverie, J.-M., Abergel, C., and Legendre, M. (2020). The DNA methylation
568 landscape of giant viruses. *Nat Commun* 11, 2657. doi: 10.1038/s41467-020-16414-2.
- 569 Jian, H., Yi, Y., Wang, J., Hao, Y., Zhang, M., Wang, S., et al. (2021). Diversity and distribution of viruses
570 inhabiting the deepest ocean on Earth. *ISME J* 15, 3094–3110. doi: 10.1038/s41396-021-00994-y.
- 571 Jurasz, H., Pawłowski, T., and Perlejewski, K. (2021). Contamination Issue in Viral Metagenomics: Problems,
572 Solutions, and Clinical Perspectives. *Frontiers in Microbiology* 12. Available at:
573 <https://www.frontiersin.org/articles/10.3389/fmicb.2021.745076> [Accessed December 9, 2022].
- 574 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for
575 gene and protein annotation. *Nucleic Acids Research* 44, D457–D462. doi: 10.1093/nar/gkv1070.
- 576 Kapitonov, V. V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the
577 National Academy of Sciences* 103, 4540–4545. doi: 10.1073/pnas.0600833103.
- 578 Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements
579 in Performance and Usability. *Mol Biol Evol* 30, 772–780. doi: 10.1093/molbev/mst010.
- 580 Khalil, J. Y. B., Andreani, J., and La Scola, B. (2016). Updating strategies for isolating and discovering giant
581 viruses. *Current Opinion in Microbiology* 31, 80–87. doi: 10.1016/j.mib.2016.03.004.
- 582 Koonin, E. V., and Krupovic, M. (2017). Polintons, virophages and transpovirons: a tangled web linking viruses,
583 transposons and immunity. *Current Opinion in Virology* 25, 7–15. doi: 10.1016/j.coviro.2017.06.008.
- 584 Krasovec, M., Vancaester, E., Rombauts, S., Bucchini, F., Yau, S., Hemon, C., et al. (2018). Genome Analyses of
585 the Microalga *Picochlorum* Provide Insights into the Evolution of Thermotolerance in the Green Lineage.
586 *Genome Biol Evol* 10, 2347–2365. doi: 10.1093/gbe/evy167.
- 587 Krupovic, M., and Koonin, E. V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution.
588 *Nat Rev Microbiol* 13, 105–115. doi: 10.1038/nrmicro3389.
- 589 La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a
590 unique parasite of the giant mimivirus. *Nature* 455, 100–104. doi: 10.1038/nature07218.
- 591 Lawrence, J., and Suttle, C. (2004). Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its
592 implications for bloom termination. *Aquat. Microb. Ecol.* 37, 1–7. doi: 10.3354/ame037001.
- 593 Luo, E., Eppley, J. M., Romano, A. E., Mende, D. R., and DeLong, E. F. (2020). Double-stranded DNA
594 viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J*
595 14, 1304–1315. doi: 10.1038/s41396-020-0604-8.
- 596 Marshall, O. J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time
597 PCR. *Bioinformatics* 20, 2471–2472. doi: 10.1093/bioinformatics/bth254.
- 598 Mikheenko, A., Saveliev, V., and Gurevich, A. (2016). MetaQUAST: evaluation of metagenome assemblies.
599 *Bioinformatics* 32, 1088–1090. doi: 10.1093/bioinformatics/btv697.
- 600 Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., and Aylward, F. O. (2020). Widespread
601 endogenization of giant viruses shapes genomes of green algae. *Nature* 588, 141–145. doi:
602 10.1038/s41586-020-2924-2.
- 603 Nerlich, A., von Orlow, M., Rontein, D., Hanson, A. D., and Dörmann, P. (2007). Deficiency in phosphatidylserine
604 decarboxylase activity in the psd1 psd2 psd3 triple mutant of *Arabidopsis* affects phosphatidylethanolamine
605 accumulation in mitochondria. *Plant Physiol* 144, 904–914. doi: 10.1104/pp.107.095414.

- 606 Pagarete, A., Grébert, T., Stepanova, O., Sandaa, R.-A., and Bratbak, G. (2015). Tsv-N1: A Novel DNA Algal Virus
607 that Infects *Tetraselmis striata*. *Viruses* 7, 3937–3953. doi: 10.3390/v7072806.
- 608 Palermo, C. N., Shea, D. W., and Short, S. M. (2021). Analysis of Different Size Fractions Provides a More
609 Complete Perspective of Viral Diversity in a Freshwater Embayment. *Applied and Environmental*
610 *Microbiology* 87, e00197-21. doi: 10.1128/AEM.00197-21.
- 611 Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with
612 Profiles instead of a Distance Matrix. *Mol Biol Evol* 26, 1641–1650. doi: 10.1093/molbev/msp077.
- 613 Redrejo-Rodríguez, M., and Salas, M. L. (2014). Repair of base damage and genome maintenance in the nucleo-
614 cytoplasmic large DNA viruses. *Virus Research* 179, 12–25. doi: 10.1016/j.virusres.2013.10.017.
- 615 Roux, S., Chan, L.-K., Egan, R., Malmstrom, R. R., McMahon, K. D., and Sullivan, M. B. (2017). Ecogenomics of
616 virophages and their giant virus hosts assessed through time series metagenomics. *Nat Commun* 8, 858. doi:
617 10.1038/s41467-017-01086-2.
- 618 Sheng, Y., Wu, Z., Xu, S., and Wang, Y. (2022). Isolation and Identification of a Large Green Alga Virus (Chlorella
619 Virus XW01) of Mimiviridae and Its Virophage (Chlorella Virus Virophage SW01) by Using Unicellular
620 Green Algal Cultures. *J Virol* 96, e0211421. doi: 10.1128/jvi.02114-21.
- 621 Short, S. M., Staniewski, M. A., Chaban, Y. V., Long, A. M., and Wang, D. (2020). Diversity of Viruses Infecting
622 Eukaryotic Algae. *Current Issues in Molecular Biology*, 29–62. doi: 10.21775/cimb.039.029.
- 623 Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data
624 integration and network visualization. *Bioinformatics* 27, 431–432. doi: 10.1093/bioinformatics/btq675.
- 625 Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of
626 massive data sets. *Nature Biotechnology* 35, 1026–1028. doi: 10.1038/nbt.3988.
- 627 Suttle, C. A., and Wilhelm, S. W. (1999). Viruses and Nutrient Cycles in the Sea. 49, 8.
- 628 Tahlan, K., Park, H. U., Wong, A., Beatty, P. H., and Jensen, S. E. (2004). Two Sets of Paralogous Genes Encode
629 the Enzymes Involved in the Early Stages of Clavulanic Acid and Clavam Metabolite Biosynthesis in
630 *Streptomyces clavuligerus*. *Antimicrob Agents Chemother* 48, 930–939. doi: 10.1128/AAC.48.3.930-
631 939.2004.
- 632 Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-
633 scale analysis of protein functions and evolution. *Nucleic Acids Res* 28, 33–36.
- 634 Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new
635 capabilities and interfaces. *Nucleic Acids Res* 40, e115–e115. doi: 10.1093/nar/gks596.
- 636 Van Etten, J. L., Burbank, D. E., Xia, Y., and Meints, R. H. (1983). Growth cycle of a virus, PBCV-1, that infects
637 Chlorella-like algae. *Virology* 126, 117–125. doi: 10.1016/0042-6822(83)90466-X.
- 638 Wilson, W. H., Van Etten, J. L., and Allen, M. J. (2009). The Phycodnaviridae: The Story of How Tiny Giants Rule
639 the World. *Curr Top Microbiol Immunol* 328, 1–42.
- 640 Yau, S., Caravello, G., Fonvieille, N., Desgranges, É., Moreau, H., and Grimsley, N. (2018). Rapidity of Genomic
641 Adaptations to Prasinovirus Infection in a Marine Microalga. *Viruses* 10, 441. doi: 10.3390/v10080441.
- 642 Yau, S., Lauro, F. M., DeMaere, M. Z., Brown, M. V., Thomas, T., Raftery, M. J., et al. (2011). Virophage control
643 of antarctic algal host-virus dynamics. *Proceedings of the National Academy of Sciences* 108, 6163–6168.
644 doi: 10.1073/pnas.1018221108.

645 Yutin, N., Shevchenko, S., Kapitonov, V., Krupovic, M., and Koonin, E. V. (2015). A novel group of diverse
646 Polinton-like viruses discovered by metagenome analysis. *BMC Biology* 13, 95. doi: 10.1186/s12915-015-
647 0207-4.

648 Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses:
649 Clusters of orthologous genes and reconstruction of viral genome evolution. *Virology* 6, 223. doi:
650 10.1186/1743-422X-6-223.

651 Zhang, W., Zhong, H., Lu, H., Zhang, Y., Deng, X., Huang, K., et al. (2018). Characterization of Ferredoxin-
652 Dependent Biliverdin Reductase PCYA1 Reveals the Dual Function in Retrograde Bilin Biosynthesis and
653 Interaction With Light-Dependent Protochlorophyllide Oxidoreductase LPOR in *Chlamydomonas*
654 *reinhardtii*. *Front Plant Sci* 9, 676. doi: 10.3389/fpls.2018.00676.

655

656 **Acknowledgments**

657 The authors thank the members of the VASCO2 consortium and the company COLDEP for their support and
658 permission to access the cultures. We also acknowledge support by the Institut Français de Recherche pour
659 l'Exploitation de la Mer (IFREMER) and work done by Angélique Gobet. Including the Ifremer Palavas-les-
660 Flots raceway facility and GeT-PLaGe platform (Genotoul). We also wish to specifically acknowledge our late
661 co-author, Dr. Christelle Desnues, who brought compassion, excellence, and inspiration to research during her
662 life.

663

664 **Author Contributions**

665 Conceptualization, C.D., G.B., E.E.C., and S.M.-B.; Methodology, C.D., G.B., S.M.-B., E.E.C.; Software,
666 E.E.C., and G.B.; Validation, formal analyses, and investigation, G.B., C.D. and E.E.C.; Resources, C.D. and
667 G.B.; Data Curation, E.E.C. and G.B.; Writing—Original Draft Preparation, E.E.C.; Writing—Review &
668 Editing, G.B., C.D., S.M.-B., T.M.P, and E.E.C.; Visualization, E.E.C.; Supervision, C.D., G.B., and S.M.-B.;
669 Project Administration, C.D., and G.B.; Funding Acquisition, G.B., and C.D. All authors have read and agreed
670 to the published version of the manuscript.

671

672 **Funding**

673 E.E.C, G.B, and C.D received funding from the European Union's Horizon 2020 research and innovation
674 programme under the Marie Skłodowska-Curie grant agreement No713750, with the financial support of the
675 Regional Council of Provence-Alpes-Côte d'Azur and with the financial support of the A*MIDEX (n° ANR-
676 11-IDEX-0001-02), funded by the Investissements d'Avenir project funded by the French Government, managed
677 by the French National Research Agency (ANR). The Phycovir project leading to this publication has received
678 funding from Excellence Initiative of Aix-Marseille University—A*MIDEX, a French “Investissements
679 d'Avenir” program. Part of the sequencing was conducted by the U.S. Department of Energy Joint Genome
680 Institute under CSP Proposal: 10.46936/10.25585/60000928, a DOE Office of Science User Facility

681 (<https://ror.org/04xml1d337>), is supported by the Office of Science of the U.S. Department of Energy operated
682 under Contract No. DE-AC02-05CH11231.

683

684 **Data Availability**

685 Data are available through and hosted by the NCBI SRA portal under the BioProject I PRJNA751746 (MiSeq
686 ultraviromes), and by the JGI Genome Portal under the JGI Proposal Id: 504991 (HiSeq megaviromes).

687

688 **Supplementary Materials**

689 Table S1: Assembly statistics, Table S2: functional annotation of viral contigs, Table S3: ASV counts, Table S4:
690 information on predicted viral regions in *Picochlorum* genomes. Figure S1: Fluorescent activated cell sorting of
691 megaviral populations, Figure S2: Broad taxonomic classification of contigs for all sequencing dataset, Figure
692 S3: Putative polinton-like viruses (PLVs) recovered from HRAP metagenomes, Figure S4: Putative
693 *Lavidaviridae* assembled from HRAP assemblies, Figure S5: Principle component analyses, Figure S6:
694 Hierarchical clustering of potential hosts and viruses. Figure S7. Hierarchical clustering of potential hosts and
695 polinton-like viruses (PL). Figure S8: Putative *Nucleocytoviricota* phylogeny based on the DNA polymerase B
696 gene, Figure S9. Putative *Nucleocytoviricota* phylogeny based on the ATPase gene. Figure S10. Dotplot
697 alignments of *Phycodnaviridae* contigs containing a PolB gene.

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

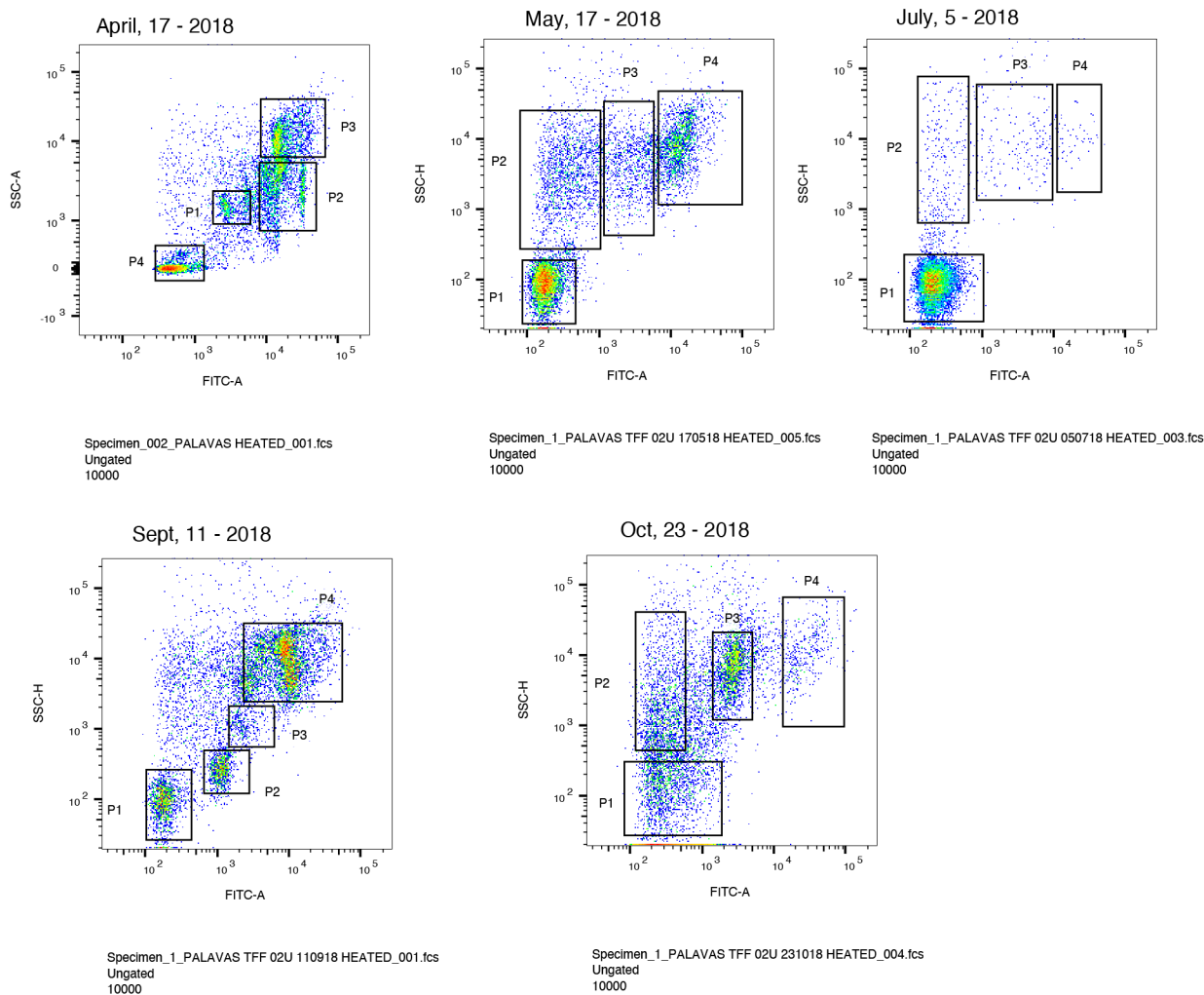
721

722

723

724 Supplemental Figures

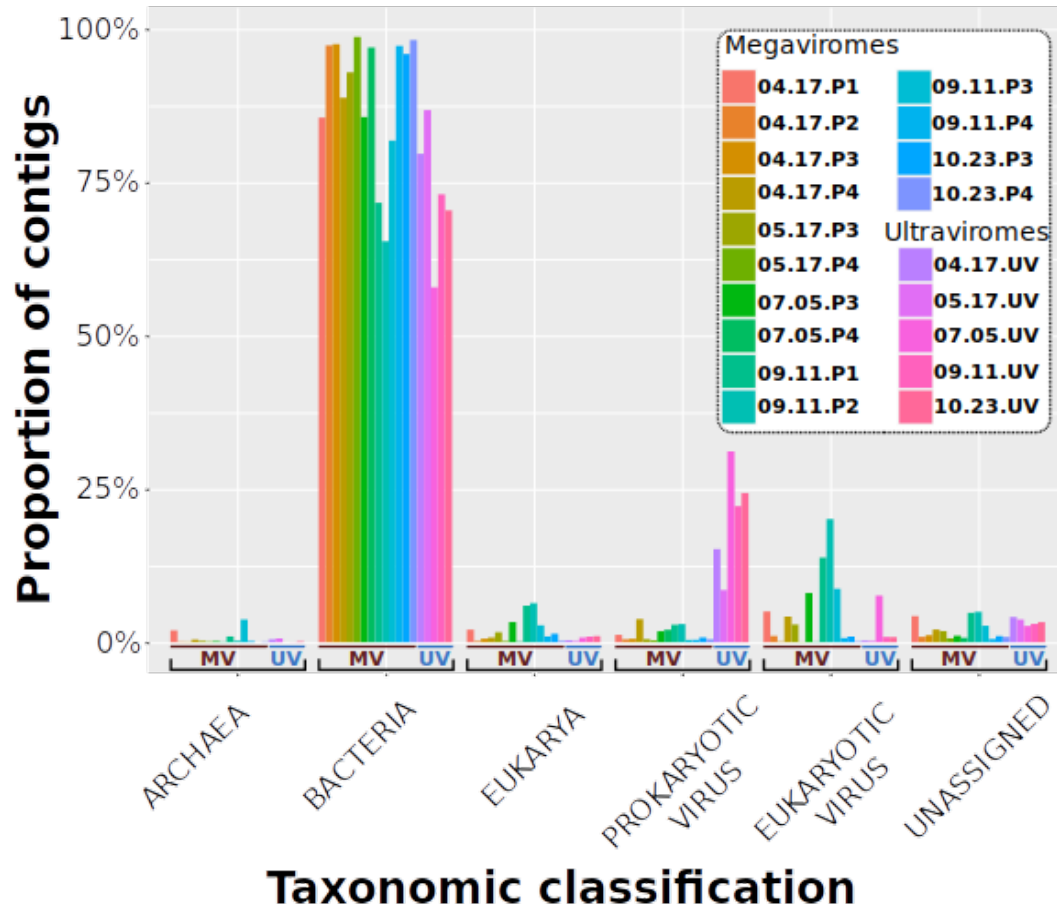
725



726

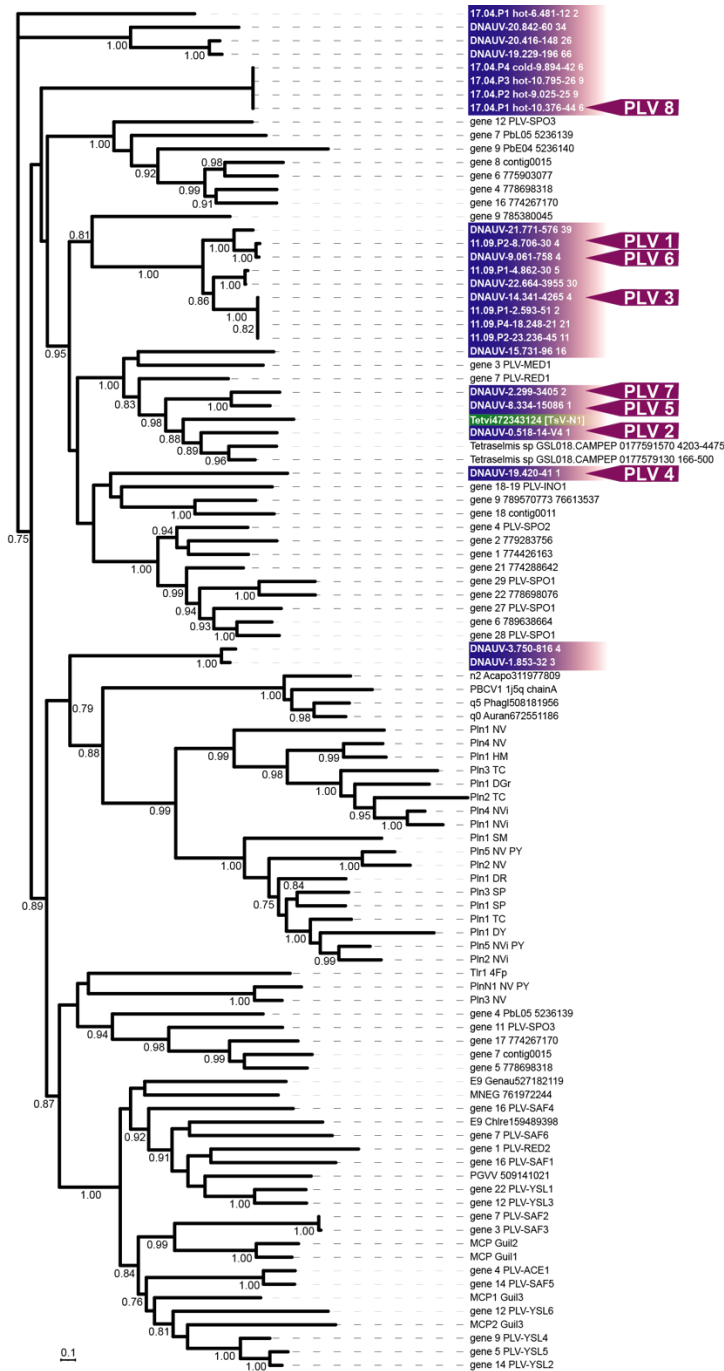
727

728 **Figure S1.** Fluorescent activated cell sorting by side scatter (SSC) and FITC (fluorescein
729 isothiocyanate; excitation and emission spectrum peak wavelengths of ~495 nm and ~519 nm).
730 Sample dates are displayed above each sample, and populations (“P”) are indicated. The sized
731 fraction for the sample water is between 0.2 μm –1.2 μm (labelled with SYBR-Green).
732



733
734
735
736
737
738
739
740
741
742

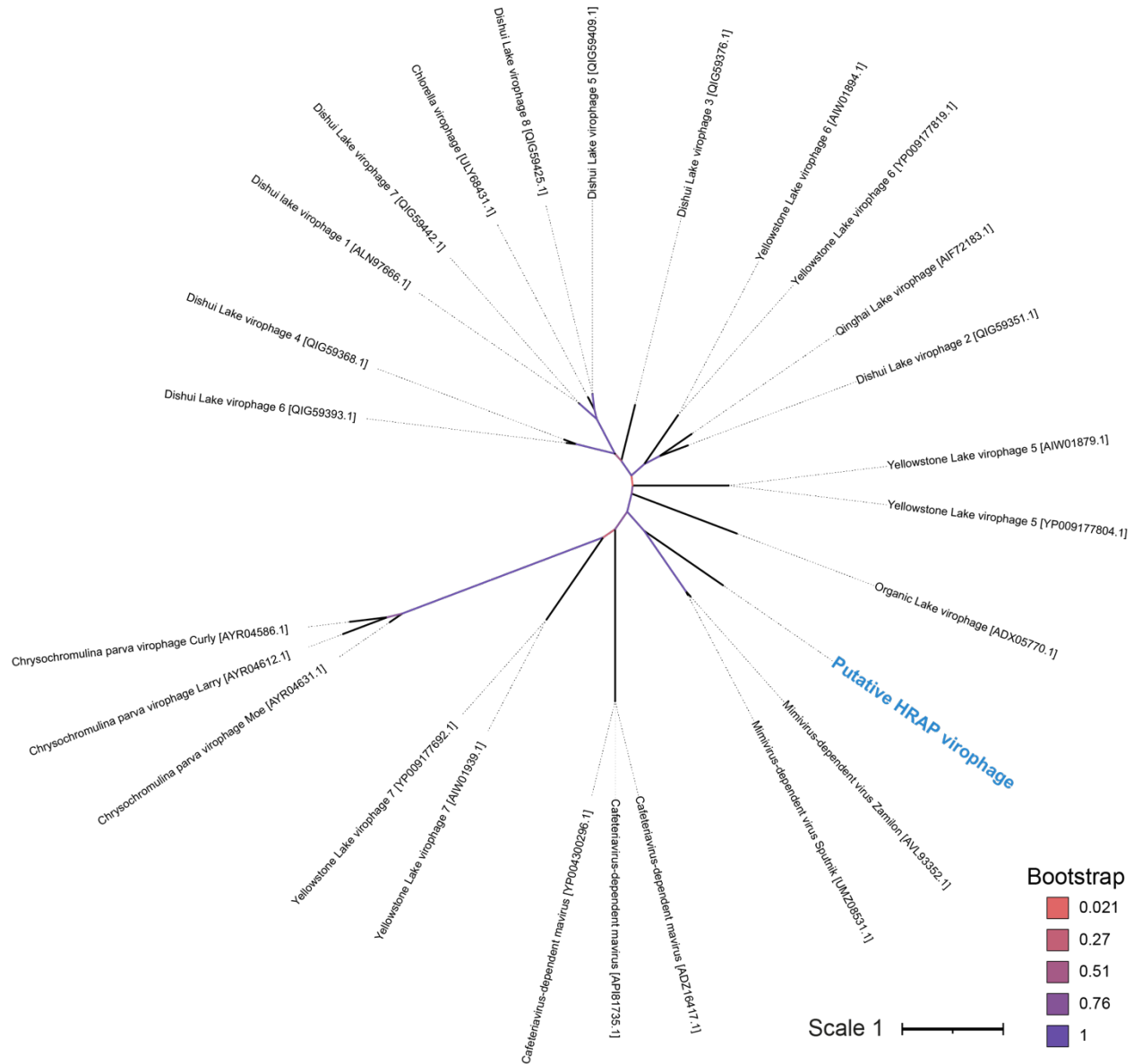
S2. Broad taxonomic classification of contigs above 2 kb for all megaviromes sequenced and a combined ultravirome sequencing dataset.



743
744

745 **Figure S3.** Putative polinton-like viruses (PLVs) recovered from HRAP metagenomes (indicated
746 in red). The PLV TsV-N1 is indicated and the green line covers the PLVs that are clustering with
747 it phylogenetically. The phylogeny based on the MCP gene. PLVs used in qPCR tracking are
748 indicated in blue. Pol 8 is excluded due to the sequence dissimilarity relative to the rest of the
749 tracked PLVs. Bootstrap values above 0.70 are displayed. All MCP (excluding those in red) were
750 extracted from public databases (see Chapter 3 for more detailed methods).

751



752

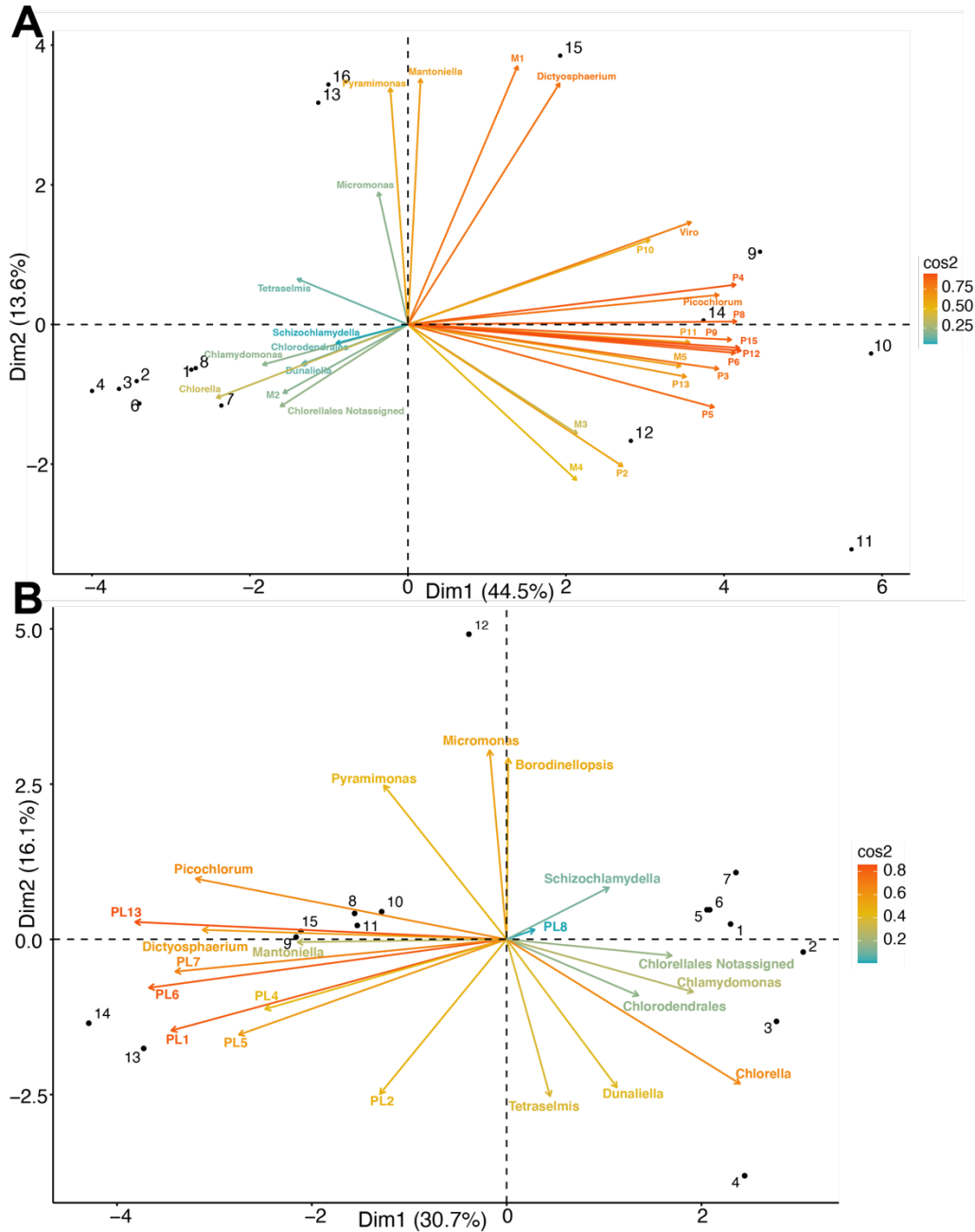
753 **Figure S4.** Putative *Lavidaviridae* (in blue) assembled from HRAP assemblies. Phylogeny is

754 based on major capsid protein sequenced retrieved from NCBI GenBank. Bootstrap probabilities

755 (1000 replicates) are displayed.

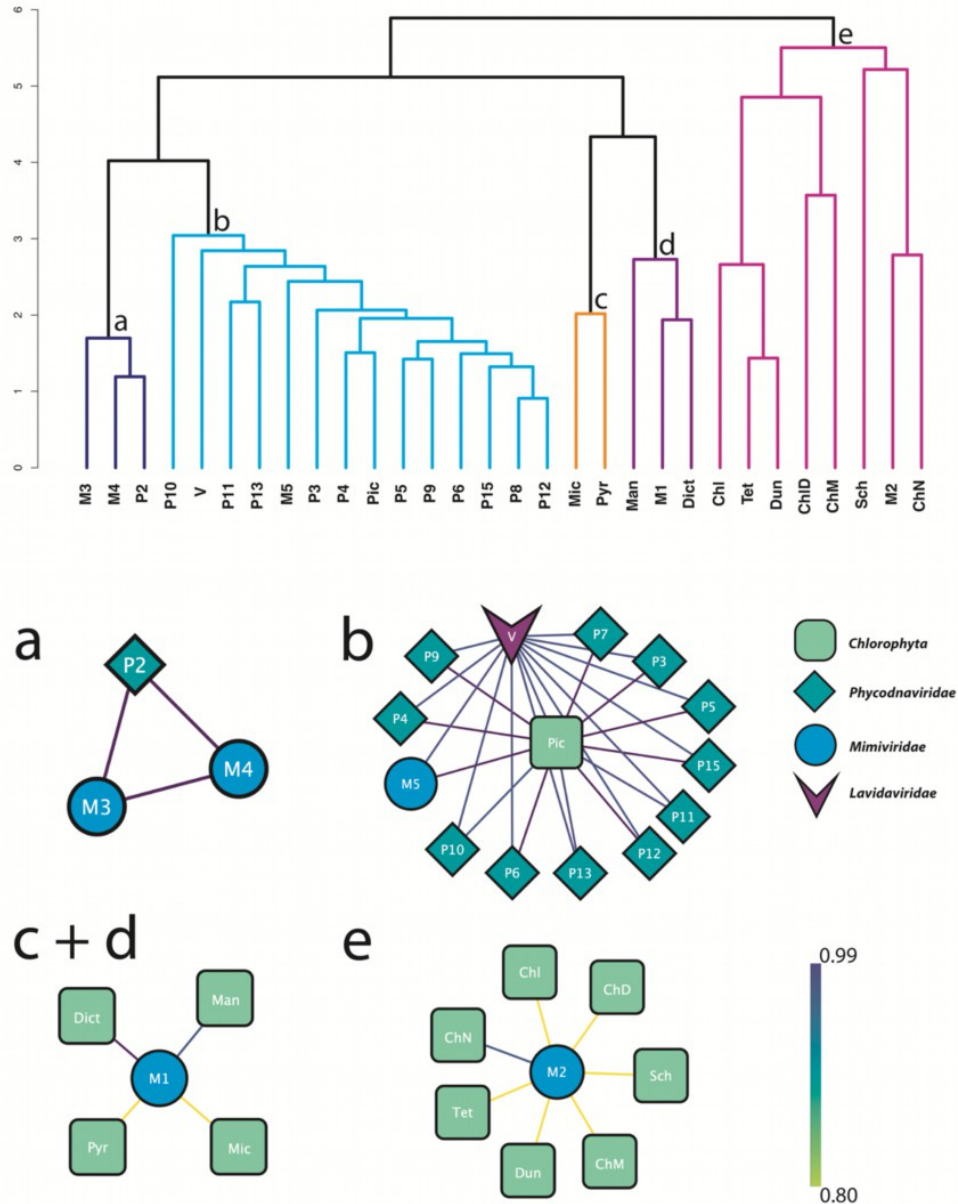
756

757



758
759

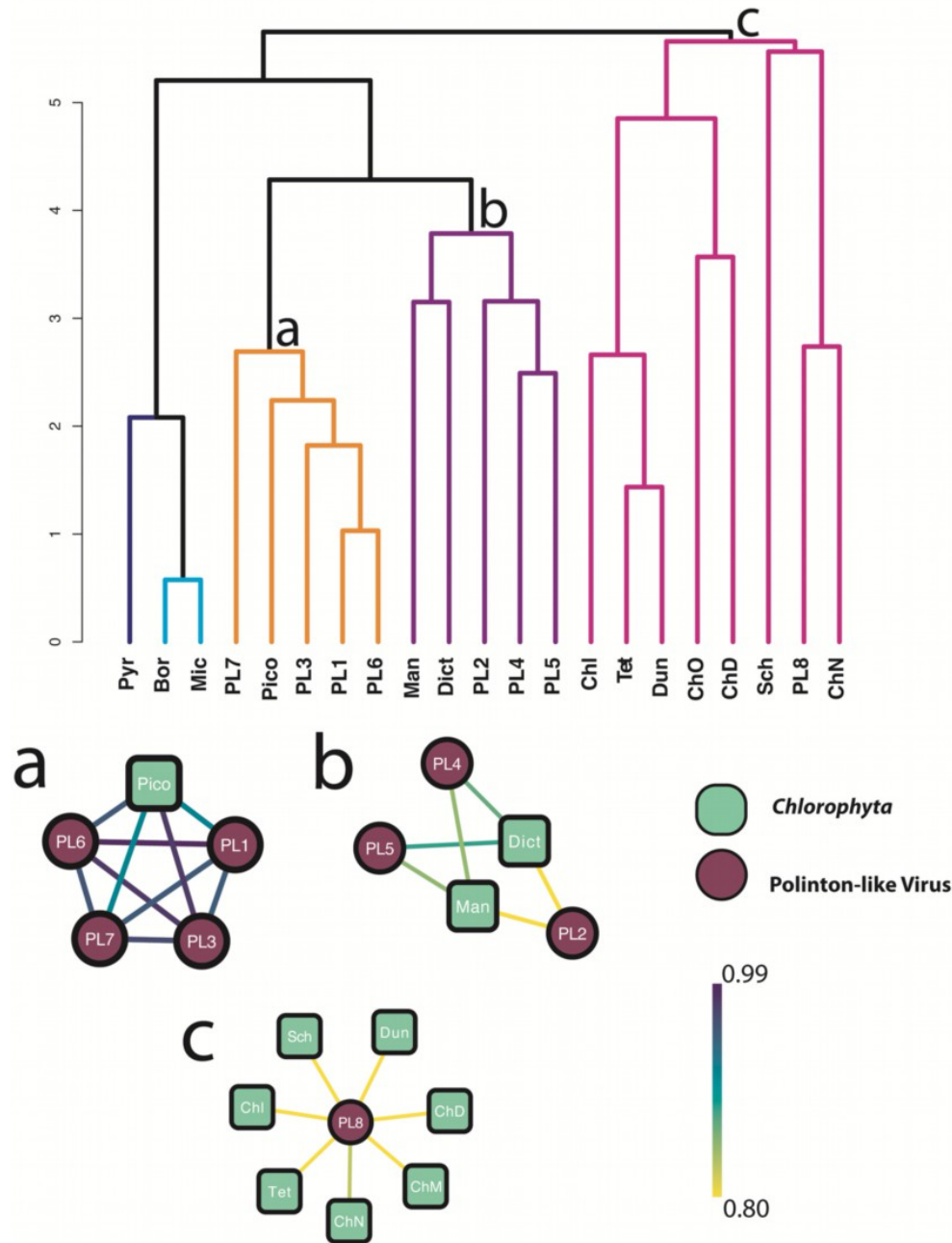
760 **Figure S5.** Principle component analyses using combined and normalised qPCR data and 18S
761 ASVs (i.e. metabarcoding) on tracked putative viruses and potential alga (Chlorophyta) hosts for
762 **(A)** *Mimiviridae* (M), *Phycodnaviridae* (P), and viroplage (Viro), and **(B)** polinton-like viruses
763 (PL). Cos2 reports the strength of the principle component for the observations (i.e. virus or
764 potential hosts), where a higher value depicts a stronger relationship between them or a “good
765 representation”. The vector length of each observations represents the contribution they make to
766 the ordination. Dates are represented by numbered objects, where 2017 is composed of 1–5
767 (April is 1 and 2, May is 3, June is 4 and 5), and 2018 is composed of 6–16 (April is 6, May is 7–
768 9, August is 9, September Is 10–13, and October is 14–16).
769



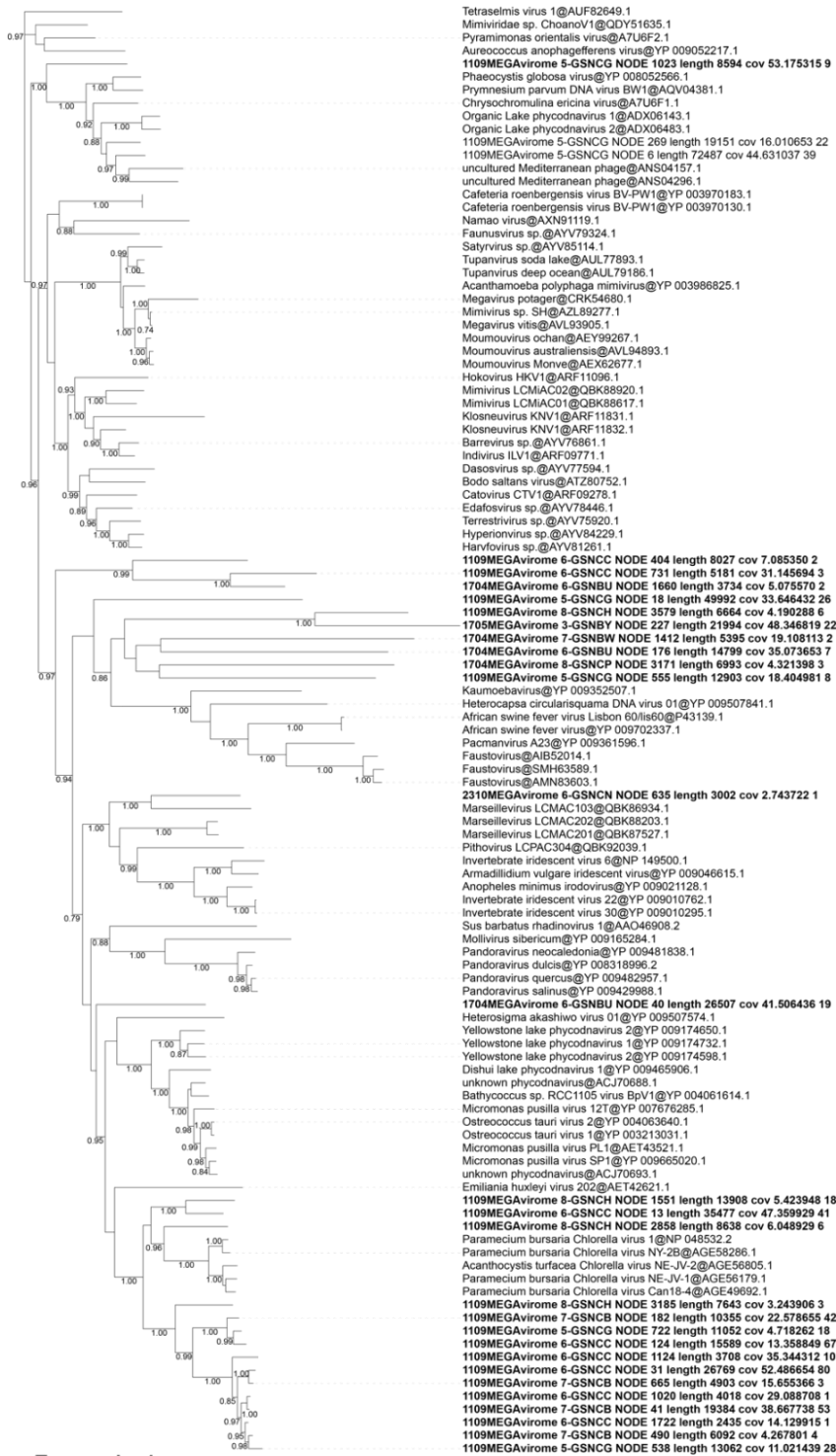
770

771 **Figure S6.** Hierarchical clustering (SIMPROF, $\alpha = 0.05$) of potential hosts (Chlorophyta) and
 772 viruses of interest; *Phycodnaviridae* (P), *Mimiviridae* (M), and virophage (V). Predicted
 773 groupings are depicted (a–e) using a network visualisation. Darker colours represented a higher
 774 correlation between group members, where 1 is an exact correlation and 0 is no correlation.
 775 Chlorophyta species are abbreviated as follows: Pic; *Picochlorum*, Dict; *Dictyosphaerium*, Man;
 776 *Mantoniella*, Pyr; *Pyramimonas*, Mic; *Micromonas*, Chl; *Chlorella*, ChD; *Chlorodendrales*,
 777 ChN; *Chlorella* not assigned, Sch; *Schizochlamydeella*, Tet; *Tetraselmis*, Dun; *Dunaliella*, and
 778 ChM; *Chlamydomonas*.

779

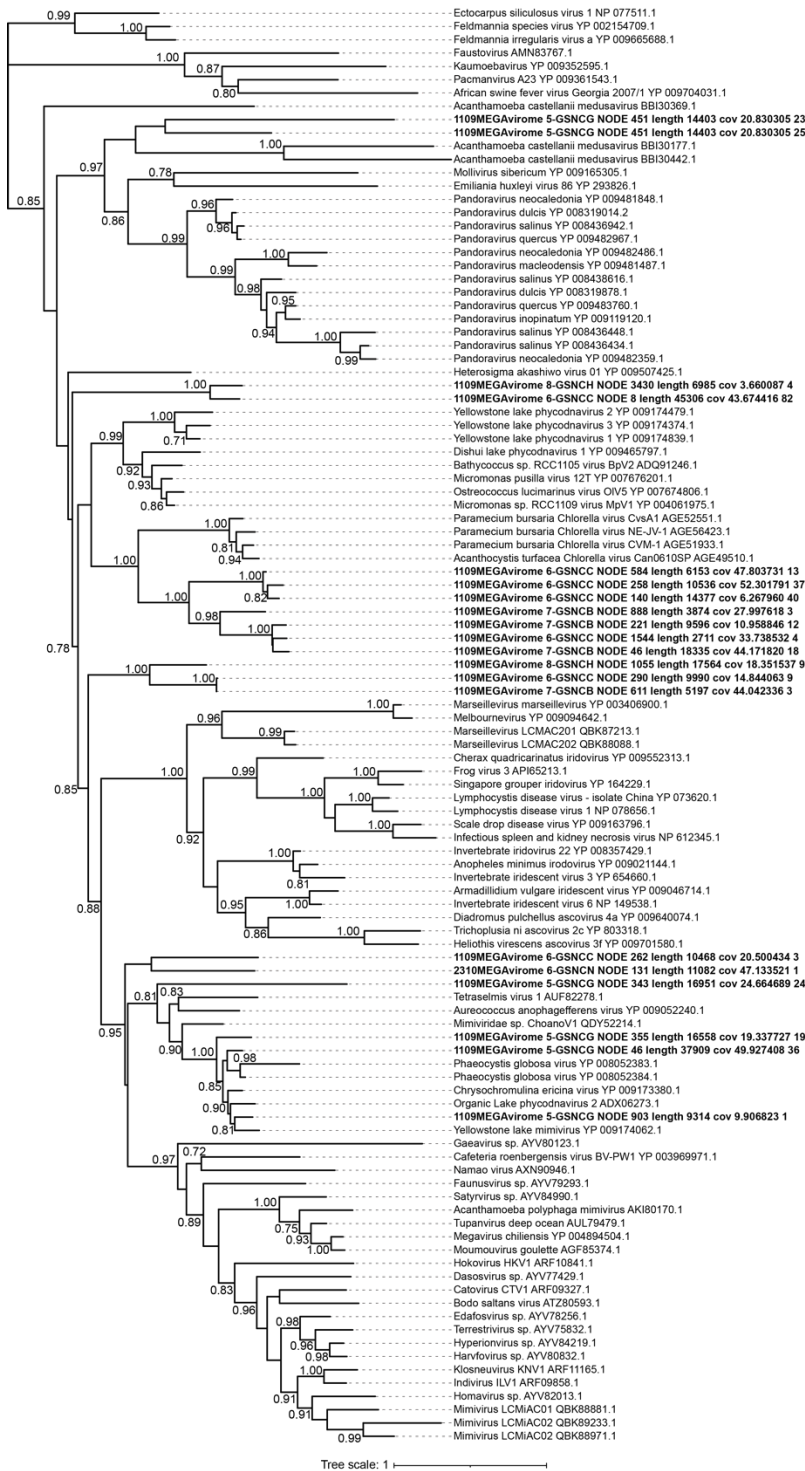


780
 781 **Figure S7.** Hierarchical clustering (SIMPROF, $\alpha = 0.05$) of potential hosts (Chlorophyta) and
 782 viruses of interest; polinton-like viruses (PL). Predicted groupings are depicted (a–e) using a
 783 network visualisation. Darker colours represented a higher correlation between group members,
 784 where 1 is an exact correlation and 0 is no correlation. Chlorophyta species are abbreviated as
 785 follows: Pic; *Picochlorum*, Dict; *Dictyosphaerium*, Man; *Mantoniella*, Sch; *Schizochlamyde*,
 786 Dun; *Dunaliella*, Chl; *Chlorella*, ChD; *Chlorodendrales*, Tet; *Tetraselmis*, ChM;
 787 *Chlamydomonas*, and ChN; *Chlorella* not assigned.
 788



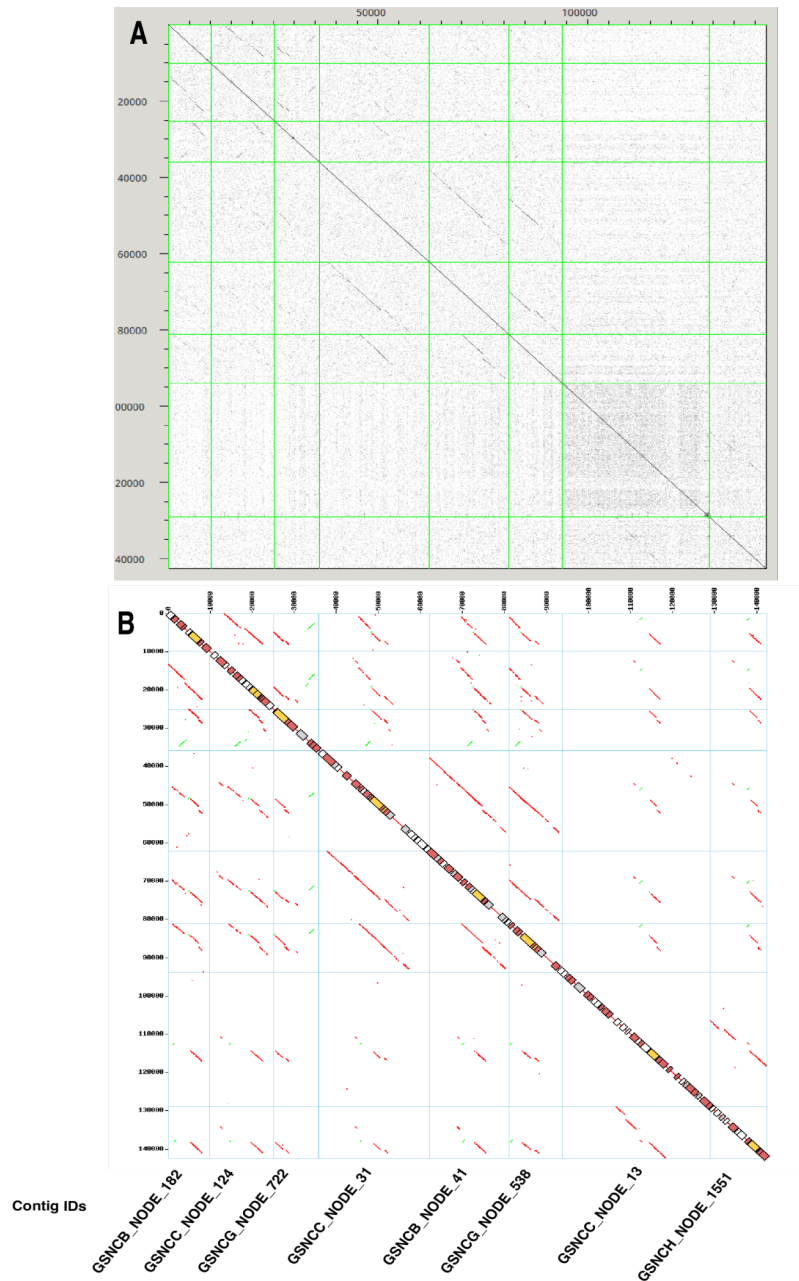
789
790
791
792
793
794
795

Figure S8. Putative *Nucleocytoviricota* phylogeny based on the DNA polymerase B gene (polB). Reference sequences were downloaded by NCBI GenBank (accession numbers as indicated) and sequenced recovered from the HRAP are included (in bold). Bootstrap values above 0.70 are displayed.



796
797
798
799

Figure S9. Putative *Nucleocytoviricota* phylogeny based on the ATPase gene. Reference sequences were downloaded by NCBI GenBank (accession numbers as indicated) and sequenced recovered from the HRAP are included (in bold). Bootstrap values above 0.70 are displayed.



800

801 **Figure S10.** Dotplot alignments of *Phycodnaviridae* contigs containing a PolB gene.

802 (A) 8 x 8 dotplot alignments of contig nucleotide sequences. (B) Dotplot representation of
803 contig's translated nucleotide sequence matches identified by TBLASTX (red for forward
804 matches, green for reverse matches). Position of the PolB gene on individual contigs is shown by
805 yellow rectangles on the plot diagonal. Additional ORFs with a significant match to viruses or
806 cellular organisms in TrEMBL are shown by red and grey rectangles respectively, and ORFs >
807 150 codons without match are shown with open rectangles. Contig orders are identical along the
808 x- and y-axis in A and B. Only contigs greater than 10 kb were used in the analysis. Altogether, a
809 residual gene collinearity is generally observed between *Phycodnaviridae* contigs, with only very
810 few inversions (green matches), however nucleotide sequence conservation is frequently
811 interrupted outside of coding sequences indicating that intergenic sequences have diverged to the
812 point that no significant similarity is yet detectable.