



HAL
open science

A high rate algal pond hosting a dynamic community of RNA viruses

Emily E Chase, Sonia Monteil-Bouchard, Felana H Andrianjakarivony,
Angélique Gobet, Christelle Desnues, Guillaume Blanc

► To cite this version:

Emily E Chase, Sonia Monteil-Bouchard, Felana H Andrianjakarivony, Angélique Gobet, Christelle Desnues, et al.. A high rate algal pond hosting a dynamic community of RNA viruses. *Viruses*, In press, 10.3390/v13112163 . hal-03394586v1

HAL Id: hal-03394586

<https://hal.science/hal-03394586v1>

Submitted on 22 Oct 2021 (v1), last revised 5 Oct 2022 (v2)

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 Article

2 **A high rate algal pond hosting a dynamic community of RNA**
3 **viruses**4 **Emily E. Chase**^{1,2*}, **Sonia Monteil-Bouchard**¹, **Angélique Gobet**³, **Felana H. Andrianjakarivony**^{1,2}, **Christelle**
5 **Desnues**^{1,2, *} and **Guillaume Blanc**^{1*}.6 ¹ Microbiologie Environnementale Biotechnologie, Institut Méditerranéen d'Océanologie, Campus de Luminy, 163 Avenue de Luminy, 13009
7 Marseille, France8 ² Institut hospitalo-universitaire (IHU) Méditerranée infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France9 ³ MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Sète, France10 * Correspondence: E.E.C.; emily.chase@mio.osupytheas.fr, C.D.; Christelle.DESNUES@univ-amu.fr, G.B.;
11 guillaume.blanc@mio.osupytheas.fr

Abstract: Despite a surge of RNA virome sequencing in recent years, there are still many RNA viruses to uncover – as indicated by the relevance of viral dark matter to RNA virome studies (i.e. putative viruses that do not match to taxonomically identified viruses). This study explores a unique site, a high rate algal pond (HRAP) for culturing industrially microalgae, to elucidate new RNA viruses. The importance of viral-host interactions in aquatic systems are well documented, and the ever-expanding microalgae industry is no exception. As the industry becomes a more important source of sustainable plastic manufacturing, producer of cosmetic pigments and alternative protein sources, and a means of CO₂ remediation in the face of climate change, studying microalgal viruses becomes a vital practice for proactive management of microalgal cultures at the industrial level. This study provides evidence of RNA microalgal viruses persisting in a CO₂ remediation pilot project HRAP and uncovers the diversity of the RNA virosphere contained within it. Evidence shows that family *Marnaviridae* is cultured in the basin, alongside other potential microalgal infecting viruses (e.g. family *Narnaviridae*, family *Totitoviridae*, and family *Yueviridae*). Finally, we demonstrate that the RNA viral diversity of the HRAP is temporally dynamic across two successive culturing seasons.

Citation: Chase, E.E.; Monteil-Bouchard, S.; Gobet, A.; Andrianjakarivony, F.H.; Desnues, C.; Blanc, G. A high rate algal pond hosting a dynamic community of RNA viruses. *Viruses* **2021**, *13*, x. <https://doi.org/10.3390/xxxxx>

Keywords: microalgae; *Marnaviridae*; community diversity; community dynamics

Academic Editor(s):

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

RNA viruses persist in a range of environments from soils [1] to seas [2,3], from the Arctic [4] to Antarctic [5], and lake waters [6,7]. The ubiquitous nature of viruses in general has been acknowledged broadly [3], however RNA viruses are historically overlooked, and it is now accepted that they may rival or exceed the amount of DNA viruses in the ocean environment based on a study of coastal waters [8]. It is unknown if this holds true spatially and temporally across other marine microbial ecosystems but does challenge previously held beliefs about the environmental abundance and importance of RNA viruses in comparison to DNA viruses. From around the time of this acknowledgement to the present day there has been a considerable spike in both the number of viral RNA metagenomic studies, and also the defining of new RNA viruses by the International Committee on Taxonomy of Viruses (ICTV) [9]. This growing number of metagenomic studies benefits from increasing sequencing depths of new technologies, alongside new bioinformatic tools that are better equipped for assembling RNA viruses, including viral populations from already publicly available metagenomic datasets and studies [9]. Along with their considerable geographical range (i.e. ubiquity), RNA viruses also infect a wide range of hosts including plants [10], vertebrates [11,12], invertebrates [13,14], fungi [15], bacteria [16], archaea [17,18], and unicellular eukaryotes including microalgae [19]. The RNA viruses of unicellular eukaryotes, the so-called

48 “protists” (when excluding fungi), are especially interesting given that many protists
49 have remained in their mostly aquatic environment instead of transitioning to primarily
50 terrestrial environments alongside other aforementioned hosts, consequentially,
51 permitting the maintenance of potentially ancient viral lineages [20]. Although the
52 prominence of viral “dark matter” (i.e. putative viral sequences that do not align to any
53 classified virus in current databases) is widespread among viruses in general [21], the
54 identification and study of protist viruses (including microalgal viruses) can shed an
55 important light on eukaryotic virus lineages [22] and expose previously unexplored (or
56 understudied) viral diversity.

57 As an overview of diversity, RNA viruses exist in forms of single stranded positive
58 [(+)ssRNA] and single stranded negative sense [(-)ssRNA] (e.g. Group IV, and Group V),
59 and double stranded(dsRNA; e.g. Group III) [23]. Among the (+)ssRNA is the order
60 Picornavirales [23], which includes 8 families as of 2019 [23]. Historically, members of
61 the order Picornavirales are composed of one post-translationally modified polyprotein
62 (with the exception of family *Dicistroviridae*), similarly structured capsids and helicases,
63 and an RNA dependent RNA polymerase (RdRp) [24], however there are many RNA
64 viruses defined as “picorna-like”, thereby not possessing all of these exact attributes or a
65 clear classification to date. At present, the importance of general viral interactions with
66 microalgae in the world’s oceans is becoming clearer after RNA viral studies being
67 overshadowed by studies of marine DNA viruses, specifically in the instance of the
68 family *Marnaviridae* (order Picornavirales), which infect microalgae.

69 Among early studies of family *Marnaviridae*, the virus HaRNAV was characterized
70 and shown to infect the toxic bloom alga *Heterosigma akashiwo* [25]. HaRNAV is among
71 other early studies of RNA viruses infecting microalgae including RsetRNAV infecting
72 *Rhizosolenia setigera* [26], and viruses infecting *Chaetoceros* spp. [27–29]. In comparison to
73 the family *Picornaviridae* (another family in the order Picornavirales), which infect many
74 economically important animals (e.g. cattle, birds, pigs), and humans [30], there are
75 substantially less studies focused on family *Marnaviridae*. Microalgae are an industrially
76 and economically important group of organisms with a variety of uses and potential
77 uses relating to cosmetics [31], food and nutritional resources [32], biopharmaceuticals
78 [33], renewable energies (i.e. biofuels) [34], wastewater treatment and CO₂ remediation
79 [35]. These uses and their potential have linked microalgae with a series of “high-value”
80 products [36]. Microalgal cultivation technologies are developed and reviewed with
81 sustainability as an important pillar (detailed more below) [37]. Given the role of
82 primary producers in regulation of marine nutrients [38] and geochemical cycles [3] it
83 has been suggested that microalgae have potential as tools for carbon sequestering [39],
84 an important action for tackling climate change [40]. We should consider how viral
85 studies can provide important information for future and current technological
86 advancements in the age of sustainability (e.g. microalgal technologies). Given that
87 elucidating new viruses and furthering our understanding of them is a major goal of the
88 virology field, we must seek unique environments to further our studies. With this
89 consideration we turn our primary focus to photosynthetic protists, the microalgae,
90 which contain a potentially unexplored diversity of RNA viruses with economic
91 importance. Viruses of microalgae, including the understudied *Marnaviridae*, have been
92 labelled as a clear biological pollutant of microalgae culturing by cell infection [41],
93 thereby creating a need for further studies specifically in the context of microalgae
94 intensive culturing practices. Furthermore, viruses are important factors known to shape
95 microbial ecosystems in general [42], so known infectants of microalgae should be
96 studied in these culturing systems.

97 In consideration of the knowledge gaps within RNA viruses, especially in the
98 *Marnaviridae* family, we carried out a study with the motive of characterizing new RNA
99 viruses by sampling an industrial microalgal culturing basin; a high rate algal pond
100 (HRAP). Our goals are to (1) provide a complete view of the RNA viral diversity of the
101 system, (2) identify new RNA viral populations, especially those that could be infecting
102 microalgae, and (3) explore specific RNA viral population dynamics in the system to
103 detect changes over time. We achieve goals (1) and (2) by a metagenomic study and (3)
104 by a quantitative RT-PCR (RT-qPCR) approach.

105 2. Materials and Methods

2.1. Experimental design and sampling

Water samples for analysis were taken from a pilot microalgal polyculture in a partially open-to-air (i.e. open system) 160m² HRAP (for details see [43]) basin system cultivating microalgae for CO₂ remediation in the IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer) marine station at Palavas-les-flots, France. Pure CO₂ with a final pH of 7.5 was systematically pumped into the HRAP for the uptake of CO₂ by these large-scale cultivations of microalgae as part of a green initiative for capturing carbon emissions expelled by industrial practices. This system featured natural colonization where seawater was pumped from the Mediterranean Sea (Plage du Prévost area) (i.e. inoculated using non-specific microalgae inoculant), and filtered through a 100 µm sand filter before entering into the HRAP. The system was restarted in the same way after each microalgal die-off (signaled by a dark green culture becoming translucent). System restarts, termed a new "run", after microalgal die-offs occurred on 17 July 2017 and May 28, July 5, and September 20 of 2018. Raw water samples (1 L) were taken for subsequent filtration and metagenomic sequencing (i.e. viral diversity studies) on April 17, May 17, July 5, September 11, and October 23 of 2018. Additional samples (50 mL) were also taken for subsequent filtration and nucleic acid extraction for RT-qPCR viral target tracking (i.e. viral community dynamic studies) throughout the basin culturing season (approximately spring to autumn) of 2017 and 2018 (detailed further below and see Supplementary Figure 1 for a timeline of sampling and HRAP runs).

2.2. RNA extraction, cDNA library preparation, and sequencing

Raw water samples (1 L) were clarified by centrifugation at 1,500 g at room temperature (RT). The supernatant (800 mL) was serially filtrated at 5 µm with a Millipore Millex-SX filter, 1.2 µm with a Minisart NML Syringe Filter (Surfactant-free Cellulose Acetate) and 140 mL of the 0.2 µm eluate was concentrated into 5 mL (viral suspension) through 100 kDa TFF (Spectrum Labs PES MicroKos). Two subsamples of 1 mL were digested with a nuclease cocktail consisting in 100 µL 10x Turbo DNAase buffer (Invitrogen), 40U Turbo DNase (Invitrogen), 18U RNase A, 125U benzonase, 100U exonuclease I, for one hour at 37°C, followed by storage at -80°C. For extraction, 3.75 mL of Trizol LS was added to 1.25 mL of the previously digested mixture, followed by homogenization and incubation for 5 minutes at RT. Chloroform (1 mL) was added and the mix was incubated at RT for 3 minutes. After centrifugation at 12,000 g (15 minutes, 4°C) the aqueous phase was recuperated and another Trizol (3x volume)/chloroform (1.25x volume) was performed. The aqueous phase containing viral RNAs was then purified and cleaned with three columns of the Zymo Research RNA Clean & Concentrator Kit 25 using manufacturer's instructions. The eluates (50 µL each) containing the purified RNAs were pooled (150 µL total) and 40U RNaseOUT (Invitrogen) was added. RNAs were again cleaned with the Zymo Research RNA Clean & Concentrator Kit 25, 40U RNaseOUT (Invitrogen) was added to the eluate (40 µL) and finally stored at -80°C.

To achieve a library of cDNA from low concentrations of RNA we carried out a reverse transcriptase and Klenow method [44,45], with a preliminary step of denaturation at 65°C for 5 minutes, with a cooling immediately afterwards. Additionally, at the end of the reverse transcriptase reaction we incubated at 94°C for 2 minutes and cooled the cDNA before the Klenow reaction. We extended the Klenow incubation step by 30 minutes (1 hour total at 37°C). An Ampurebead (Agencourt) purification was done on the cDNA, followed by an amplification with UP primers [45]. A purification was performed by NucleoFast Ultrafiltration PCR cleanup. Concentration and quality of the sample were checked by Picogreen and Agilent DNA 7500 procedures respectively. Finally, a Nextera XT Library Prep Kit (2x250bp) was used for Illumina MiSeq sequencing preparation following manufacturer's instructions and sequencing was carried out.

2.3. Quality control of reads and contig assembly of metagenomes

Paired-end Illumina MiSeq reads (2x250 bp) were preliminarily run through fastQC [46] and results were reviewed visually to check for any overrepresented sequences, per base N content, sequencing quality scores, sequence duplications, etc. Trimming of

163 Nextera adaptor sequences and low-quality sequences were performed by Trimmomatic
164 v0.36 [47], followed by a secondary run of fastQC. Sequencing assemblies were
165 produced for each metagenomic sample's reads separately, and also as a combination of
166 all ultravirome RNA reads (i.e. April, May, July, September, October) using the `-rn`aviral
167 option of SPAdes v3.15.0 [48]. Sequencing and assembly statistics on the sequencing
168 results of each sample were calculated by QUAST [49] and are reported in Table S1.
169 Reads were mapped back to the assemblies to estimate contig read coverage and observe
170 any cross-sample similarities using HISAT2 [50].

171 2.4. Taxonomic assignment of contigs and annotation

172 To study the diversity of each metagenomic sample we first performed a
173 nucleotide-to-protein alignment [51] search using MMseqs2 [52] against the National
174 Centre for Biotechnology Information (NCBI) non-redundant protein database (NR) [53]
175 using an e-value cut off of 10^{-05} . Each contig with a significant match was then attributed
176 to the same putative taxonomic clade (i.e. this method cannot define a definitive
177 taxonomy on its own) as its best match. These results were then used to filter contigs by
178 likely viral hits for inspection of the viral diversity of each sample.

179 All assembled contigs were run through EMBOSS `getorf` [54] to extract and
180 translate open reading frames (ORF) of at least 30 codons. The resulting protein
181 sequences were then searched against the NCBI NR database, and SWISS-PROT to
182 provide a functional annotation with hits corresponding to an e-value cut off of 10^{-05} .
183 Additionally, predicted proteins were assessed with InterProScan [55] for further
184 functional annotations. More specifically, contigs greater than 6Kb are presented with
185 their InterProScan [55] predicted domains annotated alongside a reference genome. The
186 reference genomes are a curated selection of *Marnaviridae* genomes accessed through
187 NCBI GenBank based on best matches of putative ORFs from the said contigs.

188 Contigs representing 99% of reads from the combined assembly of all sample reads
189 were further investigated using the previous HISAT2 [50] reads mapping data. The
190 number of reads mapping from each metagenomic sample back to the combined
191 assembly (i.e. each contig) were normalized by dividing the total reads per the said
192 sample and multiplying by 1 million. These data were used to calculate the proportion
193 of reads mapping to specific taxonomic classifications, and also the absence of
194 classification (i.e. a contig with no hit to a database).

195 2.5. Alpha and beta diversity by k-mer counts produced from metagenome contigs

196 Alpha diversity indices were calculated using all contigs above 250 bp individually
197 for each month's metagenome sample (i.e. April, May, July, September, October) and for
198 the combined reads assembly. Contigs were processed through MerCat [56] using k-
199 mers of 21bp in length. All k-mers with counts at 5 or more were assessed in the indices.
200 The following indices are reported: Shannon's diversity index [57], Simpson's diversity
201 index [58], and a richness estimator (Chao1). Bray-Curtis dissimilarity [59] (i.e. beta
202 diversity) was also calculated between these same metagenome assemblies, using Simka
203 [60] with a k-mer length of 21 bp.

204 2.6. Phylogenetic reconstruction of the RNA-dependent RNA polymerase (RdRp) domain

205 To elucidate all useful RdRp proteins from the NCBI NR database (May 2020) all
206 protein sequences assigned as order Picornavirales were downloaded using NCBI E-
207 utilities commands (`esearch`, `efetch`, and `epost`) [61], collapsed into 90% similarity by `cd-`
208 `hit` [62], processed using HMMER tool kits (`hmmsearch`, `hmmcompress`, and `hmmsearch`)[63]
209 against Pfam [64] to identify protein domains (e.g. conserved parts of proteins). This
210 step was necessary for identifying RdRp within Picornavirales because they encode
211 polyproteins (typically one in all families excluding family Discistroviridae possessing
212 two), which are long proteins that undergo post translation modifications into subunits
213 [65], consequently cut-offs between the RdRp and other proteins are difficult to pinpoint
214 and at this scale require a consistent and objective pipeline for identification. From this
215 processing, we identified and extracted 975 RdRp domains using a method adapted
216 from previous research [66]. Putative proteins predicted from our RNA ultravirome
217 MiSeq data were then searched against the 975 RdRp domain database, followed by this
218 same protein domain finding procedure that identified 23 RdRp domains within the

219 order Picornavirales. These 998 RdRp domain amino acid sequences were aligned using
220 mafft v7.487 [67], gaps were adjusted with a custom TrimAl [68] (i.e. remove gaps
221 occurring in 10% of sequences as long as at least 70% of the sequence remains), and this
222 alignment was then used to produce a maximum likelihood tree with 1000 bootstrap
223 replicates by IQTREE [69] using a general matrix substitution model (LG [70]) with
224 empirical codon frequencies, and visualized using iTol v4 [71] with modifications.

225 2.7. qPCR design for viral population tracking

226 Predicted proteins with viral functional annotations were used to select contigs
227 assembled from metagenomic data as individual targets for population dynamics
228 monitoring via RT-qPCR. Target sequences were used to produce primer pairs through
229 a Primer3 [72] Python script (Supplementary Material S1). PerlPrimer v1.2.3 [73] was
230 used to check for primer-primer and self-interactions with a Gibbs free energy cut-off of
231 -6kcal/mol to indicate whether the reaction is stable enough to reduce qPCR efficiency.
232 All primer pairs were tested by in silico PCR, and the resulting amplicons were
233 processed through mfold [74] to test for secondary structures significant enough to
234 reduce qPCR efficiency (Gibbs free energy cut-off of -10kcal/mol).

235 Aforementioned raw water sample aliquots (including from samples destined for
236 metagenomic sequencing) were filtered at 5.0 μm , then 0.2 μm in preparation for
237 extraction. Nucleic acid extractions were carried out on 200 μL of filtrate using a High
238 Pure Viral Nucleic Acid Kit (Roche) by manufacturer's instructions except an elution
239 volume of 40 μL was used. DNA was digested using a Turbo DNase Kit (ThermoFisher),
240 with a reaction termination using a 0.5M EDTA. Reverse transcriptase to acquire cDNA
241 was carried out using SuperScript™ VILO™ cDNA Synthesis Kit (ThermoFisher
242 Scientific) by manufacturer's instructions. RT-qPCR Reactions contained 0.2 μM of
243 forward and reverse primer, 1 μL template, and 5 μL SsoAdvanced Universal SYBR 2x
244 Green Supermix, supplemented with molecular biology grade (MBG) water to 10 μL
245 total volume. The qPCR program was run as follows: 5 minutes at 95°C, then 45 cycles of
246 10 seconds at each of 95°C, 60°C, and 72°C, then finalized with a melt curve to check for
247 number of products (starting at 60°C and going stepwise at 5 seconds for each 0.5°C to
248 97°C). Amplicon sizes were verified on 2% agarose gels in 1% Tris-acetate-EDTA (TAE)
249 buffer against a 100bp (BenchTop pGEM DNA Marker) ladder. Reactions were run in
250 triplicate as on one 96 well plate (all samples) for each primer pair (e.g. virus target).
251 Results were manually checked to confirm (1) amplification in the real-time
252 visualization, (2) number of products as indicated by the melt curve, (2) cycle of
253 quantification (Cq) values. A mean Cq value was calculated for each triplicate of sample
254 reactions, the number of reaction cycles (45) minus the Cq value (termed the "inverse
255 Cq") and reported.

256 3. Results

257 3.1. Taxonomic classification by database "best hit", sequencing coverage by contig, and domain 258 prediction.

259 For April, May, July, September, October, and the all reads (i.e. all reads from each
260 month combined and then assembled) assembly the proportions of contigs with a
261 database hit (any classification, virus or otherwise) were 72%, 82%, 81%, 32%, 44%, and
262 71% respectively. Of these hits the coverage of viral assignments was 14%, 8%, 10%,
263 31%, 54%, and 15%. Across the 2018 monthly metagenomic samples the proportion of
264 database viruses labelled as "unclassified" (i.e. any contigs without classification to the
265 level of family were consider "unclassified") accounts for over 50% of the entire
266 sample's hits (i.e. database hits) in all months and our combined assembly (Figure 1). Of
267 these three categories, unclassified members of order Picornavirales appear in three
268 months, and the *Marnaviridae* appear alongside them in two of these three months. These
269 are potential hits of viruses infecting microalgae in our samples. Other notable hits
270 include taxa that are known to infect fungi and oomycetes (family *Narnaviridae* [15];
271 family *Totiviridae* [75]), plants (family *Tombusviridae* [76]; family *Solemoviridae*), family
272 *Birnaviridae* infecting non-mammalian vertebrates and insects, most importantly
273 including salmon and chickens [77], a small family with species infecting freshwater
274 isopods (family *Yueviridae* [14]), and bacteriophages (family *Leviviridae* [78]). These
275 known hosts are not a complete picture of the potential viruses of these taxa however,

and recent evidence has suggested that other RNA virus families can infect microalgae, including the families *Narnaviridae*, *Totiviridae*, and *Yueviridae* that were found in the HRAP [79]. In each of these families there are distinct hits occurring in one month exclusively. Additionally, 2 hits matching family *Phycodnaviridae*, a dsDNA group of nucleocytoplasmic large DNA viruses (NCLDVs [80]) for which microalgae serve as natural hosts, are detected (these include hits encoding a putative RNA polymerase large subunit and an RNA ligase with polynucleotide kinase domain). These are likely abundant enough in the basin to contribute substantial mRNA (i.e. transcriptionally active during infection [81] or encapsidated in the virion) for being detected in RNA viral targeted samples. Finally, a hit to family *Circoviridae* (a single stranded (ss) DNA virus family [82]) is detected as well, with a low similarity (28%) to a putative capsid protein.

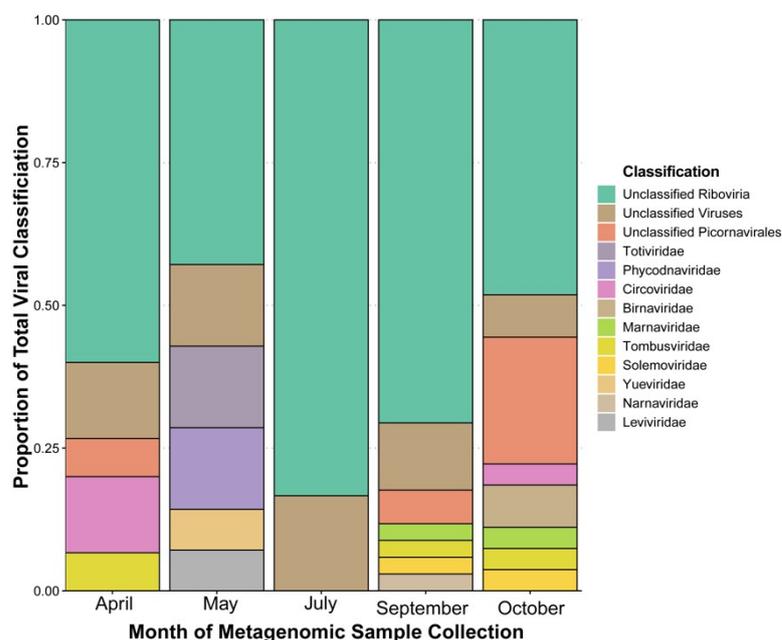


Figure 1. Proportion of contig viral classification (to family when possible) by each metagenomic sample month in 2018. Classifications are based on previously described database alignments by Blast. Only results of the Domain viruses are included. Total number of viral classifications by month were 15 for April, 14 for May, 18 for July, 34 for September, 27 for October.

When observing the proportion of reads for each classified contig (including cellular organisms and contigs without a classification (i.e. no match to a database; Figure 2)), we can attempt to infer the most relatively abundant classifications from a sequencing data perspective. In the month of April, a single contig with a *Circoviridae* hit recruits most (77%) of the reads, for May both unclassified Riboviria and “unclassified virus” hits recruit the majority of reads. For the months of July, September, and October, unclassified members of order Picornavirales dominate the read recruitments. This classification is based on only two putative full-length contigs (two viruses in theory) that recruits nearly half of the September reads each. Only one of these two contigs recruits reads in the July dataset (60%), whereas the other contig recruits most of the October reads (98%) (see Table S2).

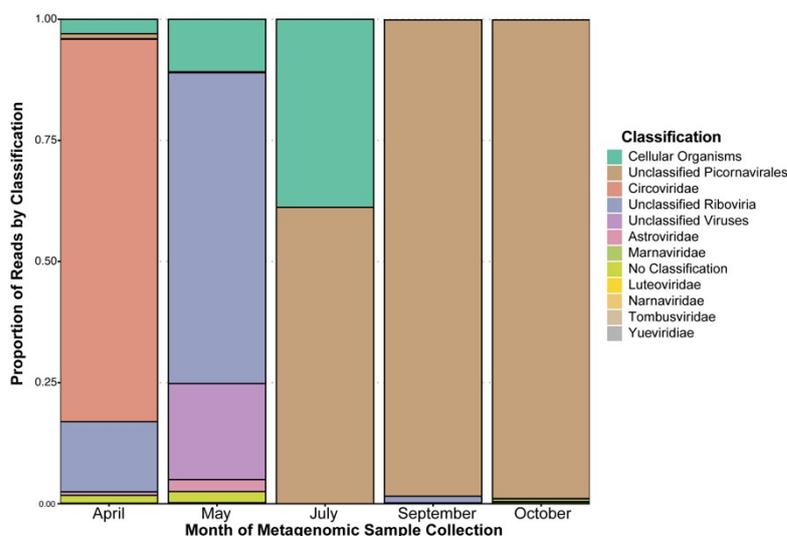


Figure 2. Proportion of normalised reads mapping to each classification by each metagenomic sampling month. Classifications were given to contigs assembled from all samples’ reads that accounted for at least 99% of all reads in each sample by month. These contigs included 26 viral best hits, 12 with no hit, and 496 with a cellular organism top hit. Note: due to the combining of all reads from each sample month for the assembled contigs presented in this figure additional classifications were permitted when compared to Figure 1.

Finally, our domain searching method applied to contigs over 6Kb uncovered several domains matching that of family *Marnaviridae* (see references genomes), including their relative order of appearance along the contig (Figure S2). Of six contigs that met the size requirement only four had adequate reference genomes available and included domains uncovered through InterProScan [55].

3.2. Alpha and beta diversity indices, and cross sample coverage

Four intra-community diversity (i.e. alpha diversity) indices were run on k-mer based data from our assembled contigs (Figure 3). Using k-mer “species” (i.e. each unique k-mer of *n* length is treated as a single species) instead of taxonomic species, therefore freeing the data from the limitations of database dependent classification where sequencing gaps in RNA classification exist and an abundance of viruses remain as “unclassified”. A consequence of this limitation is that diversity is only quantified if the virus is “known” through a database hit, whereas using k-mers species permits an interpretation of diversity that is free from information loss when there are no hits to a database for a contig. In theory, k-mer based alpha diversity uses reoccurring k-mers as an indication of shared homology among contigs, and thusly a potential shared taxonomy.

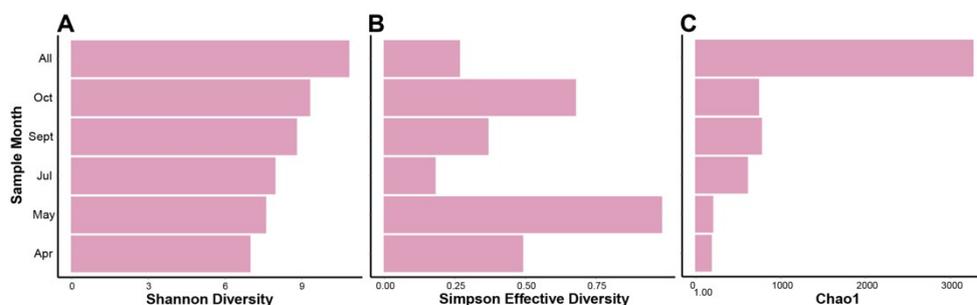
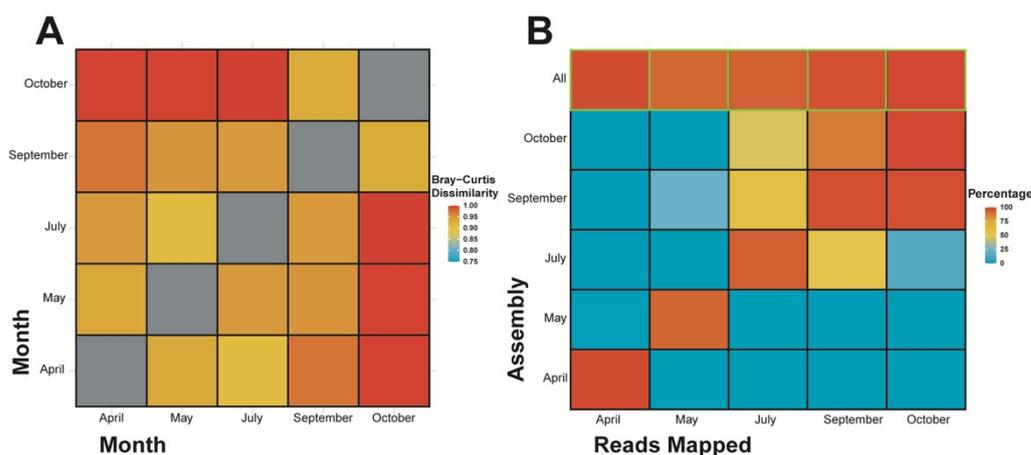


Figure 3. Alpha diversity indices of Illumina MiSeq metagenomic assemblies (contigs) as calculated by MerCat on k-mers of 21bp, with a minimum count of 5 for each “k-mer species”. Indices include (A) Shannon’s diversity index, (B) Simpson’s diversity index, and finally (C) the Chao1 k-mer measures the number of k-mer species qualifying based on minimum count parameter. Sample months are indicated. The sample denoted “All” is an assembly of all reads from each Illumina MiSeq dataset combined.

334 The Shannon diversity index (Fig. 3 A) indicates an increase in k-mer species
 335 diversity from April to October, where unsurprisingly the combined assembly from all
 336 samples' reads (denoted "All") gives the highest amount of diversity. As the microalgal
 337 growing season progresses, despite microalgal die-offs occurring, the diversity in the
 338 HRAP's viral fractions increases overtime based on k-mer species. This pattern is not
 339 reflected exactly in the Simpson diversity of k-mer species on the same samples,
 340 however it is increasing over a two-month period (April, May) and then again after a
 341 sudden decrease (July, September, October). Shannon and Simpson diversity indices are
 342 calculated in different ways (see [83]), therefore a more in-depth explanation of these
 343 results will occur in the discussion section. For the Chao1 number of estimated species
 344 (Fig. 3C) we see, as expected, the highest estimate of species is in the combined reads
 345 sample. Additionally, both April and May have a lower number of estimated species,
 346 whereas July, September, and October are all relatively higher in comparison to
 347 April/May that have similar values. Based on Chao1 there is an increase in the number
 348 of k-mer species from April/May to July/September/October months.

349 An inter-community species diversity (i.e. beta diversity) comparison using Bray-
 350 Curtis dissimilarity on k-mer species (Figure 4A) indicates that the distribution and
 351 composition of k-mer species between samples differs considerably (all values are over
 352 0.90), however October and September are more similar when compared against other
 353 samples, additionally October individually contains more differences compared to
 354 April/May/July than September to April/May/July. July is most similar to May,
 355 whereas May and April are most similar to each other. A second interspecies
 356 comparison was done using reads mapped to contigs (Fig. 4B), whereby the proportion
 357 of reads from each sample mapped to assembled contigs from each sample indicates the
 358 similarity or dissimilarity between samples. When considering all reads, and not just k-
 359 mer species (at a count of 5 or more) from contig assemblies as the Bray-Custis index
 360 did, October and September maintain strong similarities to one another, July has a
 361 stronger overlap with both October and September when compared with May and
 362 April, and finally May and April do not have strong similarity to one another. This last
 363 point being in contradiction to the k-mer based assessment (Fig. 4 A). The "all" contigs
 364 represent the assembly constructed from all samples' reads and simply shows that a
 365 high percentage of the reads from each sample were recruited in the assembly of the
 366 combined reads contigs.



367
 368 **Figure 4.** A comparison between samples done using (A) Bray-Curtis dissimilarity beta diversity of
 369 Illumina Miseq metagenomic assemblies (contigs) as calculated by Simka on k-mers of 21bp,
 370 and the (B) percentage of reads mapped from each metagenomic sample (indicated by month of
 371 sampling) to assemblies (i.e. contigs) made from their corresponding sample month. Additionally,
 372 an assembly from all reads combined (i.e. all metagenomic samples) was produced and reads from
 373 each individual month were mapped back to the said assembled contigs (denoted by "All" and
 374 highlighted in green).

375 3.3. Phylogenetic tree (RdRp) of order Picornavirales

By processing nearly 39000 sequences of published Picornavirales polyproteins and our assembled contigs we were able to form a fairly robust tree in terms of taxonomy (Figure 5). From our dataset, putative RdRp domains occur in families *Picornaviridae*, *Polycipiviridae*, *Soliniviridae*, *Secoviridae*, and *Marnaviridae*. Recovered RdRp domains from our data include 9 unique sequences clustering with *Marnaviridae*, providing further evidence that this group of RNA viruses infecting microalgae not only persists in the HRAP where cultured microalgae would serve as hosts, but also as multiple *Marnaviridae* species.

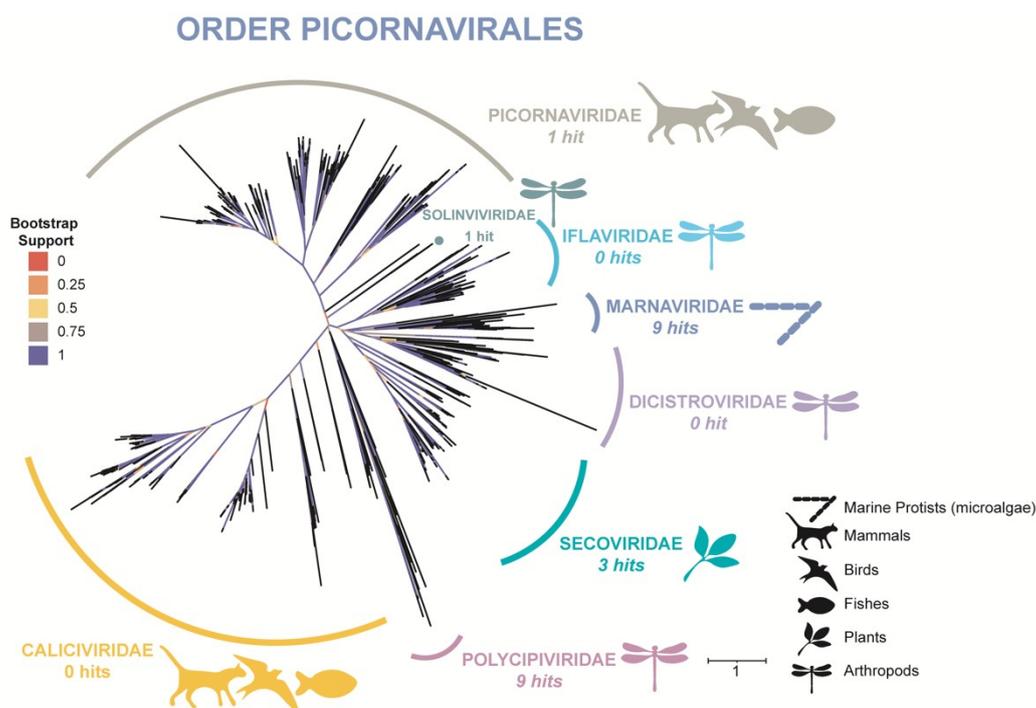


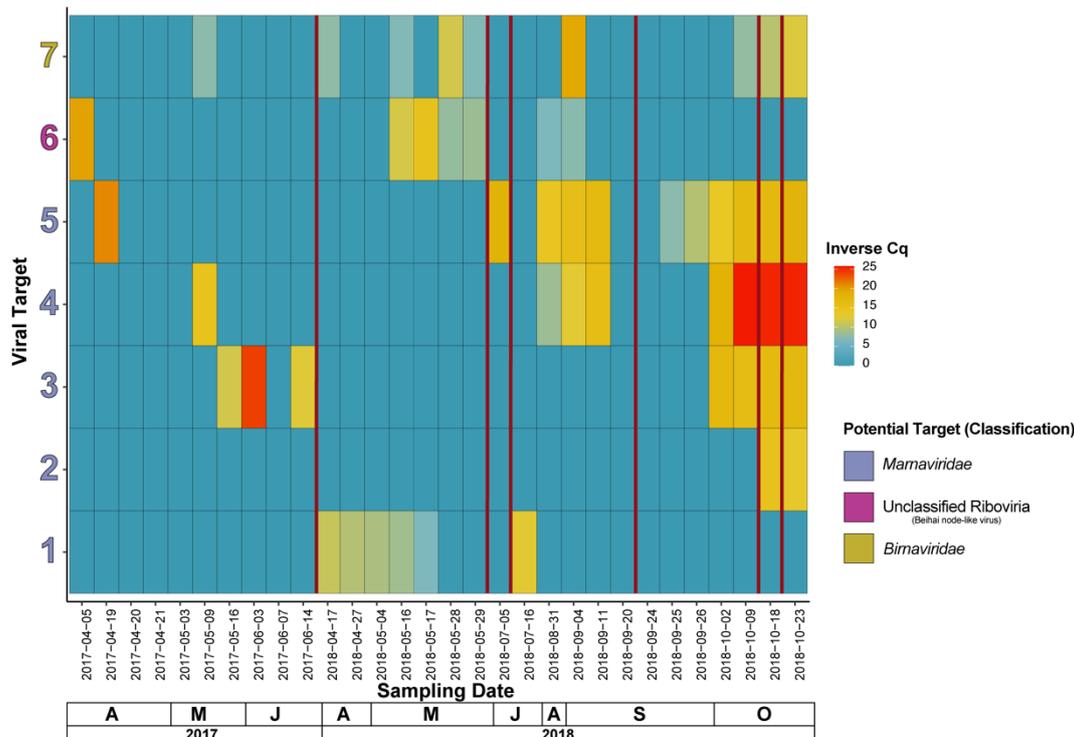
Figure 5. Maximum-likelihood phylogeny of RNA-dependent RNA polymerase (RdRp) of order Picornavirales based on RdRp sequences retrieved from metagenomic assembly results and the NCBI GenBank database. Bootstrap support is indicated by color, and number of RdRp hits extracted from our study are indicated alongside family classifications identified. Common hosts of the virus families are depicted. Newly sequenced RdRp sequences are included in Supplementary Note 2, with accession numbers.

3.4. qPCR of potential *Marnaviridae* and a *Rotifera* sp. virus

In total 7 RNA viruses of interest (see Table S3 for primer and putative viral target information), which were identified through taxonomic classification of contigs from the combined reads dataset, were tracked using an RT-qPCR technique. Viruses 1 to 5 in Figure 6 are potential members of *Marnaviridae*, with targets 1, 2, and 4 having a hit to a virus with an *Aurantiochytrium* sp. host, target 3 with a hit to a virus with *Cylindrotheca closterium* as a host, and target 5 with an unknown host. Additionally, target 6 has a top hit with a “Beihai noda-like virus”, and target 7 with a member of family *Birnaviridae* known to infect a member of phylum Rotifera and has been previously isolated before in the HRAP area in an unrelated study [84]. Two (targets 4 and 5) of the potential *Marnaviridae* virus targets share a similar pattern where infections appear to occur between the microalgae die-offs of 5 July 2018 and 20 September 2018, with persistence before, during, and after the die-offs associated with 22 October 2018. Target 2 appears only on 23 October 2018 and beyond, and finally target 1 exhibits a completely distinct pattern compared to the other potential *Marnaviridae* in which it sustains a relatively low presence during April to mid-May of 2018 and appears once thereafter. Interestingly, targets 3, 4, and 5 are all of potential *Marnaviridae*, have the most similar pattern (to a lesser extent target 3), and do not share the same host associated with their top hit. Target 6, the “Beihai noda-like virus” shows a sudden appearance and slow disappearance before the late May 2018 crash, and finally target 7, the likely *Rotifera* sp. virus comes and goes with low levels of amplification, appearing a couple of times with

411
412
413
414

a relatively higher number of copies, but disappears just as suddenly. Interestingly, there are instances in 2017 dates where targets (exclusively sequenced by metagenomics from 2018 samples) are amplifying in 2017 samples also, often with a Cq implying a relatively high copy number.



415

416

417

418

419

420

Figure 6. Occurrence and relative intra-species (i.e. viral population) abundance over a time series of 2017 and 2018 sampling dates, including metagenomic samples, using an RT-qPCR assay. An inverse of the quantitation cycle (Cq) is represented (45 cycles minus the Cq value), where a higher value indicates a higher relative abundance. The period between two samples where a microalgal culture die-off occurred in the basin are represented by vertical red lines (5 in total). Further information on RNA viral targets (1-7) is included in Table S3.

421

422

423

424

425

426

427

428

429

Specifically, to the putative *Marnaviridae* targets (1-5) we do see variation in viral dynamics. Indeed, targets 3 to 5 are present from late September to the end of our sampling in late October, but they are not necessarily present in the same periods (with the exception of targets 4 and 5 that are quite similar). As an example, target 3 appears in October of 2018, only, but is also present in May and June of 2017, although we do not have coverage of October 2017, we can at least see that the dynamics have not repeated exactly in 2018 as they were in 2017, and therefore are not annual in our study. Target 1 has a different pattern than other putative *Marnaviridae* viruses being primarily present only in April and May of 2018.

430

431

432

433

434

435

436

437

438

4. Discussion

Our study documents the RNA viral community of an HRAP during periods of microalgae growth over two years by (1) exploring the taxonomic diversity to infer potential hosts, (2) using the RdRp gene to uncover *Marnaviridae* through a phylogenetic study, and (3) measuring dynamics of select viral targets appearing in the system over time. This study offers a unique view of an HRAP RNA virome over time that has not yet been conducted before, and can offer new insights for future microalgal cultivation, specifically in an industrial context, and in cultures which are exposed to the environment (i.e. open systems).

439

440

4.1. Taxonomic classification and a phylogenetic study provide evidence of microalgal infecting RNA viruses within the HRAP

441 Classification of viruses, in general, using metagenomics is made difficult by
442 inconsistent coverage of taxonomic groups in databases causing what is known as “viral
443 dark matter” where viral sequences return no hits (i.e. no alignments) to major databases
444 hosting viral sequences [21]. In the case of RNA viruses, much is left unknown still and
445 there are calls for more culturing and more metagenomic sequencing to advance our
446 understanding and taxonomic classification system, despite the number of studies
447 conducted to date [9]. This is especially important in lieu of the ICTV’s [23,85] recent
448 acceptance of sequence-based classification of viruses, which has shown to be quite
449 accurate through a study aimed to reproduce current viral taxonomy using gene
450 signatures and genome organization [86]. In all but October of our metagenomic
451 samples, the majority of hits are to non-viral classifications (e.g. “cellular organism”)
452 despite the sample water being processed in the laboratory for targeting viral nucleic
453 acids. Our viral dark matter ranged from 18% to 68% of the total number of contigs, with
454 September harboring the highest number of contigs without matches to databases. When
455 considering all non-viral hits and dark matter (e.g. “cellular organisms” are non-viral
456 hits) also this range becomes 69% to 92%, which is not far from the ranges found in other
457 studies (40% to 90%; [21]). Given this, we must approach our conclusions regarding the
458 classified diversity in the HRAP with caution and understand it cannot reflect the true
459 and complete taxonomic diversity of these samples, and therefore the HRAP itself.

460 With the above points in mind, we can draw some conclusions from our taxonomic
461 analyses. Foremost, it is clear that unclassified viruses of Riboviria, and order
462 Picornavirales are important to this HRAP community throughout the culturing period
463 of 2018. Of course, in regard to Riboviria, it simply means that the viruses contain either
464 RdRp, or the RdDp (RNA dependent DNA polymerase) of retroviruses. Many of these
465 hits are to an invertebrate focused virus study [14], which includes hosts from phyla
466 Mollusca and Arthropoda, but remain unclassified or “picorna-like” viruses. These are
467 unsurprising considering the water source (i.e. Mediterranean Sea) would contain
468 Mollusca and Arthropoda species, and the open-faced nature of the HRAP in general.
469 Among these unclassified hits is also one to a *Sclerophthora macrospora* (Oomycete) virus,
470 which is a chimeric virus (RNA-DNA hybrid) found only in our October sample (e-
471 value = 1.9×10^{-16} , 57% coverage, 29% identity based on a putative RdRp sequence). With
472 these data we observe that the HRAP is hosting a wealth of viral diversity contained in
473 unclassified and understudied taxonomies. This includes the potential of interesting
474 chimeric viruses that are considered a rare event and with a plausible association to
475 microalgae as they are found in samples enriched for *Bathycoccus* [87]. Of course, within
476 the order Picornavirales is the family *Marnaviridae*, and with hits to members of this
477 microalgae-infecting family appearing in September and October metagenomes (2018)
478 alongside assembled RNA contigs from the HRAP resembling the domain inclusion and
479 domain order of known *Marnaviridae* and “marna-like” viruses (see Figure S2), we can
480 also speculate that some of these unclassified Picornavirales could be infecting the
481 microalgae of the HRAP as well. A recent study [79] of 570 transcriptomes from a wide
482 diversity of marine protists revealed divergent RNA viruses considered “marna-like”
483 due to being most closely related to *Marnaviridae* spp. yet being relatively divergent
484 from the few species of *Marnaviridae* in genomic databases, the authors conclude that
485 this is due to the diversity of *Marnaviridae* not yet being extensively covered and that
486 these “marna-like” viruses are in fact *Marnaviridae* spp. This lends more evidence to the
487 presence of RNA microalgal viruses in HRAP, which too had just been yet to be
488 uncovered, in other words the discovery of new *Marnaviridae* in a system culturing
489 microalgae is quite plausible. Furthermore, In the same study [79] the authors used RdRp
490 sequence and structure to identify novel microalgal RNA viruses and described them
491 phylogenetically. Several families of RNA viruses were newly considered to be capable
492 of infecting microalgae (based on closest relatives), and among them are the
493 aforementioned families *Narnaviridae*, *Totiviridae* and *Yueviridae*. *Narnaviridae* -like
494 species were found in samples of Bacillariophyta culture, and *Totiviridae* -like species
495 were inferred to infect Bacillariophyta, Haptophyta, *Chromeraceae*, *Dinophyceae*, and
496 Rhodophyta. Finally, a *Yueviridae* -like species was inferred to infect Bacillariophyta also.
497 Although we cannot infer host(s) through metagenomics alone, there are many RNA
498 viruses contained within the study samples that could be infecting microalgae. Overall

our taxonomic results reflect that of other RNA metagenomic studies and reviews in that members of order Picornvirales are common in aquatic settings [6, 9, 19].

By taking a normalized read mapping approach to investigate each of the 2018 metagenomic samples we can observe what viral contigs are recruiting a substantial number of reads, however, we should take into account sequencing biases produced by the process of sequencing. Although there are less studies on the subject of RNA virome sequencing biases, there is evidence of ssDNA viruses being recovered at higher proportions than dsDNA viruses [88,89] for example. Among RNA viromes there is evidence that sequencing preparation kit choice can change the profile of dominant species in wastewater samples and exclude some species that are identified in the alternative preparation kit [90]. More specific to RNA viruses that are generally A-T rich [91], this coupled with evidence that when using Nextera XT preparation kits the lower the G-C content in a region the stronger of a sequencing bias is produced [92], this is at least partially attributed to the secondary structure of RNA [93]. Given these studies, our inferences of dominant or “most important” viruses in this HRAP study have to be interpreted with caution. Nonetheless, based on our read mapping and classification it is clear the viral diversity contained in the HRAP is changing across the culturing timeframe of 2018. July, September, and October featured a shared dominant viral classification (unclassified Picornvirales) that accounts for between 60% and 97% of read recruitment. Picornvirales are common in metagenomic studies of water samples, for example they were in high abundance (up to a relative abundance of 90% or more) in samples collected from groundwater in Saudi Arabia (Wadi Fatimah reservoir) [94], were up to 97% of sequence matches in a coastal water sample in Canada (English Bay, British Columbia) [95], and ultimately are considered as “widespread in the world’s oceans” [96]. Given this, it is not surprising to find that Picornvirales are dominant in read recruitment for at least 3 of the 5 metagenomes (Fig 2). In regard to the *Circoviridae* hit accounting for 77% of read recruitment in April, lake metagenomic studies have previously identified *Circoviridae* hits in RNA viromes, but concluded it was a contamination based on the presence of *Circoviridae* also in the DNA viromes [97]. An alternative explanation is that these viruses could be in an intermediary step of infection (i.e. mRNA), therefore transcriptionally active [81], given that they are ssDNA viruses and not RNA viruses [82]. Additionally, family *Astroviridae* appearing in May, are known to infect both mammals and birds [98], therefore if the natural host of this specific virus was a bird it is quite possible it could have entered into the HRAP by infected birds in the area given that it is partially open to the environment. Interestingly, contigs without hits to known databases feature low level of read recruitment, despite the seemingly high level of viral dark matter. For this case our taxonomic hits profile and our read recruitment coverage studies tell conflicting stories, however, it seems plausible that sequencing and preparation bias could contribute to the problem of understudied and unclassified viruses, alongside other factors, thereby the most “unknown” viruses are experiencing the least read recruitment in our study. Despite the aforementioned caveat regarding sequencing and preparation biases, reporting presence of the viruses identified in this study is still valid and informative.

Finally, an RdRp tree of families within the order Picornvirales were chosen for our focus because one of the seven families of Picornvirales has ample evidence of infecting microalgae, the family *Marnaviridae*. These infections have been confirmed by isolation in laboratory cultures [99], and not only through *in silico* analyses. With consideration to our phylogenetic tree, based on a domain searching method, we have provided further evidence of *Marnaviridae* (and thusly microalgal infecting viruses) to occur in this basin. This domain searching method was both useful in extracting RdRp domains, and in identifying putative members of *Marnaviridae*.

4.2. Database-independent approaches to quantifying diversity show changes between and among temporal based samples of the HRAP

Due to the presence of viral dark matter in our study, we chose to use a k-mer based approach for measuring the alpha and beta diversity of our samples. Specifically, we chose k-mer species created from contigs and not reads to avoid library preparation and sequencing related biases, and sequencing errors contained in raw read data. Since these are k-mer species, and not taxonomic units, we cannot easily compare our results

557 with taxonomy based alpha diversity indices of other studies but can make observations
558 between and among our own samples. Shannon diversity is a measure of
559 entropy/disorder in a sample that is relatively sensitive to the number of rare species
560 present, whereas Simpson diversity has less bias in reference to the number of species
561 sampled (e.g. k-mer species) [83]. In other words, Simpson gives less weight to rare
562 species and more to common species, therefore more influenced by taxa evenness than
563 Shannon. Simplified, Simpson's index considers the probability that two species being
564 drawn randomly from a sample will be the same, where a higher number relates to a
565 low probability and therefore higher diversity [83]. In the context of our study, this
566 indicates that in May where the index is high, it is relatively unlikely you will draw two
567 of the same k-mer species, and thus the evenness among k-mers and the number of them
568 are more consistent and higher respectively than in July for example. Given that
569 Simpson diversity is less influenced by rare species it suggests that the number of rare or
570 relatively uncommon k-mer species changes with time and among runs in the HRAP,
571 but when less weight is placed on said species the diversity (Simpson) changes more
572 drastically from month to month. Overall there was an increases in number of different
573 k-mer species and "disorder" (i.e. Shannon) and probability of "drawing" two of the
574 same species of k-mers (i.e. Simpson), but when giving more weight to evenness (i.e.
575 Simpson) we also see "spikes" in diversity in two different seasons (May; spring,
576 October; autumn) where the increase is less gradual. These spikes, and the gradual
577 change, do not appear tightly coordinated with die-offs in the HRAP. Overall our alpha
578 diversity measures indicate that the HRAP diversity is changing throughout the
579 HRAP runs.

580 In the context of our beta diversity results, the diversity is different between the
581 samples, with the dissimilarity being especially apparent between October and other
582 months except September. There is a similar pattern when assessing similarities between
583 samples using reads mapping, however, July also recruits over ~50% to September and
584 October, whereas with beta diversity it's most similar (i.e. least dissimilar) to May. Based
585 on this study's RT-qPCR results (discussed more below), there is evidence of viruses
586 appearing and disappearing over time, so we cannot assume viruses appear according
587 to season/month alone, and furthermore we cannot exclude the fact that die-offs and
588 restarts, therefore many different runs, of the culture are occurring in the HRAP. In fact,
589 a study focusing on a subset of dsDNA viruses (NCLDVs) showed temporal variation
590 that did not follow a year to year seasonality in viral diversity [100], and another study
591 did provide evidence of strong seasonality in a subset of marine bacteriophages
592 (*Myoviridae*) that was linked to host seasonality [101]. Although both of these were DNA
593 virus based studies using metabarcoding, and ours is one of RNA and metagenomics, it
594 is likely we cannot lump all viruses into one pattern of dynamics based on taxonomy
595 and/or Baltimore classification and thusly can compare viral community dynamics
596 between DNA and RNA viruses carefully to some extent. Conclusively, viral dynamics
597 can come in many different patterns. As for the restarting of the cultures, these could
598 easily be "resetting" the diversity with each new run, considering that the water source
599 is from the Mediterranean Sea, and the HRAP is a small and contained body of water we
600 likely would see different dynamics when comparing the two environments. If the
601 HRAP takes on a different trajectory in terms of diversity and dynamics than its own
602 water source, it would not be surprising to see a sudden shift in community diversity
603 when the culture is restarted, and a new run occurs. This is a plausible hypothesis
604 because a variety of factors affecting viral populations would be changed from water
605 source to the HRAP including temperature, UV exposure, nutrient cycling, particulates,
606 grazing by predators [102], etc. The HRAP is a constrained environment, likely selecting
607 for different organisms in comparison to the Mediterranean Sea where the water is
608 sourced. Although, the full study is much more complicated by this, as evident by the
609 three restarts (i.e. four separate runs) occurring through the September and October
610 samples resulting in October being more diverse than September. Ultimately, it is
611 unsurprising our results based on proportion of reads mapped by classification and
612 reads similarity between samples provide similar conclusions as they are both based on
613 raw read data.

614 Results from these database independent approaches to studying diversity in
615 metagenomics have to be taken with caution. Using reads based approaches we are no
616 doubt affected by biases introduced before and during sequencing, and with k-mer

617 species counting we are using a relatively new method that is quite strict (k-mer species
618 are considered different if one out of 21 nucleotides are different) and in this case it may
619 be currently difficult to apply to metagenomes targeting taxa with high mutation rates –
620 such are the case for RNA viruses [103]. Logically, to improve this method an approach
621 with less strict parameters (e.g. permitting n number of nucleotides to be different) on
622 contig based k-mers could improve it, and also looking at protein-based data where k-
623 mers are created from amino acid sequences because substitutions could occur less
624 frequently at the amino acid level than at the nucleotide level. Additionally, an obstacle
625 of protein-based k-mer species is deciding what proteins to use and what not to use. For
626 example, should proteins included have a specified degree of certainty in their
627 prediction, should only taxonomically informative proteins (e.g. RdRp, capsid protein,
628 RdDp, etc.) be included exclusively? These adjustments to the process are not mutually
629 exclusive and conducting them in parallel might be necessary. These approaches
630 suggested above are not available as computational tools at the time of writing, and if
631 made available a variety of testing would need to occur for determining best practices,
632 which is outside the scope of this study.

633 In general, it is not surprising that the diversity of the HRAP overall (i.e. the
634 assembly using combined reads from all metagenomic samples) is higher than that of
635 each month alone due to the changes witnessed in the HRAP overtime, but it does
636 suggest that the diversity of the HRAP is better captured over time and not as a single
637 time point. Conclusively, we insist the RNA viral diversity using metagenomics is better
638 captured temporally in newly explored environments than as single time points.

639 4.3. Viral dynamics (RT-qPCR based) showcase a variety of different dynamics patterns among 640 putative *Marnaviridae* spp., and other viruses of interest

641 Our RT-qPCR methods in coordination with microalgae die-off dates has the
642 potential to inform us of what viruses could be contributing or be responsible for these
643 sudden die-offs. We used this technique to follow viruses of interest within the HRAP.
644 We primarily focused on the *Marnaviridae* because of their host relationship with
645 microalgae. However, the association between the appearance of these targets and the
646 die-offs is not clear. We are also limited by our sampling, where the die-offs in May and
647 July of 2018 lack the coverage that we gained by sampling done in September and
648 October 2018 around their specific die-offs. We cannot completely rule out the impact of
649 viruses on the system, however because our tracking was limited and other viruses (e.g.
650 other RNA viruses identified in this study, and DNA viruses not studied here) could
651 also be contributing. The effects of viral infection on microalgae may be that of several
652 populations and other viruses in unison. Regardless, there is evidence of viruses
653 potentially infecting microalgae appearing and disappearing temporally in the HRAP.
654 Interestingly, three of the potential *Marnaviridae* targets appear in the 2017 samples
655 (April, May, or June), however we do not see a gradual increase and decrease in the
656 relative population (inferred by RT-qPCR Cq), that appears in the 2018 dynamics. These
657 sporadic appearances could be amplification of similar strains or related viruses of the
658 2018 *Marnaviridae* targets, so-called “viral quasispecies”. Overall, these Picornavirale
659 dynamics results are not unlike results obtained in a study of two Picornavirale strains
660 among a three-lake system in America (Finger Lakes, New York), where the strains
661 exhibited different dynamics when compared to each other and also compared across
662 the three lakes at the same sampling times [104]. Members of order Picornavirales (and
663 picorna-like viruses) appear to exhibit a variety of dynamics patterns, and it is
664 unsurprising two sequential sampling years may permit different dynamics specifically
665 given the number of runs occurring in the HRAP and the timing of them in 2017 vs 2018.

666 Aside from the targets potentially infecting microalgae, the dynamics of the
667 Rotifera virus are quite interesting. The relatively common *Brachionus* sp. are known
668 predators in HRAP/microalgal culturing practices that are open faced and outdoors
669 [105] and are documented in the Mediterranean Sea along France [106]. Given that this
670 Rotifera virus strain has such a high certainty to be infecting Rotifera spp. (e-value = 0,
671 100% coverage, 99% identity), we hypothesize that Rotifera spp. are being reintroduced,
672 infected, and dying off or their populations is being dramatically changed by new HRAP
673 runs (after microalgal die-offs). We assume that the amplicon is a reflection of the
674 relative amount of the viral copies/viruses and therefore the Rotifera spp. host(s)

675 appearing in the basin, with this we might conclude a brief outbreak of Rotifera spp.
676 occurred in early September 2018, and reoccurred and persisted in October of 2018. We
677 have also assumed that this virus is a component in the regulation of the Rotifera spp.
678 likely appearing in the HRAP, but we cannot exclude the importance of factors like
679 intra-species competition and abiotic factors [100,107]. Although this virus was present
680 before and around some of the culture die-offs, and we suppose Rotifera spp. were
681 present alongside it, we cannot suggest if Rotifera spp. play a role in the die-offs in the
682 HRAP because tracking of this microalgal grazer is not reported in this study.

683 5. Conclusions

684 In this study we have identified several putative RNA viruses alongside evidence to
685 suggest that they could be infecting microalgae, specifically in the case of *Marnaviridae*,
686 which has been uncovered through putative taxonomic classification, a phylogenetic
687 study, and tracked in the HRAP using RT-qPCR. We cannot conclude that RNA viruses
688 are alone responsible for the die-offs experienced in the HRAP considering we did not
689 investigate other factors (e.g. Rotifera grazers mentioned above, algalicidal bacteria, other
690 viruses not targeted in our study laboratory methods), however we have provided
691 evidence of viruses likely infecting the microalgae being cultured in the HRAP. We have
692 provided an overview of the taxonomic diversity of viruses contained in the HRAP,
693 albeit limited by database completeness, alongside database independent approaches to
694 quantifying the basin diversity. Lastly, we glimpsed into the dynamics of some viruses
695 of interest.

696 In a broad sense, from our work we can conclude that future studies of RNA
697 viruses should aim to classify environmental viruses more thoroughly alongside efforts
698 to improve database-independent methods of studying diversity in viruses, specifically
699 in environments not previously sampled. Advancing viral taxonomy and classification is
700 on-going and will take time, thus improving our database independent approaches in
701 the near future could be quite informative for viral studies similar to this. Finally, we
702 have approached studying viral dynamics with a RT-qPCR approach, highlighting the
703 array of viral dynamics patterns that can occur temporally and concluding that
704 individual viruses can behave quite differently.

705 **Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1, Note
706 S1: python Primer3 script, Table S1: metagenomic sample assembly statistics, Table S2: reads from
707 each sample mapped to the metagenomic assembly (i.e. contigs) of merged sample reads, Table S3:
708 minimum information on the putative RT-qPCR targets and primers.

709 **Author Contributions:** Conceptualization, C.D., G.B., E.E.C., and S.B.; Methodology, C.D., G.B.,
710 S.B., E.E.C.; Software, E.E.C., F.H.A., and G.B.; Validation, formal analyses, and investigation, G.B.,
711 C.D. and E.E.C.; Resources, C.D. and G.B.; Data Curation, E.E.C. and G.B.; Writing – Original Draft
712 Preparation, E.E.C.; Writing – Review & Editing, G.B., C.D., S.B., F.H.A., A.G., and E.E.C.;
713 Visualization, E.E.C.; Supervision, C.D., G.B., and S.B.; Project Administration, C.D., and G.B.;
714 Funding Acquisition, G.B., and C.D. All authors have reviewed and approved this manuscript.

715 **Funding:** E.E.C, G.B, and C.D received funding from the European Union's Horizon 2020 research
716 and innovation programme under the Marie Skłodowska-Curie grant agreement No713750, with
717 the financial support of the Regional Council of Provence- Alpes-Côte d'Azur and with the
718 financial support of the A*MIDEX (n° ANR- 11-IDEX-0001-02), funded by the Investissements
719 d'Avenir project funded by the French Government, managed by the French National Research
720 Agency (ANR). The Phycovir project leading to this publication has received funding from
721 Excellence Initiative of Aix-Marseille University - A*MIDEX, a French "Investissements d'Avenir"
722 program. A.G. acknowledges support by the Institut Français de Recherche pour l'Exploitation de
723 la Mer (IFREMER).

724 **Institutional Review Board Statement:** Not applicable.

725 **Informed Consent Statement:** Not applicable.

726 **Data Availability Statement:** Data are available through and hosted by the NCBI SRA portal
727 under the BioProject I PRJNA751746, with submission ID SUB10083298.

728 **Acknowledgments:** The authors thank the members of the VASCO2 consortium and the company
729 COLDEP for their support and permission to access the cultures. Specifically, we would like to
730 thank Cyrille Przybyła, Sébastien Triplet, Thibault Geoffroy, Amandine Galès, and Maxime
731 Petitjean for their contributions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Starr, E.P.; Nuccio, E.E.; Pett-Ridge, J.; Banfield, J.F.; Firestone, M.K. Metatranscriptomic Reconstruction Reveals RNA Viruses with the Potential to Shape Carbon Cycling in Soil. *Proc Natl Acad Sci USA* 2019, 116, 25900–25908, doi:10.1073/pnas.1908291116.
2. Lang, A.S.; Rise, M.L.; Culley, A.I.; Steward, G.F. RNA Viruses in the Sea. *FEMS Microbiol Rev* 2009, 33, 295–323, doi:10.1111/j.1574-6976.2008.00132.x.
3. Suttle, C.A. Viruses in the Sea. *Nature* 2005, 437, 356–361, doi:10.1038/nature04160.
4. Lu, G.; Ye, Z.-X.; He, Y.-J.; Zhang, Y.; Wang, X.; Huang, H.-J.; Zhuo, J.-C.; Sun, Z.-T.; Yan, F.; Chen, J.-P.; et al. Discovery of Two Novel Negevirus in a Dungfly Collected from the Arctic. *Viruses* 2020, 12, 692, doi:10.3390/v12070692.
5. Miranda, J.A.; Culley, A.I.; Schvarcz, C.R.; Steward, G.F. RNA Viruses as Major Contributors to Antarctic Virioplankton: RNA Viruses in the Antarctic. *Environ Microbiol* 2016, 18, 3714–3727, doi:10.1111/1462-2920.13291.
6. Djikeng, A.; Kuzmickas, R.; Anderson, N.G.; Spiro, D.J. Metagenomic Analysis of RNA Viruses in a Fresh Water Lake. *PLOS ONE* 2009, 4, e7264, doi:10.1371/journal.pone.0007264.
7. Skvortsov, T.; de Leeuwe, C.; Quinn, J.P.; McGrath, J.W.; Allen, C.C.R.; McElarney, Y.; Watson, C.; Arkhipova, K.; Lavigne, R.; Kulakov, L.A. Metagenomic Characterisation of the Viral Community of Lough Neagh, the Largest Freshwater Lake in Ireland. *PLoS ONE* 2016, 11, e0150361, doi:10.1371/journal.pone.0150361.
8. Steward, G.F.; Culley, A.I.; Mueller, J.A.; Wood-Charlson, E.M.; Belcaid, M.; Poisson, G. Are We Missing Half of the Viruses in the Ocean? *The ISME Journal* 2013, 7, 672–679, doi:10.1038/ismej.2012.121.
9. Greninger, A.L. A Decade of RNA Virus Metagenomics Is (Not) Enough. *Virus Research* 2018, 244, 218–229, doi:10.1016/j.virusres.2017.10.014.
10. Elena, S.F.; Bedhomme, S.; Carrasco, P.; Cuevas, J.M.; de la Iglesia, F.; Lafforgue, G.; Lalić, J.; Pròsper, À.; Tromas, N.; Zwart, M.P. The Evolutionary Genetics of Emerging Plant RNA Viruses. *MPMI* 2011, 24, 287–293, doi:10.1094/MPMI-09-10-0214.
11. Shi, M.; Lin, X.-D.; Chen, X.; Tian, J.-H.; Chen, L.-J.; Li, K.; Wang, W.; Eden, J.-S.; Shen, J.-J.; Liu, L.; et al. The Evolutionary History of Vertebrate RNA Viruses. *Nature* 2018, 556, 197–202, doi:10.1038/s41586-018-0012-7.
12. Zhang, Y.-Z.; Wu, W.-C.; Shi, M.; Holmes, E.C. The Diversity, Evolution and Origins of Vertebrate RNA Viruses. *Curr Opin Virol* 2018, 31, 9–16, doi:10.1016/j.coviro.2018.07.017.
13. Ryabov, E.V. Invertebrate RNA Virus Diversity from a Taxonomic Point of View. *J Invertebr Pathol* 2017, 147, 37–50, doi:10.1016/j.jip.2016.10.002.
14. Shi, M.; Lin, X.-D.; Tian, J.-H.; Chen, L.-J.; Chen, X.; Li, C.-X.; Qin, X.-C.; Li, J.; Cao, J.-P.; Eden, J.-S.; et al. Redefining the Invertebrate RNA Virosphere. *Nature* 2016, 540, 539–543, doi:10.1038/nature20167.
15. Hillman, B.I.; Cai, G. The Family Narnaviridae: Simplest of RNA Viruses. *Advances in Virus Research* 2013, 86, 28.
16. Callanan, J.; Stockdale, S.R.; Shkoporov, A.; Draper, L.A.; Ross, R.P.; Hill, C. Expansion of Known SsRNA Phage Genomes: From Tens to over a Thousand. *Science Advances* 2020, 6, eaay5981, doi:10.1126/sciadv.aay5981.
17. Krupovic, M.; Cvirkaite-Krupovic, V.; Iranzo, J.; Prangishvili, D.; Koonin, E.V. Viruses of Archaea: Structural, Functional, Environmental and Evolutionary Genomics. *Virus Res* 2018, 244, 181–193, doi:10.1016/j.virusres.2017.11.025.
18. Bolduc, B.; Wirth, J.F.; Mazurie, A.; Young, M.J. Viral Assemblage Composition in Yellowstone Acidic Hot Springs Assessed by Network Analysis. *The ISME Journal* 2015, 9, 2162–2177, doi:10.1038/ismej.2015.28.
19. Sadeghi, M.; Tomaru, Y.; Ahola, T. RNA Viruses in Aquatic Unicellular Eukaryotes. *Viruses* 2021, 13, 362, doi:10.3390/v13030362.
20. Dolja, V.V.; Koonin, E.V. Metagenomics Reshapes the Concepts of RNA Virus Evolution by Revealing Extensive Horizontal Virus Transfer. *Virus Res* 2018, 244, 36–52, doi:10.1016/j.virusres.2017.10.020.
21. Krishnamurthy, S.; Wang, D. Origins and Challenges of Viral Dark Matter. *Virus Research* 2017, 239, doi:10.1016/j.virusres.2017.02.002.
22. Koonin, E.V.; Wolf, Y.I.; Nagasaki, K.; Dolja, V.V. The Big Bang of Picorna-like Virus Evolution Antedates the Radiation of Eukaryotic Supergroups. *Nature Reviews Microbiology* 2008, 6, 925–939, doi:10.1038/nrmicro2030.
23. Lefkowitz, E.J.; Dempsey, D.M.; Hendrickson, R.C.; Orton, R.J.; Siddell, S.G.; Smith, D.B. Virus Taxonomy: The Database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research* 2018, 46, D708–D717, doi:10.1093/nar/gkx932.
24. Le Gall, O.; Christian, P.; Fauquet, C.M.; King, A.M.Q.; Knowles, N.J.; Nakashima, N.; Stanway, G.; Gorbalenya, A.E. Picornavirales, a Proposed Order of Positive-Sense Single-Stranded RNA Viruses with a Pseudo-T = 3 Virion Architecture. *Arch Virol* 2008, 153, 715–727, doi:10.1007/s00705-008-0041-x.
25. Tai, V.; Lawrence, J.E.; Lang, A.S.; Chan, A.M.; Culley, A.I.; Suttle, C.A. Characterization of HARNAV, A Single-Stranded RNA Virus Causing Lysis of *Heterosigma akashiwo* (Raphidophyceae). *Journal of Phycology* 2003, 39, 343–352, doi:10.1046/j.1529-8817.2003.01162.x.
26. Nagasaki, K.; Tomaru, Y.; Katanozaka, N.; Shirai, Y.; Nishida, K.; Itakura, S.; Yamaguchi, M. Isolation and Characterization of a Novel Single-Stranded RNA Virus Infecting the Bloom-Forming Diatom *Rhizosolenia Setigera*. *Applied and Environmental Microbiology* 2004, 70, 704–711, doi:10.1128/AEM.70.2.704-711.2004.
27. Shirai, Y.; Tomaru, Y.; Takao, Y.; Suzuki, H.; Nagumo, T.; Nagasaki, K. Isolation and Characterization of a Single-Stranded RNA Virus Infecting the Marine Planktonic Diatom *Chaetoceros tenuissimus* Meunier. *Appl. Environ. Microbiol.* 2008, 74, 4022–4027, doi:10.1128/AEM.00509-08.
28. Tomaru, Y.; Takao, Y.; Suzuki, H.; Nagumo, T.; Nagasaki, K. Isolation and Characterization of a Single-Stranded RNA Virus Infecting the Bloom-Forming Diatom *Chaetoceros Socialis*. *Applied and Environmental Microbiology* 2009, 75, 2375–2381, doi:10.1128/AEM.02580-08.

- 796 29. Tomaru, Y.; Nagasaki, K. Diatom Viruses. In *The Diatom World*; Seckbach, J., Kociolek, P., Eds.; Cellular Origin, Life in
797 Extreme Habitats and Astrobiology; Springer Netherlands: Dordrecht, 2011; pp. 211–225 ISBN 978-94-007-1327-7.
- 798 30. Alexandersen, S.; Knowles, N.J.; Belsham, G.J.; Dekker, A.; Nfon, C.; Zhang, Z.; Koenen, F. Picornaviruses. In *Diseases of*
799 *Swine*; John Wiley & Sons, Ltd, 2019; pp. 641–684 ISBN 978-1-119-35092-7.
- 800 31. Morocho-Jácome, A.L.; Ruscinc, N.; Martinez, R.M.; de Carvalho, J.C.M.; Santos de Almeida, T.; Rosado, C.; Costa, J.G.;
801 Velasco, M.V.R.; Baby, A.R. (Bio)Technological Aspects of Microalgae Pigments for Cosmetics. *Appl Microbiol Biotechnol*
802 2020, 104, 9513–9522, doi:10.1007/s00253-020-10936-x.
- 803 32. Barkia, I.; Saari, N.; Manning, S.R. Microalgae for High-Value Products Towards Human Health and Nutrition. *Marine Drugs*
804 2019, 17, 304, doi:10.3390/md17050304.
- 805 33. Barolo, L.; Abbriano, R.M.; Commault, A.S.; George, J.; Kahlke, T.; Fabris, M.; Padula, M.P.; Lopez, A.; Ralph, P.J.; Pernice, M.
806 Perspectives for Glyco-Engineering of Recombinant Biopharmaceuticals from Microalgae. *Cells* 2020, 9, 633,
807 doi:10.3390/cells9030633.
- 808 34. Qari, H.; Rehan, M.; Nizami, A.-S. Key Issues in Microalgae Biofuels: A Short Review. *Energy Procedia* 2017, 142, 898–903,
809 doi:10.1016/j.egypro.2017.12.144.
- 810 35. Kamyab, H.; Chelliapan, S.; Kumar, A.; Rezania, S.; Talaiekhosani, A.; Khademi, T.; Rupani, P.F.; Sharma, S. Microalgal
811 Biotechnology Application Towards Environmental Sustainability. In *Application of Microalgae in Wastewater Treatment:*
812 *Volume 2: Biorefinery Approaches of Wastewater Treatment*; Gupta, S.K., Bux, F., Eds.; Springer International Publishing:
813 Cham, 2019; pp. 445–465 ISBN 978-3-030-13909-4.
- 814 36. Chew, K.W.; Yap, J.Y.; Show, P.L.; Suan, N.H.; Juan, J.C.; Ling, T.C.; Lee, D.-J.; Chang, J.-S. Microalgae Biorefinery: High Value
815 Products Perspectives. *Bioresource Technology* 2017, 229, 53–62, doi:10.1016/j.biortech.2017.01.006.
- 816 37. Vuppaladadiyam, A.K.; Prinsen, P.; Raheem, A.; Luque, R.; Zhao, M. Sustainability Analysis of Microalgae Production
817 Systems: A Review on Resource with Unexploited High-Value Reserves. *Environ. Sci. Technol.* 2018, 52, 14031–14049,
818 doi:10.1021/acs.est.8b02876.
- 819 38. McGlathery, K.J.; Sundbäck, K.; Anderson, I.C. The Importance of Primary Producers For Benthic Nitrogen And Phosphorus
820 Cycling. In *Estuarine Nutrient Cycling: The Influence of Primary Producers: The Fate of Nutrients and Biomass*; Nielsen, S.L.,
821 Banta, G.T., Pedersen, M.F., Eds.; Aquatic Ecology Book Series; Springer Netherlands: Dordrecht, 2004; pp. 231–261 ISBN 978-
822 1-4020-3021-5.
- 823 39. Singh, U.B.; Ahluwalia, A.S. Microalgae: A Promising Tool for Carbon Sequestration. *Mitig Adapt Strateg Glob Change* 2013,
824 18, 73–95, doi:10.1007/s11027-012-9393-3.
- 825 40. Banerjee, I.; Dutta, S.; Pohrmen, C.B.; Verma, R.; Singh, D. Microalgae-Based Carbon Sequestration to Mitigate Climate
826 Change and Application of Nanomaterials in Algal Biorefinery. 2020, 8.
- 827 41. Wang, H.; Zhang, W.; Chen, L.; Wang, J.; Liu, T. The Contamination and Control of Biological Pollutants in Mass Cultivation
828 of Microalgae. *Bioresource Technology* 2013, 128, 745–750, doi:10.1016/j.biortech.2012.10.158.
- 829 42. Suttle, C.A. Marine Viruses — Major Players in the Global Ecosystem. *Nature Reviews Microbiology* 2007, 5, 801–812,
830 doi:10.1038/nrmicro1750.
- 831 43. Galès, A.; Triplet, S.; Geoffroy, T.; Roques, C.; Carré, C.; Le Floch, E.; Lanfranchi, M.; Simier, M.; Roque d’Orbcastel, E.;
832 Przybyla, C.; et al. Control of the PH for Marine Microalgae Polycultures: A Key Point for CO2 Fixation Improvement in
833 Intensive Cultures. *Journal of CO2 Utilization* 2020, 38, 187–193, doi:10.1016/j.jcou.2020.01.019.
- 834 44. Froussard, P. A Random-PCR Method (RPCR) to Construct Whole CDNA Library from Low Amounts of RNA. *Nucleic Acids*
835 *Res* 1992, 20, 2900.
- 836 45. Monteil-Bouchard, S.; Temmam, S.; Desnues, C. Protocol for Generating Infectious RNA Viromes from Complex Biological
837 Samples. In *The Human Virome: Methods and Protocols*; Moya, A., Pérez Brocal, V., Eds.; Methods in Molecular Biology;
838 Springer: New York, NY, 2018; pp. 25–36 ISBN 978-1-4939-8682-8.
- 839 46. Brown, J.; Pirrung, M.; McCue, L.A. FQC Dashboard: Integrates FastQC Results into a Web-Based, Interactive, and Extensible
840 FASTQ Quality Control Tool. *Bioinformatics* 2017, 33, 3137–3139, doi:10.1093/bioinformatics/btx373.
- 841 47. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 2014, 30,
842 2114–2120, doi:10.1093/bioinformatics/btu170.
- 843 48. Antipov, D.; Raiko, M.; Lapidus, A.; Pevzner, P.A. MetaviralSPAdes: Assembly of Viruses from Metagenomic Data.
844 *Bioinformatics* 2020, 36, 4126–4129, doi:10.1093/bioinformatics/btaa490.
- 845 49. Mikheenko, A.; Saveliev, V.; Gurevich, A. MetaQUAST: Evaluation of Metagenome Assemblies. *Bioinformatics* 2016, 32,
846 1088–1090, doi:10.1093/bioinformatics/btv697.
- 847 50. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-Based Genome Alignment and Genotyping with HISAT2 and
848 HISAT-Genotype. *Nature Biotechnology* 2019, 37, 907–915, doi:10.1038/s41587-019-0201-4.
- 849 51. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J Mol Biol* 1990, 215, 403–
850 410, doi:10.1016/S0022-2836(05)80360-2.
- 851 52. Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets.
852 *Nature Biotechnology* 2017, 35, 1026–1028, doi:10.1038/nbt.3988.
- 853 53. NCBI Resource Coordinators Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res*
854 2018, 46, D8–D13, doi:10.1093/nar/gkx1095.
- 855 54. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 2000, 16, 276–
856 277.
- 857 55. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al.
858 InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* 2014, 30, 1236–1240,
859 doi:10.1093/bioinformatics/btu031.

- 860 56. White III, R.A.; Panyala, A.; Glass, K.; Colby, S.; Glaesemann, K.R.; Jansson, C.; Jansson, J.K. MerCat: A Versatile k-Mer
861 Counter and Diversity Estimator for Database-Independent Property Analysis Obtained from Metagenomic and/or
862 Metatranscriptomic Sequencing Data. 2017, doi:10.7287/peerj.preprints.2825v1.
- 863 57. Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* 1948, 27, 379–423,
864 doi:10.1002/j.1538-7305.1948.tb01338.x.
- 865 58. Simpson, E.H. Measurement of Diversity. *Nature* 1949, 163, 688–688, doi:10.1038/163688a0.
- 866 59. Bray, J.R.; Curtis, J.T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* 1957,
867 27, 325–349, doi:https://doi.org/10.2307/1942268.
- 868 60. Benoit, G. Simka: Fast Kmer-Based Method for Estimating the Similarity between Numerous Metagenomic Datasets.
869 Available online: <https://hal.inria.fr/hal-01231795>.
- 870 61. Kans, J. Entrez Direct: E-Utilities on the Unix Command Line; National Center for Biotechnology Information (US), 2020;
- 871 62. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data.
872 *Bioinformatics* 2012, 28, 3150–3152, doi:10.1093/bioinformatics/bts565.
- 873 63. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER Web Server: 2018 Update. *Nucleic Acids Res* 2018,
874 46, W200–W204, doi:10.1093/nar/gky448.
- 875 64. Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.;
876 Sonnhammer, E.L.L.; et al. The Pfam Protein Families Database. *Nucleic Acids Res* 2004, 32, D138–D141,
877 doi:10.1093/nar/gkh121.
- 878 65. Rueckert, R.R.; Wimmer, E. Systematic Nomenclature of Picornavirus Proteins. *J Virol* 1984, 50, 957–959.
- 879 66. Vlok, M.; Lang, A.S.; Suttle, C.A. Application of a Sequence-Based Taxonomic Classification Method to Uncultured and
880 Unclassified Marine Single-Stranded RNA Viruses in the Order Picornavirales. *Virus Evolution* 2019, 5, vez056,
881 doi:10.1093/ve/vez056.
- 882 67. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and
883 Usability. *Mol Biol Evol* 2013, 30, 772–780, doi:10.1093/molbev/mst010.
- 884 68. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. TrimAl: A Tool for Automated Alignment Trimming in Large-Scale
885 Phylogenetic Analyses. *Bioinformatics* 2009, 25, 1972–1973, doi:10.1093/bioinformatics/btp348.
- 886 69. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New
887 Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 2020, 37,
888 1530–1534, doi:10.1093/molbev/msaa015.
- 889 70. Le, S.Q.; Gascuel, O. An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution* 2008, 25,
890 1307–1320, doi:10.1093/molbev/msn067.
- 891 71. Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Research* 2019,
892 47, W256–W259, doi:10.1093/nar/gkz239.
- 893 72. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New Capabilities and
894 Interfaces. *Nucleic Acids Res* 2012, 40, e115–e115, doi:10.1093/nar/gks596.
- 895 73. Marshall, O.J. PerlPrimer: Cross-Platform, Graphical Primer Design for Standard, Bisulphite and Real-Time PCR.
896 *Bioinformatics* 2004, 20, 2471–2472, doi:10.1093/bioinformatics/bth254.
- 897 74. Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res* 2003, 31, 3406–3415,
898 doi:10.1093/nar/gkg595.
- 899 75. Ghabrial, S.A.; Nibert, M.L. Victorivirus, a New Genus of Fungal Viruses in the Family Totiviridae. *Arch Virol* 2009, 154, 373–
900 379, doi:10.1007/s00705-008-0272-x.
- 901 76. Stuart, G.; Moffett, K.; Bozarth, R.F. A Whole Genome Perspective on the Phylogeny of the Plant Virus Family Tombusviridae.
902 *Arch Virol* 2004, 149, doi:10.1007/s00705-004-0298-7.
- 903 77. Delmas, B.; Attoui, H.; Ghosh, S.; Malik, Y.S.; Mundt, E.; Vakharia, V.N.; ICTV Report Consortium ICTV Virus Taxonomy
904 Profile: Birnaviridae. *Journal of General Virology* 2019, 100, 5–6, doi:10.1099/jgv.0.001185.
- 905 78. Bollback, J.P.; Huelsenbeck, J.P. Phylogeny, Genome Evolution, and Host Specificity of Single-Stranded RNA Bacteriophage
906 (Family Leviviridae). *J Mol Evol* 2001, 52, 117–128, doi:10.1007/s002390010140.
- 907 79. Charon, J.; Murray, S.; Holmes, E.C. Revealing RNA Virus Diversity and Evolution in Unicellular Algae Transcriptomes.
908 bioRxiv 2021, 2021.05.09.443335, doi:10.1101/2021.05.09.443335.
- 909 80. Wilson, W.H.; Van Etten, J.L.; Allen, M.J. The Phycodnaviridae: The Story of How Tiny Giants Rule the World. *Curr Top*
910 *Microbiol Immunol* 2009, 328, 1–42.
- 911 81. Zeigler Allen, L.; McCrow, J.P.; Ininbergs, K.; Dupont, C.L.; Badger, J.H.; Hoffman, J.M.; Ekman, M.; Allen, A.E.; Bergman, B.;
912 Venter, J.C. The Baltic Sea Virome: Diversity and Transcriptional Activity of DNA and RNA Viruses. *mSystems* 2017, 2,
913 doi:10.1128/mSystems.00125-16.
- 914 82. Breitbart, M.; Delwart, E.; Rosario, K.; Segalés, J.; Varsani, A. ICTV Virus Taxonomy Profile: Circoviridae. *J Gen Virol* 2017, 98,
915 1997–1998, doi:10.1099/jgv.0.000871.
- 916 83. Daly, A.J.; Baetens, J.M.; De Baets, B. Ecological Diversity: Measuring the Unmeasurable. *Mathematics* 2018, 6, 119,
917 doi:10.3390/math6070119.
- 918 84. Gibrat, J.-F.; Mariadassou, M.; Boudinot, P.; Delmas, B. Analyses of the Radiation of Birnaviruses from Diverse Host Phyla
919 and of Their Evolutionary Affinities with Other Double-Stranded RNA and Positive Strand RNA Viruses Using Robust
920 Structure-Based Multiple Sequence Alignments and Advanced Phylogenetic Methods. *BMC Evolutionary Biology* 2013, 13,
921 154, doi:10.1186/1471-2148-13-154.
- 922 85. Simmonds, P.; Adams, M.J.; Benkó, M.; Breitbart, M.; Brister, J.R.; Carstens, E.B.; Davison, A.J.; Delwart, E.; Gorbalenya, A.E.;
923 Harrach, B.; et al. Virus Taxonomy in the Age of Metagenomics. *Nature Reviews Microbiology* 2017, 15, 161–168,
924 doi:10.1038/nrmicro.2016.177.

- 925 86. Aiewsakun, P.; Simmonds, P. The Genomic Underpinnings of Eukaryotic Virus Taxonomy: Creating a Sequence-Based
926 Framework for Family-Level Virus Classification. *Microbiome* 2018, 6, 38, doi:10.1186/s40168-018-0422-7.
- 927 87. Roux, S.; Enault, F.; Bronner, G.; Vaulot, D.; Forterre, P.; Krupovic, M. Chimeric Viruses Blur the Borders between the Major
928 Groups of Eukaryotic Single-Stranded DNA Viruses. *Nat Commun* 2013, 4, 2700, doi:10.1038/ncomms3700.
- 929 88. Parras-Moltó, M.; Rodríguez-Galet, A.; Suárez-Rodríguez, P.; López-Bueno, A. Evaluation of Bias Induced by Viral
930 Enrichment and Random Amplification Protocols in Metagenomic Surveys of Saliva DNA Viruses. *Microbiome* 2018, 6, 119,
931 doi:10.1186/s40168-018-0507-3.
- 932 89. Brinkman, N.E.; Villegas, E.N.; Garland, J.L.; Keely, S.P. Reducing Inherent Biases Introduced during DNA Viral Metagenome
933 Analyses of Municipal Wastewater. *PLOS ONE* 2018, 13, e0195350, doi:10.1371/journal.pone.0195350.
- 934 90. Pérez-Cataluña, A.; Cuevas-Ferrando, E.; Randazzo, W.; Sánchez, G. Bias of Library Preparation for Virome Characterization
935 in Untreated and Treated Wastewaters. *Science of The Total Environment* 2021, 767, 144589,
936 doi:10.1016/j.scitotenv.2020.144589.
- 937 91. Kustin, T.; Stern, A. Biased Mutation and Selection in RNA Viruses. *Molecular Biology and Evolution* 2021, 38, 575–588,
938 doi:10.1093/molbev/msaa247.
- 939 92. Sato, M.P.; Ogura, Y.; Nakamura, K.; Nishida, R.; Gotoh, Y.; Hayashi, M.; Hisatsune, J.; Sugai, M.; Takehiko, I.; Hayashi, T.
940 Comparison of the Sequencing Bias of Currently Available Library Preparation Kits for Illumina Sequencing of Bacterial
941 Genomes and Metagenomes. *DNA Research* 2019, 26, 391–398, doi:10.1093/dnares/dsz017.
- 942 93. Price, A.; Garhyan, J.; Gibas, C. The Impact of RNA Secondary Structure on Read Start Locations on the Illumina Sequencing
943 Platform. *PLoS ONE* 2017, 12, e0173023, doi:10.1371/journal.pone.0173023.
- 944 94. Zaouri, N.; Jumart, M.R.; Cheema, T.; Hong, P.-Y. Metagenomics-Based Evaluation of Groundwater Microbial Profiles in
945 Response to Treated Wastewater Discharge. *Environmental Research* 2020, 180, 108835, doi:10.1016/j.envres.2019.108835.
- 946 95. Culley, A.I. Metagenomic Analysis of Coastal RNA Virus Communities. *Science* 2006, 312, 1795–1798,
947 doi:10.1126/science.1127404.
- 948 96. Vlok, M.; Lang, A.S.; Suttle, C.A. Marine RNA Virus Quasispecies Are Distributed throughout the Oceans. *mSphere* 2019, 4,
949 doi:10.1128/mSphereDirect.00157-19.
- 950 97. López-Bueno, A.; Rastrojo, A.; Peiró, R.; Arenas, M.; Alcamí, A. Ecological Connectivity Shapes Quasispecies Structure of
951 RNA Viruses in an Antarctic Lake. *Mol Ecol* 2015, 24, 4812–4825, doi:10.1111/mec.13321.
- 952 98. Fernández-Correa, I.; Truchado, D.A.; Gomez-Lucia, E.; Doménech, A.; Pérez-Tris, J.; Schmidt-Chanasit, J.; Cadar, D.; Benítez,
953 L. A Novel Group of Avian Astroviruses from Neotropical Passerine Birds Broaden the Diversity and Host Range of
954 Astroviridae. *Scientific Reports* 2019, 9, 9513, doi:10.1038/s41598-019-45889-3.
- 955 99. Salomon, P.S.; Imai, I. Pathogens of Harmful Microalgae. In *Ecology of Harmful Algae*; Granéli, E., Turner, J.T., Eds.;
956 Ecological Studies; Springer Berlin Heidelberg, 2006; Vol. 189, pp. 271–282 ISBN 978-3-540-32209-2.
- 957 100. Negreiros, N.F.; Santos-Wisniewski, M.J. dos; Santos, R.M. dos; Rocha, O. The Influence of Environmental Factors on the
958 Seasonal Dynamics and Composition of Rotifera in the Sapucaí River Arm of Furnas Reservoir, MG, Brazil. *Biota Neotrop.*
959 2010, 10, 173–182, doi:10.1590/S1676-06032010000400023.
- 960 101. Pagarete, A.; Chow, C.-E.T.; Johannessen, T.; Fuhrman, J.A.; Thingstad, T.F.; Sandaa, R.A. Strong Seasonality and Interannual
961 Recurrence in Marine Myovirus Communities. *Appl Environ Microbiol* 2013, 79, 6253–6259, doi:10.1128/AEM.01075-13.
- 962 102. Mojica, K.D.A.; Brussaard, C.P.D. Factors Affecting Virus Dynamics and Microbial Host–Virus Interactions in Marine
963 Environments. *FEMS Microbiol Ecol* 2014, 89, 495–515, doi:10.1111/1574-6941.12343.
- 964 103. Duffy, S. Why Are RNA Virus Mutation Rates so Damn High? *PLoS Biol* 2018, 16, e3000003, doi:10.1371/journal.pbio.3000003.
- 965 104. Hewson, I.; Bistolas, K.S.I.; Button, J.B.; Jackson, E.W. Occurrence and Seasonal Dynamics of RNA Viral Genotypes in Three
966 Contrasting Temperate Lakes. *PLoS One* 2018, 13, e0194419, doi:10.1371/journal.pone.0194419.
- 967 105. Abou-Shanab, R.A.I.; Singh, M.; Rivera-Cruz, A.; Power, G.; Bagby-Moon, T.; Das, K. Effect of *Brachionus Rubens* on the
968 Growth Characteristics of Various Species of Microalgae. *Electronic Journal of Biotechnology* 2016, 22, 68–74,
969 doi:10.1016/j.ejbt.2016.06.005.
- 970 106. Delpy, F.; Pagano, M.; Blanchot, J.; Carlotti, F.; Thibault-Botha, D. Man-Induced Hydrological Changes, Metazooplankton
971 Communities and Invasive Species in the Berre Lagoon (Mediterranean Sea, France). *Marine Pollution Bulletin* 2012, 64, 1921–
972 1932, doi:10.1016/j.marpolbul.2012.06.020.
- 973 107. Conde-Porcuna, J.M. Relative Importance of Competition with *Daphnia* (Cladocera) and Nutrient Limitation on *Anuraeopsis*
974 (*Rotifera*) Population Dynamics in a Laboratory Study. *Freshwater Biology* 2000, 44, 423–430, doi:10.1046/j.1365-
975 2427.2000.00582.x.