Spatio-temporal variability of airborne viruses over sub-Antarctic South Georgia: Effects of sampling device and proximity of the ocean

Ritam Das[1], Lucie Malard[2,3], David A. Pearce[2,4], Janina Rahlff[1,5,6,#]

[1]Aero-Aquatic Virus Research Group, Faculty of Mathematics and Computer Science, Friedrich Schiller University Jena, Jena, Germany

[2]Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle-upon-Tyne, United Kingdom

[3]Department of Ecology and Evolution, University of Lausanne, Switzerland

[4]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, United Kingdom

[5]Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Jena, Germany

[6]Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Department of Biology and Environmental Science, Linnaeus University, Kalmar, Sweden

#corresponding author: Janina.rahlff@uol.de

Key words: phage, sea, marine, virome, aerosol, Coriolis

Abstract

Studying airborne viral diversity in pristine, remote environments like the sub-Antarctic island South Georgia provides crucial insights into viral ecology and their role in sustaining unique ecosystems. Viruses influence microbial dynamics and nutrient cycles, which are vital for ecological balance and long-term ecosystem sustainability. We explored the community composition of airborne viral operational taxonomic units (vOTUs) of two sites in South Georgia, using various sampling devices and viral metagenomics. The Coriolis μ device (wet collection) was the most effective, yielding 30 viral scaffolds. Two-thirds of the scaffolds were from sea-level samples, indicating that location impacts viral diversity. Protein-based clustering of 39 vOTUs revealed similarities of 15 with known marine viruses, suggesting oceanic influence on the island's airborne viral community. Genes related to UV damage protection and photosynthesis from two

airborne vOTUs were widely distributed in the major oceans, emphasizing the potential resilience in changing climates. Host predictions indicated associations with bacterial genera like *Rickettsia*, *Myroides*, and Bacteroidota. Some vOTUs matched viruses from extremophiles, indicating adaptations to harsh environments. This study provides a baseline for understanding the complexity and sustainability of airborne viral communities in remote ecosystems, underscoring the need for continued monitoring in the face of environmental change.

Sustainability Statement

Antarctic microbes have potentially unique adaptations to extreme cold and harsh conditions. Viruses that infect these microbes may play a pivotal role in shaping these adaptations and thus understanding virus-host interactions is vital for preserving the unique microbial biodiversity of Antarctica. Moreover, investigating the dynamics of viral aerosolization in and around Antarctica is important for understanding the global dispersal of viruses and their potential impacts on distant ecosystems. Monitoring microbial viruses in Antarctica is essential for tracking community changes over time. Since Antarctica is often considered one of the last pristine environments on Earth, understanding the role of microbial viruses is relevant for developing conservation strategies to protect this ecosystem from potential anthropogenic impacts, including unintentional introduction of foreign microorganisms and viruses. Our research aligns with Sustainable Development Goals Life below water (SDG14) and Life on land (SDG15).

Introduction

The cryosphere encompassing ice sheets, glaciers, and sea ice, covers approximately 20% of Earth's surface, representing climatically sensitive ecosystems [1]. One such isolated and pristine ecosystem is the southernmost and fifth-largest continent Antarctica, which is surrounded by the Southern Ocean' Circumpolar Current. With its low temperatures, enveloped by a thick ice sheet and featuring harsh environmental conditions, Antarctica plays a critical role in our planet's climate and ocean systems. Governed by the Antarctic Treaty System [2], which ensures the continent's preservation for peaceful research and environmental protection, Antarctica embodies international cooperation. The separation of Antarctica from the rest of the Earth almost 33 million years ago followed by a subsequent cooling, has led to the evolution of a diverse cold-adapted microbiota [3]. Human interventions in this pristine environment have caused the near extinction of

58 several native inhabitants of this continent (such as the Antarctic fur seals) and brought forward
59 alien species (such as springtails)[4] from the mainland [5]. In addition to man-made contributions,
60 transmission via aerial dispersal can be a major route for the invasion of new species [6]. Few studies
61 have investigated airborne bacteria in the Antarctic region [7, 8, 9, 10], and shown that Antarctica
62 indeed has a complex microbial diversity, which is primarily seeded from the atmosphere [3]. Yet,
63 there is a notable absence of research on airborne viruses [11], which is rather astonishing considering
64 that these biological entities are ubiquitous and reach global numbers of 10^{31} viral particles [12]. One
65 reason for the absence of viral investigations is that DNA yields from aerosols are usually low due
66 to the very limited biomass, which precludes metagenomic studies that would otherwise identify
67 viruses. Instead, 16S rRNA gene surveys (metabarcoding) involving a PCR step for amplification
68 are often applied [13, 14], which are sufficient if only prokaryotes are targeted. Viruses do not disperse
69 in the air as free particles; instead, they attach to soil dust or marine organic aggregates, forming
70 aerosols. The deposition rates of viruses in the atmosphere, which can be up to approximately 460
71 times higher than those of bacteria, are found to be positively correlated with organic aerosols [15].
72 This suggests that viruses may have longer residence times in the atmosphere, likely due to their
73 smaller size. Consequently, the extended time viruses spend in the atmosphere increases the
74 likelihood of their dispersal over long distances.
75 Sea spray aerosols (SSA), which are formed by the primary emission from the ocean, are the key
76 component by mass of marine aerosols and play a crucial role in the Earth's climate system [16]. For
77 example, a virus infecting *Emiliania huxleyi* can retain infectivity and induce coccolith shedding
78 and emission to the SSA, thereby acting as a regulator of the mass of marine aerosols [17]. Therefore,
79 through aerosol-mediated long-distance transport, viruses could constitute an ecological and
80 evolutionary driving force not only in marine ecosystems but also in adjoining ones. This could
81 occur through rain-washout events, where these aerosolized particles are removed from the
82 atmosphere through direct sedimentation, thereby allowing airborne viruses to interact with
83 microbes on the ground. This affects the abundance, composition, and distribution of microbes in
84 the recipient ecosystems [15].
85 In a study based on Antarctic surface snow, similar viral populations were found several hundred
86 to thousands of kilometres apart across Antarctica and overlaps between snow and marine viral
87 populations were revealed [18]. Flavobacteria from Antarctic snow were found to carry adaptive
88 immunity in the form of clustered regularly interspaced short palindromic repeats (CRISPR)-Cas

systems, indicating that defence mechanisms against mobile genetic elements are present [19]. Otherwise, investigations of viral diversity and lifestyle have been conducted on microbial viruses from Antarctica's lakes [20], sea ice [21, 22], under the ice shelf [23], soils [24], from cryoconite holes [25] and marine ecosystems including the Southern Ocean [26, 27]. Recent work demonstrated that viral diversity in the Southern Ocean around the Western Antarctic Peninsula is complex and uncovered both diversity and novelty among viruses. Alongside tailed dsDNA bacteriophages of the class Caudoviricetes, abundant eukaryotic viruses including polinton-like viruses, nucleocytoplasmic large DNA viruses (NCLDV), and virophages were found [28]. We propose that monitoring viruses in this and other isolated (but human-affected) places could help track changes in viral community over time, determine their stability and resilience and assess the impact of human activities on microbial communities. Moreover, the relative increase in global temperature and impact on glacial life forms in such extreme environments presents a unique opportunity to understand ecosystem function, which could help us shape a sustainable 'One World' for the future.

In this study, we used viral metagenomics to investigate the spatial-temporal variability of airborne viruses on a sub-Antarctic island, collected over seven days using different air samplers and a rain collector. We expected the surrounding ocean to be a major contributing source for atmospheric viruses in bioaerosols but also sought homologs of viral proteins from South Georgia in other oceans, to which viruses are probably dispersed or from which they are recruited.

Materials and methods

Field sampling and filtration

Air and rain samples were collected in South Georgia (latitude -54.4°, longitude -36.5°), which is ~2000 km from Antarctica, from October 15 to October 21 2021 at two different sampling sites: King Edwards Point (KEP) and Deadman's Cairn at the Lewis Pass (DMP). These sites represent two different sampling heights of 0 and ~200 m above sea level, respectively. Site KEP (54° 17' 02 'S, 36° 29' 75 'W) was influenced by the sea, whereas DMP was located at altitude (54° 16' 05 'S, 36° 30' 20 'W, Fig. 1f). Three different air samplers were used, namely the "wet" Coriolis ® μ air sampler, "dry" Coriolis Compact air sampler (both Bertin Technologies SAS, Montigny-le-Bretonneux, France), a Whatman Vacuum Pump attached to a Sartorius 250 ml filter holder collector ("pump"), and for rain sampling, a plastic funnel attached to a 1 L sterile collection

118 bottle ("collector"). At the DMP site, only the wet and dry Coriolis samplers could be used as they
119 function on batteries. The sampling time varied between 1 and 6 hours when using the Coriolis
120 instruments with high flow rates, whether in wet or dry conditions, and extended up to 22 hours
121 when air was collected using the pump. Rain collectors were used for 15 hours (Table S1). We
122 extracted hourly wind speed, wind direction, pressure and air temperature for the sampling period
123 (Table S1) from the Grytviken Automatic Weather Station. For each sample, the corresponding
124 values of each variable were averaged to cover each sampling duration, allowing us to obtain
125 average environmental conditions for each sample.

126 **DNA extraction, sequencing**

127 Cellulose nitrate membrane filters with air samples were sliced into 0.5 mm wide strips using a
128 sterilized scalpel to maximize the filters' exposure to the cell lysis buffer during DNA extraction.
129 For samples collected in nuclease-free water using the Coriolis wet device, the collection cones
130 were vortexed (SciQuip, Shropshire, UK) before being filtered through 0.2 µm pore size cellulose
131 nitrate membrane filters (GE Healthcare Life Sciences, Chicago, IL, USA) using a Welch WOBL
132 vacuum pump (Welch, Mt. Prospect, IL, USA) for 10 minutes. Samples obtained via the Coriolis
133 Compact sampler (the dry device) were resuspended in 15 mL of DNA/RNAase-free water,
134 manually shaken, and vortexed before undergoing the same filtration process. All pre-processing
135 was conducted in a Class II cabinet, and the vacuum pump, tubing, and filtration apparatus were
136 sterilized with 70% ethanol before use. DNA extraction from air samples was performed using the
137 Qiagen PowerSoil kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.
138 DNA extracts were submitted to NU-OMICS sequencing facility at Northumbria University,
139 Newcastle, for Illumina NextSeq shotgun metagenomics sequencing. DNA libraries were prepared
140 using the Nextera XT DNA Library Prep Kit, following the manufacturer's instructions.
141 Sequencing was conducted on an Illumina NextSeq system (Illumina Inc., USA) with V2.5 300-
142 cycle chemistry.

143 **Metagenomic analysis**

144 Sequencing reads went through adapter trimming using BBduk within BBTools [29], and afterwards
145 Sickle v.1.33 [30] was run in paired-end mode and '-t sanger' setting. Taxonomic profiling of
146 microbes was performed using the tool mOTUs v.3.0.2 on trimmed reads with options '-A' (reports

147 full taxonomy) '-c' (reports counts) '-M' (to save intermediate marker gene cluster count)
148 including a separate run to retrieve unassigned taxa. MOTUs employs universal, protein-coding,
149 and single-copy phylogenetic marker gene sequences [31, 32] The trimmed reads were subsequently
150 assembled using metaSPAdes v.3.15.5. [33] and metaviral SPAdes [34]. Assemblies were combined
151 and viral scaffolds were predicted by running VIBRANT v.1.2.1 [35], VirSorter v2 with setting '--
152 include-groups "dsDNAphage,ssDNA,NCLDV,lavidaviridae"' and default score [36], and geNomad
153 v.1.3 [37]. Viral scaffolds detected by the different tools were combined and filtered to a minimum
154 length of 3000 bp. The number of assembled viral scaffolds by sampling device was summed up.
155 As the number of viruses and the genome size were very low, assemblies across all read files
156 (respective forward and reverse reads of all samples merged) were performed using the above tools
157 and additionally MEGAHIT v.1.2.9 [38]. The resulting viral scaffolds were combined with the
158 previous ones, and a size cut-off of 5000 bp was applied. Viruses were clustered using VIRIDIC
159 v.1.0 r3.6 [39], and only one viral operational taxonomic unit (vOTU) of each species cluster
160 (demarcation threshold = 95%) was used in downstream analysis, and CheckV v.1.0.1 [40] was run.
161 To determine if a vOTU was present in a sample, reads were mapped back to vOTUs using Bowtie
162 2 v.2.3.5.1 [41] with settings '--mp 1,1 --np 1 --rdg 0,1 --rfg 0,1 --score-min L,0,-0.1' [42], which ensure
163 that only reads with 90% sequence identity map to the vOTU. To follow guidelines of Roux,
164 Emerson, Eloe-Fadrosh and Sullivan [43], a vOTU was considered present if at least 75% of the
165 scaffold was covered with reads, which was checked using the 'calcopo.rb' script [18]. Mean
166 coverages of vOTU were determined using the '04_01calc_coverage_v3.rb' script [44]. Gene calling
167 was performed using Prodigal v.2.6.3 [45] and for viral clustering, vConTACT2 v.0.9.19 [46] was run
168 on the vOTU proteins together with those from the 1August2023 viral reference database
169 (https://github.com/RyanCook94/inphared/tree/a330daa635cd3c78843d470668cb22ff842960e4)
170 derived from INPHARED [47]. Results were compiled using graphanalyzer v.1.5.1 [48] and a network
171 was built in Cytoscape v.3.9.0 [49]. PhaMer [50] and PhaGCN [51] were used to discriminate phages
172 from non-phages, and to determine the taxonomic rank of a phage at family level, respectively.
173 and VirClust webtool [52] with default settings was run on vOTUs and related genomes, that were
174 identified by ViPTree v.4.0 [53]. ViPTree was run with dsDNA prokaryotic and eukaryotic hosts as
175 reference viruses. The guanine/cytosine (GC) content of vOTUs was determined using EMBOSS
176 v. 5.0.0 [54]. Annotations of vOTUs were performed using DRAM-v v.1.4.6. [55]. vOTUs were
177 BLASTed against the IMG/VR Viral Nucleotide Database v.4 with an e-value cut-off of 1e-5 [56].

178 Domain analysis of the proteins was performed using NCBI Conserved Domain Search website
179 [57]. A BLASTp analysis for selected amino acids sequences from two vOTUs was run against the
180 Ocean Gene Atlas [58, 59] using the Tara Oceans Microbiome Reference Genome Catalog v1 OM-
181 RGC_v1 and an e-value cut-off of 1e-10. Photosynthetically active radiation (PAR) was chosen as
182 the environmental variable based on which the spatial distribution of the protein homologs of
183 vOTU_35_ORF14 and vOTU_39_ORF2 were observed, because the protein homologs are
184 functionally related to PAR or ultraviolet radiation, which both correlate with global solar radiation
185 [60]. The phylogenetic tree acquired from the Ocean Gene Atlas was edited using the Interactive
186 Tree of Life (iTOL) v6 website[61]. Overlaps of vOTUs between stations were visualized with Venn
187 Diagrams made with the Ugent webtool (https://bioinformatics.psb.ugent.be/webtools/Venn/).
188 Host predictions were performed using iPHoP v.1.2.0 and the database Sept_2021_pub [62].

189 **Diversity analysis and statistics**

190 Community and diversity analysis of vOTUs was performed using phyloseq v. 1.48.0 [63] within the
191 R programming environment v.4.4.0 [64]. Plotting and statistical analysis of Shannon-Wiener index
192 was performed in Graphpad Prism 10.

193 **Results:**

194 **Viral community composition is affected by sampling site, collection device, and daily**
195 **fluctuations**

196 Different sampling devices collected different numbers of viral genome fragments. From the 60
197 assembled vOTUs >3 kb length, most vOTUs (30) were assembled from samples collected with
198 the wet Coriolis and only one vOTU from the dry Coriolis (Figure 1a, collection devices depicted
199 in Figure 1c). Seven and 22 vOTUs were assembled after collection with the pump and the
200 collector, respectively. Of the 60 assembled vOTUs, 43 and 17 vOTUs were collected at sea
201 level/KEP and altitude/DMP, respectively (Figure 1d). From these assembled scaffolds and the
202 cross-assembled ones, 39 vOTUs >5 kb were identified and further analyzed. Mapping reads to
203 these vOTUs showed that 15 unique vOTUs were in samples taken with the collector, another
204 unique 15 in samples from wet Coriolis, and 9 vOTUs occurring in samples obtained by both
205 devices, suggesting that these two sampling methods complement each other well for virus
206 sampling (Figure 1b). Mapping further revealed that 11 vOTUs were shared between both

sampling sites, while 19 were only present at sea-level/KEP and 9 were detected at the altitude/DMP site (Figure 1e).

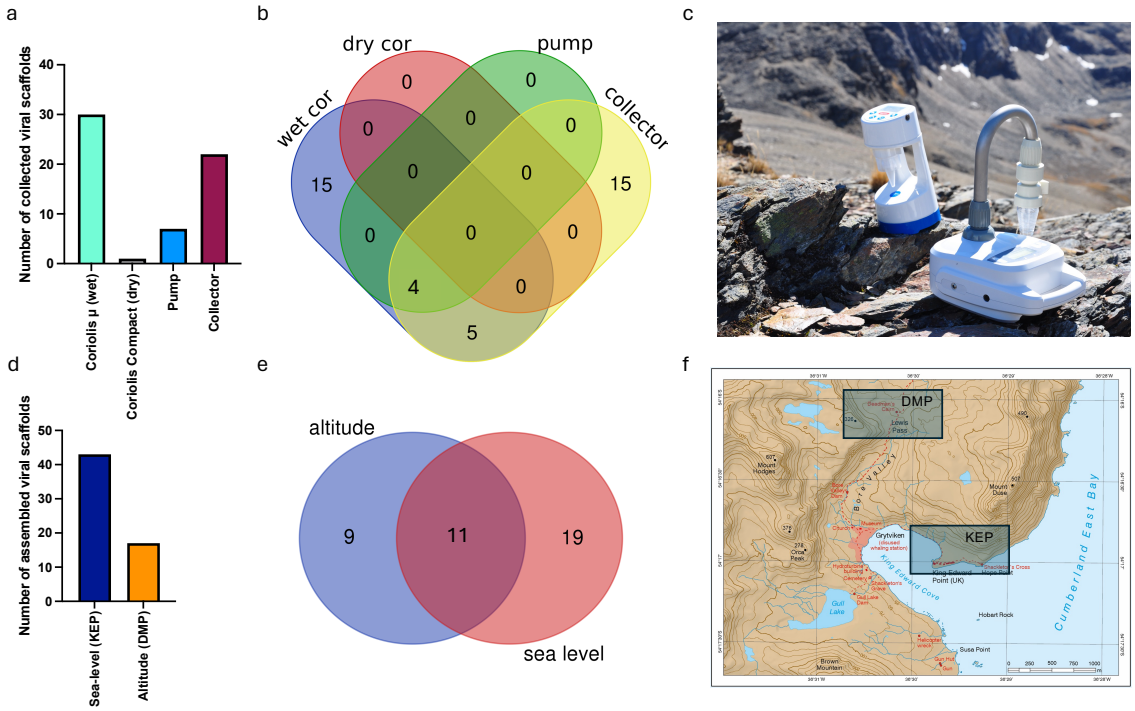


Figure 1: Airborne virus sampling with different collection device from two sampling sites in South Georgia. *a) Number of collected viral scaffolds >3000 bp per sampling device; for description of sampling devices, please see main text. b) Venn diagram showing overlapping vOTUs for different sampling devices based on read mapping. c) Image of Coriolis µ and Coriolis COMPACT sampler in the Antarctic landscape (Image: David Pearce). d) Number of collected viral scaffolds >2 kb by sampling site, e) Venn diagram showing overlapping vOTUs (>5 kb) between sampling sites based on read mapping. f) Map of the sampling site "South Georgia and The Shackleton Crossing" by British Antarctic Survey, 2021, 1:200,000 and 1:40,000 scale, BAS (Misc 12), Cambridge, British Antarctic Survey. Reproduced with permission.*

The vOTUs were clustered into viral clusters (VCs) by vConTACT2 corresponding to genus level clusters, while "singletons" and "outliers" may represent novel viruses. Relative abundance analysis of VCs showed that VC_1229_0 (containing vOTU_12 & vOTU_13), vOTU_24, and vOTU_30 were most abundant and prevalent in the community (Figure 2a, Table S2). Based on

Virsorter2, VC_1229_0 might contain virophage vOTUs, family *Lavidaviridae*, vOTU24 belongs to order Ortervirales (single-stranded RNA viruses replicating through a DNA intermediate), and vOTU_30 was classified as Bamfordvirae by geNomad, which includes NCLDV. Shannon-Wiener diversity was highest for vOTUs sampled with the collector at sea-level/KEP (2.9 and 3.0) compared to the wet Coriolis at altitude/DMP (2.1 and 2.3) and compared to the wet Coriolis and pump at sea-level (0.6 – 1.9) and thus did not show a clear Shannon-Wiener diversity difference between the two sites as for bacteria (see below) but may indicate a joint effect of the sampling device and site on alpha-diversity (Figure 2b).

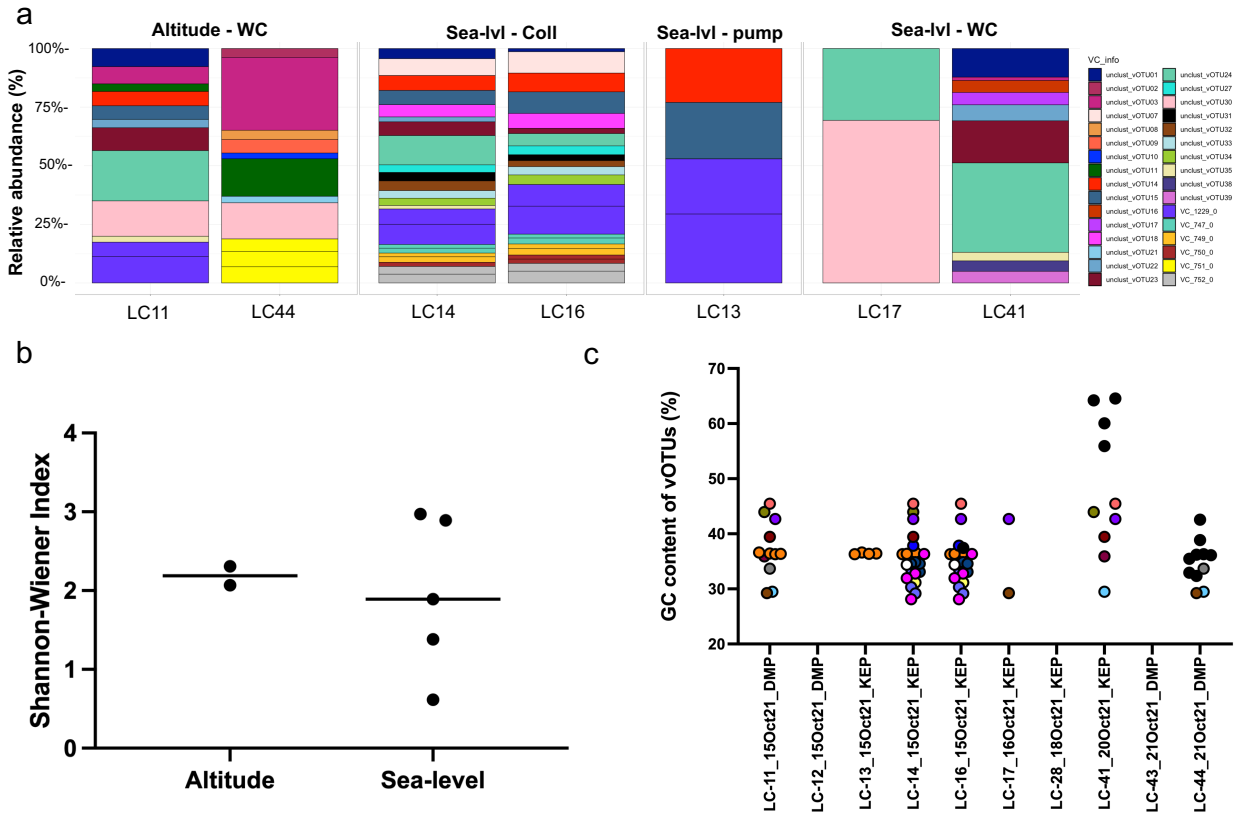


Figure 2: Viral cluster and vOTU distribution and relation to guanine/cytosine (GC) content. *a) Relative abundance of viral clusters (VC) and unclustered viral operational taxonomic units (vOTUs) across stations. b) Shannon-Wiener index for viruses separated by study site. c) GC content of vOTUs versus sampling date and site. If vOTUs were found in several samples based on read mapping, they have the same color; if they are unique for a sample, they are black. DMP=Deadman's Point, KEP=King Edward Point, Sea/lvl= Sea-level, WC= wet Coriolis, LC refers to internal sample identification.*

239

240 During sampling on October 15th, 2021, both the pump and the collector captured some of the
241 most abundant vOTUsBased on read mapping, vOTUs and VCs reoccur and co-occur on different
242 days and on both sites, but also new vOTUs appear. Rain samples (LC-14 & LC-16) contained
243 some vOTUs that were also recovered from air on October 15th, suggesting that these overlapping
244 vOTUs were washed out by falling rain from the atmosphere or resulted from raindrops that were
245 collected with the aerosol sampler. Some viral clusters were only detected in the rainwater, namely
246 VC_752_0, VC_750_0, VC_749_0, VC_747_0 (all predicted Caudoviricetes), which could
247 indicate that they originate from a higher altitude. In a previous, non-polar study, viral genomes
248 retrieved from aerosols and rainwater had a significantly higher GC base content compared to
249 marine viruses [65], which is why the GC content of the Antarctic air vOTUs was further explored.
250 Notably, on one day (LC-41), where viruses were sampled at sea level with the wet Coriolis, four
251 vOTUs present in the sample based on read mapping had a GC content of mean ± standard
252 deviation = 61.8% ± 4.0 (n = 4, Figure 2c, Table S3).

253 **Protein-based clustering revealed similarities to known marine viruses**

254 According to classifications by geNomad, Adintoviridae (3), Caudoviricetes (15) including
255 Crassvirales (3), Imitervirales (1), Bamfordvirae (2), unclassified (10) or not determined by the
256 tool (8) were found. VirSorter2 predicted dsDNA phages (13), lavidaviridae (9), NCLDV (7) and
257 not determined (10, Table S3). Depending on analysis tool, evidence for the presence of 13 – 17
258 phages among the 39 vOTUs was found (Figure 3a & 3b, Table S3). Four phages were predicted
259 to be family Demerecviridae and two others as family Straboviridae (Table S3).

260 Protein-based clustering using vConTACT2 showed similarities of South Georgia vOTUs to
261 known marine phages, vOTU_07 shared protein similarities with *Nonlabens* phage P12024L
262 (GenBank, JQ823123, Figure 3a&b), which originates from coastal seawater [66]. vOTU_36 and
263 vOTU_37 detected in rainwater formed VC_747_0 sharing protein similarities with
264 Flavobacterium phage vB_FspP_elemoA_8-9C (GenBank, MT497073). Based on VipTree,
265 vOTU_35 was related to oceanic, widespread Cyanophage S-TIM5 (GenBank, NC_019516) [67, 68]
266 and estuarine Synechococcus phage S-CBWM1 (GenBank, NC_048106, Figure 4a) [69], vOTU_28,

267 vOTU_29, vOTU_36 and vOTU_37 were related to different Flavobacterium phages of the elemo
268 and immuto group from the Baltic Sea [70], and vOTU_27 to different Cellulophaga phages, which
269 were isolated from the Baltic Sea [71] and North Sea [72]. In total, 15 vOTUs were related to viruses
270 having marine hosts or marine origin themselves (Figure 3a, Table S4) and provide evidence that
271 the Antarctic air viral community was influenced by both the land and the sea.

272 Functional annotations of the 39 vOTUs are listed in Table S5. vOTU_19 being related to
273 Nonlabens phages P12024S (GenBank, NC_018271) and P12024L (GenBank, NC_018272)
274 (Figure 3a) had the most number of ORFs (19) encoding for proteins involved in host lytic
275 machinery such as holin (ORF6) and peptidase (ORF9) [73]. vOTU_11 encoded for a Pao
276 retrotransposon peptidase being homologous to aspartic proteinases expressed by retroposons and
277 retroviruses (Pfam database, PF05380)[74]. Sequence alignment and tBLASTx analysis of the
278 vOTUs indicated potential similarities with viruses from different ecosystems. We found several
279 protein coding genes of vOTU_07 to show ~50% sequence identity with phages P12024S
280 (GenBank, NC_018271) and P12024L (GenBank, NC_018272) (Figure 4a) infecting the marine
281 host *Persicivirga* of the family *Flavobacteriaceae* [66]. Additionally, we found protein similarities
282 of vOTU_17 and vOTU_39 with viruses that infect extremophilic hosts such as the archaeon
283 *Haloarcula californiae* isolated from salt brines (DSMZ, DSM 8905) [75] and *Microbacterium*
284 *radiodurans* NRRL B-24799, a UV radiation-tolerant bacterium [76], respectively (Figure 4b).

285 Hierarchical and phylogenetic clustering of viral core proteins highlighted distinct protein
286 clustering into eight different clades (clusters 1, 2, 6, 10-12, 14 and 20) out of the predicted 22
287 (Figure 4c, Table S6). These clusters in the hierarchical tree demonstrate protein overlaps and
288 constitute similar phylogenetic relations as observed in the ViPtree analysis (Figures 3a & 3b).
289 Clusters 2, 6, and 11 have the greatest number of members (04) while cluster 20 comprised of just
290 two members shared the greatest number of common core genes. This cluster is formed from
291 vOTU_36 and vOTU_37, being related to Flavobacterium phage vB_FspP_elemoA_8-9C
292 (GenBank, MT497073) in vConTACT2. Overall, we find high diversity of proteome in the
293 analyzed vOTUs, with proteins overlapping among scaffolds being phylogenetically similar.

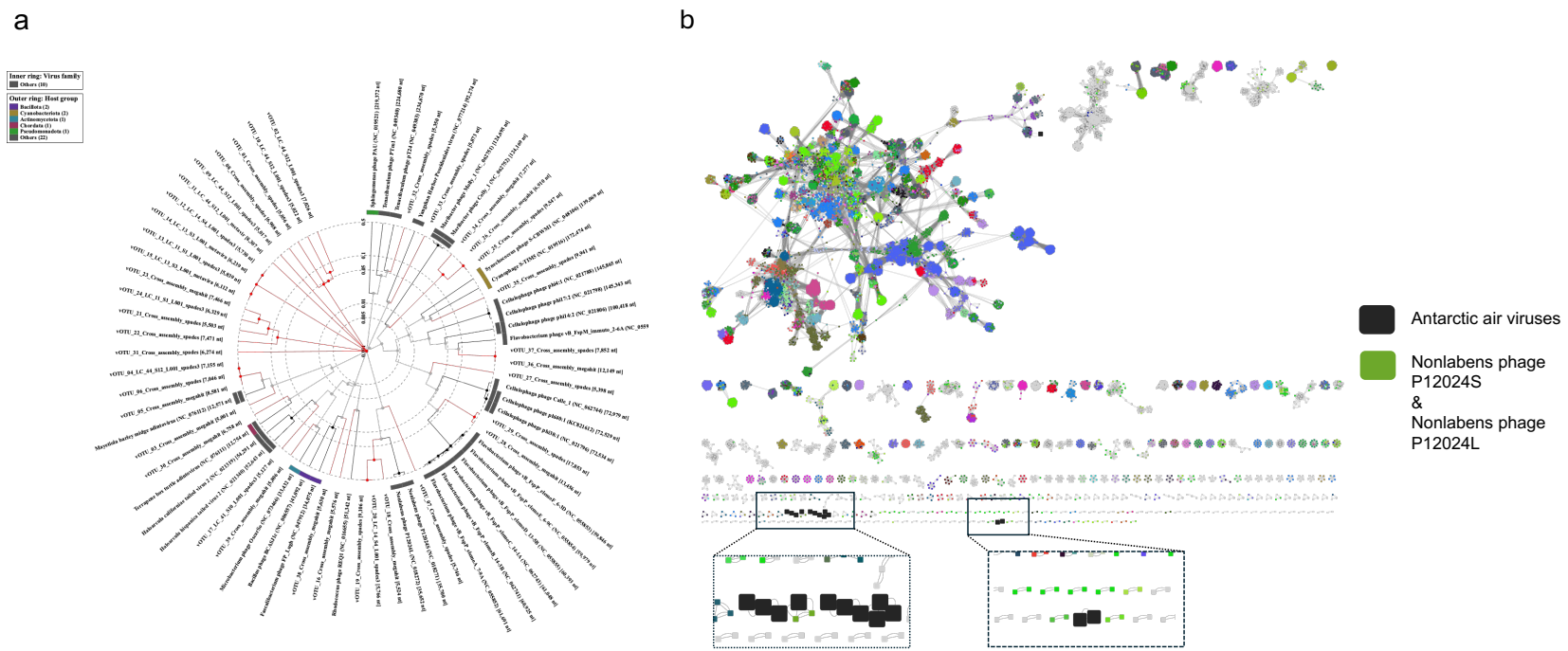
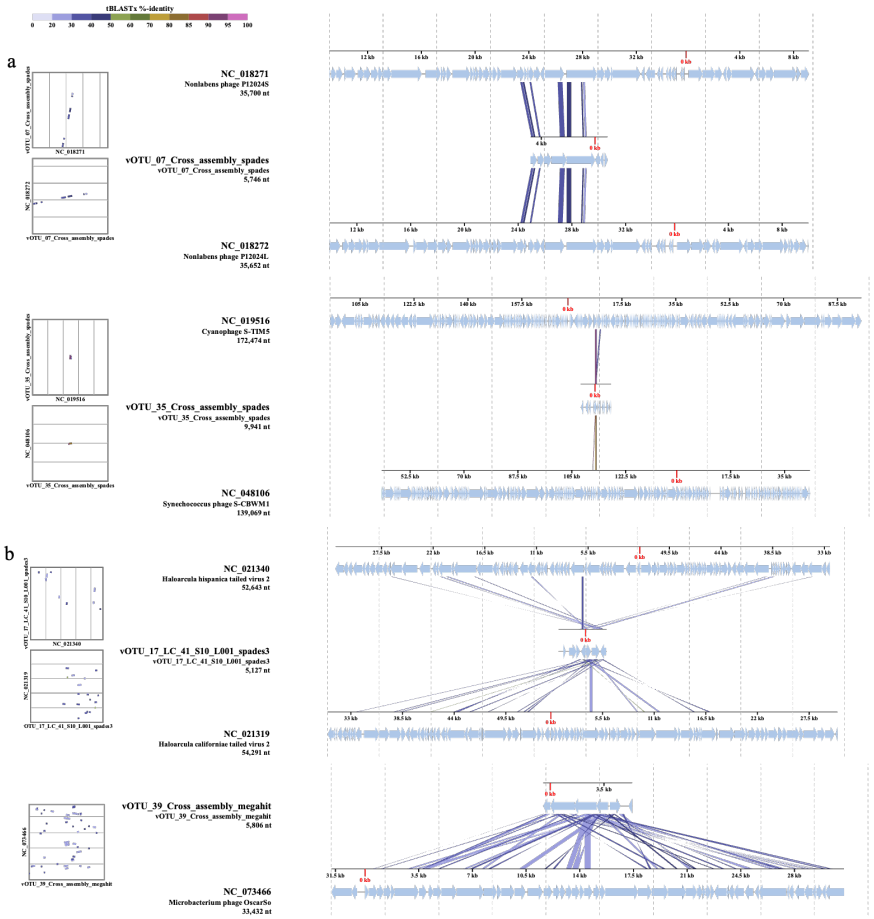


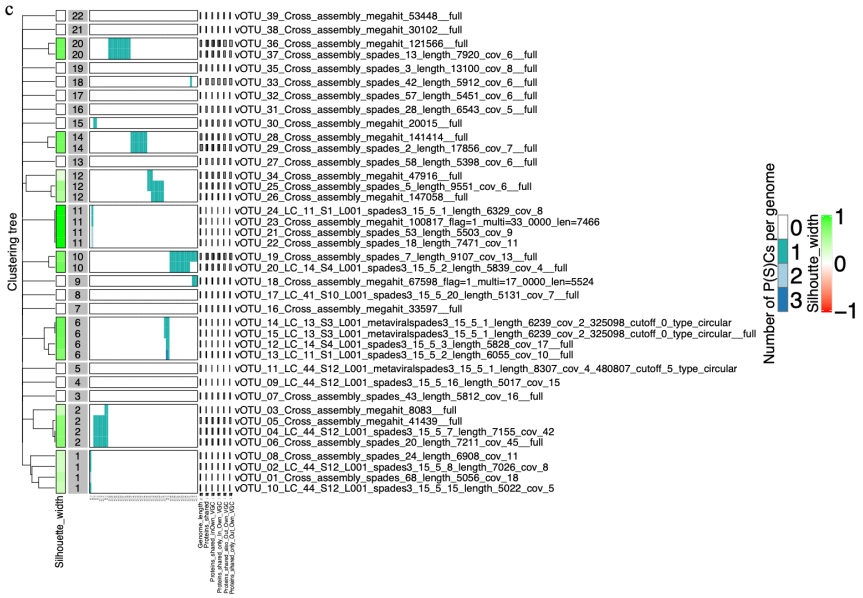
Figure 3: Viral protein-based clustering. *a) Proteomic tree of 39 viral operational taxonomic units (red branches) and their related viruses constructed using ViPTree. b) Protein-sharing network of Antarctic vOTUs (big black squares) forming viral clusters with Nonlabens phage P12024S and P12024L and with each other.*

299 **Spatial distribution of protein homologs in the major oceans**

300 Further analyses were performed to reveal biogeographic distribution of two predicted viral
301 proteins. vOTU_35 sharing protein identity with the Cyanophage S-TIM5 (Figure 4a) encoded the
302 photosynthetic reaction centre protein D1 with a psbA domain (vOTU_35_ORF14) (Figure 5a).
303 The D1 and D2 reaction centre proteins form a heterodimer responsible for the establishment of
304 photosystem II in cyanobacteria [77, 78], which is the most light-sensitive complex of cyanobacteria
305 being affected by photosynthetically active radiation (PAR) [79]. Another protein from vOTU_39
306 sharing similarity with a phage infecting the extremophilic UV-tolerant host *Microbacterium*
307 *radiodurans* [76] carried a gene cluster encoding for impB/mucB/samB with a DinP domain having
308 functions in the bacterial SOS (DNA damage) response and thus has relevance in UV protection
309 (vOTU_39_ORF2) (Pfam database, PF00817) [80] (Figure 5a). To understand the distribution of
310 these protein sequences and the possible origin of the encoding vOTUs, we explored the
311 biogeography of the protein sequence based on similarities with the Ocean Gene Atlas. The
312 homologs of photosynthetic reaction centre protein D1 (vOTU_35_ORF14) were widely detected
313 in several Tara Oceans sampling locations at the surface water layer (SRF) at varying PAR ranging
314 from approximately 0.07 mol quanta m^{-2} day^{-1} to 41.3 mol quanta m^{-2} day^{-1} as compared to the
315 mesopelagic zone (MES), deep chlorophyll maximum (DCM), and the marine epipelagic mixed
316 layer (MIX) (Figure 5b, Table S7). A total of 2625 hits were obtained for this protein and the
317 number of abundance measures was 192236. Taxonomic distribution of these homologs indicated
318 high abundance of this protein in viral genomes (37%) (Figure S1). The UV protection related
319 protein (vOTU_39_ORF2) was mostly abundant in the SRF as compared to the MES, DCM, and
320 MIX layers, however, the distribution across the sampling locations was rather sparse as compared
321 to vOTU_35_ORF14 (Figure 5b, Table S8). Only a total of 59 hits were obtained for this protein
322 and the number of abundance measures was 1924. Viral homologs of impB/mucB/samB family
323 protein could not be found. The closest homolog predicted by the Ocean Gene Atlas for this protein
324 is the impB/mucB/samB and DinP domain containing protein of a bacterium, *Vulcanimicrobium*
325 *alpinum* (GenBank, WP_317995911.1)[81], which has been isolated from a fumarolic ice cave of the
326 volcano Mount Erebus in Antarctica (Figure S2). Additionally, homologs of both proteins were
327 present at the sampling location TARA_076, which is closer to the sampling site of our study than
328 most other TARA Ocean stations (Figure 5c).



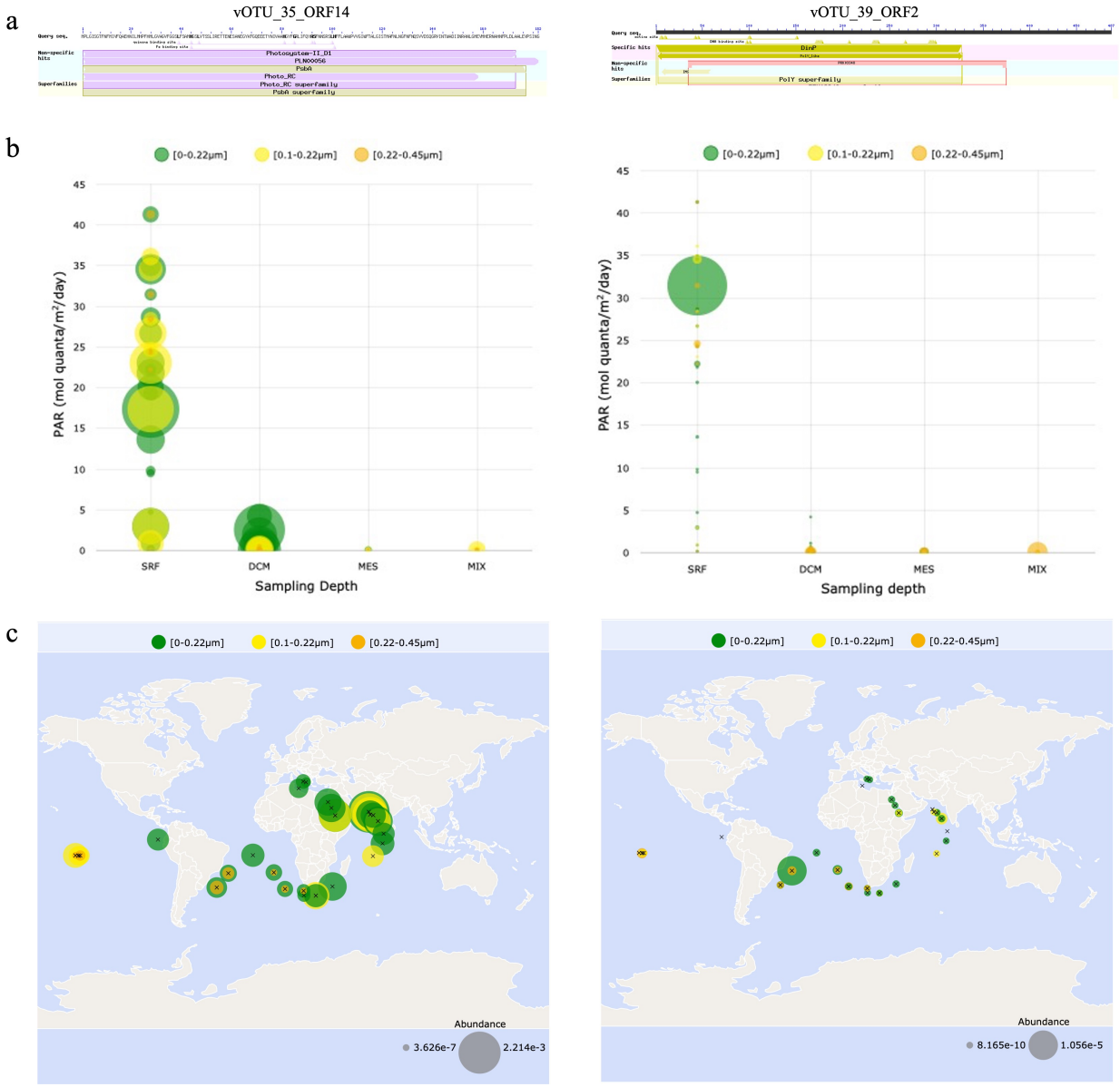
329



330

331 ***Figure 4: Sequence alignment and proteome clustering of annotated vOTUs.*** *a) Sequence*
332 *alignment of vOTUs with marine viruses. vOTU_07 demonstrate nucleotide identity with two*

Nonlabens phages. Sequence alignment of vOTU_35 indicate nucleotide identity with two marine cyanophages. b) Sequence alignment of vOTUs from extremophilic hosts. vOTU_17 demonstrate tBLASTx identity with sequences of two archaeal viruses. Sequence alignment of vOTU_39 shows tBLASTx identity with a high genomic GC content containing actinobacteriophage OscarSo (Genbank, OP434449.1) infecting Microbacterium radiodurans. c) Viral proteome clustering of all vOTUs with Virclust indicates eight different clades (clusters 1, 2, 6, 10-12, 14 and 20) out of the predicted 22.



341 ***Figure 5: Distribution and abundance of viral protein sequence homologs in the major oceans.***
342 *a) The amino acid sequence of Photosynthetic reaction centre protein D1 derived from vOTU_35*
343 *and impB/mucB/samB derived from vOTU_39 based on BLASTp to the Tara Oceans Microbiome*
344 *Reference Genome Catalog v1 OM-RGC_v1 as obtained from Ocean Gene Atlas. b) the*
345 *abundance of the protein homolog for "photosystem" and "DinP" in three different filtered*
346 *fractions and across different sampling depths (SRF = surface water, DCM = deep chlorophyll*
347 *maximum, MES = mesopelagic zone, MIX = mixed layer) and in relation to photosynthetically*
348 *active radiation (PAR). c) Distribution of proteins across stations of the Tara Ocean expedition.*
349 *Size of bubble indicates abundance in b and c.*

350 **Microbial community composition and host assignments**

351 For the 39 vOTUs, only three hosts could be predicted, which were the genus *Rickettsia*
352 (confidence score 90.8, phylum Pseudomonadota) for vOTU_34, *Myroides* for vOTU18
353 (confidence score 90.9, phylum Bacteroidota) and UBA9320 for vOTU_37 (confidence score 90.9,
354 phylum Bacteroidota), the latter being a metagenome assembled genome from the Central Arctic
355 [82]. Microbial community profiling (mainly prokaryotes, with few reads to Basidiomycota) showed
356 that the air community was dominated by Actinobacteria, Proteobacteria (now Pseudomonadota)
357 and unassigned taxa. More unassigned taxa were found in altitude/DMP samples (up to 80%
358 relative abundance) than at sea-level/KEP (up to 18% relative abundance, Table S9, Figure S3).
359 BLASTing vOTUs to the IMG/VR database revealed hits to orders Ortervirales (vOTU_01,
360 vOTU_23, vOTU_24) and Algavirales (vOTU_03), to the class Caudoviricetes (vOTU_07,
361 vOTU_12 – vOTU_16, vOTU_18, vOTU_19, vOTU_20, vOTU_25 – vOTU_29, vOTU_33,
362 vOTU_34 – vOTU_37, vOTU_39), and to order Priklausovirales (vOTU_31). Predicted hosts
363 were of the phyla Actinobacteriota (vOTU_12 – vOTU_15, Alphaproteobacteria (vOTU_39),
364 Saprospiraceae (vOTU_36 & vOTU_37) and Spirochaetota (vOTU_12 – vOTU_15). Most hits
365 were to marine, aqueous, seawater ecosystems from various parts of the world (Table S10).

366

367

368

369 **Discussion**

370 **Aerosolization of different viral taxonomic ranks, site effects, and biogeographic**
371 **distribution**

372 Our data have shown that the air viral community composition over South Georgia was influenced
373 by the sampling site (proximity to the sea) and the type of air sampling device used. We found
374 typical eukaryotic viral groups like NCLDV, which often get infected by dsDNA viral phages, so-
375 called virophages, which we also found evidence for and which have been previously observed in
376 the Southern Ocean close to the Antarctica continent [28] and to influence algal host-virus
377 interactions in a meromictic Antarctic lake [83]. About half of all vOTUs corresponded to phages.
378 Overall, this shows that viruses of various taxonomic ranks become airborne, and we here provide
379 first baseline data on the viral communities in this region, day-to-day fluctuations and influences
380 by the marine environment

381 More vOTUs were assembled from and were present based on read mapping at the close-to-the-
382 sea KEP site suggesting previous aerosolization of marine viruses. An earlier study in the Arctic
383 region has explored viral distribution and adaptation at the air-sea interface and showed that virus
384 distribution across the Arctic might be facilitated by aerosolization of viruses residing in the sea-
385 surface microlayer, the uppermost 1 mm surface layer in contact with the atmosphere [84]. Viruses
386 can be emitted into aerosols from the surface microlayer [85] and viruses are active in the Antarctic
387 microlayer [86]. As both sites were located on a remote island in the middle of the Southern Ocean
388 and far away from major terrestrial influences, such small-scale local differences in marine
389 contributions to the viral community were not expected as both sites are eventually close to the
390 sea. Several vOTUs shared proteins with known cyanophages and phage isolates infecting typical
391 marine heterotrophic bacteria and additionally, typical marine prokaryotes such as
392 Thaumarchaeota (*Nitrosopumilus* sp.), were part of the microbial community (Figure S3).

393 High GC carrying vOTUs known to occur in air and rain samples from Swedish air [65] were found
394 on one sampling day in the data set explored here. They were derived from a sample on a day
395 where particularly low wind speeds were recorded, sunny weather, and a wind direction of 110
396 degrees (east-southeast) direction, whereas during all other samplings wind came from a westward
397 direction (Table S1). High GC viruses shared protein identities with Microbacterium phage

398 OscarSo (vOTU_39), which has a GC content of 69.2% (https://phagesdb.org/phages/OscarSo/),
399 and with two archaeal viruses of the halophilic archaeon *Haloarcula* (vOTU_17). OscarSo has
400 been isolated from *Microbacterium radiodurans* being resistant to UV radiation and originating
401 from sand of the Gobi Desert [76]. Hence, our findings provide some evidence that air viruses with
402 high GC content could have extremophilic hosts, letting us speculate that such viruses have a non-
403 marine origin and stem from the higher atmosphere or terrestrial influences.

404 The distance between the two sampling sites DMP and KEP is approximately 2 km and hence our
405 data overall confirm short-range dispersal of Antarctic viruses with bioaerosols, with likely
406 influences from sea spray and wet precipitation as it was raining during some of the sampling
407 (Table S1). Interestingly, the alpha diversity between both sites was not significantly different for
408 the viruses. We speculate that viral communities are less influenced by a study site, because longer
409 residence times and lower deposition velocities in air, since they are smaller [15, 87]. Therefore, they
410 can be dispersed further and can become part of another remote community more easily.

411 Proteomic alignment and classification of the vOTUs aligned with the results of the phylogenetic
412 and clustering analysis. We found sequence identity at the amino acid level of some vOTUs with
413 viruses of typical marine bacteria and archaea. We identified two different proteins encoded by
414 vOTU_35_ORF14 and vOTU_39_ORF2, which shared several homologs widely distributed in the
415 surface water layer of the major oceans (such as the Pacific, Atlantic, and Indian Ocean),
416 supporting a probable marine origin. What is speculative yet essential to note is the functionality
417 of these proteins and their involvement in modulating a host's metabolism. Phages present in
418 diverse ecological habitats must be highly adaptive; this could involve the placement of structural
419 genes in regions of above-average GC content of the genome [88], encoding specialized proteins
420 promoting survival and extremotolerance of their hosts or reverting to lysogenic or
421 pseudolysogenic life cycles [89, 90]. A typical trait of marine phages is the augmentation of the host
422 metabolism or enhancing viral fitness by the acquisition and expression of auxiliary metabolic
423 genes (AMGs) [91]. Considering that phage-derived psbA has already been demonstrated to support
424 photosynthetic activity of their hosts in surface water [92], and the potential of the impB/mucB/samB
425 gene cluster to support microbial growth in harsh environments [93], we hypothesize that
426 vOTU_35_ORF14 and vOTU_39_ORF2 could potentially be AMGs supporting the replication of
427 these viruses in the marine and atmospheric environments, respectively. This might be needed as

428 both environments have challenging conditions, e.g., surface water in the open ocean is
429 oligotrophic [94], and the atmosphere is prone to desiccation and solar and UV radiation effects [95].
430

431 **Methodological challenges of sampling air viruses for metagenomics**

432 Several viruses remained unclustered by vConTACT2 or just clustered with other Antarctic viruses
433 from this study, indicating the presence of novel viruses. Still the overall viral biomass in these air
434 samples was low, a common challenge in air microbiology [96, 97]. This led us to lower the usual 10
435 kb length cut-off for vOTUs proposed by the standard viromics guidelines [43] to accept smaller, 5
436 kb viral scaffolds to be used for viral presence/absence determination, which might be more
437 adequate for other low and ultra-low biomass ecosystems such as clean rooms or space equipment
438 [98, 99]. However, allowing smaller fragments can also lead to more uncertainties in various
439 downstream predictions and increases the possibility for false-positive predictions. Because
440 airborne viruses are present in low abundance, the detection of a vOTU in a sample is more likely
441 to be affected by limited coverage compared to high biomass ecosystems, making it difficult to
442 accurately interpret overlapping communities and day-to-day diversity.

443 The type of sampling device used had an influence on the viral findings. The dry Coriolis sampled
444 few viruses (just one got assembled) and none were found by mapping reads from dry Coriolis
445 samples to vOTUs retrieved from other samplers. According to the manufacturer's information,
446 the Coriolis Compact has a flow rate of 50 L min^{-1} and is described as a device capable of virus
447 sampling but is not mentioned in the literature for this purpose. The wet Coriolis (Coriolis µ) has
448 been used extensively for viral air sampling [100, 101]. It has a higher flow rate of up to 300 L min^{-1}
449 and collects particles into a liquid and thus might be more efficient than the dry Coriolis. The liquid
450 might be more suitable for conserving the viral particles. Both Coriolis samplers have a size range
451 of 500 nm – 10 µm for collected particles, which would exclude most viruses. As we have no
452 information on the particle size of the sampled viruses here, it remains speculative if small viruses
453 were not sampled or were only collected if being attached to larger particles. Further tests should
454 be conducted to thoroughly validate the devices used for viral sampling. This includes tests on the
455 stability of viral nucleic acids on filters and in liquid, as degradation of nucleic acids is possible
456 during long-term air sampling [102]. Instead of sampling on 0.2 µm pore-sized membranes or

Sterivex filters, viruses collected into liquid could be iron-flocculated or subjected to ultrafiltration and processed as viromes as done for aquatic samples [103], which remains to be tested in the future.

We noticed that more different viruses were sampled with the rain collector Rain and aerosol-derived viruses can be genomically very different [65], and airborne microbial communities also vary between precipitation types and with seasonal variations in Antarctica [10, 104]. This would explain why different, non-overlapping vOTUs were collected by the two devices. Therefore, it might be advisable to use a funnel collector in combination with an impinger air sampling system to obtain a more accurate representation of the air's viral community.

Acknowledgements: We acknowledge use of the HPC cluster DRACO, with tools and database kindly provided by the VEO group of Bas Dutilh at the FSU Jena. We acknowledge technical assistance by Luke Cockerton for DNA extractions and by Bingli Chai for submitting the sequence data to SRA.

Funding: JR received funding by the DFG RA3432/1-1, project number 446702140, the WBP Return Grant RA3432/1-3, project number 534276621, and the Swedish Research Council, Starting Grant ID 2023-03310_VR. Fieldwork was supported by the British Antarctic Survey, Project CASS196. Data handling was enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

Author Contributions: RD: data curation, formal analysis, investigation, methodology, visualization, validation, writing original draft, writing review & editing; LM: field sampling, data acquisition, writing review & editing; DAP: conceptualization, field sampling, writing review & editing; JR: conceptualization, formal analysis, funding acquisition, project administration, supervision, visualization, validation, writing original draft, writing review & editing

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

Data Availability: Sequencing data are stored an NCBI's SRA under the Bioproject PRJNA1107129, Biosample IDs: SAMN41160173 - SAMN41160182. The vOTUs are deposited at figshare under doi: 10.6084/m9.figshare.26536105. Weather data is downloadable via

https://legacy.bas.ac.uk/cgi-bin/metdb-form-2.pl?tabletouse=U_MET.GRYTVIKEN_AWS&complex=1&idmask=.....&acct=u_met&pass=weather.

References:

1. Malard LA, Sabacka M, Magiopoulos I, Mowlem M, Hodson A, Tranter M, Siegert MJ, Pearce DA. Spatial Variability of Antarctic Surface Snow Bacterial Communities. *Front Microbiol* **10**, 461(2019). doi:10.3389/fmicb.2019.00461
2. Stokke OS, Vidas D. *Governing the Antarctic: the effectiveness and legitimacy of the Antarctic Treaty System*. Cambridge University Press (1996).
3. Pearce DA, Bridge PD, Hughes KA, Sattler B, Psenner R, Russell NJ. Microorganisms in the atmosphere over Antarctica. *FEMS Microbiol Ecol* **69**, 143-157(2009). doi:10.1111/j.1574-6941.2009.00706.x
4. Leihy RI, Peake L, Clarke DA, Chown SL, McGeoch MA. Introduced and invasive alien species of Antarctica and the Southern Ocean Islands. *Sci Data* **10**, 200(2023). doi:10.1038/s41597-023-02113-2
5. Rogers AD, Frinault BAV, Barnes DKA, Bindoff NL, Downie R, Ducklow HW, Friedlaender AS, Hart T, Hill SL, Hofmann EE, Linse K, McMahon CR, Murphy EJ, Pakhomov EA, Reygondeau G, Staniland IJ, Wolf-Gladrow DA, Wright RM. Antarctic Futures: An Assessment of Climate-Driven Changes in Ecosystem Structure, Function, and Service Provisioning in the Southern Ocean. *Ann Rev Mar Sci* **12**, 87-120(2020). doi:10.1146/annurev-marine-010419-011028
6. Pretorius I, Schou WC, Richardson B, Ross SD, Withers TM, Schmale DG, 3rd, Strand TM. In the wind: Invasive species travel along predictable atmospheric pathways. *Ecol Appl* **33**, e2806(2023). doi:10.1002/eap.2806
7. Cao Y, Yu X, Ju F, Zhan H, Jiang B, Kang H, Xie Z. Airborne bacterial community diversity, source and function along the Antarctic Coast. *Sci Total Environ* **765**, 142700(2021). doi:10.1016/j.scitotenv.2020.142700
8. Pearce DA, Hughes KA, Lachlan-Cope T, Harangozo SA, Jones AE. Biodiversity of air-borne microorganisms at Halley Station, Antarctica. *Extremophiles* **14**, 145-159(2010). doi:10.1007/s00792-009-0293-8
9. King-Miaow K, Lee K, Maki T, LaCap-Bugler D, Archer SDJ. Airborne microorganisms in Antarctica: transport, survival and establishment. *The Ecological Role of Micro-Organisms in the Antarctic Environment*, 163-196(2019). doi:10.1007/978-3-030-02786-5_8
10. Malard LA, Avila-Jimenez ML, Schmale J, Cuthbertson L, Cockerton L, Pearce DA. Aerobiology over the Southern Ocean - Implications for bacterial colonization of Antarctica. *Environ Int* **169**, 107492(2022). doi:10.1016/j.envint.2022.107492
11. Heinrichs ME, Piedade GJ, Popa O, Sommers P, Trubl G, Weissenbach J, Rahlff J. Breaking the Ice: A Review of Phages in Polar Ecosystems. *Bacteriophages: Methods and Protocols*, 31-71(2024). doi:10.1007/978-1-0716-3549-0_3

12. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* **96**, 2192-2197(1999). doi:10.1073/pnas.96.5.2192
13. Nga DDY, Nhung VH, Nhan NT, Hien TT. Study on the concentration, composition, and recovery rate of bacterial bioaerosols after rainfall in Ho Chi Minh City. *Environ Monit Assess* **196**, 295(2024). doi:10.1007/s10661-024-12442-3
14. Vishwakarma YK, Mayank, Ram K, Gogoi MM, Banerjee T, Singh RS. Bioaerosol emissions from wastewater treatment process at urban environment and potential health impacts. *J Environ Manage* **361**, 121202(2024). doi:10.1016/j.jenvman.2024.121202
15. Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME J* **12**, 1154-1162(2018). doi:10.1038/s41396-017-0042-4
16. Quinn PK, Collins DB, Grassian VH, Prather KA, Bates TS. Chemistry and related properties of freshly emitted sea spray aerosol. *Chem Rev* **115**, 4383-4399(2015). doi:10.1021/cr500713g
17. Trainic M, Koren I, Sharoni S, Frada M, Segev L, Rudich Y, Vardi A. Infection Dynamics of a Bloom-Forming Alga and Its Virus Determine Airborne Coccolith Emission from Seawater. *iScience* **6**, 327-335(2018). doi:10.1016/j.isci.2018.07.017
18. Rahlff J, Bornemann TLV, Lopatina A, Severinov K, Probst AJ. Host-Associated Phages Disperse across the Extraterrestrial Analogue Antarctica. *Appl Environ Microbiol* **88**, e0031522(2022). doi:10.1128/aem.00315-22
19. Lopatina A, Medvedeva S, Shmakov S, Logacheva MD, Krylenkov V, Severinov K. Metagenomic analysis of bacterial communities of Antarctic surface snow. *Front Microbiol* **7**, 398(2016). doi:10.3389/fmicb.2016.00398
20. Kepner RL, Jr., Wharton RA, Jr., Suttle CA. Viruses in Antarctic lakes. *Limnol Oceanogr* **43**, 1754-1761(1998). doi:10.4319/lo.1998.43.7.1754
21. Luhtanen AM, Eronen-Rasimus E, Oksanen HM, Tison JL, Delille B, Dieckmann GS, Rintala JM, Bamford DH. The first known virus isolates from Antarctic sea ice have complex infection patterns. *FEMS Microbiol Ecol* **94**, fiy028(2018). doi:10.1093/femsec/fiy028
22. Paterson H, Laybourn-Parry J. Antarctic sea ice viral dynamics over an annual cycle. *Polar Biology* **35**, 491-497(2012). doi:10.1007/s00300-011-1093-z
23. Lopez-Simon J, Vila-Nistal M, Rosenova A, De Corte D, Baltar F, Martinez-Garcia M. Viruses under the Antarctic Ice Shelf are active and potentially involved in global nutrient cycles. *Nat Commun* **14**, 8295(2023). doi:10.1038/s41467-023-44028-x
24. Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D. High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. *Appl Environ Microbiol* **80**, 6888-6897(2014). doi:10.1128/AEM.01525-14
25. Sommers P, Fontenele RS, Kringen T, Kraberger S, Porazinska DL, Darcy JL, Schmidt SK, Varsani A. Single-Stranded DNA Viruses in Antarctic Cryoconite Holes. *Viruses* **11**, 1022(2019). doi:10.3390/v11111022
26. Guixa-Boixereu N, Vaqué D, Gasol JM, Sánchez-Cámara J, Pedrós-Alió C. Viral distribution and activity in Antarctic waters. *Deep Sea Res Part II Top Stud Oceanogr* **49**, 827-845(2002). doi:10.1016/s0967-0645(01)00126-6

27. Liang Y, Bai X, Jiang Y, Wang M, He J, McMinn A. Distribution of marine viruses and their potential hosts in Prydz Bay and adjacent Southern Ocean, Antarctic. *Polar Biol* **39**, 365-378(2015). doi:10.1007/s00300-015-1787-8
28. Piedade GJ, Schön ME, Lood C, Fofanov MV, Wesdorp EM, Biggs TEG, Wu L, Bolhuis H, Fischer MG, Yutin N, Dutilh BE, Brussaard CPD. Seasonal dynamics and diversity of Antarctic marine viruses reveal a novel viral seascape. *Research Square*, (2024). doi:10.21203/rs.3.rs-3778832/v1
29. Bushnell B. BBTools software package. *URL http://sourceforge net/projects/bbmap* **578**, 579(2014)
30. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. (2011)
31. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, Hingamp P, Alves R, Costea PI, Coelho LP, Schmidt TSB, Almeida A, Mitchell AL, Finn RD, Huerta-Cepas J, Bork P, Zeller G, Sunagawa S. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* **10**, 1014(2019). doi:10.1038/s41467-019-08844-4
32. Ruscheweyh HJ, Milanese A, Paoli L, Karcher N, Clayssen Q, Keller MI, Wirbel J, Bork P, Mende DR, Zeller G, Sunagawa S. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome* **10**, 212(2022). doi:10.1186/s40168-022-01410-z
33. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824-834(2017). doi:10.1101/gr.213959.116
34. Antipov D, Raiko M, Lapidus A, Pevzner PA. Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* **36**, 4126-4129(2020). doi:10.1093/bioinformatics/btaa490
35. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90(2020). doi:10.1186/s40168-020-00867-0
36. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitua MC, Vik D, Sullivan MB, Roux S. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37(2021). doi:10.1186/s40168-020-00990-y
37. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC. Identification of mobile genetic elements with geNomad. *Nat Biotechnol*, (2023). doi:10.1038/s41587-023-01953-y
38. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676(2015). doi:10.1093/bioinformatics/btv033
39. Moraru C, Varsani A, Kropinski AM. VIRIDIC-A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses* **12**, (2020). doi:10.3390/v12111268
40. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*, (2020). doi:10.1038/s41587-020-00774-7
41. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359(2012). doi:10.1038/nmeth.1923

42. Nilsson E, Bayfield OW, Lundin D, Antson AA, Holmfeldt K. Diversity and host interactions among virulent and temperate Baltic Sea *Flavobacterium* phages. *Viruses* **12**, 158(2020). doi:10.3390/v12020158
43. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817(2017). doi:10.7717/peerj.3817
44. Bornemann TLV, Esser SP, Stach TL, Burg T, Probst AJ. uBin: A manual refining tool for genomes from metagenomes. *Environ Microbiol* **25**, 1077-1083(2023). doi:10.1111/1462-2920.16351
45. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119(2010). doi:10.1186/1471-2105-11-119
46. Bolduc B, Jang HB, Doulcier G, You ZQ, Roux S, Sullivan MB. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* **5**, e3243(2017). doi:10.7717/peerj.3243
47. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman J, Jones MA, Millard A. INfrastructure for a PHAge REference Database: Identification of large-scale biases in the current collection of cultured phage genomes. *Phage* **2**, 214-223(2021). doi:10.1089/phage.2021.0007
48. Pandolfo M, Telatin A, Lazzari G, Adriaenssens EM, Vitulo N. MetaPhage: an automated pipeline for analyzing, annotating, and classifying bacteriophages in metagenomics sequencing data. *mSystems* **7**, e0074122(2022). doi:10.1128/msystems.00741-22
49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504(2003). doi:10.1101/gr.1239303
50. Shang J, Tang X, Guo R, Sun Y. Accurate identification of bacteriophages from metagenomic data using Transformer. *Brief Bioinform* **23**, bbac258(2022). doi:10.1093/bib/bbac258
51. Shang J, Jiang J, Sun Y. Bacteriophage classification for assembled contigs using graph convolutional network. *Bioinformatics* **37**, i25-i33(2021). doi:10.1093/bioinformatics/btab293
52. Moraru C. VirClust-A tool for hierarchical clustering, core protein detection and annotation of (Prokaryotic) viruses. *Viruses* **15**, 1007(2023). doi:10.3390/v15041007
53. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379-2380(2017). doi:10.1093/bioinformatics/btx157
54. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277(2000). doi:10.1016/s0168-9525(00)02024-2
55. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodriguez-Ramos J, Bolduc B, Gazitua MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**, 8883-8900(2020). doi:10.1093/nar/gkaa621
56. Camargo AP, Nayfach S, Chen IA, Palaniappan K, Ratner A, Chu K, Ritter SJ, Reddy TBK, Mukherjee S, Schulz F, Call L, Neches RY, Woyke T, Ivanova NN, Eloe-Fadrosh

663 EA, Kyrpides NC, Roux S. IMG/VR v4: an expanded database of uncultivated virus
664 genomes within a framework of extensive functional, taxonomic, and ecological
665 metadata. *Nucleic Acids Res* **51**, D733-D743(2023). doi:10.1093/nar/gkac1037
666 57. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC,
667 He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N,
668 Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. CDD: NCBI's conserved
669 domain database. *Nucleic Acids Res* **43**, D222-226(2015). doi:10.1093/nar/gku1221
670 58. Vernette C, Lecubin J, Sanchez P, Tara Oceans C, Sunagawa S, Delmont TO, Acinas SG,
671 Pelletier E, Hingamp P, Lescot M. The Ocean Gene Atlas v2.0: online exploration of the
672 biogeography and phylogeny of plankton genes. *Nucleic Acids Res* **50**, W516-
673 W526(2022). doi:10.1093/nar/gkac420
674 59. Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, Bachelerie P,
675 Rosnet T, Pelletier E, Sunagawa S, Hingamp P. The Ocean Gene Atlas: exploring the
676 biogeography of plankton genes online. *Nucleic Acids Res* **46**, W289-W295(2018).
677 doi:10.1093/nar/gky376
678 60. Escobedo JF, Gomes EN, Oliveira AP, Soares J. Ratios of UV, PAR and NIR
679 components to global solar radiation measured at Botucatu site in Brazil. *Renewable*
680 *Energy* **36**, 169-178(2011). doi:10.1016/j.renene.2010.06.018
681 61. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic
682 tree display and annotation tool. *Nucleic Acids Res* **52**, W78-W82(2024).
683 doi:10.1093/nar/gkae268
684 62. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A.
685 iPHoP: An integrated machine learning framework to maximize host prediction for
686 metagenome-derived viruses of archaea and bacteria. *PLoS Biol* **21**, e3002083(2023).
687 doi:10.1371/journal.pbio.3002083
688 63. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis
689 and graphics of microbiome census data. *PLoS One* **8**, e61217(2013).
690 doi:10.1371/journal.pone.0061217
691 64. Team RC. R: A language and environment for statistical computing, R Foundation for
692 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (2019)
693 65. Rahlff J, Esser SP, Plewka J, Heinrichs ME, Soares A, Scarchilli C, Grigioni P, Wex H,
694 Giebel HA, Probst AJ. Marine viruses disperse bidirectionally along the natural water
695 cycle. *Nat Commun* **14**, 6354(2023). doi:10.1038/s41467-023-42125-5
696 66. Kang I, Jang H, Cho J-C. Complete genome sequences of two *Persicivirga*
697 bacteriophages, P12024S and P12024L. *J Virol* **86**, 8907-8908(2012).
698 doi:10.1128/jvi.01327-12
699 67. Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. A novel lineage of
700 myoviruses infecting cyanobacteria is widespread in the oceans. *P Natl Acad Sci USA*
701 **109**, 2037-2042(2012). doi:10.1073/pnas.1115467109
702 68. Baran N, Carlson MCG, Sabehi G, Peleg M, Kondratyeva K, Pekarski I, Lindell D.
703 Widespread yet persistent low abundance of TIM5-like cyanophages in the oceans.
704 *Environ Microbiol* **24**, 6476-6492(2022). doi:10.1111/1462-2920.16210
705 69. Xu Y, Zhang R, Wang N, Cai L, Tong Y, Sun Q, Chen F, Jiao N. Novel phage-host
706 interactions and evolution as revealed by a cyanomyovirus isolated from an estuarine
707 environment. *Environ Microbiol* **20**, 2974-2989(2018). doi:10.1111/1462-2920.14326

70. Hoetzinger M, Nilsson E, Arabi R, Osbeck CMG, Pontiller B, Hutinet G, Bayfield OW, Traving S, Kisand V, Lundin D, Pinhassi J, Middelboe M, Holmfeldt K. Dynamics of Baltic Sea phages driven by environmental changes. *Environ Microbiol* **23**, 4576-4594(2021). doi:10.1111/1462-2920.15651
71. Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, Verberkmoes NC, Sullivan MB. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci U S A* **110**, 12798-12803(2013). doi:10.1073/pnas.1305956110
72. Bartlau N, Wichels A, Krohne G, Adriaenssens EM, Heins A, Fuchs BM, Amann R, Moraru C. Highly diverse flavobacterial phages isolated from North Sea spring blooms. *ISME J* **16**, 555-568(2022). doi:10.1038/s41396-021-01097-4
73. Shah S, Das R, Chavan B, Bajpai U, Hanif S, Ahmed S. Beyond antibiotics: phage-encoded lysins against Gram-negative pathogens. *Front Microbiol* **14**, 1170418(2023). doi:10.3389/fmicb.2023.1170418
74. Majorek KA, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K, Bujnicki JM. The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res* **42**, 4160-4179(2014). doi:10.1093/nar/gkt1414
75. Sencilo A, Jacobs-Sera D, Russell DA, Ko CC, Bowman CA, Atanasova NS, Osterlund E, Oksanen HM, Bamford DH, Hatfull GF, Roine E, Hendrix RW. Snapshot of haloarchaeal tailed virus genomes. *RNA Biol* **10**, 803-816(2013). doi:10.4161/rna.24045
76. Zhang W, Zhu HH, Yuan M, Yao Q, Tang R, Lin M, Yang SZ, Li ZK, Chen M. *Microbacterium radiodurans* sp. nov., a UV radiation-resistant bacterium isolated from soil. *Int J Syst Evol Microbiol* **60**, 2665-2670(2010). doi:10.1099/ijs.0.017400-0
77. Shi LX, Schroder WP. The low molecular mass subunits of the photosynthetic supracomplex, photosystem II. *Biochim Biophys Acta* **1608**, 75-96(2004). doi:10.1016/j.bbabio.2003.12.004
78. Kamiya N, Shen JR. Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7-A resolution. *Proc Natl Acad Sci U S A* **100**, 98-103(2003). doi:10.1073/pnas.0135651100
79. Pathak J, Ahmed H, Singh PR, Singh SP, Häder D-P, Sinha RP. Chapter 7 - Mechanisms of Photoprotection in Cyanobacteria. In: *Cyanobacteria* (eds Mishra AK, Tiwari DN, Rai AN). Academic Press (2019).
80. Smith BT, Walker GC. Mutagenesis and more: umuDC and the *Escherichia coli* SOS response. *Genetics* **148**, 1599-1610(1998). doi:10.1093/genetics/148.4.1599
81. Yabe S, Muto K, Abe K, Yokota A, Staudigel H, Tebo BM. *Vulcanimicrobium alpinus* gen. nov. sp. nov., the first cultivated representative of the candidate phylum "Eremiobacterota", is a metabolically versatile aerobic anoxygenic phototroph. *ISME Commun* **2**, 120(2022). doi:10.1038/s43705-022-00201-9
82. Winder JC, Boulton W, Salamov A, Eggers SL, Metfies K, Moulton V, Mock T. Genetic and Structural Diversity of Prokaryotic Ice-Binding Proteins from the Central Arctic Ocean. *Genes (Basel)* **14**, (2023). doi:10.3390/genes14020363
83. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA, Cavicchioli R. Virophage control of antarctic algal host-virus dynamics. *Proc Natl Acad Sci U S A* **108**, 6163-6168(2011). doi:10.1073/pnas.1018221108

84. Rahlff J, Westmeijer G, Weissenbach J, Antson A, Holmfeldt K. Surface microlayer-mediated virome dissemination in the Central Arctic. *Microbiome* **in press**, (2024). doi:10.1186/s40168-024-01902-0
85. Aller JY, Kuznetsova MR, Jahns CJ, Kemp PF. The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols. *J Aerosol Sci* **36**, 801-812(2005). doi:10.1016/j.jaerosci.2004.10.012
86. Vaqué D, Boras JA, Arrieta JM, Agusti S, Duarte CM, Sala MM. Enhanced viral activity in the surface microlayer of the Arctic and Antarctic Oceans. *Microorganisms* **9**, 317(2021). doi:10.3390/microorganisms9020317
87. Alsante AN, Thornton DCO, Brooks SD. Ocean Aerobiology. *Front Microbiol* **12**, 764178(2021). doi:10.3389/fmicb.2021.764178
88. Das R, Rahlff J. Phage genome architecture and GC content: Structural genes and where to find them. *bioRxiv*, 2024.2006.2005.597531(2024). doi:10.1101/2024.06.05.597531
89. Hwang Y, Rahlff J, Schulze-Makuch D, Schloter M, Probst AJ. Diverse viruses carrying genes for microbial extremotolerance in the Atacama Desert hyperarid soil. *mSystems* **6**, 10.1128/msystems.00385-00321(2021). doi:10.1128/mSystems.00385-21
90. Irby I, Broddrick JT. Microbial adaptation to spaceflight is correlated with bacteriophage-encoded functions. *Nat Commun* **15**, 3474(2024). doi:10.1038/s41467-023-42104-w
91. Heyerhoff B, Engelen B, Bunse C. Auxiliary Metabolic Gene Functions in Pelagic and Benthic Viruses of the Baltic Sea. *Front Microbiol* **13**, 863620(2022). doi:10.3389/fmicb.2022.863620
92. Sharon I, Tzahor S, Williamson S, Shmoish M, Man-Aharonovich D, Rusch DB, Yooseph S, Zeidner G, Golden SS, Mackey SR, Adir N, Weingart U, Horn D, Venter JC, Mandel-Gutfreund Y, Beja O. Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J* **1**, 492-501(2007). doi:10.1038/ismej.2007.67
93. Wang SY, Chen YP, Huang RF, Wu YL, Ho ST, Li KY, Watanabe K, Chen MJ. Subspecies classification and comparative genomic analysis of *Lactobacillus kefiranofaciens* HL1 and M1 for potential niche-specific genes and pathways. *Microorganisms* **10**, 1637(2022). doi:10.3390/microorganisms10081637
94. Goldman JC. Oceanic nutrient cycles. In: *Flows of Energy and Materials in Marine Ecosystems: Theory and Practice* (ed Fasham MYR). Springer (1984).
95. Polymenakou PN. Atmosphere: A Source of Pathogenic or Beneficial Microbes? *Atmosphere* **3**, 87-102(2012). doi:10.3390/atmos3010087
96. Hou J, Fujiyoshi S, Perera IU, Nishiuchi Y, Nakajima M, Ogura D, Yarimizu K, Maruyama F. Perspectives on Sampling and New Generation Sequencing Methods for Low-Biomass Bioaerosols in Atmospheric Environments. *J Indian Inst Sci* **103**, 1-11(2023). doi:10.1007/s41745-023-00380-x
97. Harnpicharnchai P, Pumkaeo P, Siriarchawatana P, Likhitrattanapisal S, Mayteeworakoon S, Ingsrisawang L, Boonsin W, Eurwilaichitr L, Ingsriswang S. AirDNA sampler: An efficient and simple device enabling high-yield, high-quality airborne environment DNA for metagenomic applications. *PLoS One* **18**, e0287567(2023). doi:10.1371/journal.pone.0287567
98. Highlander SK, Wood JM, Gillece JD, Folkerts M, Fofanov V, Furstenau T, Singh NK, Guan L, Seuylemezian A, Benardini JN, Engelthaler DM, Venkateswaran K, Keim PS. Multi-faceted metagenomic analysis of spacecraft associated surfaces reveal planetary

protection relevant microbial composition. *PLoS One* **18**, e0282428(2023). doi:10.1371/journal.pone.0282428

99. Zhang Y, Xin CX, Zhang LT, Deng YL, Wang X, Chen XY, Wang ZQ. Detection of Fungi from Low-Biomass Spacecraft Assembly Clean Room Aerosols. *Astrobiology* **18**, 1585-1593(2018). doi:10.1089/ast.2017.1803
100. Verreault D, Gendron L, Rousseau GM, Veillette M, Masse D, Lindsley WG, Moineau S, Duchaine C. Detection of airborne lactococcal bacteriophages in cheese manufacturing plants. *Appl Environ Microbiol* **77**, 491-497(2011). doi:10.1128/AEM.01391-10
101. Letourneau V, Gagne MJ, Vyskocil JM, Brochu V, Robitaille K, Gauthier M, Brassard J, Duchaine C. Hunting for a viral proxy in bioaerosols of swine buildings using molecular detection and metagenomics. *J Environ Sci (China)* **148**, 69-78(2025). doi:10.1016/j.jes.2023.08.017
102. Boifot KO, Skogan G, Dybwad M. Sampling efficiency and nucleic acid stability during long-term sampling with different bioaerosol samplers. *Environ Monit Assess* **196**, 577(2024). doi:10.1007/s10661-024-12735-7
103. Langenfeld K, Chin K, Roy A, Wigginton K, Duhaime MB. Comparison of ultrafiltration and iron chloride flocculation in the preparation of aquatic viromes from contrasting sample types. *PeerJ* **9**, e11111(2021). doi:10.7717/peerj.11111
104. Els N, Larose C, Baumann-Stanzer K, Tignat-Perrier R, Keuschnig C, Vogel TM, Sattler B. Microbial composition in seasonal time series of free tropospheric air and precipitation reveals community separation. *Aerobiologia* **35**, 671-701(2019). doi:10.1007/s10453-019-09606-x