Host-associated phages disperse across the extraterrestrial analogue Antarctica

Running title: Dispersal of Antarctic phage

Janina Rahlff[a, #], Till L.V. Bornemann[a], Anna Lopatina[b], Konstantin Severinov[b], Alexander J. Probst[a,c]

[a]Group for Aquatic Microbial Ecology, University of Duisburg-Essen, Department of Chemistry, Environmental Microbiology and Biotechnology (EMB), Universitätsstraße 5, 45141 Essen, Germany

[b]Institutes of Molecular Genetics and Gene Biology of the Russian Academy of Sciences, Moscow, Russia

[c]Centre of Water and Environmental Research (ZWU), University of Duisburg-Essen, Universitätsstraße 5, 45141, Essen, Germany

#Corresponding author: Janina Rahlff, Janina.rahlff@uol.de

Present address: Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Department of Biology and Environmental Science, Linnaeus University, SE-39182, Kalmar, Sweden

Key words: viruses, extreme environments, astrovirology, metagenomics, cryosphere, *Ralstonia*

Abstract

Extreme Antarctic conditions provide one of the closest analogues of extraterrestrial environments. Since air and snow samples, especially from polar regions, yield DNA amounts in the lower picogram range, binning of prokaryotic genomes is challenging and renders studying the dispersal of biological entities across these environments difficult. Here, we hypothesized that dispersal of host-associated bacteriophages (adsorbed, replicating or prophages) across the Antarctic continent can be tracked via their genetic signatures aiding our understanding of virus and host dispersal across long distances. Phage genome fragments (PGFs) reconstructed from surface snow metagenomes of three Antarctic stations were assigned to four host genomes, mainly Betaproteobacteria including *Ralstonia* spp. We reconstructed the complete genome of a temperate phage with near-complete alignment to a prophage in the reference genome of *Ralstonia pickettii* 12D. PGFs from different stations were related to each other at the genus level and matched similar hosts. Metagenomic read mapping and nucleotide polymorphism analysis revealed a wide dispersal of highly identical PGFs, 13 of which were detected in seawater from the Western Antarctic Peninsula in distance of 5538 km to the snow sampling stations. Our results suggest that host-associated phages, especially of *Ralstonia* sp., disperse over long distances despite harsh conditions of the Antarctic continent. Given that 14 phages associated with two *R. pickettii* draft genomes isolated from space equipment were identified, we conclude that *Ralstonia* phages are ideal mobile genetic elements to track dispersal and contamination in ecosystems relevant for Astrobiology.

48 **Importance**

49 Host-associated phages of the bacterium *Ralstonia* identified in snow samples can be used to
50 track microbial dispersal over thousands of kilometers across the Antarctic continent, which
51 functions as an extraterrestrial analogue because of its harsh environmental conditions. Due to
52 presence of these bacteria carrying genome-integrated prophages on space-related equipment,
53 and the here demonstrated potential for dispersal of host-associated phages, our work has
54 implications for Planetary Protection, a discipline in Astrobiology interested in preventing
55 contamination of celestial bodies with alien biomolecules or forms of life.

56

57 **Introduction**

58 Due to harsh environmental conditions and isolation by the surrounding Southern Ocean's
59 Circumpolar Current, Antarctica is considered an analogue for multipurpose space exploration
60 (1, 2). For example, its McMurdo Dry Valleys are regarded as a close terrestrial analogue to
61 Mars (3). Astrobiology model organisms found on Antarctica are highly adapted to stressful
62 conditions and comprise prokaryotes such as spore-forming *Bacilli* (4, 5), but also microfungi
63 (3, 6). Understanding endurance and dispersal of microorganisms under conditions that mimic
64 those on extraterrestrial planets, i.e., high UV radiation, low temperature, and low nutrient
65 availability, has important implications for Planetary Protection. For instance, the dispersal of
66 microbes that hitchhike to a celestial body is currently not considered in Planetary Protection,
67 a discipline in Astrobiology set out with the aim of preventing contamination of celestial bodies
68 with foreign biomolecules or forms of life.

69

70 Among the potential candidates hitchhiking spacecraft are Betaproteobacteria of the genus
71 *Ralstonia* (order *Burkholderiales*). These bacteria are able to thrive under oligotrophic
72 conditions (7) and were reported to be ubiquitously present on space-related equipment
73 including water systems of the International Space Station (ISS) (8, 9), and the Mir space station
74 (10, 11). Likewise, they belonged to the microbial inventory of Mars Odyssey and Mars
75 Phoenix lander facilities and can thus prevail under strict Planetary Protection regulations (12,
76 13). *Ralstonia pickettii* strains were found to thrive in simulated microgravity compared to
77 normal gravity (11) and demonstrated high resistance against different metal ions and UV-C
78 radiation (8).
79
80 *Ralstonia* spp., (mainly *R. pickettii*), were previously found Antarctic soils (14), Antarctic snow
81 (15-17), in snow over Tibetan Plateau Glaciers (18), and in the air of the Antarctic base
82 Concordia (19). Interestingly, this genus has been regarded as an atmospheric traveler rather
83 than being part of true snow microflora in Antarctica (15). Despite a report on bacterial activity
84 at subzero temperatures in South pole snow (20), *Ralstonia* was not among active bacterial
85 communities as inferred from cDNA-based 16S rRNA amplicon sequencing (15). This view
86 was supported by a comprehensive Antarctic surface snow microbiome study that did not detect
87 this bacterial genus (21), and by the fact that spatial variability of snow microbiomes in
88 Antarctica is high (22). Although aerial dispersal is probably the major contributor to (micro-
89)biological input to remote regions (23), the role of bioaerosol transport to microbial ecology
90 of isolated systems such as the Antarctic continent is poorly understood (24). Applying high
91 throughput sequencing approaches to study bacterial dispersal over the Antarctic continent
92 remains a considerable challenge due to the low microbial biomass of atmosphere-derived
93 samples regarding their DNA content (25), and resulting issues of recovering high quality
94 assemblies from metagenomic reads (26).
95

96 In addition to the limited knowledge about how transport via aerosols and snow across
97 Antarctica shapes microbial dispersal patterns, another open question relates to the distances
98 that microbes can cover within the atmosphere of extreme environments. A study on the
99 dispersal of airborne faecal coliforms showed a distribution over about just 175 m from a
100 sewage outfall at Rothera Research Station (Antarctic Peninsula) and thus prolonged survival
101 was considered unlikely (27). More stress-resistant microorganisms could, however, endure for
102 much longer periods. Recently, L. A. Malard et al. (22) found high abundances of spore-
103 forming *Bacilli* and suggested that long-term dispersal may seed continental Antarctic snow
104 ecosystems. However, to date, the role of aerial dispersal in shaping patterns of microbial
105 biogeography is supported by little empirical evidence (23).
106
107 Here, we follow the hypothesis that geographically widespread (28) *Ralstonia* spp. prophages
108 and/or replicating and adsorbed bacteriophages of this genus (all types further referred to as
109 "host-associated" phages) can be used to study host bacterium dispersal across the Antarctic
110 continent. Reconstructing prokaryotic genomes from samples containing low amounts of DNA,
111 including air or precipitation is challenging, as low input libraries (~1 pg) can result in problems
112 of genome binning (26). However, (pro)phage genomes or their fragments are much smaller
113 than prokaryotic genomes and thus easier to identify, track and compare. In this study, genome-
114 resolved metagenomics was applied to demonstrate dispersal of host-associated phage genome
115 fragments (PGFs) from surface snow across the Antarctic continent over hundreds to thousands
116 of kilometers. We detected PGFs belonging to the orders *Caudovirales* and *Tubulavirales* in
117 this extraterrestrial analogue and additionally show that similar genome-integrated phages are
118 frequent colonizers of space equipment. Therefore, we suggest that host-associated PGFs
119 represent a useful tool to study spatial dispersal of bacteria and their phages in extreme
120 environmental settings and further envision implications for the dispersal of microbiological

121 contaminations on spacecraft and celestial bodies that previously escaped Planetary Protection
122 measures.
123

124 **Results**

125 **Reconstruction of low coverage MAGs from Antarctic snow metagenomic data**

126 The microbial community composition of surface snow samples collected close to three Russian
127 Antarctic stations, Druzhnaja, Mirnii and Progress, based on 16S rRNA gene sequencing was
128 described earlier (16) and showed that *Ralstonia* was the most dominant organism in snow
129 collected at the Mirnii station, the second most dominant after *Janthinobacterium* at Druzhnaja,
130 and the third most dominant (after *Flavobacteria* and *Hydrogenophaga*) around Progress
131 station. We reconstructed four prokaryotic and one eukaryotic metagenome assembled genomes
132 (MAG) from the three low-biomass snow metagenomes (Figure 1A). Two MAGs related to
133 *Ralstonia* sp. and *R. pickettii* were recovered from Druzhnaja and Mirnii samples and had
134 86%/10% and 55%/0% completeness/contamination scores, respectively. MAGs of
135 *Janthinobacterium lividum* with 92%/6% scores and of *Flavobacterium micromati* with
136 96%/0% scores were recovered from Druzhnaja and Progress, respectively. We also identified
137 a MAG of a diatom (likely *Thalassiosira* sp.) in the Mirnii metagenome, which we used for
138 normalizing some of our viral analysis (see below) but was not further characterized in this
139 work. Read mapping revealed coverage scores of 5.9, 7.1, 6.7 and 14.6 for the MAGs of
140 *Ralstonia* sp., *R. pickettii, J. lividum* and *F. micromati*, respectively.

141 **Prevalent absence of CRISPR-Cas systems suggest low adaptive immunity**

142 We searched for clustered regularly interspaced short palindromic repeats (CRISPR) spacers in
143 the snow metagenome reads to link them to potential protospacers on PGFs. CRISPR arrays

144 and *cas* genes were absent from both *Ralstonia* and the *J. lividum* MAGs as determined by
145 CRISPRcasFinder (29). BLASTing of direct repeat (DR) sequences to NCBI's NR database
146 also did not indicate *Ralstonia* or *Janthinobacterium* to be the host of the CRISPR array. Solely
147 the genome of *F. micromati* contained two CRISPR arrays with four spacers each (both
148 evidence level 3 = highly likely candidates). CRISPR arrays were not detected on *R. pickettii*
149 draft genomes SSH4 and CW2 obtained from space equipment.

150

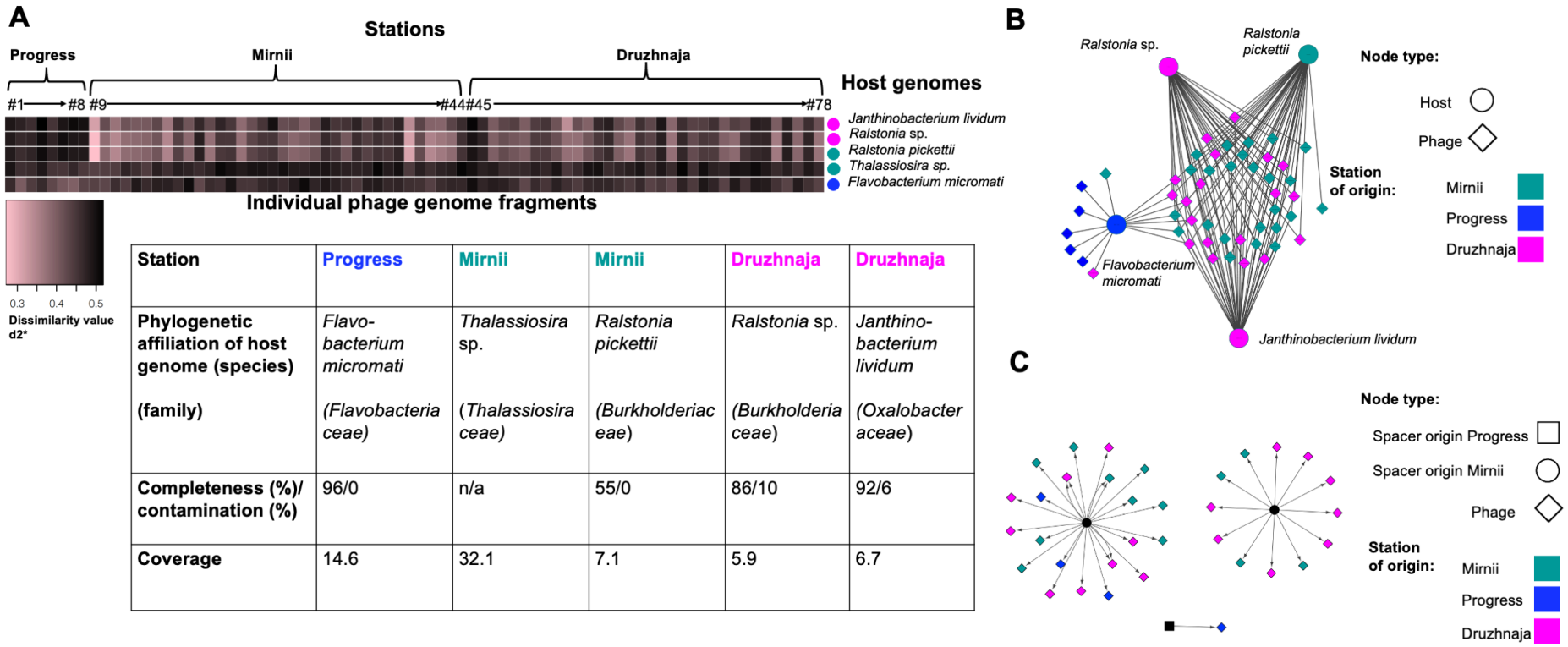
151 **PGF-host analyses suggest shared hosts of Antarctic phages**

152 A total of 26 predicted PGFs, eight of them being putative PGFs, was found in Antarctic snow
153 metagenomic data (Table S1). VIBRANT (30) additionally detected 52 PGFs, resulting in a
154 total of 78 PGFs (Figure 1A). PGFs with minimum 75% of the genome covered with reads had
155 coverages ranging between 1.9 and 17.5. Most PGFs were partial sequences according to
156 viralComplete (31) and CheckV (32), only Antarcphage10_Mi_4716 and
157 Antarcphage49_Dr_7823_circ were estimated to be of full length (Table S1), although
158 annotations of Antarcphage10_Mi_4716 in comparison to *Ralstonia* phage p12J revealed
159 missing genes and let us question its completeness (Figure 2). Out of 78 PGFs, 77 were
160 categorized as lytic by VirSorter (33) and 75 were found to be of viral or unclassified origin by
161 CheckV (Table S1). A proportion of 43.6%, 46.1% and 10.3% identified PGFs originated from
162 the Druzhnaja, Mirnii and Progress, respectively.

163 All 78 PGFs were analyzed in conjunction with reconstructed MAGs using VirHostMatcher
164 (34). We defined a dissimilarity threshold of d2* value = 0.436, which corresponds to the lowest
165 dissimilarity value (= highest similarity) for PGFs matching the MAG of *Thalassiosira* sp. from
166 the snow environment (Figure 1A) by assuming that the eukaryotic MAG does not match any
167 of our extracted prokaryote infecting PGFs. In total, 50 of the 78 PGFs matched a host MAG

168 below the defined dissimilarity threshold, based on their shared k-mer patterns. Most PGFs (41)
169 matched both MAGs of *Ralstonia* sp. and *R. pickettii* (Figure 1A). In total, 38 and 14 PGFs
170 matched *J. lividum* and *F. micromati* MAGs, respectively. Five out of eight PGFs extracted
171 from the Progress station metagenomic snow sequences matched the *F. micromati* MAG
172 recovered from this station. Seven PGF matches were shared between all prokaryotic MAGs
173 (Figure S1). However, with 30 shared matches, the three MAGs belonging to the order
174 *Burkholderiales* shared the most overlap (Figure 1B, Figure S1).

175 Matching of CRISPR spacers derived from CRISPR loci of unknown hosts (reconstructed from
176 DR sequences) revealed that spacers from Mirnii matched 19, 14 and three PGFs from
177 Druzhnaja, Mirnii and Progress, respectively, and one *Flavobacterium* sp. spacer from Progress
178 matched a PGF from that station (Figure 1C, Table S2, ≥80% similarity). Out of these 37 spacer
179 matches, three matched protospacers of PGFs of unknown hosts, and 26 were assigned to a
180 *Ralstonia* host according to VirHostMatcher or other predictions (Table S1). Since p12D (see
181 below) was among the PGF spacer targets and is a certain *Ralstonia* phage, this could indicate
182 that an unknown host belonging to the order *Burkholderiales* uses adaptive immunity against
183 viruses. We compiled evidence of host prediction for the 78 PGFs in Table S1.
184



Station	Progress	Mirnii	Mirnii	Druzhnaja	Druzhnaja
Phylogenetic affiliation of host genome (species)	*Flavo-bacterium micromati*	*Thalassiosira* sp.	*Ralstonia pickettii*	*Ralstonia* sp.	*Janthino-bacterium lividum*
(family)	*(Flavobacteria ceae)*	*(Thalassiosira ceae)*	*(Burkholderiac eae)*	*(Burkholderia ceae)*	*(Oxalobacter aceae)*
Completeness (%)/ contamination (%)	96/0	n/a	55/0	86/10	92/6
Coverage	14.6	32.1	7.1	5.9	6.7

Figure 1: Host-phage pairings based on shared k-mer frequency patterns. A) Heatmap representing dissimilarity value d2* for matches between five host MAGs and 78 PGFs derived from three Antarctic snow metagenomes from the stations Progress, Mirnii, and Druzhnaja. The PGF number corresponds to the number in the phage name, for instance PGF#78 refers to Antarcphage78_Dr_3477 (Table S1). The table shows MAG characteristics (phylogenetic affiliation, completeness, contamination, and coverage). n/a = not available **B)** Network of phage-host interactions based

190 on k-mer frequency pattern reveals strong overlap between *Ralstonia* and *Janthinobacterium lividum* infecting PGFs. Here, only PGFs matching the
191 host MAG below the defined threshold (see text) are shown. **C)** CRISPR spacer matches to PGFs. Different spacers matching to the same PGFs are
192 shown by multiple arrows. The two hosts of CRISPR arrays for Mirnii-derived spacers remain unidentified according to their direct repeat sequence,
193 whereas the Progress spacer is derived from a *Flavobacterium* sp. (Table S2).

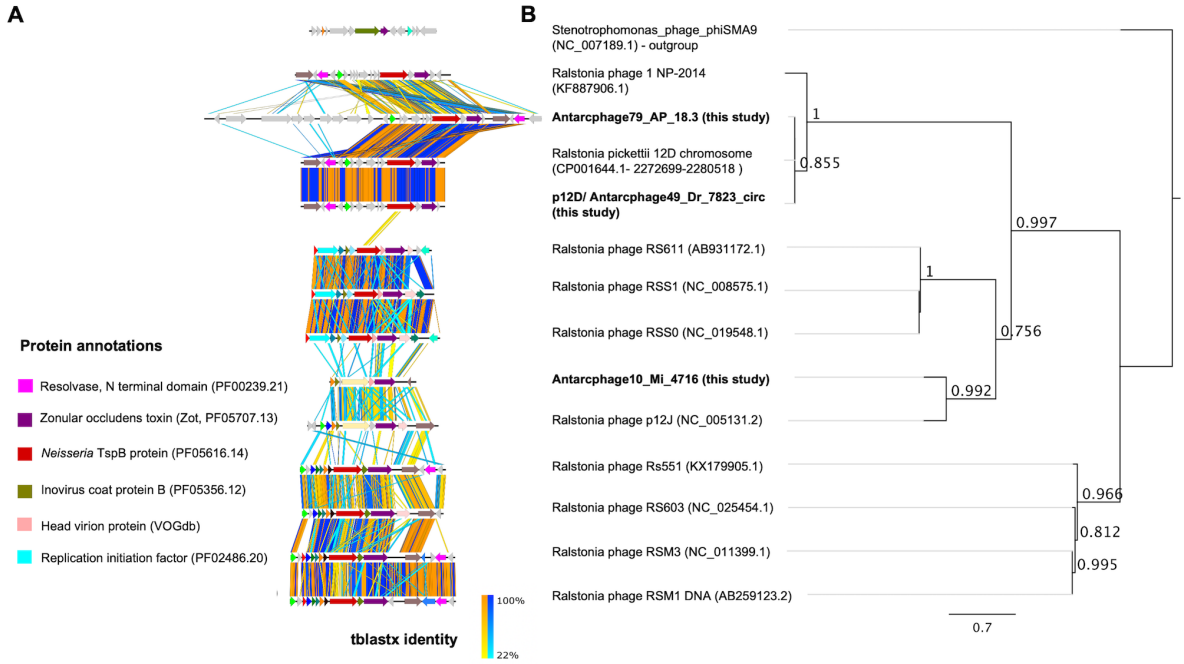
194

195 **The temperate phage p12D forms a distinct, monophyletic clade excluding most known**
196 ***Ralstonia* phages**

197 Among the 78 PGFs, we detected a circular (and thus complete) 7.8 kb PGF termed
198 Antarcphage49_Dr_7823_circ (Figure 2A, Figure S2). Since the Antarcphage49_Dr_7823_circ
199 PGF from the Druzhnaja station was found in the chromosome of a *R. pickettii* 12D strain
200 isolated from copper-contaminated sediment from a lake in Michigan with 99.9% identity
201 (accession number: CP001644.1- 2272699-2280518, Figure 2A), we here propose the name
202 p12D phage equivalent to *Ralstonia* p12J, a known phage infecting the *R. pickettii* 12J strain.
203 Antarcphage49_Dr_7823_circ (from now on referred to as p12D) only matched the MAG of
204 *Ralstonia* sp. and *R. pickettii* based on shared k-mer frequency patterns. Functional protein
205 annotations provided evidence that p12D contained a gene for a resolvase domain containing
206 protein/site-specific recombinase, likely used for integration into the host genome, as well as
207 for the zonular occludens toxin (Zot, PF05707) (Figure 2A, Figure S2, Table S3). The non-
208 toxic component of Zot at the N-terminus represents a characteristic protein in filamentous
209 phages that has been used for phage classification (35, 36). Reconstruction of a phylogenetic
210 relationships of Zot proteins from this study and respective references (Figure 2B, Figure S3)
211 confirmed a close identity of p12D to *Ralstonia* phage 1 NP-2014 and
212 Antarcphage79_WAP_18.3, all belonging to the same monophyletic clade. *Zot* of
213 Antarcphage10_Mi_4716 was phylogenetically related to filamentous *Ralstonia* phage p12J.
214 The tree shows four distinctive clusters for the Zot protein, reflecting the overall synteny of the
215 gene order of the different *Ralstonia* phages very well.
216 Annotations against UniRef100 revealed that 68.8% of all annotated Antarctic PGF proteins
217 remain hypothetical. The annotations taxonomically assigned 34.4% of all Antarctic PGF
218 proteins to either *Ralstonia* or *R. pickettii,* and 11.6 % to phages of *Janthinobacterium* or *J.*

219 *lividum*, potentially indicating lateral gene transfer between virus and host in their respective
220 evolution (37, 38) or supporting that these PGF indeed represent prophages. Three of the
221 Antarctic PGFs carried a site-specific integrase or resolvase domain, and 4.2% of genes were
222 related to phage structural proteins, e.g., head, tail, or capsid proteins (Table S4, Table S5).

223
224



225

226 **Figure 2: Phage genome comparisons, functional annotations and phylogenetic**
227 **relationship based on zonular occludes toxin (Zot, Pfam-ID: PF05707). A)** Synteny of
228 known *Ralstonia* sp. phages from NCBI and Antarctic PGFs from this study with similar coding
229 sequences (CDS) being in the same color if they occur on more than one phage genome.
230 Functional annotations performed by DRAM-v (115) are given for all colored CDS where
231 available. Vertical lines between sequences indicate regions of shared similarity shaded
232 according to tBLASTx (orange gradient for matches in the same direction or blue gradient for
233 inverted matches). Figure was created using Easyfig (119). **B)** The phylogenetic tree was built
234 using FastTree 2.1.11 (122) in Geneious 11.1.5 (112) under default settings and shows four
235 distinct clusters of Zot proteins based on their amino acid sequences (aligned with MUSCLE
236 (121)).

237 **Cross-mapping and nucleotide variations reveal dispersal patterns of MAGs and PGFs**
238 **across Antarctica**

239 Mapping of 100% identical reads from the three snow samples and the Western Antarctic
240 Peninsula (WAP, Figure 3A) seawater sample to the prokaryotic MAGs revealed that both
241 *Ralstonia* MAGs were detected at all sites (89-100% genome coverage). *J. lividum* was
242 considered absent from the Progress station (genome coverage of 62%) and detected at other
243 stations (minimum 95% covered genome). *F. micromati* was only detected in Progress (100%)
244 and Druzhnaja (96%) but considered absent from Mirnii (58.7%) and the WAP (1.8%). Cross-
245 mapping on the 78 PGFs demonstrated that PGFs derived from the Progress station could not
246 be found in the other two snow metagenomes (Figure 3B&C). By contrast, 16 different PGFs
247 from Druzhnaja were detected at Mirnii, and 27 Mirnii PGFs were found at Druzhnaja (Figure
248 3C). In addition, 4 and 9 PGFs from Mirnii and Druzhnaja were detected in the WAP dataset,
249 respectively (Figure 3 B&C), all of them having minimum 97% of their lengths covered with
250 reads from the snow metagenomes (Table S1). Of 43 PGFs that were found at both stations
251 (Mirnii and Druzhnaja) based on read mapping, 26 PGFs shared a viral cluster (VC), and 34
252 matched prokaryotic hosts based on k-mer frequencies between stations (Table S1).
253 Interestingly, nine of the 13 PGFs occurring in WAP and Mirnii/Druzhnaja samples contained
254 identical single nucleotide polymorphisms (SNPs, Figure 3D&E, Table S6) or were missing
255 common SNPs pointing towards a common phage population before dispersal led to separation.

256

257 Clustering of PGFs in VICTOR (39) and vConTACT2 (40) revealed that groups of two to four
258 PGFs could be assigned to twelve distinct viral genera-forming viral clusters (Figure 3E, Figure
259 S4 & S5, Table S1). Members of the same cluster were often recovered from different stations,
260 for example, Druzhnaja and Mirnii (Figure 3B), which are located 923 km apart (Figure 3C).
261 The phylogenomic Genome-BLAST Distance Phylogeny (GBDP) tree yielded average support

of 47% and 80% in the nucleic acid and amino acid-based analysis, respectively. OPTSIL clustering yielded 78 and 64 species and genus clusters at the nucleic acid level and 78 and 70 at the amino acid level, respectively.

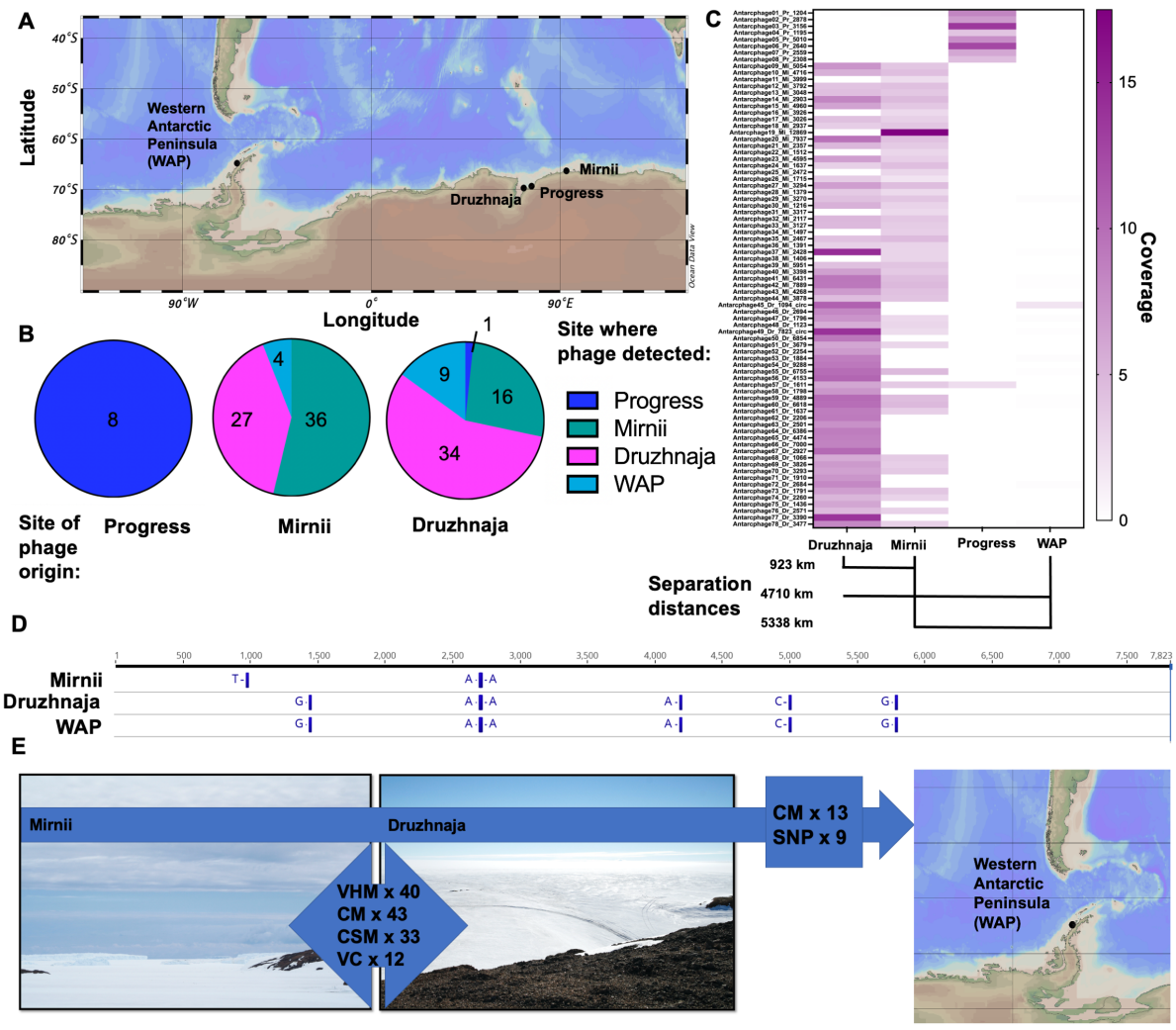
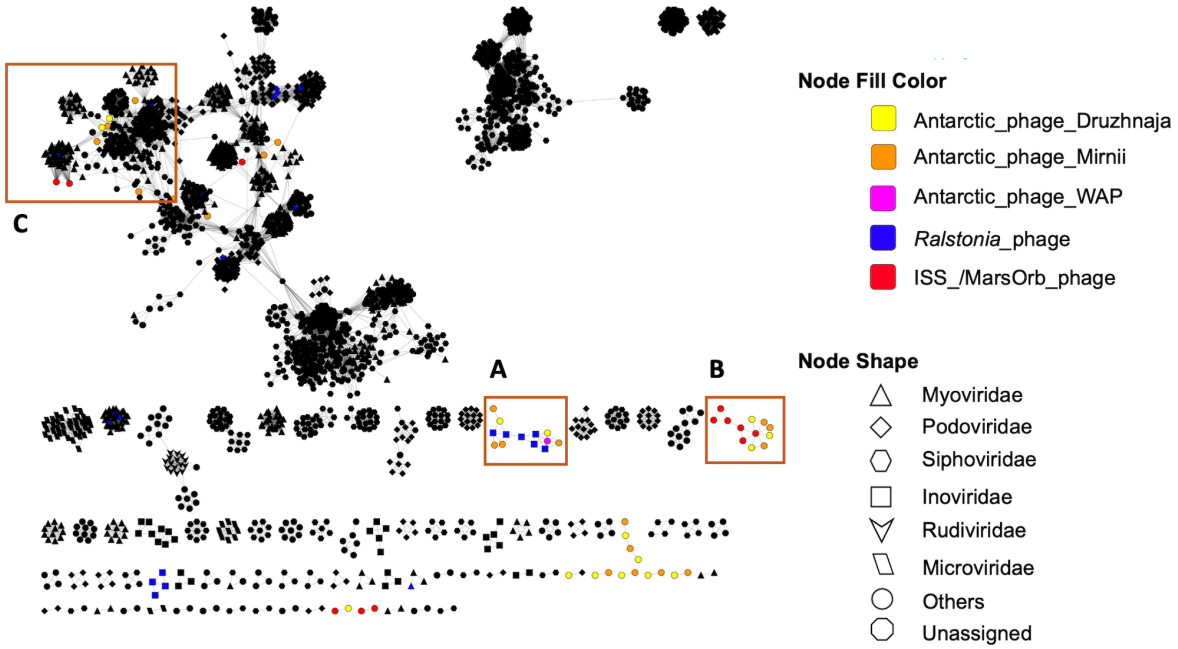


Figure 3: Evidence for dispersal of PGFs across the Antarctic continent. A) A map showing the different snow sampling stations in the East and the sampling station for seawater at the Western Antarctic Peninsula. Map was built using Ocean Data View (123). **B)** Pie charts summarizing the number of PGFs considered present at each station based on cross mapping of reads and the site of PGF assembly (phage origin). **C)** The heat map depicts the normalized coverages based on cross-mapping of reads against the 78 PGFs and separation distances between stations. White areas indicate PGFs that were not called present because they had less than 75% of scaffold length covered in mapping (with least 1-fold coverage). **D)** Single-

274 nucleotide polymorphism (SNP) analysis of p12D/Antarcphage49_Dr_7823_circ based on
275 reads from Druzhnaja, Mirnii and the WAP. This PGF was absent from the metagenome of the
276 Progress station. Further SNP analysis data can be found in Table S6. **E)** Summary figure for
277 the major dispersal route and supporting evidence. The majority of PGF disperse between
278 Mirnii and Druzhnaja, and 13 PGFs additionally occurred at the WAP based on cross-mapping
279 (CM) and SNP analysis. Values refer to the number of tested features, which include number
280 of virus clusters (VC), number of shared viruses based on CM, virus-host matches based on k-
281 mer links (VHM), CRISPR-spacer matches (CSM). A VHM=1 indicates that one PGF has
282 infected host MAGs from two stations. No VC and only one CM were observed for Progress-
283 Druzhnaja and Progress-Mirnii. Therefore, this station was excluded from the summary.

284 Many PGFs formed VCs with no genomic relatedness to any known phages from the
285 ProkaryoticViralRefSeq v94 database. Others were, however, similar to known *Ralstonia* PGFs
286 (current family *Inoviridae,* order *Tubulavirales*) and hence clustered with those, such as p12D
287 with *Ralstonia* phage 1 NP-2014, or Antarcphage10_Mi_4716 with *Ralstonia* phage PE226 and
288 *Ralstonia* phage p12J (Figure 4, Table S1). Other Antarctic PGFs shared protein clusters with
289 *Ralstonia* phages from space equipment based on vConTACT2 or were related to phages of the
290 *Caudovirales* order from the RefSeq database (Figure 4).



291

292 **Figure 4: Phage network of Antarctic PGFs derived from Druzhnaja, Mirnii, Western**
293 **Antarctic Peninsula (WAP) and space equipment (ISS/MarsOrb) clustered with viruses**
294 **of the viral RefSeq database.** Based on shared protein clusters, Antarctic PGFs from this study
295 group with known *Ralstonia* phage of the family *Inoviridae* (A), with phages obtained from
296 *Ralstonia* isolates from space equipment (B) or with known phages of the order *Caudovirales*
297 (C). Black nodes refer prokaryotic viruses other than *Ralstonia* and PGFs from this study.
298 Interactions show relatedness of genomes on viral genus or higher taxonomic level. "Others"
299 refer to other known viral families not listed in the legend. Visualization was done using
300 Cytoscape 3.8.2. (109). For details about viral clusters, please see Table S1.

301

302 In summary, our results support the idea of dispersal of host-associated phages between widely
303 separated stations because a) cross-mapping revealed presence of Mirnii PGFs at Druzhnaja
304 and vice versa, presence of a temperate phage from snow in the WAP seawater dataset, and
305 presence of the hosts *Ralstonia* and *J. lividum* in metagenomes of most stations; b) PGFs in
306 widely separated locations carry identical SNPs or lack nucleotide variations; c) some PGFs
307 from Mirnii and Druzhnaja belong to the same genus/VC; and d) Mirnii and Druzhnaja PGFs
308 share host MAGs according to their k-mer frequency patterns. These observations were mainly
309 true for PGFs affiliated to *Ralstonia* and *J. lividum* MAGs and less so for *F. micromati* PGFs,
310 which could indicate that certain bacterial species or their phages in Antarctic snow are more
311 prone to atmospheric dispersal than others. Finally, we found CRISPR spacers from two
312 unidentified hosts (represented by two distinct CRISPR DR sequences) at the Mirnii station to
313 match PGFs from Druzhnaja, Mirnii and Progress. Mapping revealed little to no abundance of
314 Progress PGFs at other stations, and no *F. micromati* MAG was detected in Mirnii and WAP
315 samples. This could be related to the low biomass and/or or insufficient sequencing coverage.
316 However, spacers from Mirnii suggest an infection history with PGF from Progress, thus
317 reflecting (past) dispersal of host and/or phage.

318

319 ***Ralstonia* phages occur across diverse ecosystems and on space equipment**

320 By BLASTing all recovered PGFs against the IMG/VR viral database (41), we found that p12D
321 shared high identity (85.9 %) with a *R. pickettii* prophage (IMG/VR v.2 scaffold ID
322 Ga0075447_10000781, genome ID: 3300006191) from a seawater metagenome of the WAP
323 (Figure 2), which was further validated by cross-mapping of WAP reads to p12D delivering
324 100% scaffold coverage with 90% identical reads. The WAP is 4710 km away from the

325 Druzhnaja station (Figure 3C). The PGF Antarcphage79_WAP_18.3 was obtained by
326 BLASTing the p12D scaffold against the WAP assembly but was assembled containing
327 genomic regions extending the actual PGF. Based on vConTACT2, it was affiliated to the same
328 viral genus as p12D. Other PGFs recovered from the Antarctic snow metagenome showed hits
329 to phages deposited at IMG/VR. Hits were related to phages originating from freshwater,
330 wastewater, groundwater, or phages that were associated with plant root microbial
331 communities, which is in accordance with the fact that *Ralstonia* sp. is a frequent
332 phytopathogen (42). Many of the matching entries in IMG/VR were identified as phages of
333 *Ralstonia* due to matching with CRISPR spacers (mostly of *R. solanacearum,* Table S7).

334

335 Altogether, 14 PGFs were found on space equipment, hereafter referred to as ISSphage (7) and
336 MarsOrbphage (7), for PGFs extracted from the *R. pickettii* strains CW2 and SSH4 draft
337 genomes from the ISS cooling system and Mars Odyssey Orbiter, respectively. Seven of these
338 PGFs were identified as lysogenic and according to protein annotations (Table S4 & S5) rely
339 on proteins of the integrase family for integration into the host chromosome. vConTACT2
340 revealed that some of the ISSphage4 formed a common genus cluster with four Antarctic PGFs,
341 MarsOrbphage5 shared protein clusters with ISSphage3, and MarsOrbphage7 formed a genus
342 cluster with several *Burkholderia* phages among other *Caudovirales* members (Figure 4, Table
343 S1).

344

345 **Discussion**

346 CRISPR-Cas systems equip bacteria and archaea with a powerful defense system against
347 invading mobile genetic elements including plasmids, phages and viruses. Only ca. 50% of
348 bacteria rely on this adaptive immune system, and only 31% of genomes from public databases
349 belonging to plant-pathogenic *Ralstonia solanacearum* contain CRISPR-Cas arrays (43). Upon

350 exposure to a virulent phage under laboratory conditions, the CRISPR array of *R. solanacearum*
351 strain CFBP2957 did not acquire new spacers from viral protospacers (43). O. S. Gonçalves et
352 al. (28) reported that in the presence of CRISPR arrays, 27.9% of CRISPR spacers from
353 *Ralstonia* genomes targeted prophage elements. In our study, two *Ralstonia* MAGs from a low
354 temperature environment were devoid of CRISPR systems, implying that representatives of this
355 genus have alternative strategies for defense against mobile genetic elements (43). However,
356 analyses on more *Ralstonia* genomes are necessary to corroborate this statement.

357 Based on shared protein clusters, some Antarctic PGFs such as Antarcphage10_Mi_4716,
358 Antarcphage48_Dr_1123, Antarcphage17_Mi_3026, and p12D clustered with known
359 *Ralstonia* phages including *Ralstonia* phage p12J (NC_005131.2), *Ralstonia* phage PE226
360 (NC_015297.1) or *Ralstonia* phage 1 NP-2014 (NC_023586.1). These are filamentous phages
361 of the *Inoviridae*, a family that has been recently called for reclassification to higher taxonomic
362 ranks (36, 44). Inoviruses (order *Tubulavirales*) typically feature circular, single stranded DNA
363 genomes of ~5-15 kb length, lead to chronic infections, and are globally abundant (36). As our
364 protocol should have only revealed double-stranded DNA, detection of ssDNA inoviruses
365 sequences shows that these must be either replicating phages or genome-integrated prophages.
366 *Zot*, which was detected on p12D, Antarcphage10_Mi_4716 and Antarcphage79_WAP_18.3 is
367 a typical gene in filamentous *Ralstonia* phages (45, 46). This phage type confers little burden
368 to its host or can even serve it, e.g., by increasing its hosts' virulence and evolutionary fitness,
369 and because virions can leave the host in a non-destructible way (35, 36). *Ralstonia* phages
370 detected in Arctic viromes were shown to transduce genomic information of cold-shock
371 proteins to their hosts (47), a clear asset for microorganisms in polar environments. However,
372 we did not find transduction of beneficial genes, which seems to be common in extreme
373 environments (48).
374

375 Some *Ralstonia* phages occur as non-integrative, episomal forms, e.g., RS603, a hybrid of
376 RSM1/3 infecting the phytopathogen *R. solanacearum* (49), whose genome lacks a resolvase
377 domain (Figure 2A), but many mesophilic *Ralstonia* also occur as lysogens (50). Lysogeny, a
378 lifestyle during which the phage genome becomes integrated into the host chromosome, is a
379 widespread phenomenon in low temperature environments (51-53) and likely attributed to
380 prolonged starvation and low activity of host cells under harsh conditions, the latter being
381 previously reported for *Ralstonia* (15). Since p12D and its counterpart from the WAP have a
382 resolvase-domain containing protein likely functioning in integration/excision during
383 lysogenisation (54, 55), we conclude that they must be temperate phages of *Ralstonia*.

384 Many, i.e., 75 of the 78 Antarctic PGFs found in this study shared little protein clusters with
385 known phages from public databases. This is certainly related to the often high diversity of
386 viruses, limited accessibility to Antarctic environments as well as a stronger focus on
387 sequencing metaviromes and viral isolates of direct human interest (56). The missing
388 relatedness is known for ssDNA viruses originating from Antarctic cryoconite holes and was
389 attributed to the isolation and extreme environmental conditions at the Antarctic continent (57).
390 In total, 45 of the 78 PGFs could be assigned to a host, which in 71% of cases was *Ralstonia*
391 (summarized in Table S1, column "host prediction").

392 Aeolian transport of viruses over polar environments, especially attached to snowflakes, has
393 been barely investigated to date. Former investigations, mainly conducted at low latitudes,
394 demonstrated intercontinental transport of microbes by winds (58), and that highly identical
395 phages can be found in distantly related areas and in various ecosystems around the globe (59-
396 62). C. M. Bellas et al. (63) recently reported the presence of near-identical phage genomes
397 being spread by up to 4000 km in cryoconite holes of Svalbard, Greenland, and the Alps. Our
398 data show that despite the isolation of the Antarctic continent, and under no consideration of
399 anthropogenic dispersal (64), bacteria and phage distribution via snow over extensive distances

400 across Antarctica is possible. Man-made dispersal seems unlikely for our samples, due to the
401 relatedness of Antarctic phages to phages from environmental sources according to database
402 hits. We cannot be certain about the nature of the PGFs (prophage or lytic phage) by
403 metagenomic predictions alone. While the *Inoviridae* fraction likely occurs as prophages or
404 episomal forms, detection of PGFs assigned to lytic categories by VIBRANT and VirSorter
405 (category 1-3) suggests that host-associated, lytic phages captured at the adsorption or infection
406 stages were present as well. The degree of uncertainty about the category of a virus presumably
407 results from many fragmented scaffolds with relatively low coverages, which were however
408 sufficient to identify shared elements of viruses between stations. Conclusions about the
409 presence/absence of hallmark genes should be drawn from more complete datasets based on
410 greater sequencing depth of samples containing more biomass. We further commend
411 experiments that involve cultivation attempts and sequencing of metaviromes (free phages) to
412 reveal the extent of lysogenic or chronic compared to lytic infection styles in Antarctic snow.
413

414 Snow PGFs carried identical SNPs with those from the seawater metagenome from the WAP
415 located 4710 km and 5338 km apart from Druzhnaja and Mirnii, respectively (Figure 3C),
416 implying long distance transport. This is further supported by the detection of their hosts,
417 *Ralstonia* and *J. lividum,* in the WAP. From seawater, a transmission route via sea spray
418 aerosols to snow that is blown over ice surfaces (65) can be assumed. Aerosolization of bacteria
419 from the sea surface is highly taxon-specific but seems to work well for *R. pickettii* (66) and
420 also viruses (67). Alternatively, a transport via aerosols to clouds and precipitating snow is
421 possible. In the latter scenario, bioaerosols including microbial cells would act as ice nucleation
422 particles (68), transferring microbes to ice clouds where they might induce their own
423 precipitation (69, 70). The abundance data support the isolation of Progress-derived PGFs and
424 their potential host *F. micromati* and point towards a decreasing gradient of abundances from

425 Druzhnaja to Mirnii to the WAP for many PGFs (Table S1). Thus, general dispersal patterns of
426 microbes across Antarctica seem governed by westward drift and are probably mediated by the
427 prevailing Southern Hemisphere westerly winds (71). On short spatial-temporal scales,
428 dispersal seems more complex and is probably shaped by multifactorial dependencies such as
429 the different potential of a species to become airborne (66), meteorological conditions or the
430 local geography. For instance, while Druzhnaja is located ~50 km into the continent, the stations
431 Mirnii and Progress are near the coast and more exposed to the sea in summer when ice breaks
432 occur.

433 We assume that the transferability of the use of PGFs to study microbial dispersal in space
434 analogues such as the Antarctic continent is likely applicable to other celestial bodies like Mars.
435 Three celestial bodies in our solar system (Mars, Europa, and Enceladus) have environmental
436 conditions that could favor microbial life (reviewed by M. G. Netea et al. (72)). Most microbial
437 isolates (85-95%) obtained from spacecraft and assembly facilities are associated with humans
438 (73). Since *Ralstonia* spp. can be human and plant pathogens (42, 74) and are able to thrive
439 under harsh and oligotrophic conditions (7), they might contribute to the transmission of viruses
440 to extraterrestrial environments, particularly via manned missions.

441 We found evidence for viral signatures associated with two *R. pickettii* strain draft genomes
442 previously obtained from space equipment of a spacecraft assembly clean room and from water
443 systems of the ISS. This result in conjunction with the result that temperate *Ralstonia* phages
444 (and other viruses) can undergo long-range dispersal in association with their hosts across the
445 extraterrestrial analogue Antarctica suggest that contaminations of space equipment with
446 particularly persistent microbes such as *Ralstonia* should receive more focus during
447 microbiological monitoring in the framework of Planetary Protection. However, the field of
448 astrovirology has so far generally found little attention (75). The contribution of lysogenic and
449 episomal phage ('hidden hitchhikers') to overall viral loads on spacecraft and associated

450 equipment has been overlooked despite early work reporting on alterations in prophage
451 induction during spaceflight (76-78), tobacco mosaic virus to survive space flight equivalent
452 proton irradiation (79), the occurrence of phages and human-related circoviruses in clean rooms
453 (80) and inoviruses on the ISS (81, 82).

454 Planetary Protection aims to prevent the spread of biological contaminants (forward
455 contamination) to space shuttles and stations as well as extraterrestrial environments of the solar
456 system. However, this policy largely ignores the potential of escaped biological contaminants
457 to heavily disperse on foreign celestial bodies once being released, for instance after crash
458 landings as happened for the Schiaparelli module of the ExoMars program in 2016 (83). Our
459 results show that host-associated PGFs are not only suitable indicators for tracking long-
460 distance dispersal in space analogues but also demonstrate that the release of contaminants that
461 previously escaped Planetary Protection measures could spread far across extraterrestrial
462 ecosystems and, in the worst-case scenario, confound future life detection missions.

463 **Material and Methods**

464

465 **Metagenomic and genomic data processing**

466 We analyzed publicly available metagenomic data sets, which correspond to Antarctic surface
467 snow collected around three Russian stations (Druzhnaja, Mirnii, Progress) and are deposited
468 at NCBI's Sequence Read Archive (SRA) as Bioproject PRJNA674475 (Table S8) and MG-
469 RAST under project accession mgp13052 including taxon abundance data. Snow sampling was
470 conducted in December 2008 and 2009 as described previously (15); in brief, a sterile plastic
471 scoop was used to sample ~10 kg corresponding to a 2-3 cm layer of surface snow across several
472 1 m^2 areas. Snow was melted over a period of 12 hours and concentrated using Pellicon
473 tangential flow filters (Millipore, Burlington, MA, USA) to a final volume of ~ 10 mL. DNA

474 extraction for these metagenomes was carried out with the DNA Blood & Tissue kit (Qiagen,
475 Hilden, Germany), whose protocol causes removal of free viruses, and resulted in DNA
476 amounts of 170 – 490 ng. Sequencing libraries were prepared with MiSeq reagent kit v.2
477 (Illumina, USA), targeting mainly dsDNA viruses. Metagenomic data of a seawater sample
478 from the WAP was also obtained from SRA (accession #SRR5591034). Raw shotgun
479 sequencing reads of the three snow metagenomes and the WAP dataset were quality-trimmed
480 using BBDuk (https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh) from the
481 BBTools package (84) and Sickle (85) resulting in read counts between 1.31 – 2.65 Mio. for
482 the three snow samples. Assembly of reads was done using MetaSPAdes version 3.13 (86), and
483 scaffolds <1 kbps were removed. Draft genome sequences of *R. pickettii* strains SSH4 and CW2
484 isolated from space equipment were obtained from Genbank accession #JFZG00000000 and
485 #JFZH00000000. Strains SSH4 and CW2 were isolated pre-flight from the surface of the Mars
486 Odyssey Orbiter during assembly and from a water sample taken in-flight from the ISS cooling
487 system, respectively (9).

488

489 **Reconstruction of microbial genomes from metagenomes and CRISPR prediction**

490 Binning of MAGs from the three stations Druzhnaja, Mirnii and Progress was performed using
491 Emergent self-organizing maps (ESOMs, (87)), ABAWACA
492 (https://github.com/CK7/abawaca) and MaxBin 2.0 (88). Aggregation of bins was performed
493 using DASTool (89) and curation of bins was done in uBin (90), also delivering contamination
494 and completeness scores (91). Read coverage of recovered MAGs was obtained from read
495 mapping using Bowtie2 (92) in sensitive mode to the individual bins, followed by mismatch
496 filtering (2% mismatch allowance, depending on read length). To investigate the dispersal of
497 MAGs, mismatch criteria were set to 0% mismatches (100% similarity) and calcopo.rb

498 (https://github.com/ProbstLab/viromics/tree/master/calcopo) was used to calculate the
499 coverage per nucleotide (breadth) and the percentage of positions in the genome covered by
500 reads. Only genomes with a least 70% breadth were considered present in the respective
501 metagenome.

502 CRISPR arrays, which represent the prokaryotic adaptive immunity, were searched in host
503 MAGs using CRISPRcasFinder and considering evidence level 3 or 4 (29). CRISPR loci consist
504 of repeats interspaced with short DNA sequences (spacers) obtained from invading mobile
505 genetic elements such as phages and thus provide a record of past infections. Since CRISPR
506 arrays might get lost during the binning processes, e.g., due to fragmentation in assembly of
507 strain variants, absence of CRISPR arrays in CRISPRcasFinder was further investigated by
508 reconstructing CRISPR systems from raw reads using Crass (93) and BLASTing the obtained
509 direct repeat (DR) sequences, which are phylogenetically well-conserved (94), against the
510 NCBI non-redundant database (release 1st March 2021) using BLASTn --short algorithm (95)
511 with subsequent filtering at 80% similarity and e-value threshold of 10e-05. DR sequences were
512 BLASTed against the MAGs and used to extract CRISPR spacers from reads using
513 MetaCRAST with settings -d 3 -l 60 -c 0.99 -a 0.99 -r (96). Spacers were matched against PGFs
514 as mentioned above for DRs. In most cases, BLASTing the DR sequences against the NR
515 database did not reveal the host's identity. Nevertheless, spacers derived from unknown hosts
516 were considered, as their matches show that targeted PGFs represent true mobile genetic
517 elements, and matches can be used to infer infection patterns between stations.
518

519 **PGF detection, host allocation and viral clustering**

520 PGFs were identified from metagenome assemblies using a combination of bioinformatic tools,
521 namely Virsorter v1 (33), VirFinder (97), CircMG (98), renamed to VRCA

522 (https://github.com/alexcritschristoph/VRCA), VOGdb (version VOG93) (99), and
523 Endmatcher (https://github.com/ProbstLab/viromics/tree/master/Endmatcher). The
524 classification of predicted PGFs as "putative viruses" and "viruses" was done as previously
525 described in the Supplementary Figure 3 of (100). VIBRANT v.1.2.1 (30) with default settings
526 was used to find additional PGFs. No length cut-off was set (101), since many known *Ralstonia*
527 phages and inoviruses have genome sizes <10 kb (36, 50), and because of the low biomass of
528 Antarctic snow samples little PGFs compared to other ecosystems were expected. VirSorter
529 and VIBRANT aided in (pro)phage detection in draft genomes of *R. pickettii* from space
530 equipment. CheckV (32) was used to determine the type of identified PGF and completeness,
531 and viralComplete (31) was applied to predict closely related phages. PGFs associated with
532 MAGs were identified by grepping the scaffold ID on the bins. Pairwise comparisons of
533 Antarctic PGFs and clustering was done using nucleic acid- and amino acid-based VICTOR
534 (39). The resulting intergenomic distances were used to infer a balanced minimum evolution
535 tree with branch support via FASTME including Subtree Pruning and Regrafting (SPR)
536 postprocessing (102) using the distance formula D0. Branch support was inferred from 100
537 pseudo-bootstrap replicates each. Trees were rooted at the midpoint (103) and visualized with
538 FigTree v1.4.4 (104). Taxon boundaries at the species and genus level were estimated with the
539 OPTSIL program (105), using the recommended clustering thresholds (39) and an F value
540 (fraction of links required for cluster fusion) of 0.5 (106).

541 Clustering of PGFs from Antarctic snow and space equipment was further substantiated via
542 vConTACT2 v.0.9.19 (40, 107) in combination with the ProkaryoticViralRefSeq database
543 (v94, (108)) followed by visualization of viral clusters (VC) in Cytoscape v. 3.8.2 (109). Virus-
544 host matches were determined using the tool VirHostMatcher (34) with d_{2*} oligonucleotide
545 frequency dissimilarity measures for a k-mer length of 6. Viral genus clusters were determined

546 using VIRIDIC (110). Venn diagrams were calculated using the VIB-ugent webtool
547 (http://bioinformatics.psb.ugent.be/webtools/Venn/).

548

549 **Mapping of reads to assembled PGFs and nucleotide polymorphism analysis**

550 We assume that presence of a PGF at two locations confirmed by read mapping represents
551 dispersal. To determine if an assembled PGF from a single sample occurred in other
552 metagenomes even if not being assembled, we performed read mapping following the
553 previously published guidelines (101): that reads should map to a PGF with at least 90% identity
554 (Bowtie2 settings as in (111)), and more than 75% of scaffold should have a coverage of at least
555 1x. To detect breadth of a PGF, we again used calcopo.rb (see above). Mean coverage of PGFs
556 was calculated using calc_coverage_v3 (https://github.com/ProbstLab/uBin-
557 helperscripts/blob/master/bin/04_01calc_coverage_v3.rb) and normalized to sequencing depth.
558 Analysis of nucleotide polymorphisms was conducted for the 13 PGFs that underwent long
559 range dispersal, i.e., PGF being present in the WAP sample and at least one of the snow
560 metagenomes based on read mapping (Table S1). Variant analysis was performed in Geneious
561 11.1.5 (112) by applying default settings to the read-mappings generated as explained above.

562

563 **Gene prediction and annotations**

564 Open reading frames on PGFs were detected using Prodigal in meta mode (113). Functional
565 and taxonomic annotations of predicted proteins of PGFs were performed by DIAMOND
566 searches with a e-value of 10e-05 (114) against FunTaxDB (90) and by using DRAM-v (115).
567 For the full-length genome of the PGF p12D, annotations were improved using HHpred against
568 PDB, Pfam, UniProt-SwissProt-viral and NCBI_Conserved Domains ((116, 117),

569 https://toolkit.tuebingen.mpg.de/tools/hhpred) with a probability threshold of 70%. Sequences
570 of PGFs were BLASTed against IMG/VR 2.0 (118) with an e-value cut-off of 10e-05 to find
571 related phages from other metagenomic datasets.
572

573 **Synteny of *Ralstonia* phage and phylogenetic comparison of the *zot* gene**

574 Synteny of known *Ralstonia* sp. phages from NCBI and all Antarctic PGFs that were identified
575 herein and carried the zonular occludens toxin (*zot*, Pfam-ID: PF05707) was performed with
576 tBLASTx comparisons using Easyfig v.2.2.5 (119) on .gbk files generated by Prokka (120). A
577 phylogenetic tree for the MUSCLE-aligned (121) amino acid and nucleic acid sequences of *zot*
578 was constructed using the FastTree (122) algorithm in Geneious 11.1.5 (112).

579

580 **Acknowledgements**

581 We acknowledge Ken Dreger for server administration and maintenance as well as Cristina
582 Moraru for sharing insights on virus taxonomy.
583

584 **Author Contribution statement**

585 J.R. designed the study, wrote the manuscript, and carried out the analyses with input from
586 T.L.V.B.; A.L. and K.S. generated the raw data and performed the sampling; A.J.P
587 conceptualized the project, provided supervision, resources, and was involved in data
588 interpretation; All authors edited drafts of the manuscript.
589

590 **Author Disclosure Statement**

For all authors, no competing financial interests exist.

Funding statement

We acknowledge funding from the German Aerospace Center (DLR) for the project DISPERS (50WB1922). J.R. was partially supported by the German Science Foundation for the project VIBOCAT (grant number DFG RA3432/1-1). A.J.P. and T.L.V.B. were supported by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen ("Nachwuchsgruppe Dr. Alexander Probst").

References:

1. Pyne SJ. 2007. The extraterrestrial Earth: Antarctica as analogue for space exploration. Space Policy 23:147-149.
2. Lugg D, Shepanek M. 1999. Space analogue studies in Antarctica. Acta Astronaut 44:693-9.
3. Onofri S, Selbmann L, Zucconi L, Pagano S. 2004. Antarctic microfungi as models for exobiology. Planet Space Sci 52:229-237.
4. Puskeppeleit M, Quintern LE, El Naggar S, Schott JU, Eschweiler U, Horneck G, Bucker H. 1992. Long-term dosimetry of solar UV radiation in Antarctica with spores of *Bacillus subtilis*. Appl Environ Microbiol 58:2355-9.
5. Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. 2000. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol Mol Biol Rev 64:548-72.

613 6. Onofri S, Barreca D, Selbmann L, Isola D, Rabbow E, Horneck G, De Vera J, Hatton
614 J, Zucconi L. 2008. Resistance of Antarctic black fungi and cryptoendolithic
615 communities to simulated space and Martian conditions. Stud Mycol 61:99-109.
616 7. McAlister MB, Kulakov LA, O'Hanlon JF, Larkin MJ, Ogden KL. 2002. Survival and
617 nutritional requirements of three bacteria isolated from ultrapure water. J Ind
618 Microbiol Biotechnol 29:75-82.
619 8. Mijnendonckx K, Provoost A, Ott CM, Venkateswaran K, Mahillon J, Leys N, Van
620 Houdt R. 2013. Characterization of the survival ability of *Cupriavidus metallidurans*
621 and *Ralstonia pickettii* from space-related environments. Microb Ecol 65:347-60.
622 9. Monsieurs P, Mijnendonckx K, Provoost A, Venkateswaran K, Ott CM, Leys N, Van
623 Houdt R. 2014. Draft genome sequences of *Ralstonia pickettii* strains SSH4 and CW2,
624 isolated from space equipment. Genome Announc 2.
625 10. Ott CM, Bruce RJ, Pierson DL. 2004. Microbial characterization of free floating
626 condensate aboard the Mir space station. Microb Ecol 47:133-6.
627 11. Baker PW, Leff L. 2004. The effect of simulated microgravity on bacteria from the
628 Mir space station. Microgravity Sci Technol 15:35-41.
629 12. Vaishampayan P, Osman S, Andersen G, Venkateswaran K. 2010. High-density 16S
630 microarray and clone library-based microbial community composition of the Phoenix
631 spacecraft assembly clean room. Astrobiology 10:499-508.
632 13. La Duc MT, Nicholson W, Kern R, Venkateswaran K. 2003. Microbial
633 characterization of the Mars Odyssey spacecraft and its encapsulation facility. Environ
634 Microbiol 5:977-85.
635 14. Lysak V, Maksimova IA, Nikitin DA, Ivanova AE, Kudinova AG, Soina VS,
636 Marfenina OE. 2018. Soil microbial communities of Eastern Antarctica. Moscow
637 Univ Biol Sci Bull 73:104-112.

638 15. Lopatina A, Krylenkov V, Severinov K. 2013. Activity and bacterial diversity of snow
639 around Russian Antarctic stations. Res Microbiol 164:949-58.
640 16. Lopatina A, Medvedeva S, Shmakov S, Logacheva MD, Krylenkov V, Severinov K.
641 2016. Metagenomic analysis of bacterial communities of Antarctic surface snow.
642 Front Microbiol 7:398.
643 17. Antony R, Mahalinganathan K, Krishnan KP, Thamban M. 2012. Microbial
644 preference for different size classes of organic carbon: a study from Antarctic snow.
645 Environ Monit Assess 184:5929-43.
646 18. Liu Y, Yao T, Jiao N, Kang S, Xu B, Zeng Y, Huang S, Liu X. 2009. Bacterial
647 diversity in the snow over Tibetan Plateau Glaciers. Extremophiles 13:411-23.
648 19. Van Houdt R, De Boever P, Coninx I, Le Calvez C, Dicasillati R, Mahillon J,
649 Mergeay M, Leys N. 2009. Evaluation of the airborne bacterial population in the
650 periodically confined Antarctic base Concordia. Microb Ecol 57:640-8.
651 20. Carpenter EJ, Lin S, Capone DG. 2000. Bacterial activity in South Pole snow. Appl
652 Environ Microbiol 66:4514-7.
653 21. Michaud L, Lo Giudice A, Mysara M, Monsieurs P, Raffa C, Leys N, Amalfitano S,
654 Van Houdt R. 2014. Snow surface microbiome on the High Antarctic Plateau (DOME
655 C). PLoS One 9:e104505.
656 22. Malard LA, Sabacka M, Magiopoulos I, Mowlem M, Hodson A, Tranter M, Siegert
657 MJ, Pearce DA. 2019. Spatial variability of Antarctic surface snow bacterial
658 communities. Front Microbiol 10:461.
659 23. Pearce DA, Alekhina IA, Terauds A, Wilmotte A, Quesada A, Edwards A,
660 Dommergue A, Sattler B, Adams BJ, Magalhaes C, Chu WL, Lau MC, Cary C, Smith
661 DJ, Wall DH, Eguren G, Matcher G, Bradley JA, de Vera JP, Elster J, Hughes KA,
662 Cuthbertson L, Benning LG, Gunde-Cimerman N, Convey P, Hong SG, Pointing SB,

663 Pellizari VH, Vincent WF. 2016. Aerobiology over Antarctica - A new initiative for
664 atmospheric ecology. Front Microbiol 7:16.
665 24. Bottos EM, Woo AC, Zawar-Reza P, Pointing SB, Cary SC. 2014. Airborne bacterial
666 populations above desert soils of the McMurdo Dry Valleys, Antarctica. Microb Ecol
667 67:120-8.
668 25. Behzad H, Gojobori T, Mineta K. 2015. Challenges and opportunities of airborne
669 metagenomics. Genome Biol Evol 7:1216-26.
670 26. Bowers RM, Clum A, Tice H, Lim J, Singh K, Ciobanu D, Ngan CY, Cheng JF,
671 Tringe SG, Woyke T. 2015. Impact of library preparation protocols and template
672 quantity on the metagenomic reconstruction of a mock microbial community. BMC
673 Genomics 16:856.
674 27. Hughes KA. 2003. Aerial dispersal and survival of sewage-derived faecal coliforms in
675 Antarctica. Atmos Environ 37:3147-3155.
676 28. Gonçalves OS, de Oliveira Souza F, Bruckner FP, Santana MF, Alfenas-Zerbini P.
677 2021. Widespread distribution of prophages signaling the potential for adaptability
678 and pathogenicity evolution of *Ralstonia solanacearum* species complex. Genomics
679 113:992-1000.
680 29. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, Rocha
681 EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of
682 CRISRFinder, includes a portable version, enhanced performance and integrates
683 search for Cas proteins. Nucleic Acids Res 46:W246-W251.
684 30. Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation
685 and curation of microbial viruses, and evaluation of viral community function from
686 genomic sequences. Microbiome 8:90.
687 31. Antipov D, Raiko M, Lapidus A, Pevzner PA. 2020. Metaviral SPAdes: assembly of
688 viruses from metagenomic data. Bioinformatics 36:4126-4129.

689 32. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2020.
690 CheckV assesses the quality and completeness of metagenome-assembled viral
691 genomes. Nat Biotechnol doi:10.1038/s41587-020-00774-7.
692 33. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal
693 from microbial genomic data. PeerJ 3:e985.
694 34. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. 2017. Alignment-free d_2^*
695 oligonucleotide frequency dissimilarity measure improves prediction of hosts from
696 metagenomically-derived viral sequences. Nucleic Acids Res 45:39-53.
697 35. Hay ID, Lithgow T. 2019. Filamentous phages: masters of a microbial sharing
698 economy. EMBO Rep 20.
699 36. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, Sharrar A, Matheus
700 Carnevali PB, Cheng JF, Ivanova NN, Bondy-Denomy J, Wrighton KC, Woyke T,
701 Visel A, Kyrpides NC, Eloe-Fadrosh EA. 2019. Cryptic inoviruses revealed as
702 pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol 4:1895-1906.
703 37. Filée J, Forterre P, Laurent J. 2003. The role played by viruses in the evolution of their
704 hosts: a view based on informational protein phylogenies. Res Microbiol 154:237-243.
705 38. Moreira D. 2000. Multiple independent horizontal transfers of informational genes
706 from bacteria to plasmids and phages: implications for the origin of bacterial
707 replication machinery. Mol Microbiol 35:1-5.
708 39. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and
709 classification of prokaryotic viruses. Bioinformatics 33:3396-3404.
710 40. Bolduc B, Jang HB, Doulcier G, You ZQ, Roux S, Sullivan MB. 2017. vConTACT:
711 an iVirus tool to classify double-stranded DNA viruses that infect Archaea and
712 Bacteria. PeerJ 5:e3243.

713 41. Prospero JM, Blades E, Mathison G, Naidu R. 2005. Interhemispheric transport of
714 viable fungi and bacteria from Africa to the Caribbean with soil dust. Aerobiologia
715 21:1-19.
716 42. Askora A, Kawasaki T, Usami S, Fujie M, Yamada T. 2009. Host recognition and
717 integration of filamentous phage phiRSM in the phytopathogen, *Ralstonia*
718 *solanacearum*. Virology 384:69-76.
719 43. da Silva Xavier A, de Almeida JCF, de Melo AG, Rousseau GM, Tremblay DM, de
720 Rezende RR, Moineau S, Alfenas-Zerbini P. 2019. Characterization of CRISPR-Cas
721 systems in the *Ralstonia solanacearum* species complex. Mol Plant Pathol 20:223-
722 239.
723 44. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn
724 JH. 2020. Global organization and proposed megataxonomy of the virus world.
725 Microbiol Mol Biol Rev 84.
726 45. Murugaiyan S, Bae JY, Wu J, Lee SD, Um HY, Choi HK, Chung E, Lee JH, Lee SW.
727 2011. Characterization of filamentous bacteriophage PE226 infecting *Ralstonia*
728 *solanacearum* strains. J Appl Microbiol 110:296-303.
729 46. Mai-Prochnow A, Hui JG, Kjelleberg S, Rakonjac J, McDougald D, Rice SA. 2015.
730 'Big things in small packages: the genetics of filamentous phage and effects on fitness
731 of their host'. FEMS Microbiol Rev 39:465-87.
732 47. Sanguino L, Franqueville L, Vogel TM, Larose C. 2015. Linking environmental
733 prokaryotic viruses and their host through CRISPRs. FEMS Microbiol Ecol 91.
734 48. Hwang Y, Rahlff J, Schulze-Makuch D, Schloter M, Probst AJ. 2021. Diverse Viruses
735 Carrying Genes for Microbial Extremotolerance in the Atacama Desert Hyperarid
736 Soil. mSystems 6:e00385-21.

737 49. Van TTB, Yoshida S, Miki K, Kondo A, Kamei K. 2014. Genomic characterization of
738 ϕRS603, a filamentous bacteriophage that is infectious to the phytopathogen Ralstonia
739 solanacearum. Microbiology and immunology 58:697-700.
740 50. Askora A, Yamada T. 2015. Two different evolutionary lines of filamentous phages in
741 *Ralstonia solanacearum:* their effects on bacterial virulence. Front Genet 6:217.
742 51. Filippova SN, Surgucheva NA, Sorokin VV, Akimov VN, Karnysheva EA, Brushkov
743 AV, Andersen D, Gal'chenko VF. 2016. Bacteriophages in Arctic and Antarctic low-
744 temperature systems. Microbiology 85:359-366.
745 52. Dziewit L, Radlinska M. 2016. Two inducible prophages of an Antarctic
746 *Pseudomonas* sp. ANT_H14 use the same capsid for packaging their genomes -
747 Characterization of a novel phage helper-satellite system. PLoS One 11:e0158889.
748 53. Filippova SN, Surgucheva NA, Kulikov EE, Sorokin VV, Akimov VN, Bej AK,
749 McKay C, Andersen D, Galchenko VF. 2013. Detection of phage infection in the
750 bacterial population of Lake Untersee (Antarctica). Microbiology 82:383-386.
751 54. Askora A, Kawasaki T, Fujie M, Yamada T. 2011. Resolvase-like serine recombinase
752 mediates integration/excision in the bacteriophage phiRSM. J Biosci Bioeng 111:109-
753 16.
754 55. Ahmad AA, Stulberg MJ, Mershon JP, Mollov DS, Huang Q. 2017. Molecular and
755 biological characterization of varphiRs551, a filamentous bacteriophage isolated from
756 a race 3 biovar 2 strain of *Ralstonia solanacearum*. PLoS One 12:e0185034.
757 56. Rodrigues RAL, Andrade A, Boratto PVM, Trindade GS, Kroon EG, Abrahao JS.
758 2017. An anthropocentric view of the virosphere-host relationship. Front Microbiol
759 8:1673.
760 57. Sommers P, Fontenele RS, Kringen T, Kraberger S, Porazinska DL, Darcy JL,
761 Schmidt SK, Varsani A. 2019. Single-stranded DNA viruses in Antarctic cryoconite
762 holes. Viruses 11.

763 58. Smith DJ, Timonen HJ, Jaffe DA, Griffin DW, Birmele MN, Perry KD, Ward PD,
764 Roberts MS. 2013. Intercontinental dispersal of bacteria and archaea by transpacific
765 winds. Appl Environ Microbiol 79:1134-9.
766 59. Short CM, Suttle CA. 2005. Nearly identical bacteriophage structural gene sequences
767 are widely distributed in both marine and freshwater environments. Appl Environ
768 Microb 71:480-486.
769 60. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus?
770 Trends Microbiol 13:278-84.
771 61. Breitbart M, Miyake JH, Rohwer F. 2004. Global distribution of nearly identical
772 phage-encoded DNA sequences. FEMS Microbiol Lett 236:249-56.
773 62. Li Y, Endo H, Gotoh Y, Watai H, Ogawa N, Blanc-Mathieu R, Yoshida T, Ogata H.
774 2019. The Earth is small for "Leviathans": Long distance dispersal of giant viruses
775 across aquatic environments. Microbes Environ 34:334-339.
776 63. Bellas CM, Schroeder DC, Edwards A, Barker G, Anesio AM. 2020. Flexible genes
777 establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems. Nat
778 Commun 11:4403.
779 64. Hughes KA, Convey P, Pertierra LR, Vega GC, Aragon P, Olalla-Tarraga MA. 2019.
780 Human-mediated dispersal of terrestrial species between Antarctic biogeographic
781 regions: A preliminary risk assessment. J Environ Manage 232:73-89.
782 65. Benninghoff W, Benninghoff A. 1985. Wind transport of electrostatically charged
783 particles and minute organisms in Antarctica, p 592-596, Antarctic nutrient cycles and
784 food webs. Springer.
785 66. Michaud JM, Thompson LR, Kaul D, Espinoza JL, Richter RA, Xu ZZ, Lee C, Pham
786 KM, Beall CM, Malfatti F, Azam F, Knight R, Burkart MD, Dupont CL, Prather KA.
787 2018. Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-
788 atmosphere mesocosm. Nat Commun 9:2017.

789 67. Aller JY, Kuznetsova MR, Jahns CJ, Kemp PF. 2005. The sea surface microlayer as a
790 source of viral and bacterial enrichment in marine aerosols. J Aerosol Sci 36:801-812.
791 68. Wilbourn EK, Thornton DCO, Ott C, Graff J, Quinn PK, Bates TS, Betha R, Russell
792 LM, Behrenfeld MJ, Brooks SD. 2020. Ice nucleation by marine aerosols over the
793 North Atlantic Ocean in late spring. J Geophys Res Atmos 125:e2019JD030913.
794 69. Amato P. 2012. Clouds provide atmospheric oases for microbes. Microbe 7:119-123.
795 70. Christner BC, Morris CE, Foreman CM, Cai R, Sands DC. 2008. Ubiquity of
796 biological ice nucleators in snowfall. Science 319:1214.
797 71. Strother SL, Salzmann U, Roberts SJ, Hodgson DA, Woodward J, Van Nieuwenhuyze
798 W, Verleyen E, Vyverman W, Moreton SG. 2015. Changes in Holocene climate and
799 the intensity of Southern Hemisphere Westerly Winds based on a high-resolution
800 palynological record from sub-Antarctic South Georgia. Holocene 25:263-279.
801 72. Netea MG, Dominguez-Andres J, Eleveld M, Op den Camp HJM, van der Meer JWM,
802 Gow NAR, de Jonge MI. 2020. Immune recognition of putative alien microbial
803 structures: Host-pathogen interactions in the age of space travel. PLoS Pathog
804 16:e1008153.
805 73. Nicholson WL, Schuerger AC, Race MS. 2009. Migrating microbes and planetary
806 protection. Trends Microbiol 17:389-92.
807 74. Coenye T, Vandamme P, LiPuma JJ. 2002. Infection by *Ralstonia* species in cystic
808 fibrosis patients: identification of *R. pickettii* and *R. mannitolilytica* by polymerase
809 chain reaction. Emerg Infect Dis 8:692-6.
810 75. Berliner AJ, Mochizuki T, Stedman KM. 2018. Astrovirology: Viruses at large in the
811 universe. Astrobiology 18:207-223.
812 76. Mattoni RHT, Keller Jr EC. 1972. Induction of lysogenic bacteria in the space. The
813 Experiments of Biosatellite II 204:309.

814 77. Mattoni RHT. 1968. Influence of spaceflight and radiation on induction of prophage
815 P-22 in *Salmonella typhimurium*. Jpn J Genet 43:465-465.
816 78. Mattoni RHT. 1968. Space-flight effects and gamma radiation interaction on growth
817 and induction of lysogenic bacteria, a preliminary report. BioScience 18:602-608.
818 79. Koike J. 1991. Fundamental questions concerning the contamination of other planets
819 with terrestrial microorganisms carried by space-probes. The Journal of Space
820 Technology and Science 7:9-14.
821 80. Weinmaier T, Probst AJ, La Duc MT, Ciobanu D, Cheng JF, Ivanova N, Rattei T,
822 Vaishampayan P. 2015. A viability-linked metagenomic analysis of cleanroom
823 environments: eukarya, prokaryotes, and viruses. Microbiome 3:62.
824 81. Mora M, Wink L, Kogler I, Mahnert A, Rettberg P, Schwendner P, Demets R, Cockell
825 C, Alekhova T, Klingl A, Krause R, Zolotariof A, Alexandrova A, Moissl-Eichinger
826 C. 2019. Space Station conditions are selective but do not alter microbial
827 characteristics relevant to human health. Nat Commun 10:3990.
828 82. Pavletić B, Runzheimer K, Siems K, Koch S, Cortesão M, Ramos-Nascimento A,
829 Moeller R. 2022. Spaceflight virology: What do we know about viral threats in the
830 spaceflight environment? Astrobiology doi:10.1089/ast.2021.0009.
831 83. Aboudan A, Colombatti G, Bettanini C, Ferri F, Lewis S, Van Hove B, Karatekin O,
832 Debei S. 2018. ExoMars 2016 Schiaparelli module trajectory and atmospheric profiles
833 reconstruction. Space Science Reviews 214:1-31.
834 84. Bushnell B. 2014. BBTools software package. URL http://sourceforge
835 net/projects/bbmap 578:579.
836 85. Joshi N, Fass J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming
837 tool for FastQ files (Version 1.33)[Software].
838 86. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new
839 versatile metagenomic assembler. Genome Res 27:824-834.

840 87. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield
841 JF. 2009. Community-wide analysis of microbial genome sequence signatures.
842 Genome Biol 10:R85.
843 88. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning
844 algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics
845 32:605-7.
846 89. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF.
847 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and
848 scoring strategy. Nat Microbiol 3:836-843.
849 90. Bornemann TLV, Esser SP, Stach L, Burg T, Probst AJ. 2020. uBin – a manual
850 refining tool for metagenomic bins designed for educational purposes. bioRxiv
851 doi:10.1101/2020.07.15.204776
852 91. Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, Hug LA,
853 Burstein D, Emerson JB, Thomas BC, Banfield JF. 2017. Genomic resolution of a
854 cold subsurface aquifer community provides metabolic insights for novel microbes
855 adapted to high CO_2 concentrations. Environ Microbiol 19:459-474.
856 92. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat
857 Methods 9:357-9.
858 93. Skennerton CT, Imelfort M, Tyson GW. 2013. Crass: identification and reconstruction
859 of CRISPR from unassembled metagenomic data. Nucleic Acids Res 41:e105.
860 94. Mojica FJ, Diez-Villasenor C, Soria E, Juez G. 2000. Biological significance of a
861 family of regularly spaced repeats in the genomes of Archaea, Bacteria and
862 mitochondria. Mol Microbiol 36:244-6.
863 95. Altschul SF. 2001. BLAST algorithm. e LS.
864 96. Moller AG, Liang C. 2017. MetaCRAST: reference-guided extraction of CRISPR
865 spacers from unassembled metagenomes. PeerJ 5:e3788.

866 97. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel k-mer
867 based tool for identifying viral sequences from assembled metagenomic data.
868 Microbiome 5:69.
869 98. Crits-Christoph A, Gelsinger DR, Ma B, Wierzchos J, Ravel J, Davila A, Casero MC,
870 DiRuggiero J. 2016. Functional interactions of archaea, bacteria and viruses in a
871 hypersaline endolithic community. Environ Microbiol 18:2064-77.
872 99. Marz M, Beerenwinkel N, Drosten C, Fricke M, Frishman D, Hofacker IL, Hoffmann
873 D, Middendorf M, Rattei T, Stadler PF, Topfer A. 2014. Challenges in RNA virus
874 bioinformatics. Bioinformatics 30:1793-9.
875 100. Rahlff J, Turzynski V, Esser SP, Monsees I, Bornemann TLV, Figueroa-Gonzalez PA,
876 Schulz F, Woyke T, Klingl A, Moraru C, Probst AJ. 2021. Lytic archaeal viruses
877 infect abundant primary producers in Earth's crust. Nat Commun 12:4642.
878 101. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics:
879 an in silico evaluation of metagenome-enabled estimates of viral community
880 composition and diversity. PeerJ 5:e3817.
881 102. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: A comprehensive, accurate, and
882 fast distance-based phylogeny inference program. Mol Biol Evol 32:2798-800.
883 103. Farris JS. 1972. Estimating phylogenetic trees from distance matrices. The American
884 Naturalist 106:645-668.
885 104. Rambaut A. 2006. FigTree, a graphical viewer of phylogenetic trees and as a program
886 for producing publication-ready figures.
887 105. Göker M, Garcia-Blazquez G, Voglmayr H, Telleria MT, Martin MP. 2009. Molecular
888 taxonomy of phytopathogenic fungi: a case study in *Peronospora*. PLoS One 4:e6319.
889 106. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde
890 C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy T, Pati A, Ivanova NN,
891 Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk HP. 2014. Complete genome

892 sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a
893 proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 9:2.
894 107. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR,
895 Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic
896 assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing
897 networks. Nat Biotechnol 37:632-639.
898 108. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource.
899 Nucleic Acids Res 43:D571-7.
900 109. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N,
901 Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated
902 models of biomolecular interaction networks. Genome Res 13:2498-504.
903 110. Moraru C, Varsani A, Kropinski AM. 2020. VIRIDIC-A novel tool to calculate the
904 intergenomic similarities of prokaryote-infecting viruses. Viruses 12.
905 111. Nilsson E, Bayfield OW, Lundin D, Antson AA, Holmfeldt K. 2020. Diversity and
906 host interactions among virulent and temperate Baltic Sea *Flavobacterium* phages.
907 Viruses 12:158.
908 112. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S,
909 Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.
910 2012. Geneious Basic: an integrated and extendable desktop software platform for the
911 organization and analysis of sequence data. Bioinformatics 28:1647-9.
912 113. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
913 prokaryotic gene recognition and translation initiation site identification. BMC
914 Bioinformatics 11:119.
915 114. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using
916 DIAMOND. Nature Methods 12:59-60.

917 115. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P,
918 Narrowe AB, Rodriguez-Ramos J, Bolduc B, Gazitua MC, Daly RA, Smith GJ, Vik
919 DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling
920 microbial metabolism to automate the curation of microbiome function. Nucleic Acids
921 Res 48:8883-8900.
922 116. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding
923 J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit
924 with a new HHpred server at its core. J Mol Biol 430:2237-2243.
925 117. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein
926 homology detection and structure prediction. Nucleic Acids Res 33:W244-8.
927 118. Paez-Espino D, Roux S, Chen IA, Palaniappan K, Ratner A, Chu K, Huntemann M,
928 Reddy TBK, Pons JC, Llabres M, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2019.
929 IMG/VR v.2.0: an integrated data management and analysis system for cultivated and
930 environmental viral genomes. Nucleic Acids Res 47:D678-D686.
931 119. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer.
932 Bioinformatics 27:1009-10.
933 120. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
934 30:2068-9.
935 121. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time
936 and space complexity. BMC Bioinformatics 5:113.
937 122. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-
938 likelihood trees for large alignments. PLoS One 5:e9490.
939 123. Schlitzer R. 2015. Ocean Data View. https://odv.awi.de/.

940