

1 Phages actively challenge niche communities in the Antarctic
2 soils

3

4 Oliver K.I. Bezuidt^{a*}, Pedro Humberto Lebre^{a*}, Rian Pierneef^{f^{a,b}}, Carlos León-
5 Sobrino^a, Evelien M. Adriaenssens^c, Don A. Cowan^a, Yves Van de Peer^{a,d,e} and
6 Thulani P. Makhalanyane^{a#}

7

8 ^aCentre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics
9 and Microbiology, University of Pretoria, Lynnwood Road, Pretoria, 0028, South
10 Africa, ^b Biotechnology Platform, Agricultural Research Council, Pretoria, 0002, South
11 Africa, ^c Quadram Institute Bioscience, Norwich Research Park, NR4 7HQ Norwich,
12 United Kingdom, ^dDepartment of Plant Biotechnology and Bioinformatics, Ghent
13 University, 9052 Ghent, Belgium, ^eCenter for Plant Systems Biology, VIB, 9052
14 Ghent, Belgium

15

16 Running Head: host-phage interactions in Antarctica

17

18 Oliver K.I. Bezuidt and Pedro H Lebre contributed equally to this work. Author order
19 was determined on the basis of seniority.

20

21 #Address correspondence to Thulani P. Makhalanyane,

22 thulani.makhalanyane@up.ac.za

23

24 **Abstract**

25 By modulating the structure, diversity and trophic outputs of microbial communities,
26 phages play crucial roles in many biomes. In oligotrophic polar deserts, the effects of
27 katabatic winds, constrained nutrients and low water availability are known to limit
28 microbial activity. Although phages may substantially govern trophic interactions in
29 cold deserts, relatively little is known regarding the precise ecological mechanisms.
30 Here, we provide the first evidence of widespread antiphage innate immunity in
31 Antarctic environments using metagenomic sequence data from hypolith
32 communities as model systems. In particular, immunity systems such as DISARM
33 and BREX are shown to be dominant systems in these communities. Additionally, we
34 show a direct correlation between the CRISPR-cas adaptive immunity and the
35 metavirome of hypolith communities, suggesting the existence of dynamic host-
36 phage interactions. In addition to providing the first exploration of immune systems in
37 cold deserts, our results suggest that phages actively challenge niche communities
38 in Antarctic polar deserts. We provide evidence suggesting that the regulatory role
39 played by phages in this system is an important determinant of bacterial host
40 interactions in this environment.

41

42 **Importance**

43 In Antarctic environments, the combination of both abiotic and biotic stressors results
44 in simple trophic levels dominated by microbiomes. Although the past two decades
45 have revealed substantial insights regarding the diversity and structure of
46 microbiomes, we lack mechanistic insights regarding community interactions and
47 how phages may affect these. By providing the first evidence of widespread
48 antiphage innate immunity, we shed light on phage-host dynamics in Antarctic niche
49 communities. Our analyses reveal several antiphage defense systems including
50 DISARM and BREX, which appear to dominate in cold desert niche communities. In
51 contrast, our analyses revealed that genes, which encode antiphage adaptive
52 immunity were under-represented in these communities suggesting lower infection
53 frequencies in cold edaphic environments. We propose that by actively challenging
54 niche communities, phages play crucial roles in the diversification of Antarctic
55 communities.

56

57

58 **Introduction**

59 Antarctic terrestrial environments including open soils, permafrost and the
60 surface and interior of rocks, are typically oligotrophic and dominated by
61 psychrophilic and psychrotolerant microbial communities (1-4). It has been
62 suggested that the extreme abiotic pressures of the environment such as
63 temperature, desiccation stress and UV radiation are dominant drivers of both the
64 diversity and function of cold-adapted bacterial communities in terrestrial polar
65 deserts (5-7). Similarly, biotic interactions such as competition, symbioses, horizontal
66 gene transfer (HGT) and predation have also been shown to play a role in the
67 distribution and diversity of microbial communities in these soil ecosystems (8-10).
68 The presence of viruses, including bacteriophages, in these cold hyper-arid desert
69 soils potentially adds an additional layer of complexity to the microbial system, but
70 the extent to which phage-host interactions play a role in shaping community
71 compositions and processes in cold desert soil niches remains a matter of
72 speculation (11, 12).

73 Antarctic desert hypolithic communities, in particular, have been shown to
74 contain substantial viral populations, dominated by tailed bacteriophages of the order
75 *Caudovirales* (11, 13-15). Micro-array analysis of lithic niches identified an even
76 greater phage diversity, including signatures of RNA bacteriophages of the family
77 *Leviviridae*, ssDNA phage of the family *Microviridae* and non-tailed dsDNA
78 tectiviruses (16). Together, these observations suggest that phages may play an
79 important role of in microbial community structures and functions.

80 The presence of active bacteriophages in a microbial community inevitably
81 leads to the evolution of specialized bacterial defensive measures (17), and a
82 diverse range of bacterial defense mechanisms against parasitic phages have been

83 identified (18, 19). These include adaptive immunity elements, such as the CRISPR-
84 Cas systems, and innate immunity mechanisms, such as restriction-modification
85 (RM) and toxin-antitoxin abortive infection (Abi) systems (18). Recent pangenomics
86 studies have also identified novel defense systems that are widely distributed across
87 bacterial taxa and are thought to play a role in anti-phage resistance (20-23). These
88 include the bacteriophage exclusion (BREX) system, coded by a 4-8 gene cluster,
89 that provides resistance to *Siphoviridae* and *Myoviridae* tailed phages by inhibition of
90 phage DNA replication (21), and other less well characterized systems such as the
91 Thoeris, Shedu and Gabija elements that increase bacterial host resistance to
92 specific groups of phages (22).

93 Combining the valuable evidence on phage diversity and prevalence in polar
94 desert soils, we hypothesize that phage-host interactions play an important role in
95 shaping the structure of edaphic microbial communities in these environments. To
96 test our hypothesis, we assess the known bacterial defense systems in
97 metagenomic sequence data derived from niche Antarctic hypolith community. We
98 were able to link some of these data to specific phage genomes and propose that
99 phages play an active role in shaping the immunity of Antarctic soil microbial
100 communities.

101

102 **Results**

103 **The distribution of anti-phage defense mechanisms shows an abundance of** 104 **innate immunity genes**

105 The distribution of antiphage defense systems in the metagenome was
106 determined by mapping defense genes against the taxonomically assigned contigs.
107 In total, 24,941 defense genes were detected, comprising 1.2% of the entire

108 metagenome gene count. Approximately 40% of these were found in contigs
109 attributed to unknown phyla. The general distribution of defense genes across known
110 phyla was consistent with the relative abundance of each phylum in the metagenome
111 (Figure 1A, Table S4). Proteobacteria harbored the highest number of anti-phage
112 genes (5289 genes, 1.1% of total gene count for this phylum), followed by
113 Actinobacteria (3808, 0.9% total gene count) and Bacteroidetes (2128, 1.08% of total
114 contig count). RM, DISARM and BREX systems were the most abundant systems in
115 the metagenome, contributing 67.6% of the total gene hits for anti-phage defense
116 systems. On the other side of the spectrum, the defense systems Shedu, Hachiman
117 and CRISPR-type 2 were present at relatively low abundances, and therefore had
118 little apparent contribution to the global defense system distribution. The average
119 contribution of defense genes to the total gene count per phyla was 1.8%, with
120 Deferribacteres and Candidatus Tectomicrobia as outliers. However, it is important
121 to note that these phyla represent a very small portion of the metagenome, and
122 therefore the possibility that the high percentage of defense genes is biased toward
123 the low gene count for these phyla cannot be disregarded.

124 Analysis of the relative contribution of each defense system within each
125 phylum also showed that genes belonging to the RM, DISARM, and BREX systems
126 were the main contributors across the majority of phyla (Figure 1B). The recently
127 discovered Zorya system was predominantly represented in the phyla
128 Gemmatimonadetes, Bacteroidetes, Planctomycetes, Proteobacteria and
129 Verrumicrobia, while CRISPR systems showed the highest contribution in
130 Cyanobacteria and Euryarchaeota. Interestingly, non-canonical anti-phage systems
131 represented more than 50% of the defense systems identified for all phyla aside from

132 Euryarchaeota, with Verrucomicrobia, Planctomycetes and Acidobacteria possessing
133 the highest distribution of non-canonical defense genes.

134

135 **Innate immunity is dominated by BREX and DISARM genes**

136 As highlighted above, anti-phage systems across phyla in the hypolith
137 metagenome were dominated by non-canonical innate systems. Further analysis of
138 the distribution of defense genes revealed that anti-phage systems in the majority of
139 phyla were dominated by BREX and DISARM genes. The two systems together
140 accounted for 33.4% of defense genes, compared to 31.7% genes belonging to
141 canonical RM systems.

142 A total of 3758 genes for the DISARM system were identified. These included
143 the Class I marker gene *drmD* (449 counts, 11.9% of DISARM genes), which
144 encodes the SNF2-like helicase (23), as well as the Class II marker gene *drmA*
145 (1020, 17.1% of DISARM genes), which encodes a protein with a putative helicase
146 domain (23). Similarly, a total of 4598 genes representing all BREX types were
147 identified in the metagenome. Interestingly, the most abundant gene from this
148 system found in the metagenome, *pgW* (2640, 57.4% of BREX genes), which codes
149 for a serine/threonine kinase, is specific to the type 2 BREX system, also called the
150 Pgl system (21). By comparison, of the 7908 RM genes found in the metagenome,
151 the most abundant is *hsdM* (1423, 18% of RM genes), a type I DNA methylase
152 responsible for the protection of host DNA (24). In fact, more than 50% of RM
153 defense genes were attributed to type I RM systems.

154 The third non-canonical system representing more than 10% of the anti-phage
155 defense systems in a subset of the phyla, the Zorya system, included a total of 2411
156 genes in the metagenome. The majority of these were homologous to the two genes

157 that make up a proton channel, *zorA* and *zorB*. This is a common feature in all types
158 of Zorya system and is thought to cause depolarization of the membrane upon
159 infection (22).

160

161 **Type I CRISPR-Cas genes comprise the bulk of anti-phage adaptive immunity** 162 **genes**

163 In total, 2234 CRISPR-cas genes were identified in 1601 contigs by searching
164 for shared sequence similarities against the CDD database. A substantial proportion
165 of all classified CRISPR-cas loci (71.4%) belonged to type I CRISPR-Cas systems,
166 followed by type III (18.5%) and type II (10.2%) (Table S5). While the abundance of
167 Cas I-B loci sequences in the public databases suggests that the Cas-I mechanism
168 is the most common in both bacteria and archaea (20 and 30% of total CRISPR loci
169 (25), less than 3% of these loci were present in our composite metagenome (Table
170 S5, Figure 2). Surprisingly, CRISPR-cas loci linked to Types I-C and I-E were the
171 most prevalent, at 24.1% and 12.9% of classified CRISPR-cas loci, respectively.
172 Another subtype identified at higher relative abundances than previously reported
173 (25) was I-U, at 10.76% of classified *cas* loci. This subtype is characterized by the
174 marker GSU0054 domain, which was the fourth most abundant *cas* CDD overall
175 (108 occurrences) after *cas4*, *cas1*, and *cas2*.

176 **Phage presence in the niche community is correlated with the CRISPR arrays**

177 CRISPR arrays represent the history of infection by invading DNA (e.g.
178 phages, plasmids (26, 27), and a study of their composition and frequencies
179 provides insights into phage-host interactions in an ecological context (28). A total of
180 878 CRISPR arrays harboring 10,292 spacers were identified in the metagenome,
181 with an average length of 36 protospacers per array (Figure S1A). CRISPR array

182 sizes ranged from 2 to 249, with the majority (83.5% of total arrays) falling between 2
183 and 18 protospacers per array (Figure S1B).

184 The distribution of CRISPR array sizes in the metagenome was compared to
185 data collected from a ground-water microbiome (29), to compare the array size
186 distributions from environments with potentially different phage-host dynamics (11).
187 The results show that CRISPR arrays in the hypolith metagenome exhibited a
188 smaller and narrower size range, compared to the ground-water community
189 metagenome (Figure 3). This suggests the existence of distinct phage infection
190 frequencies between the different environments; i.e., lower infection frequencies in
191 the cold edaphic community.

192 In addition to using the CRISPR array as a tool for understanding infection
193 history, the viral population in the Antarctic soil community was also assessed by
194 assembly of the metavirome. A total of 793 contigs was assembled from the
195 metagenomic sequence data using VirSorter (30). Taxonomic annotation of these
196 contigs, using a database of viral reference genomes (31), unambiguously assigned
197 645 of these as viral, 560 of which were further assigned to the order of tailed
198 phages *Caudovirales*. Within this order, the majority of contigs were assigned to
199 Siphoviridae (52%), followed by unclassified *Caudovirales* (14%), and viruses with
200 no assigned family (13%) (Figure S2). To access the correlation between the viral
201 contigs and the CRISPR arrays, spacers from the metagenome were matched to
202 both the VirSorter contigs and a set of contigs from environmental datasets (IMG/VR
203 (32), which allowed for the taxonomic assignment of 394 (3.8% of total number of
204 spacers) CRISPR-cas spacers (Figure S3). The resulting similarity network (Figure
205 4) showed that all 73 VirSorter phage contigs included in the network (red nodes)
206 matched to CRISPR-cas spacers (grey nodes), suggesting that a substantial fraction

207 (11.3%) of the identified viral population had a history of infection *in situ* in the host
208 population, and may therefore be actively involved in shaping the adaptive immunity
209 of the microbial community. In addition, several distinct clusters showed matches
210 between a single VirSorter contig and several spacers, suggesting these viral contigs
211 are common infection agents.

212 Functional analysis using eggNOG showed the presence of genes that
213 facilitate infection such as genes that code for chitinases, which are involved in the
214 degradation of the protective biofilm (33), as well as a AntA/AntB antirepressor
215 gene, thought to be involved in phage anti-immunity (34) (Figure 5). In addition, the
216 eggNOG functional analysis of the 645 VirSorter viral contigs also revealed the
217 presence of genes contributing to phage virulence (Table S2), the most abundant of
218 which encode for methyltransferases, which are actively involved in the evasion of
219 the R.M systems (35). This result suggests the possibility of an evolutionary pressure
220 for the phages to develop evasion mechanisms against their hosts, which further
221 hints at active phage-host dynamics in these long-enduring Antarctic hypoliths.

222

223 **Discussion**

224 Due to the relatively simple trophic structures in cold desert systems,
225 including Antarctic soils, cryptic microbial communities are considered to be
226 important drivers of local ecosystem services (36). However, the extent to which
227 these communities are influenced by phages remains largely unexplored. Such
228 interactions may shape the diversification and community interactions in cold desert
229 systems. Qualitative surveys of Antarctic metaviromics have reported a high diversity
230 of viruses associated with microbial communities of open soils, and cryptic niches
231 (12, 13). Evidence, albeit limited, that Antarctic soil phages exist predominantly in a

232 lysogenic rather than lytic lifestyle (14), has led to suggestions that the functional role
233 of phages in this spatially restricted, water-constrained desert soil niche may be
234 limited (11).

235 The results presented in this study provide the first evidence of interaction
236 between phage and hosts in this psychrophilic edaphic environment. This is most
237 evident in the correlation between the metavirome of the hypolith community and the
238 CRISPR-arrays, which suggest the active evolution of the adaptive immune system
239 against local viral threats. This idea of community adaption to local phage threat is
240 further implied by the positive correlation between the CRISPR arrays and viruses
241 extracted from local soils. In fact, a previous study (37) has already suggested that
242 recruitment from surrounding soils plays an important role in the development of
243 hypoliths, and this might also be extended to the recruitment of phages from the
244 surrounding ecosystem. Another indication of active interaction between phage and
245 host is suggested by the presence of several methyltransferases in the
246 metagenome-assembled viral contigs, which are a hallmark of viral evasion against
247 native host RM systems (31, 35). Other genes found in this virome contigs include
248 genes specifically involved in the degradation of biofilm matrices and evasion against
249 RM systems, further suggesting that there is a complex network of interactions at
250 play between phages and their hosts in the hypolithic environment.

251

252 While the metagenomics data analysed in this study does not give a direct
253 indication of the temporal scale of the phage-host interactions occurring in the
254 hypolith, the short sizes of CRISPR-array sizes in the hypolith metagenome suggest
255 a low frequency of infection. This low frequency is further hinted at when comparing
256 the hypolith CRISPR-array sizes with those of a more fluid and homogenous

257 environment, where viral-host interactions are assumed to be a frequent occurrence
258 (29). Together, these results imply a model for viral-host interactions in hypoliths
259 that follows the 'static-step-static' development model suggested by Pointing et al.
260 (38), driven by the stochastic and intermittent nature of rain events in such water-
261 limited ecosystems. A surprising result from this study is the prevalence of non-
262 canonical innate immunity systems, the most prominent of which are the BREX and
263 DISARM systems. While these two systems have been shown to be widespread in
264 bacteria using a pan-genomic dataset (21, 23), the present study represent the first
265 evidence for the prevalence of these systems in ecological samples. As such, this
266 result implies that non-canonical innate immunity is more important for anti-phage
267 microbial community defense than previously thought and should therefore be the
268 focus for future studies into innate immunity in the ecological context. There are also
269 indications from the hypolith metagenome that the prevalence of non-canonical
270 innate immunity over traditional RM and Abi system for defense against phages is
271 related to the adaptation of the hypolith communities to specific local viral
272 populations. For instance, the Zorya system, the third most prevalent non-canonical
273 immunity system in the metagenome, is hypothesized to operate similarly to the Abi
274 system (22). In turn, Zorya systems provide resistance against a limited range of
275 phages, including the ssDNA family *Microviridae* (22), which has been shown to be
276 prevalent in Antarctic aquatic and soil niches (39).

277

278 **Conclusion**

279 Together, these results are not consistent with the suggestion that the
280 constraints of the environment, such as low temperatures, low a_w and resulting very
281 limited capacity for inter-particle diffusion, lead to extremely localized phage-host

282 interactions (11). Rather, the data are suggestive of a dynamic and continual
283 interaction between host and phage. Nevertheless, inter-particle communication and
284 exchange may be limited to brief periods when bulk liquid water is present, after
285 snow melt, for example. Furthermore, the low metabolic rates (the inevitable
286 consequence of Arrhenius effects (temperature dependence of reaction rates) in cold
287 environments) should also limit the rates at which phages can replicate and
288 propagate, further limiting the frequency of interactions with their hosts (40). We
289 suggest that the localized nature of host-phage interactions in the hypolith niche and
290 the limited inter-particle communication, where bacterial hosts are not frequently
291 challenged by novel phage threats, leads to a reliance of microbial communities on
292 innate immunity as the primary defense against phage infection. The smaller sizes of
293 CRISPR arrays in the Antarctic soil metagenome sequences compared to those from
294 a temperate aquatic environment, and the under-representation of CRISPR systems,
295 give further credence to the temporally sporadic interaction between phages and
296 their hosts. Nevertheless, the correlation between the metavirome and the CRISPR-
297 cas arrays, together with the presence of bacteriophage evasion genes in the
298 metavirome, suggest that phage-host interactions within the hypolith community are
299 a dynamic process that leads to co-evolution of both phages and hosts. We therefore
300 suggest that phages play a hitherto underestimated role in driving the evolution of
301 Antarctic soil microbial communities by shaping their collective immunity.

302

303 **Materials and Methods**

304 **Sample collection, DNA extraction and metagenomic sequencing**

305 The sample collection, DNA extraction and metagenomic sequencing
306 protocols used in this study have been described previously (41). Briefly, a total of 50
307 samples were collected from hypolithic niches in the Antarctica Miers Valley (GPS
308 78°09'36.0"S 164°06'00.0"E) and stored in sterile Whirl-Pak bags (Nasco
309 International, Fort Atkinson, WI, USA) at -20 °C. Metagenomic DNA was extracted
310 from each sample using a PowerSoil DNA isolation kit (MO BIO, Carlsbad, CA,
311 USA), and the purified DNA was pooled before further processing. Purified DNA was
312 sheared into fragments of approximately 300 bps and further purified from 1%
313 agarose gels. Subsequent sequencing was performed using Illumina HiSeq-2000
314 paired-end technology (2 x 101 bp), and the resulting reads were trimmed and
315 assembled as described below.

316

317 **Metagenome assembly and taxonomical annotation**

318 Metagenomic DNA sequence data were quality-filtered by trimmomatic
319 version 0.36 using a phred cut-off > 30 (42). The assembly of high-quality reads from
320 the metagenome sequence dataset was conducted using the IDBA-UD tool (43) and
321 contig lengths were extended (scaffolded) using SSPACE Basic (43). The statistics
322 for the assembly of the metagenome are presented in Table S1. Contigs were
323 taxonomically assigned using the MEGAN v6 pipeline (44) with the NCBI taxonomy
324 database for taxon ID assignment.

325

326 **Detection of the innate and adaptive defense systems**

327 Metagenomic contigs were used for functional gene predictions using prodigal
328 v2.50, with the `-meta` parameter implementation (45). Predicted genes were
329 subsequently screened for domain similarity with known defense systems against

330 the conserved domain database (CDD) of clusters of orthologous groups (COGs)
331 and protein families (Pfam) using rps-blast (E value < 1e-02) (33). These results
332 were manually filtered for the identification of phage-specific defense systems, which
333 include restriction-modification (R.M), bacteriophage exclusion (BREX), abortive
334 infection (Abi), defense island system associated with restriction-modification
335 (DISARM), and other recently identified systems using a refined list of COG and
336 Pfam position-specific score matrices (PSSMs) for marker genes in these systems
337 (21-23, 46). A list of the marker genes used in this study can be found in Table S2.
338 Additionally, defense genes that could not be clustered into a specific system were
339 classified as ambiguous as were not considered for subsequent analysis (Table S3).

340 ORFs predicted using prodigal v2.50 were queried against the CDD database
341 for the presence of putative CRISPR-cas genes (47), using delta-blast at a cutoff E
342 value of 1e-03. Multi-gene cas modules were identified as those having multiple cas
343 annotated genes with ≤ 5 ORF spacings. Type and subtype classifications were
344 assigned following the updated classification set by Makarova et al. (25).

345

346 **Phage genome identification and CRISPR spacer matching**

347 Antarctic hypolith phage genomes were identified from the assembled
348 metagenome using VirSorter (30) on the iVirus platform hosted by Cyverse (48),
349 using the virome database and the microbial decontamination option. Only
350 predictions of categories 1, 2, 4 and 5 were used (phages and prophages identified
351 with the “pretty sure” and “quite sure” qualification). Additional phage environmental
352 phage contigs were downloaded from the IMG/VR database version 2018-07-01_4
353 (32) and used for the network construction. Taxonomic assignment of assembled
354 contigs was performed by using the DIAMOND blastx function with a viral database

355 downloaded from the NCBI Viral Genomes Resource and e-value set to 1e-5. ORFs
356 of VirSorter contigs were predicted using Prodigal v2.50 (31, 49) with the virus
357 genomes setting and annotated using eggNOG-mapper v1 (50) with the DIAMOND
358 option and the EggNOG v4.5.1 database (51). Annotation were visualized with the
359 ApE v2.0.55 plasmid editor (<http://jorgensen.biology.utah.edu/wayned/apel/>).

360 The CRISPR recognition tool (CRT) v1.2 was used with the default settings to
361 search for CRISPR arrays in the hypolith metagenome (52). The identified spacers in
362 the arrays were matched with the VirSorter phage database and the IMG/VR
363 database using blastn of the BLAST+ suite with the following parameters: -
364 qcov_hsp_perc 80 -task blastn -dust no -soft_masking false (53). Spacer matches of
365 > 90% sequence identity for the VirSorter genomes and >95% identity for the
366 IMG/VR genomes were exported and visualized as a network in Cytoscape (54),
367 where the nodes are spacers (grey) or genomes (blue = IMG/VR; red = VirSorter)
368 and the edges blastn matches.

369

370

371 **Acknowledgements**

372 We are grateful to the National Research Foundation (NRF) (Grant ID
373 118981, the South African National Antarctic Programme (SANAP 110717), and the
374 University of Pretoria for funding. TPM also wishes to acknowledge the Fulbright
375 Visiting Scholar Program for providing sabbatical funding. EMA gratefully
376 acknowledges the support of the Biotechnology and Biological Sciences Research
377 Council (BBSRC); this research was funded by the BBSRC Institute Strategic
378 Programme Gut Microbes and Health BB/R012490/1 and its constituent project(s)
379 BBS/E/F/000PR10353 and BBS/E/F/000PR10356

380

381 **References**

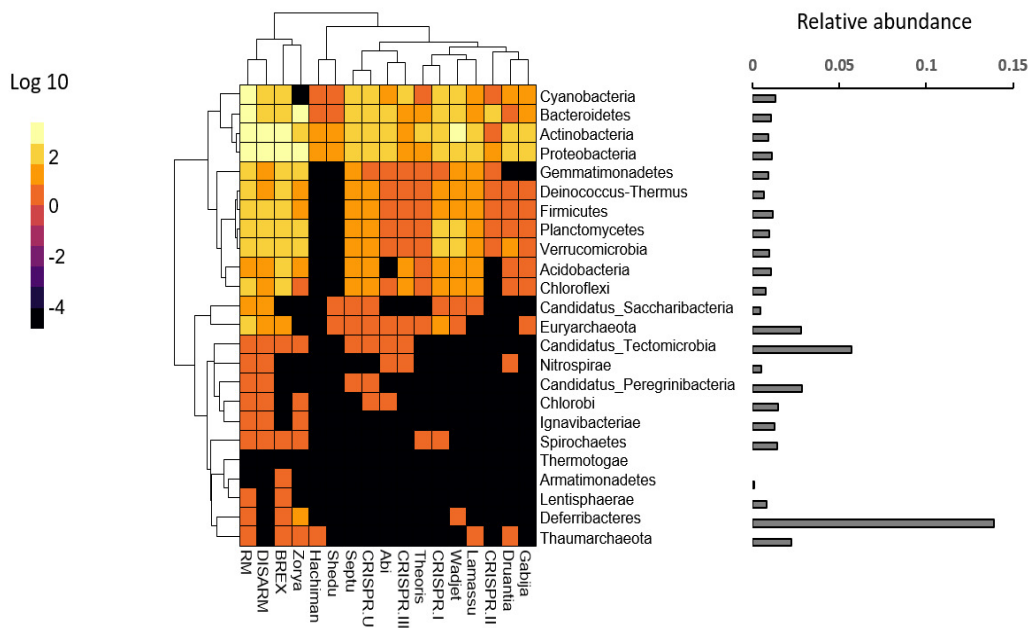
- 382 1. Makhalanyane TP, Van Goethem MW, Cowan DA. 2016. Microbial diversity and
383 functional capacity in polar soils. *Current Opinion in Biotechnology* 38:159-166.
- 384 2. Pearce DA. 2012. Extremophiles in Antarctica: Life at low temperatures, p 87-
385 118. *In* Stan-Lotter H, Fendrihan S (ed), *Adaption of Microbial Life to*
386 *Environmental Extremes*. Springer Vienna.
- 387 3. Casanueva A, Tuffin M, Cary C, Cowan DA. 2010. Molecular adaptations to
388 psychrophily: the impact of 'omic' technologies. *Trends in microbiology* 18:374-
389 81.
- 390 4. de Scally SZ, Makhalanyane TP, Frossard A, Hogg ID, Cowan DA. 2016. Antarctic
391 microbial communities are functionally redundant, adapted and resistant to
392 short term temperature perturbations. *Soil Biology & Biochemistry* 103:160-170.
- 393 5. de los Rios A, Wierzechos J, Sancho LG, Ascaso C. 2004. Exploring the physiological
394 state of continental Antarctic endolithic microorganisms by microscopy. *FEMS*
395 *microbiology ecology* 50:143-52.
- 396 6. Yergeau E, Kowalchuk GA. 2008. Responses of Antarctic soil microbial
397 communities and associated functions to temperature and freeze-thaw cycle
398 frequency. *Environmental microbiology* 10:2223-35.
- 399 7. Cowan DA, Makhalanyane TP, Dennis PG, Hopkins DW. 2014. Microbial ecology
400 and biogeochemistry of continental Antarctic soils. *Frontiers in Microbiology*
401 5:154.
- 402 8. Hogg ID, Cary SC, Convey P, Newsham KK, O'Donnell AG, Adams BJ, Aislabie J,
403 Frati F, Stevens MI, Wall DH. 2006. Biotic interactions in Antarctic terrestrial
404 ecosystems: Are they a factor? *Soil biology & biochemistry* 38:3035-3040.
- 405 9. Caruso T, Hogg ID, Nielsen UN, Bottos EM, Lee CK, Hopkins DW, Cary SC, Barrett
406 JE, Green TA, Storey BC. 2019. Nematodes in a polar desert reveal the relative
407 role of biotic interactions in the coexistence of soil animals. *Communications*
408 *biology* 2:63.

- 409 10. Lambrechts S, Willems A, Tahon G. 2019. Uncovering the uncultivated majority
410 in antarctic soils: toward a synergistic approach. *Frontiers in microbiology*
411 10:242.
- 412 11. Zablocki O, Adriaenssens EM, Cowan D. 2016. Diversity and ecology of viruses in
413 hyperarid desert soils. *Appl Environ Microbiol* 82:770-777.
- 414 12. Adriaenssens EM, Kramer R, Van Goethem MW, Makhalanyane TP, Hogg I, Cowan
415 DA. 2017. Environmental drivers of viral community composition in Antarctic
416 soils identified by viromics. *Microbiome* 5.
- 417 13. Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D.
418 2014. High-level diversity of tailed phages, eukaryote-associated viruses, and
419 virophage-like elements in the metaviromes of antarctic soils. *Appl Environ*
420 *Microbiol* 80:6888-6897.
- 421 14. Williamson KE, Radosevich M, Smith DW, Wommack KE. 2007. Incidence of
422 lysogeny within temperate and extreme soil environments. *Environmental*
423 *microbiology* 9:2563-2574.
- 424 15. Wei ST, Lacap-Bugler DC, Lau MC, Caruso T, Rao S, de los Rios A, Archer SK, Chiu
425 JM, Higgins C, Van Nostrand JD. 2016. Taxonomic and functional diversity of soil
426 and hypolithic microbial communities in Miers Valley, McMurdo Dry Valleys,
427 Antarctica. *Frontiers in microbiology* 7:1642.
- 428 16. Wei ST, Higgins CM, Adriaenssens EM, Cowan DA, Pointing SB. 2015. Genetic
429 signatures indicate widespread antibiotic resistance and phage infection in
430 microbial communities of the McMurdo Dry Valleys, East Antarctica. *Polar*
431 *Biology* 38:919-925.
- 432 17. Gómez P, Buckling A. 2011. Bacteria-phage antagonistic coevolution in soil.
433 *Science* 332:106-109.
- 434 18. Stern A, Sorek R. 2011. The phage - host arms race: shaping the evolution of
435 microbes. *Bioessays* 33:43-51.
- 436 19. Koskella B, Brockhurst MA. 2014. Bacteria-phage coevolution as a driver of
437 ecological and evolutionary processes in microbial communities. *FEMS*
438 *microbiology reviews* 38:916-931.
- 439 20. Makarova KS, Wolf YI, Koonin EV. 2013. Comparative genomics of defense
440 systems in archaea and bacteria. *Nucleic acids research* 41:4360-4377.
- 441 21. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak - Amikam Y, Afik
442 S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in
443 microbial genomes. *The EMBO journal* 34:169-183.
- 444 22. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R.
445 2018. Systematic discovery of antiphage defense systems in the microbial
446 pangenome. *Science* 359:eaar4120.
- 447 23. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, Doron S, Sorek R.
448 2018. DISARM is a widespread bacterial defence system with broad anti-phage
449 activities. *Nature microbiology* 3:90.
- 450 24. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. 2013. Type I restriction enzymes
451 and their relatives. *Nucleic acids research* 42:20-44.
- 452 25. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou
453 R, Brouns SJ, Charpentier E, Haft DH. 2015. An updated evolutionary
454 classification of CRISPR-Cas systems. *Nature Reviews Microbiology* 13:722.
- 455 26. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E.
456 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive
457 bacterial immunity system. *Nature communications* 3:945.

- 458 27. Fineran PC, Charpentier E. 2012. Memory of viral infections by CRISPR-Cas
459 adaptive immune systems: acquisition of new information. *Virology* 434:202-
460 209.
- 461 28. Vale PF, Little TJ. 2010. CRISPR-mediated phage resistance and the ghost of
462 coevolution past. *Proceedings of the Royal Society B: Biological Sciences*
463 277:2097-2103.
- 464 29. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, Thomas BC,
465 Banfield JF. 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas
466 viral defence systems. *Nature communications* 7:10613.
- 467 30. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal
468 from microbial genomic data. *PeerJ* 3:e985.
- 469 31. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2014. NCBI viral genomes resource.
470 *Nucleic acids research* 43:D571-D577.
- 471 32. Paez-Espino D, Chen I-MA, Palaniappan K, Ratner A, Chu K, Szeto E, Pillay M,
472 Huang J, Markowitz VM, Nielsen T. 2016. IMG/VR: a database of cultured and
473 uncultured DNA Viruses and retroviruses. *Nucleic acids research*:gkw1030.
- 474 33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
475 search tool. *Journal of molecular biology* 215:403-410.
- 476 34. Ravin NV, Svarchevsky AN, Dehò G. 1999. The anti - immunity system of phage -
477 plasmid N15: identification of the antirepressor gene and its control by a small
478 processed RNA. *Molecular microbiology* 34:980-994.
- 479 35. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms.
480 *Nature Reviews Microbiology* 8:317.
- 481 36. Cary SC, McDonald IR, Barrett JE, Cowan DA. 2010. On the rocks: the
482 microbiology of Antarctic Dry Valley soils. *Nature Reviews Microbiology* 8:129-
483 38.
- 484 37. Makhallanyane TP, Valverde A, Lacap DC, Pointing SB, Tuffin MI, Cowan DA. 2013.
485 Evidence of species recruitment and development of hot desert hypolithic
486 communities. *Environmental Microbiology Reports* 5:219-224.
- 487 38. Pointing SB, Warren-Rhodes KA, Lacap DC, Rhodes KL, McKay CP. 2007.
488 Hypolithic community shifts occur as a result of liquid water availability along
489 environmental gradients in China's hot and cold hyperarid deserts.
490 *Environmental microbiology* 9:414-24.
- 491 39. Daniel AdC, Pedrós-Alió C, Pearce DA, Alcamí A. 2016. Composition and
492 interactions among bacterial, microeukaryotic, and T4-like viral assemblages in
493 lakes from both polar zones. *Frontiers in microbiology* 7:337.
- 494 40. D'Amico S, Collins T, Marx JC, Feller G, Gerday C. 2006. Psychrophilic
495 microorganisms: challenges for life. *EMBO Rep* 7:385-389.
- 496 41. Guerrero LD, Vikram S, Makhallanyane TP, Cowan DA. 2017. Evidence of
497 microbial rhodopsins in Antarctic Dry Valley edaphic systems. *Environmental*
498 *Microbiology* 19:3755-3767.
- 499 42. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for
500 Illumina sequence data. *Bioinformatics* 30:2114-2120.
- 501 43. Peng Y, Leung HC, Yiu S-M, Chin FY. 2012. IDBA-UD: a de novo assembler for
502 single-cell and metagenomic sequencing data with highly uneven depth.
503 *Bioinformatics* 28:1420-1428.
- 504 44. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2010. Scaffolding pre-
505 assembled contigs using SSPACE. *Bioinformatics* 27:578-579.

- 506 45. Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic
507 data. *Genome research* 17:377-386.
- 508 46. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P,
509 Moineau S, Mojica FJ, Wolf YI, Yakunin AF. 2011. Evolution and classification of
510 the CRISPR–Cas systems. *Nature Reviews Microbiology* 9:467.
- 511 47. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK,
512 Geer RC, Gonzales NR. 2016. CDD/SPARCLE: functional classification of proteins
513 via subfamily domain architectures. *Nucleic acids research* 45:D200-D203.
- 514 48. Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB. 2017. iVirus:
515 facilitating new insights in viral ecology with software and community data sets
516 imbedded in a cyberinfrastructure. *The ISME journal* 11:7.
- 517 49. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
518 prokaryotic gene recognition and translation initiation site identification. *BMC*
519 *bioinformatics* 11:119.
- 520 50. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C,
521 Bork P. 2017. Fast genome-wide functional annotation through orthology
522 assignment by eggNOG-mapper. *Molecular biology and evolution* 34:2115-2122.
- 523 51. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T,
524 Mende DR, Sunagawa S, Kuhn M. 2015. eggNOG 4.5: a hierarchical orthology
525 framework with improved functional annotations for eukaryotic, prokaryotic
526 and viral sequences. *Nucleic acids research* 44:D286-D293.
- 527 52. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P.
528 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered
529 regularly interspaced palindromic repeats. *BMC bioinformatics* 8:209.
- 530 53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
531 2009. BLAST+: architecture and applications. *BMC bioinformatics* 10:421.
- 532 54. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N,
533 Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for
534 integrated models of biomolecular interaction networks. *Genome research*
535 13:2498-2504.
- 536

A



B

