Phages Actively Challenge Niche Communities in Antarctic Soils

Oliver K. I. Bezuidt,[a] Pedro Humberto Lebre,[a] Rian Pierneef,[a,b] Carlos León-Sobrino,[a] Evelien M. Adriaenssens,[c] Don A. Cowan,[a] Yves Van de Peer,[a,d,e] Thulani P. Makhalanyane[a]

aCentre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa
bBiotechnology Platform, Agricultural Research Council, Pretoria, South Africa
cQuadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom
dDepartment of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
eCenter for Plant Systems Biology, VIB, Ghent, Belgium

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

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

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

KEYWORDS Antarctic soils, archaea, bacteria, hypoliths, phages, viromics

Antarctic terrestrial environments, including open soils, permafrost, and the surface and interior of rocks, are typically oligotrophic and dominated by psychrophilic and psychrotolerant microbial communities (1–4). It has been suggested that the

extreme abiotic pressures of the environment, such as temperature, desiccation stress, and UV radiation, are dominant drivers of both the diversity and function of cold-adapted bacterial communities in terrestrial polar deserts (5–7). Similarly, biotic inter-actions, such as competition, symbioses, horizontal gene transfer (HGT), and predation, have also been shown to play a role in the distribution and diversity of microbial communities in these soil ecosystems (8–10). The presence of viruses, including bac-teriophages (henceforth termed phages), in these cold hyper-arid desert soils poten-tially adds an additional layer of complexity to the microbial system, but the extent to which phage-host interactions play a role in shaping community compositions and processes in cold desert soil niches remains a matter of speculation (11, 12).

Hypoliths, photosynthetic assemblages found below translucent rocks, are widely distributed across hot and cold deserts (13–15). In these arid environments, the absence of plants and larger organisms makes these niche communities, with relatively simple trophic structures, ideal systems for understanding microbial community dynamics (16). There is an increasing body of literature providing detailed insights regarding species richness, composition, and functional diversity of these communities (14, 17–25). These studies have provided clear evidence that hypoliths are essential drivers of functional processes in deserts, including carbon and nitrogen cycling in these ecosystems (26, 27). For instance, a recent study which reconstructed bacterial genomes from Antarctic hypoliths showed that these communities harbor xanthorhodopsin-like proteins and Na-pumping-like rhodopsins (28). Yet, fundamental ecological questions regarding the role played by viruses as drivers of diversification and trophic cycling in hypoliths remain unclear.

Antarctic desert hypolithic communities, in particular, have been shown to contain substantial viral populations, dominated by tailed bacteriophages of the order *Caudovirales* (11, 29–31). Microarray analysis of lithic niches identified an even greater phage diversity, including signatures of RNA bacteriophages of the family *Leviviridae*, ssDNA phage of the family *Microviridae*, and nontailed dsDNA tectiviruses (32). Together, these observations suggest that phages may play an important role in microbial community structures and functions.

The presence of active bacteriophages in a microbial community inevitably leads to the evolution of specialized bacterial defensive measures (33), and a diverse range of bacterial defense mechanisms against parasitic phages have been identified (34, 35). These include adaptive immunity elements, such as the CRISPR-Cas systems, and innate immunity mechanisms, such as restriction-modification (RM) and toxin-antitoxin abor-tive infection (Abi) systems (34). Recent pangenomic studies have also identified novel defense systems that are widely distributed across bacterial taxa and are thought to play a role in antiphage resistance (36–39). These include the bacteriophage exclusion (BREX) system, encoded by a 4 to 8 gene cluster, that provides resistance to *Siphoviridae* and *Myoviridae* tailed phages by inhibition of phage DNA replication (37), and other less well characterized systems such as the Thoeris, Shedu, and Gabija elements that increase bacterial host resistance to specific groups of phages (38).

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

RESULTS

The distribution of antiphage defense mechanisms shows a diversity of innate immunity genes. The distribution of antiphage defense systems in the metagenome was determined by mapping defense genes against the taxonomically assigned con-tigs. In total, 24,941 defense genes were detected, comprising 1.2% of the entire metagenome gene count. Approximately 40% of these were found in contigs attrib-
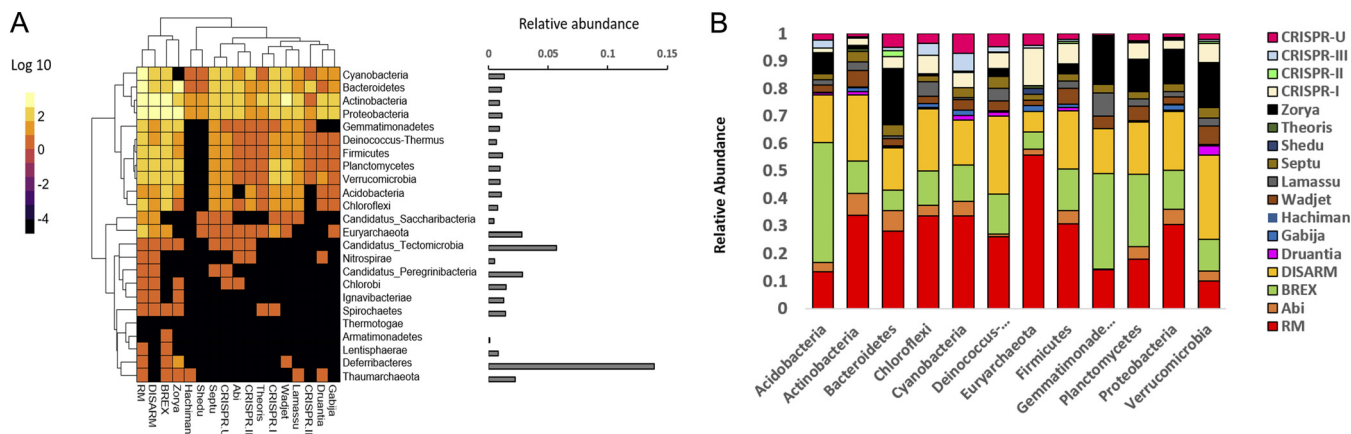
FIG 1 The relationship between relative abundances of taxa and defense systems. (A) The distribution of defense genes across known phyla in the metagenome. The heatmap includes all phyla which represented more than 1% of total scaffolds in the metagenome. The color scaling illustrates absolute abundance values after $\log_{10}$-transformation. The bar plot represents the percentage of defense genes detected relative to total gene count for each of the phyla represented in the heatmap. (B) Relative abundance of antiphage defense systems across different phyla. The relative abundance of each defense system is expressed as the ratio between the number of genes for that system and the total number of defense genes for the phylum. The top 12 phylum (representing 99% of contigs in the metagenome) are represented in the barplot.

uted to unknown phyla. The general distribution of defense genes across known phyla was consistent with the relative abundance of each phylum in the metagenome (Fig. 1A, Table S1). *Proteobacteria* harbored the highest number of antiphage genes (5,289 genes, 1.1% of total gene count for this phylum), followed by *Actinobacteria* (3,808, 0.9% of total gene count) and *Bacteroidetes* (2,128, 1.08% of total contig count). RM, DISARM, and BREX systems were the most abundant systems in the metagenome, contributing 67.6% of the total gene hits for antiphage defense systems. On the other side of the spectrum, the defense systems Shedu, Hachiman, and CRISPR-type 2 were present at relatively low abundances, and therefore had little apparent contribution to the global defense system distribution. The average contribution of defense genes to the total gene count per phyla was 1.8%, with *Deferribacteres* and "*Candidatus* Tecto-microbia" as outliers. However, it is important to note that these phyla represent a very small portion of the metagenome, and therefore the possibility that the high percentage of defense genes is biased toward the low gene count for these phyla cannot be disregarded.

Analysis of the relative contribution of each defense system within each phylum also showed that genes belonging to the RM, DISARM, and BREX systems were the main contributors across the majority of phyla (Fig. 1B). The recently discovered Zorya system was predominantly represented in the phyla *Gemmatimonadetes*, *Bacteroidetes*, *Planctomycetes*, *Proteobacteria*, and *Verrumicrobia*, while CRISPR systems showed the highest contribution in *Cyanobacteria* and *Euryarchaeota*. Interestingly, noncanonical antiphage systems represented more than 50% of the defense systems identified for all phyla aside from *Euryarchaeota*, with *Verrucomicrobia*, *Planctomycetes*, and *Acidobacteria* possessing the highest distribution of noncanonical defense genes.

Innate immunity is dominated by BREX and DISARM genes. As highlighted above, antiphage systems across phyla in the hypolith metagenome were dominated by noncanonical innate systems. Further analysis of the distribution of defense genes revealed that antiphage systems in the majority of phyla were dominated by BREX and DISARM genes. The two systems together accounted for 33.4% of defense genes, compared to 31.7% of genes belonging to canonical RM systems.

A total of 3,758 genes for the DISARM system were identified. These included the Class I marker gene *drm*D (449 counts, 11.9% of DISARM genes), which encodes the SNF2-like helicase (39), as well as the Class II marker gene *drm*A (1,020 counts, 17.1% of DISARM genes), which encodes a protein with a putative helicase domain (39). Similarly, a total of 4,598 genes representing all BREX types were identified in the metagenome.
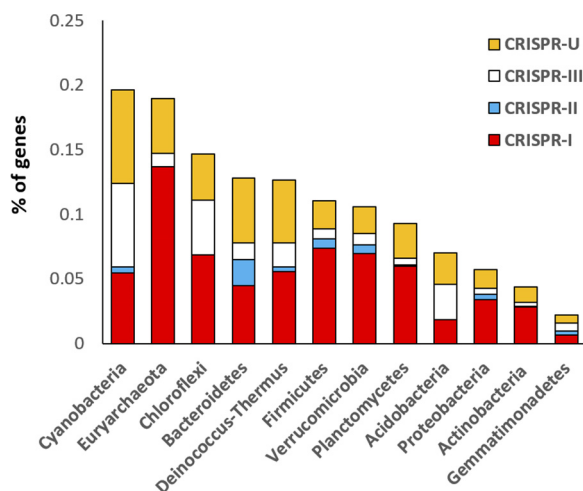
FIG 2 The relative abundances of the three CRISPR-Cas types as well as unclassified defense types across the 12 most dominant phyla. Relative abundances were determined as counts of CRISPR-*cas* genes compared to the total number of genes per phylum.

Interestingly, the most abundant gene from this system found in the metagenome, *pgl*W (2,640 counts, 57.4% of BREX genes), which codes for a serine/threonine kinase, is specific to the type 2 BREX system, also called the Pgl system (37). By comparison, of the 7,908 RM genes found in the metagenome, the most abundant is *hsdM* (1,423 counts, 18% of RM genes), which encodes a type I DNA methylase responsible for the protection of host DNA (40). In fact, more than 50% of RM defense genes were attributed to type I RM systems.

The third noncanonical system representing more than 10% of the antiphage defense systems in a subset of the phyla, the Zorya system, included a total of 2,411 genes in the metagenome. The majority of these were homologous to the two genes that encode elements of a proton channel, *zor*A and *zor*B. This is a common feature in all types of Zorya system and is thought to cause depolarization of the membrane upon infection (38).

Type I CRISPR-*cas* genes comprise the bulk of antiphage adaptive immunity genes. In total, 2,234 CRISPR-*cas* genes were identified in 1,601 contigs by searching for shared sequence similarities against the conserved domains database (CDD). A substantial proportion of all classified CRISPR-*cas* loci (71.4%) belonged to type I CRISPR-Cas systems, followed by type III (18.5%) and type II (10.2%) (Table S2). While the abundance of *cas* I-B loci sequences in the public databases suggests that the Cas-I mechanism is the most common in both bacteria and archaea (20% and 30% of total CRISPR loci, respectively) (41), less than 3% of these loci were present in our composite metagenome (Fig. 2, Table S2). Surprisingly, CRISPR-*cas* loci linked to Types I-C and I-E were the most prevalent, at 24.1% and 12.9% of classified CRISPR-*cas* loci, respectively. Another subtype identified at higher relative abundances than previously reported (41) was I-U, at 10.76% of classified *cas* loci. This subtype is characterized by the marker GSU0054 domain, which was the fourth most abundant *cas* CDD overall (108 occurrences) after *cas4*, *cas1*, and *cas2*.

Phage presence in the niche community is correlated with the CRISPR arrays. CRISPR arrays represent the history of infection by invading DNA (e.g., phages and plasmids) (42, 43), and a study of their composition and frequencies provides insights into phage-host interactions in an ecological context (44). A total of 878 CRISPR arrays harboring 10,292 spacers were identified in the metagenome, with an average length of 36 protospacers per array (Fig. S1A). CRISPR array sizes ranged from 2 to 249, with the majority (83.5% of total arrays) falling between 2 and 18 protospacers per array (Fig. S1B).
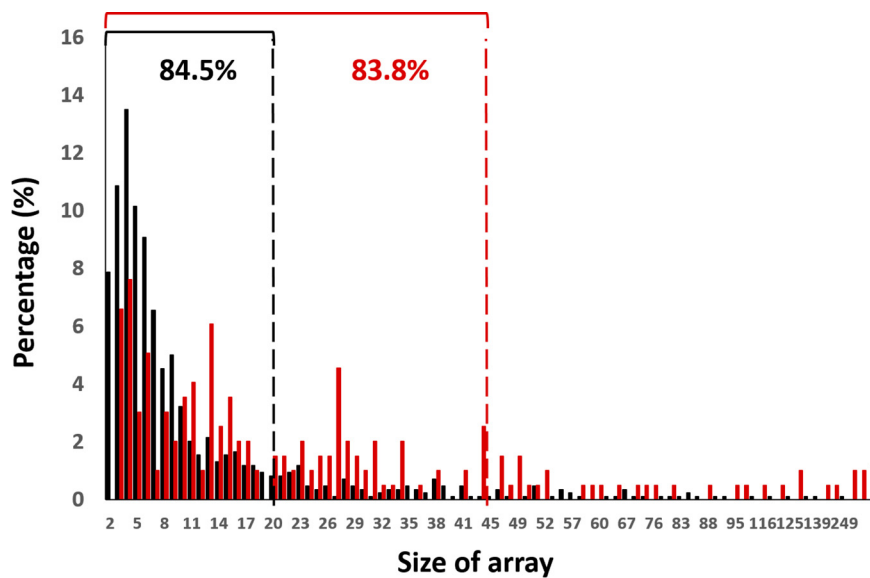
FIG 3 The distribution of CRISPR array sizes in the hypolith metagenome and a groundwater metagenome. Spacers for the hypolith and groundwater metagenomes are colored as red and black, respectively. Array size distributions in each metagenome are expressed as percentage of total number of arrays detected for each metagenome. Cutoffs below 85% are indicated by black and red dashed lines for both the hypolith and groundwater metagenomes, respectively, and represent the size range containing the majority (≤85%) of the array population.

The distribution of CRISPR array sizes in the metagenome was compared to data collected from a groundwater microbiome (45) to compare the array size distributions from environments with potentially different phage-host dynamics (11). The results show that CRISPR arrays in the hypolith metagenome exhibited a smaller and narrower size range compared to the groundwater community metagenome (Fig. 3). This suggests the existence of distinct phage infection frequencies between the different environments, i.e., with lower infection frequencies in the cold edaphic community.

In addition to using the CRISPR array as a tool for understanding infection history, the viral population in the Antarctic soil community was also assessed by assembly of the metavirome. A total of 793 contigs was assembled from the metagenomic sequence data using VirSorter (46). Taxonomic annotation of these contigs, using a database of viral reference genomes (47), unambiguously assigned 645 of these as viral, 560 of which were further assigned to the order of tailed phages, the *Caudovirales*. Within this order, the majority of contigs were assigned to *Siphoviridae* (52%), followed by unclassified *Caudovirales* (14%), and viruses with no assigned family (13%) (Fig. S2). To access the correlation between the viral contigs and the CRISPR arrays, spacers from the metagenome were matched to both the VirSorter contigs and a set of contigs from environmental data sets (IMG/VR) (48), which allowed for the taxonomic assignment of 394 (3.8% of total number of spacers) CRISPR-*cas* spacers (Fig. S3). The resulting similarity network (Fig. 4) showed that all 73 VirSorter phage contigs included in the network (red nodes) matched to CRISPR-*cas* spacers (gray nodes). This result suggests that a substantial fraction (11.3%) of the identified viral population had a history of infection *in situ* in the host population, and may therefore be actively involved in shaping the adaptive immunity of the microbial community. In addition, several distinct clusters showed matches between a single VirSorter contig and several spacers, suggesting these viral contigs are common infection agents.

Functional analysis using eggNOG showed the presence of genes that facilitate infection, such as genes that code for chitinases, which are involved in the degradation of the protective biofilm (49), as well as an AntA/AntB antirepressor gene, thought to be involved in phage anti-immunity (50) (Fig. 5). In addition, the eggNOG functional analysis of the 645 VirSorter viral contigs also revealed the presence of genes contrib-
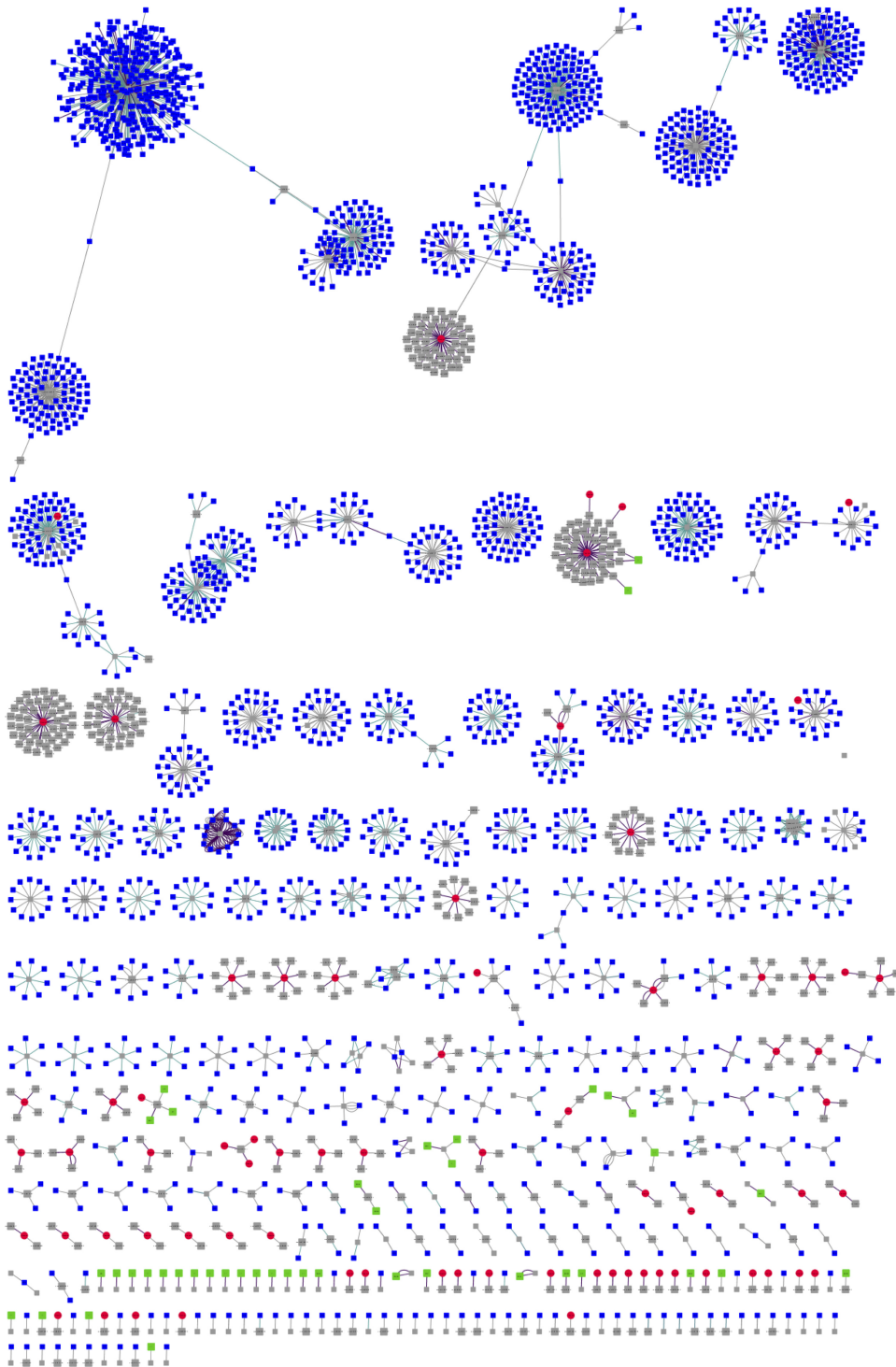
FIG 4 Network visualization of the relationship between the hypolith virome, CRISPR-spacers, and reference data sets. The nodes in the network are CRISP-R spacers (gray), hypolith metavirome (red), IMG/VR viromes (blue), and Antarctic soil viromes (green). Positive alignments between data sets are represented by the connections (edges) between nodes, with the connection distance being inversely proportional to the alignment strength.

uting to phage virulence (Table 1), the most abundant of which encode methyltransferases, which are actively involved in the evasion of the RM systems (51). This result suggests the possibility of an evolutionary pressure for the phages to develop evasion mechanisms against their hosts, which further hints at active phage-host dynamics in these long-enduring Antarctic hypoliths.
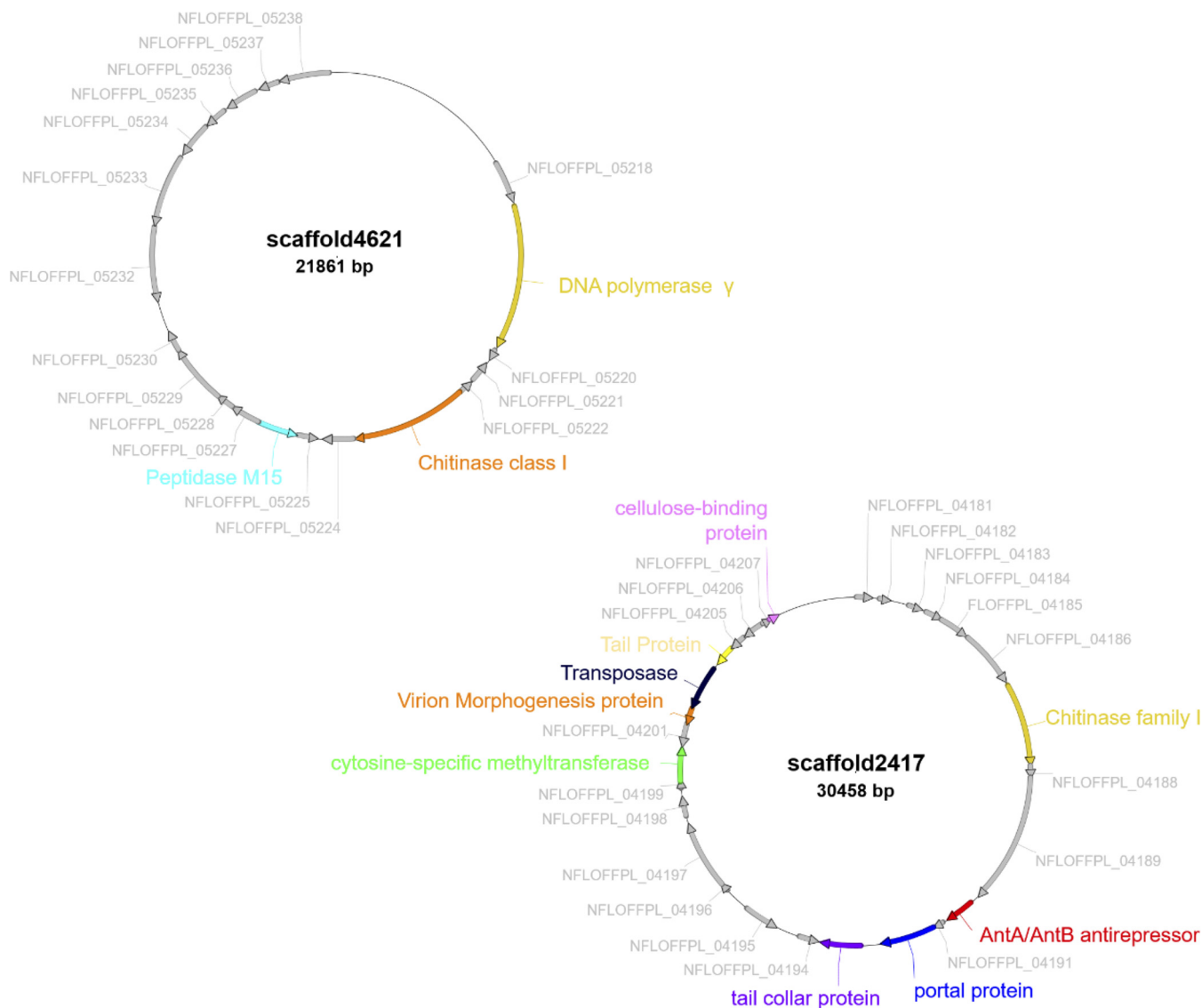
FIG 5 Gene architecture of the VirSorter contigs with matches to multiple CRISPR array spacers. These contigs were selected based on the presence of virulence and immunity evasion genes. Genes with known functions are highlighted to distinguish from hypothetical genes (gray).

DISCUSSION

Due to the relatively simple trophic structures in cold desert systems, including Antarctic soils, cryptic microbial communities are considered to be important drivers of local ecosystem services (52). However, the extent to which these communities are influenced by phages remains largely unexplored. Such interactions may shape the diversification and community interactions in cold desert systems. Qualitative surveys of Antarctic metaviromics have reported a high diversity of viruses associated with microbial communities of open soils, and cryptic niches (12, 29). Evidence, albeit limited, that Antarctic soil phages exist predominantly in a lysogenic rather than lytic lifestyle (30) has led to suggestions that the functional role of phages in this spatially restricted, water-constrained desert soil niche may be limited (11).

The results presented in this study provide the first evidence of interaction between phage and hosts in this psychrophilic edaphic environment. This is most evident in the correlation between the metavirome of the hypolith community and the CRISPR-arrays, which suggest the active evolution of the adaptive immune system against local viral threats. This idea of community adaption to local phage threat is further implied by the

TABLE 1 The number of genes linked to the contigs[a]

eggNOG annotations	No. of genes
DNA methylase	50
Adenine-specific methyltransferase	28
Cytosine-specific DNA methylase	11
Chitinase class I	6
Virulence-associated protein E	2
Repressor protein	1
Antirestriction ArdA family protein	1
Type 12 methyltransferase	1
Virulence-associated protein	1
Methyltransferase, type 11	1
Methyltransferase FkbM family	1

[a]eggNOG was used for VirSorter contig annotations, which were then linked to the virulence phenotypes. The number of genes for each category is shown in descending order.

positive correlation between the CRISPR arrays and viruses extracted from local soils. In fact, a previous study (19) has already suggested that recruitment from surrounding soils plays an important role in the development of hypoliths, and this might also be extended to the recruitment of phages from the surrounding ecosystem. Another indication of active interaction between phage and host is suggested by the presence of several methyltransferases in the metagenome-assembled viral contigs, which are a hallmark of viral evasion against native host RM systems (47, 51). Other genes found in this set of virome contigs includes genes specifically involved in the degradation of biofilm matrices and evasion against RM systems, further suggesting a complex network of interactions at play between phages and their hosts in the hypolithic environment.

While the metagenomics data analyzed in this study does not give a direct indication of the temporal scale of the phage-host interactions occurring in the hypolith, the short sizes of CRISPR-array sizes in the hypolith metagenome suggest a low frequency of infection. This low frequency is further hinted at when comparing the hypolith CRISPR-array sizes with those of a more fluid and homogenous environment, where viral-host interactions are assumed to be a frequent occurrence (45). Together, these results imply a model for viral-host interactions in hypoliths that follows the "static-step-static" development model suggested by Pointing et al. (53), driven by the stochastic and intermittent nature of rain events in such water-limited ecosystems. A surprising result from this study is the prevalence of noncanonical innate immunity systems, the most prominent of which are the BREX and DISARM systems. While these two systems have been shown to be widespread in bacteria using a pangenomic data set (37, 39), the present study represents the first evidence for the prevalence of these systems in ecological samples. As such, this result implies that noncanonical innate immunity is more important for antiphage microbial community defense than previously thought, and should therefore be the focus for future studies into innate immunity in the ecological context. There are also indications from the hypolith metagenome that the prevalence of noncanonical innate immunity over traditional RM and Abi systems for defense against phages is related to the adaptation of the hypolith communities to specific local viral populations. For instance, the Zorya system, the third most prevalent noncanonical immunity system in the metagenome, is hypothesized to operate similarly to the Abi system (38). In turn, Zorya systems provide resistance against a limited range of phages, including the ssDNA family *Microviridae* (38), which has been shown to be prevalent in Antarctic aquatic and soil niches (54).

Conclusion. Together, these results are not consistent with the suggestion that the constraints of the environment, such as low temperatures, low a_w, and resulting limited capacity for interparticle diffusion, lead to extremely localized phage-host interactions (11). Rather, the data are suggestive of a dynamic and continual interaction between host and phage. Nevertheless, interparticle communication and exchange may be limited to brief periods when bulk liquid water is present, such as after snow melt, for

example. Furthermore, the low metabolic rates (the inevitable consequence of Arrhenius effects [temperature dependence of reaction rates] in cold environments) should also limit the rates at which phages can replicate and propagate, further limiting the frequency of interactions with their hosts (55). We suggest that the localized nature of host-phage interactions in the hypolith niche and the limited interparticle communication, where bacterial hosts are not frequently challenged by novel phage threats, leads to a reliance of microbial communities on innate immunity as the primary defense against phage infection. The smaller sizes of CRISPR arrays in the Antarctic soil metagenome sequences compared to those from a temperate aquatic environment, and the underrepresentation of CRISPR systems, give further credence to the temporally sporadic interaction between phages and their hosts. Nevertheless, the correlation between the metavirome and the CRIPR-*cas* arrays, together with the presence of bacteriophage evasion genes in the metavirome, suggest that phage-host interactions within the hypolith community are a dynamic process that leads to coevolution of both phages and hosts. We therefore suggest that phages play a hitherto underestimated role in driving the evolution of Antarctic soil microbial communities by shaping their collective immunity.

MATERIALS AND METHODS

Sample collection, DNA extraction, and metagenomic sequencing. The sample collection, DNA extraction, and metagenomic sequencing protocols used in this study have been described previously (28). Briefly, a total of 50 samples were collected from hypolithic niches in the Antarctica Miers Valley (GPS 78°09′36.0″S 164°06′00.0″E) and stored in sterile Whirl-Pak bags (Nasco International, Fort Atkinson, WI, USA) at −20°C. Metagenomic DNA was extracted from each sample using a Power Soil DNA isolation kit (MO BIO, Carlsbad, CA, USA), and the purified DNA was pooled before further processing. Purified DNA was sheared into fragments of approximately 300 bp and further purified from 1% agarose gels. Subsequent sequencing was performed using Illumina HiSeq-2000 paired-end technology (2 × 101 bp), and the resulting reads were trimmed and assembled as described below.

Metagenome assembly and taxonomical annotation. Metagenomic DNA sequence data were quality filtered by trimmomatic version 0.36 using a phred cutoff of >30 (56). The assembly of high-quality reads from the metagenome sequence data set was conducted using the IDBA-UD tool (57) and contig lengths were extended (scaffolded) using SSPACE Basic (57). The statistics for the assembly of the metagenome are presented in Table S3. We estimated diversity and coverage using Nonpareil v3.0 (see Fig. S4) (58). To determine the relative abundances of bacteria, archaea, and viruses, reads were taxonomically assigned using kaiju v1.7.3 (Fig. S5) (59). Contigs were taxonomically assigned using the MEGAN v6 pipeline (60) with the NCBI taxonomy database for taxon ID assignment.

Detection of the innate and adaptive defense systems. Metagenomic contigs were used for functional gene predictions using prodigal v2.50, with the –meta parameter implementation (61). Predicted genes were subsequently screened for domain similarity with known defense systems against the conserved domain database (CDD) of clusters of orthologous groups (COGs) and protein families (Pfams) using RPS-BLAST (E value <1e-02) (49). These results were manually filtered for the identification of phage-specific defense systems, which include restriction-modification (RM), bacteriophage exclusion (BREX), abortive infection (Abi), defense island system associated with restriction-modification (DISARM), and other recently identified systems using a refined list of COG and Pfam position-specific score matrices (PSSMs) for marker genes in these systems (37–39, 62). A list of the marker genes used in this study can be found in Table S2. Defense genes that could not be clustered into specific systems were classified as ambiguous and were not considered for subsequent analysis.

Open reading frames (ORFs) predicted using prodigal v2.50 were queried against the CDD database for the presence of putative CRISPR-*cas* genes (63), using DELTA-BLAST at a cutoff E value of 1e-03. Multigene *cas* modules were identified as those having multiple *cas* annotated genes with ≤5 ORF spacings. Type and subtype classifications were assigned following the updated classification set by Makarova et al. (41).

Phage genome identification and CRISPR spacer matching. Antarctic hypolith phage genomes were identified from the assembled metagenome using VirSorter (46) on the iVirus platform hosted by Cyverse (64), using the virome database and the microbial decontamination option. Only predictions of categories 1, 2, 4, and 5 were used (phages and prophages identified with the "pretty sure" and "quite sure" qualification). Additional phage environmental phage contigs were downloaded from the IMG/VR database version 2018-07-01_4 (48) and used for the network construction. Taxonomic assignment of assembled contigs was performed by using the DIAMOND blastx function with a viral database downloaded from the NCBI Viral Genomes Resource and E value set to 1e-5. ORFs of VirSorter contigs were predicted using Prodigal v2.50 (47, 65) with the virus genomes setting and annotated using eggNOG-mapper v1 (66) with the DIAMOND option and the EggNOG v4.5.1 database (67). Annotations were visualized with the ApE v2.0.55 plasmid editor (http://jorgensen.biology.utah.edu/wayned/ape/).

The CRISPR recognition tool (CRT) v1.2 was used with the default settings to search for CRISPR arrays in the hypolith metagenome (68). The identified spacers in the arrays were matched with the VirSorter phage database and the IMG/VR database using blastn of the BLAST+ suite with the following

parameters: -qcov_hsp_perc 80 -task blastn -dust no -soft_masking false (69). Spacer matches of >90% sequence identity for the VirSorter genomes and >95% identity for the IMG/VR genomes were exported and visualized as a network in Cytoscape (70), where the nodes are spacers (gray) or genomes (blue = IMG/VR; red = VirSorter) and the edges blastn matches.

Sequence data availability. All sequence data linked to this study have been deposited in the NCBI SRA under accession number SRR3471615.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, DOCX file, 0.1 MB.
FIG S2, DOCX file, 0.1 MB.
FIG S3, DOCX file, 0.1 MB.
FIG S4, DOCX file, 0.1 MB.
FIG S5, DOCX file, 0.1 MB.
TABLE S1, DOCX file, 0.01 MB.
TABLE S2, DOCX file, 0.01 MB.
TABLE S3, DOCX file, 0.01 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Makhalanyane TP, Van Goethem MW, Cowan DA. 2016. Microbial diversity and functional capacity in polar soils. Curr Opin Biotechnol 38: 159–166. https://doi.org/10.1016/j.copbio.2016.01.011.

2. Pearce DA. 2012. Extremophiles in Antarctica: life at low temperatures, p 87–118. In Stan-Lotter H, Fendrihan S (ed), Adaption of microbial life to environmental extremes. Springer, Vienna, Austria. https://doi.org/10.1007/978-3-211-99691-1_5.

3. Casanueva A, Tuffin M, Cary C, Cowan DA. 2010. Molecular adaptations to psychrophily: the impact of 'omic' technologies. Trends Microbiol 18:374–381. https://doi.org/10.1016/j.tim.2010.05.002.

4. de Scally SZ, Makhalanyane TP, Frossard A, Hogg ID, Cowan DA. 2016. Antarctic microbial communities are functionally redundant, adapted and resistant to short term temperature perturbations. Soil Biology & Biochemistry 103:160–170. https://doi.org/10.1016/j.soilbio.2016.08.013.

5. de los Rios A, Wierzchos J, Sancho LG, Ascaso C. 2004. Exploring the physiological state of continental Antarctic endolithic microorganisms by microscopy. FEMS Microbiol Ecol 50:143–152. https://doi.org/10.1016/j.femsec.2004.06.010.

6. Yergeau E, Kowalchuk GA. 2008. Responses of Antarctic soil microbial communities and associated functions to temperature and freeze-thaw cycle frequency. Environ Microbiol 10:2223–2235. https://doi.org/10.1111/j.1462-2920.2008.01644.x.

7. Cowan DA, Makhalanyane TP, Dennis PG, Hopkins DW. 2014. Microbial ecology and biogeochemistry of continental Antarctic soils. Front Microbiol 5:154. https://doi.org/10.3389/fmicb.2014.00154.

8. Hogg ID, Cary SC, Convey P, Newsham KK, O'Donnell AG, Adams BJ, Aislabie J, Frati F, Stevens MI, Wall DH. 2006. Biotic interactions in Antarctic terrestrial ecosystems: are they a factor? Soil Biology & Biochemistry 38:3035–3040. https://doi.org/10.1016/j.soilbio.2006.04.026.

9. Caruso T, Hogg ID, Nielsen UN, Bottos EM, Lee CK, Hopkins DW, Cary SC, Barrett JE, Green TGA, Storey BC, Wall DH, Adams BJ. 2019. Nematodes in a polar desert reveal the relative role of biotic interactions in the coexistence of soil animals. Commun Biol 2:63. https://doi.org/10.1038/s42003-018-0260-y.

10. Lambrechts S, Willems A, Tahon G. 2019. Uncovering the uncultivated majority in Antarctic soils: toward a synergistic approach. Front Microbiol 10:242. https://doi.org/10.3389/fmicb.2019.00242.

11. Zablocki O, Adriaenssens EM, Cowan D. 2016. Diversity and ecology of viruses in hyperarid desert soils. Appl Environ Microbiol 82:770–777. https://doi.org/10.1128/AEM.02651-15.

12. Adriaenssens EM, Kramer R, Van Goethem MW, Makhalanyane TP, Hogg I, Cowan DA. 2017. Environmental drivers of viral community composition in Antarctic soils identified by viromics. Microbiome 5. https://doi.org/10.1186/s40168-017-0301-7.

13. Valverde A, Makhalanyane TP, Seely M, Cowan DA. 2015. Cyanobacteria drive community composition and functionality in rock-soil interface communities. Mol Ecol 24:812–821. https://doi.org/10.1111/mec.13068.

14. Chan Y, Lacap DC, Lau MC, Ha KY, Warren-Rhodes KA, Cockell CS, Cowan DA, McKay CP, Pointing SB. 2012. Hypolithic microbial communities: between a rock and a hard place. Environ Microbiol 14:2272–2282. https://doi.org/10.1111/j.1462-2920.2012.02821.x.

15. Pointing SB, Chan Y, Lacap DC, Lau MC, Jurgens JA, Farrell RL. 2009. Highly specialized microbial diversity in hyper-arid polar desert. Proc Natl Acad Sci U S A 106:19964–19969. https://doi.org/10.1073/pnas.0908274106.

16. Makhalanyane TP, Valverde A, Gunnigle E, Frossard A, Ramond J-B, Cowan DA. 2015. Microbial ecology of hot desert edaphic systems. FEMS Microbiol Rev 39:203–221. https://doi.org/10.1093/femsre/fuu011.

17. Bahl J, Lau MC, Smith GJ, Vijaykrishna D, Cary SC, Lacap DC, Lee CK, Papke RT, Warren-Rhodes KA, Wong FK, McKay CP, Pointing SB. 2011. Ancient origins determine global biogeography of hot and cold desert cyanobacteria. Nat Commun 2:163. https://doi.org/10.1038/ncomms1167.

18. Caruso T, Chan Y, Lacap DC, Lau MC, McKay CP, Pointing SB. 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. ISME J 5:1406–1413. https://doi.org/10.1038/ismej.2011.21.

19. Makhalanyane TP, Valverde A, Lacap DC, Pointing SB, Tuffin MI, Cowan DA. 2013. Evidence of species recruitment and development of hot

desert hypolithic communities. Environ Microbiol Rep 5:219–224. https://doi.org/10.1111/1758-2229.12003.

20. Makhalanyane TP, Pointing SB, Cowan DA. 2014. Lithobionts: cryptic and refuge niches, p 163–179. In Cowan DA (ed), Antarctic terrestrial microbiology: physical and biological properties of Antarctic soils. Springer, Vienna, Austria.

21. Adesioye F, Makhalanyane T, Vikram S, Sewell B, Schubert W, Cowan D. 2018 6FKX: crystal structure of an acetyl xylan esterase from a desert metagenome. https://doi.org/10.2210/pdb6fkx/pdb.

22. Le PT, Makhalanyane TP, Guerrero LD, Vikram S, Van de Peer Y, Cowan DA. 2016. Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts. Genome Biol Evol 8:2737–2747. https://doi.org/10.1093/gbe/evw189.

23. Vikram S, Guerrero LD, Makhalanyane TP, Le PT, Seely M, Cowan DA. 2016. Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. Environ Microbiol 18:1875–1888. https://doi.org/10.1111/1462-2920.13088.

24. Van Goethem MW, Makhalanyane TP, Cowan DA, Valverde A. 2017. Cyanobacteria and Alphaproteobacteria may facilitate cooperative interactions in niche communities. Front Microbiol 8:2099. https://doi.org/10.3389/fmicb.2017.02099.

25. Van Goethem MW, Makhalanyane TP, Valverde A, Cary SC, Cowan DA. 2016. Characterization of bacterial communities in lithobionts and soil niches from Victoria Valley. FEMS Microbiol Ecol 92:fiw051. https://doi.org/10.1093/femsec/fiw051.

26. Cowan DA, Sohm JA, Makhalanyane TP, Capone DG, Green TGA, Cary SC, Tuffin IM. 2011. Hypolithic communities: important nitrogen sources in Antarctic desert soils. Environ Microbiol Rep 3:581–586. https://doi.org/10.1111/j.1758-2229.2011.00266.x.

27. Chan Y, Van Nostrand JD, Zhou J, Pointing SB, Farrell RL. 2013. Functional ecology of an Antarctic Dry Valley. Proc Natl Acad Sci U S A 110:8990–8995. https://doi.org/10.1073/pnas.1300643110.

28. Guerrero LD, Vikram S, Makhalanyane TP, Cowan DA. 2017. Evidence of microbial rhodopsins in Antarctic Dry Valley edaphic systems. Environ Microbiol 19:3755–3767. https://doi.org/10.1111/1462-2920.13877.

29. Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D. 2014. High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. Appl Environ Microbiol 80:6888–6897. https://doi.org/10.1128/AEM.01525-14.

30. Williamson KE, Radosevich M, Smith DW, Wommack KE. 2007. Incidence of lysogeny within temperate and extreme soil environments. Environ Microbiol 9:2563–2574. https://doi.org/10.1111/j.1462-2920.2007.01374.x.

31. Wei STS, Lacap-Bugler DC, Lau MCY, Caruso T, Rao S, de Los Rios A, Archer SK, Chiu JMY, Higgins C, Van Nostrand JD, Zhou J, Hopkins DW, Pointing SB. 2016. Taxonomic and functional diversity of soil and hypolithic microbial communities in Miers Valley, McMurdo Dry Valleys, Antarctica. Front Microbiol 7:1642. https://doi.org/10.3389/fmicb.2016.01642.

32. Wei ST, Higgins CM, Adriaenssens EM, Cowan DA, Pointing SB. 2015. Genetic signatures indicate widespread antibiotic resistance and phage infection in microbial communities of the McMurdo Dry Valleys, East Antarctica. Polar Biol 38:919–925. https://doi.org/10.1007/s00300-015-1649-4.

33. Gómez P, Buckling A. 2011. Bacteria-phage antagonistic coevolution in soil. Science 332:106–109. https://doi.org/10.1126/science.1198767.

34. Stern A, Sorek R. 2011. The phage-host arms race: shaping the evolution of microbes. Bioessays 33:43–51. https://doi.org/10.1002/bies.201000071.

35. Koskella B, Brockhurst MA. 2014. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol Rev 38:916–931. https://doi.org/10.1111/1574-6976.12072.

36. Makarova KS, Wolf YI, Koonin EV. 2013. Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41:4360–4377. https://doi.org/10.1093/nar/gkt157.

37. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J 34:169–183. https://doi.org/10.15252/embj.201489455.

38. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the

39. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, Doron S, Sorek R. 2018. DISARM is a widespread bacterial defence system with broad antiphage activities. Nat Microbiol 3:90–98. https://doi.org/10.1038/s41564-017-0051-0.

40. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. 2014. Type I restriction enzymes and their relatives. Nucleic Acids Res 42:20–44. https://doi.org/10.1093/nar/gkt847.

41. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. 2015. An updated evolutionary classification of CRISPR–Cas systems. Nat Rev Microbiol 13:722–736. https://doi.org/10.1038/nrmicro3569.

42. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 3:945. https://doi.org/10.1038/ncomms1937.

43. Fineran PC, Charpentier E. 2012. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. Virology 434:202–209. https://doi.org/10.1016/j.virol.2012.10.003.

44. Vale PF, Little TJ. 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. Proc Biol Sci 277:2097–2103. https://doi.org/10.1098/rspb.2010.0055.

45. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, Thomas BC, Banfield JF. 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun 7:10613. https://doi.org/10.1038/ncomms10613.

46. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.

47. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. Nucleic Acids Res 43:D571–D577. https://doi.org/10.1093/nar/gku1207.

48. Paez-Espino D, Chen I-MA, Palaniappan K, Ratner A, Chu K, Szeto E, Pillay M, Huang J, Markowitz VM, Nielsen T, Huntemann M, K Reddy TB, Pavlopoulos GA, Sullivan MB, Campbell BJ, Chen F, McMahon K, Hallam SJ, Denef V, Cavicchioli R, Caffrey SM, Streit WR, Webster J, Handley KM, Salekdeh GH, Tsesmetzis N, Setubal JC, Pope PB, Liu W-T, Rivers AR, Ivanova NN, Kyrpides NC. 2017. IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. Nucleic Acids Res : 45:D457–D465. https://doi.org/10.1093/nar/gkw1030.

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

50. Ravin NV, Svarchevsky AN, Dehò G. 1999. The anti-immunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol Microbiol 34:980–994. https://doi.org/10.1046/j.1365-2958.1999.01658.x.

51. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317–327. https://doi.org/10.1038/nrmicro2315.

52. Cary SC, McDonald IR, Barrett JE, Cowan DA. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat Rev Microbiol 8:129–138. https://doi.org/10.1038/nrmicro2281.

53. Pointing SB, Warren-Rhodes KA, Lacap DC, Rhodes KL, McKay CP. 2007. Hypolithic community shifts occur as a result of liquid water availability along environmental gradients in China's hot and cold hyperarid deserts. Environ Microbiol 9:414–424. https://doi.org/10.1111/j.1462-2920.2006.01153.x.

54. Daniel AC, Pedrós-Alió C, Pearce DA, Alcamí A. 2016. Composition and interactions among bacterial, microeukaryotic, and T4-like viral assemblages in lakes from both polar zones. Front Microbiol 7:337. https://doi.org/10.3389/fmicb.2016.00337.

55. D'Amico S, Collins T, Marx J-C, Feller G, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. EMBO Rep 7:385–389. https://doi.org/10.1038/sj.embor.7400662.

56. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

57. Peng Y, Leung HC, Yiu S-M, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven

depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.

58. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. 2018. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems 3:e00039-18. https://doi.org/10.1128/mSystems.00039-18.

59. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257. https://doi.org/10.1038/ncomms11257.

60. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579. https://doi.org/10.1093/bioinformatics/btq683.

61. Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. Genome Res 17:377–386. https://doi.org/10.1101/gr.5969107.

62. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and classification of the CRISPR–Cas systems. Nat Rev Microbiol 9:467–477. https://doi.org/10.1038/nrmicro2577.

63. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129.

64. Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB. 2017. iVirus: facilitating new insights in viral ecology with software and community

data sets imbedded in a cyberinfrastructure. ISME J 11:7–14. https://doi.org/10.1038/ismej.2016.89.

65. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.

66. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.

67. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.

68. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics 8:209. https://doi.org/10.1186/1471-2105-8-209.

69. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

70. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.