Diverse Viruses Carrying Genes for Microbial Extremotolerance in the Atacama Desert Hyperarid Soil

Yunha Hwang,[a,b,c] Janina Rahlff,[b*] Dirk Schulze-Makuch,[c,d,e] Michael Schloter,[f] Alexander J. Probst[b]

aDepartment of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

bGroup for Aquatic Microbial Ecology, Environmental Microbiology and Biotechnology, Department of Chemistry, University of Duisburg-Essen, Essen, Germany

cCenter for Astronomy and Astrophysics, Technische Universität Berlin, Berlin, Germany

dGerman Research Centre for Geosciences, Section Geomicrobiology, Potsdam, Germany

eDepartment of Experimental Limnology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Stechlin, Germany

fResearch Unit for Comparative Microbiome Analysis, Helmholtz Zentrum München, German Research Center for Environmental Health, Oberschleißheim, Germany

ABSTRACT Viruses play an essential role in shaping microbial community structures and serve as reservoirs for genetic diversity in many ecosystems. In hyperarid desert environments, where life itself becomes scarce and loses diversity, the interactions between viruses and host populations have remained elusive. Here, we resolved host-virus interactions in the soil metagenomes of the Atacama Desert hyperarid core, one of the harshest terrestrial environments on Earth. We show evidence of diverse viruses infecting a wide range of hosts found in sites up to 205 km apart. Viral genomes carried putative extremotolerance features (i.e., spore formation proteins) and auxiliary metabolic genes, indicating that viruses could mediate the spread of microbial resilience against environmental stress across the desert. We propose a mutualistic model of host-virus interactions in the hyperarid core where viruses seek protection in microbial cells as lysogens or pseudolysogens, while viral extremotolerance genes aid survival of their hosts. Our results suggest that the host-virus interactions in the Atacama Desert soils are dynamic and complex, shaping uniquely adapted microbiomes in this highly selective and hostile environment.

IMPORTANCE Deserts are one of the largest and rapidly expanding terrestrial ecosystems characterized by low biodiversity and biomass. The hyperarid core of the Atacama Desert, previously thought to be devoid of life, is one of the harshest environments, supporting only scant biomass of highly adapted microbes. While there is growing evidence that viruses play essential roles in shaping the diversity and structure of nearly every ecosystem, very little is known about the role of viruses in desert soils, especially where viral contact with viable hosts is significantly reduced. Our results demonstrate that diverse viruses are widely dispersed across the desert, potentially spreading key stress resilience and metabolic genes to ensure host survival. The desertification accelerated by climate change expands both the ecosystem cover and the ecological significance of the desert virome. This study sheds light on the complex virus-host interplay that shapes the unique microbiome in desert soils.

KEYWORDS auxiliary metabolic genes, viromics, metagenomics, viral dispersal, extremophiles, virus-host interactions

Viruses are considered the most abundant biological entities on Earth (1), with high genomic diversity (2) and an expanding ecological and biogeochemical importance. Viruses, particularly bacteriophages, shape microbial community turnover and composition (3, 4), nutrient cycling (5, 6), and microbial evolution (7, 8) in marine (9) and freshwater (10) environments. Progress in soil virome studies is lagging compared to those in marine and gut microbiome systems (11–13), mainly due to the difficulties

of isolating viruses from heterogeneous and complex soil environments (14). However, recent metagenomic approaches revealed diverse soil viruses in high abundance (15) which play significant roles in carbon processing (16–18) and other types of nutrient turnover (19, 20). Even less explored are viruses in extreme soil environments, where life itself becomes scarce in biomass and low in biodiversity (21–23). Understanding the abundance and diversity of viruses as well as their interactions with extremotolerant microbes in environments can highlight the unique roles viruses may play in driving the adaptation of their hosts and reveal the dispersal and diversification of viruses in sparsely populated and harsh environments.

Hyperarid desert soils are unique terrestrial environments, where low water availability limits proliferation and diversification of life. The biota that permanently inhabits these environments is often limited to a few bacterial and archaeal phyla. Recent studies of warm (i.e., the Namib Desert and Sahara Desert) and cold (i.e., Antarctic soil) hyperarid desert viromes have revealed abundant viruses of diverse lineages and sizes, with lysogenic and pseudolysogenic viruses being more prevalent than lytic viruses in warm deserts (21, 24–26). With little water availability and extended periods of drought, hyperarid desert soils present a distinct model for studying viral persistence and dispersal. In these ecosystems, viral mobility is limited compared to that in aquatic environments, in which both viruses and hosts freely diffuse (14).

The Atacama Desert is one of the harshest environments on Earth, with its hyperarid core experiencing extreme desiccation with a mean annual precipitation of <2 mm (27). The surface soil of the Atacama hyperarid core generally contains $<1\%$ (by weight) water and experiences high daily UV radiation (30 Js^2m^{-2}) (28), extreme diurnal temperature fluctuations ($\sim60°C$) (29), and additional osmotic pressure from the accumulation of salts (28, 30). Scarce populations of highly adapted microbial communities consisting of *Actinobacteria*, *Firmicutes*, *Chloroflexi* (28, 30, 31), and, more recently discovered, *Thaumarchaeota* (29) were found to inhabit soils of the Atacama hyperarid core. However, very little is known about viruses from these desert soil microbiomes. Crits-Christoph et al. (32) identified viral sequences and their potential hosts in halite endoliths of the Atacama. Additionally, Uritskiy et al. (33) detected transcriptionally active viruses potentially infecting *Halobacteria* also inhabiting halite salt nodules in a salar located in the Atacama Desert. These niche halite host-virus relationships highlight the need to characterize the impact of viruses in broad desert soils that represent one of the largest and rapidly expanding terrestrial ecosystems on the planet ($\sim35\%$ of the Earth's land surface [34]). In our previous study (28), we also detected viruses in the Atacama Desert soils of the hyperarid core using read-based analyses of metagenomes; however, no information exists regarding host-virus interactions, dispersal, or the potential function of these viruses.

To understand the diversity and ecological impact of viruses inhabiting hyperarid soils, we investigated viral genomes assembled from soil metagenomes of the Atacama hyperarid core. We identified host-virus interactions, innate and adaptive host immunity elements, and phylogenetic diversity of viruses across geographically distant sampling locations. We analyzed putative extremotolerance genes and auxiliary metabolic genes (AMGs) found in the predicted viral sequences, providing evidence for a complex trade-off between viral predation and viral delivery of extremotolerance genes to microbes inhabiting harsh hyperarid desert soils.

RESULTS

Soil metagenomes of the Atacama hyperarid core feature heterogeneous viromes. We predicted viral scaffolds in 11 assembled metagenomes (4.1 Gbp in total) from three different boulder fields (Lomas Bayas [L], Maria Elena [M], and Yungay [Y]), which were previously studied for the impact of boulder cover on the soil microbiome, uncovering highly adapted microbes sheltered below the boulders of expansive boulder fields in the Atacama Desert hyperarid core (29). The aforementioned study compared the microbiomes in two different soil compartments (below the boulder [B] and control, i.e., exposed soil adjacent to the boulder [C]), for which we kept the designations consistent in
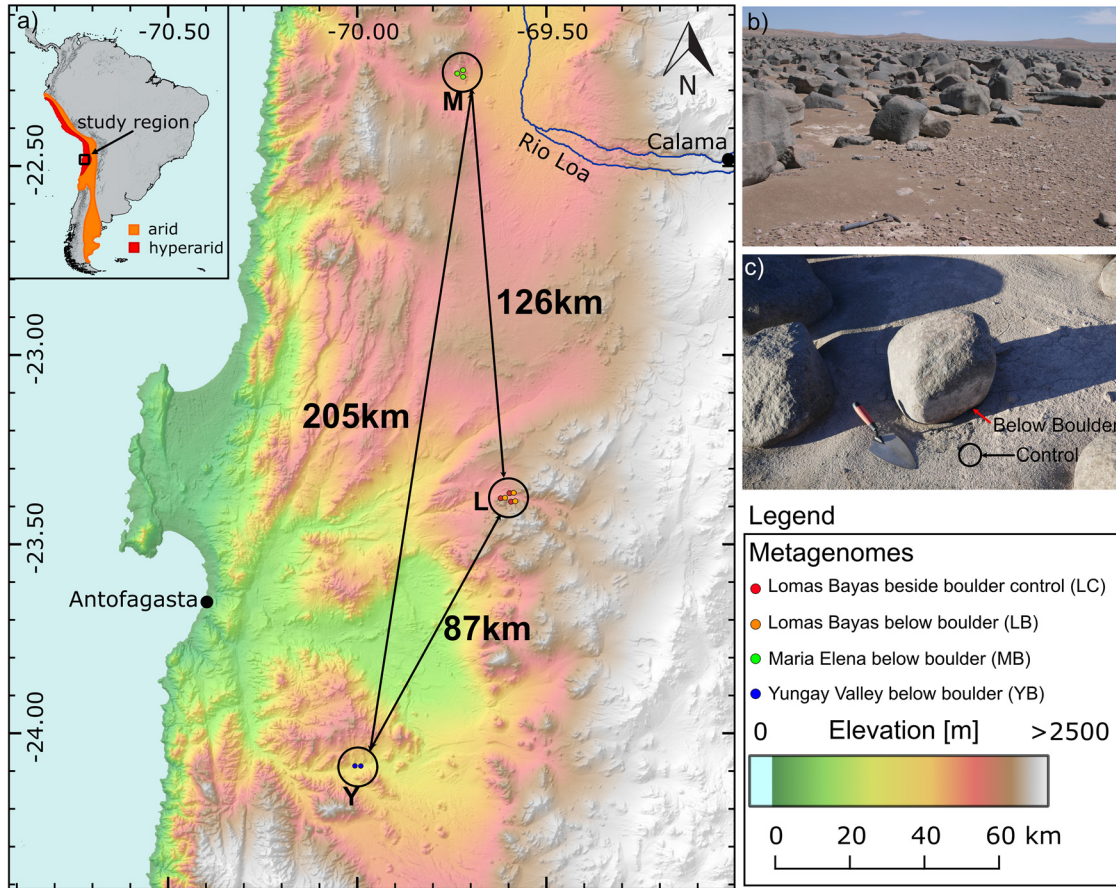
FIG 1 Sampling location information. (a) Map of the sampling locations M (Maria Elena), L (Lomas Bayas), and Y (Yungay Valley). Distances between sampling locations are shown. The number of metagenome samples per location is shown in circles. (b) Yungay boulder field. (c) A sampled boulder in Maria Elena boulder field. Red and black arrows indicate sample types B and C, respectively. The map was made using ASTER Global Digital Elevation Model (104).

this paper (for a map of the sampling locations, see Fig. 1; for the analysis workflow, see Fig. S1). In total, 6,809 of 707,509 examined scaffolds were predicted to be viral. After quality and length filtering, we identified 86 viral scaffolds (referred to here as viral genomes) with lengths of >0 kb forming 84 viral "populations" (dereplicated at 99% identity). In detail, VirSorter (35) predicted 79 viral genomes, while VIBRANT (36) predicted 37, including 30 overlapping between the two tools. The average length of the predicted viral genomes was 32.7 kbp (± 29.5 kbp), with the longest being 177 kbp and smallest being 10 kbp. The average G+C content of viral genomes was 58.7% (±10.3%), and the average coding density was 91.0% (±4.2%). The viral genomes were of varying quality: 9.30% "complete," 10.5% "high quality," 17.4% "medium quality," 61.6% "low quality," and 1.16% "not determined" according to CheckV (37). From these 84 viral populations, eight were predicted to be lysogenic. An overview of the viral population genomes can be found in Table S1a.

We observed a large degree of heterogeneity in the viral populations between samples, in terms of both alpha and beta diversity. Figure 2a shows the relative abundances of viral populations based on the total number of reads normalized across samples. The mapping-based coverage of viral populations identified in this study varied significantly, ranging between 4.5 and 5,075, and the total abundance of these viral populations varied up to 65-fold between samples. Notably, we did not detect prevalence of lysogenic viruses in L and M sites, while YB samples exhibited higher relative abundance (~79%) of lysogenic viruses. Samples collected from L sites ($n = 6$) had higher
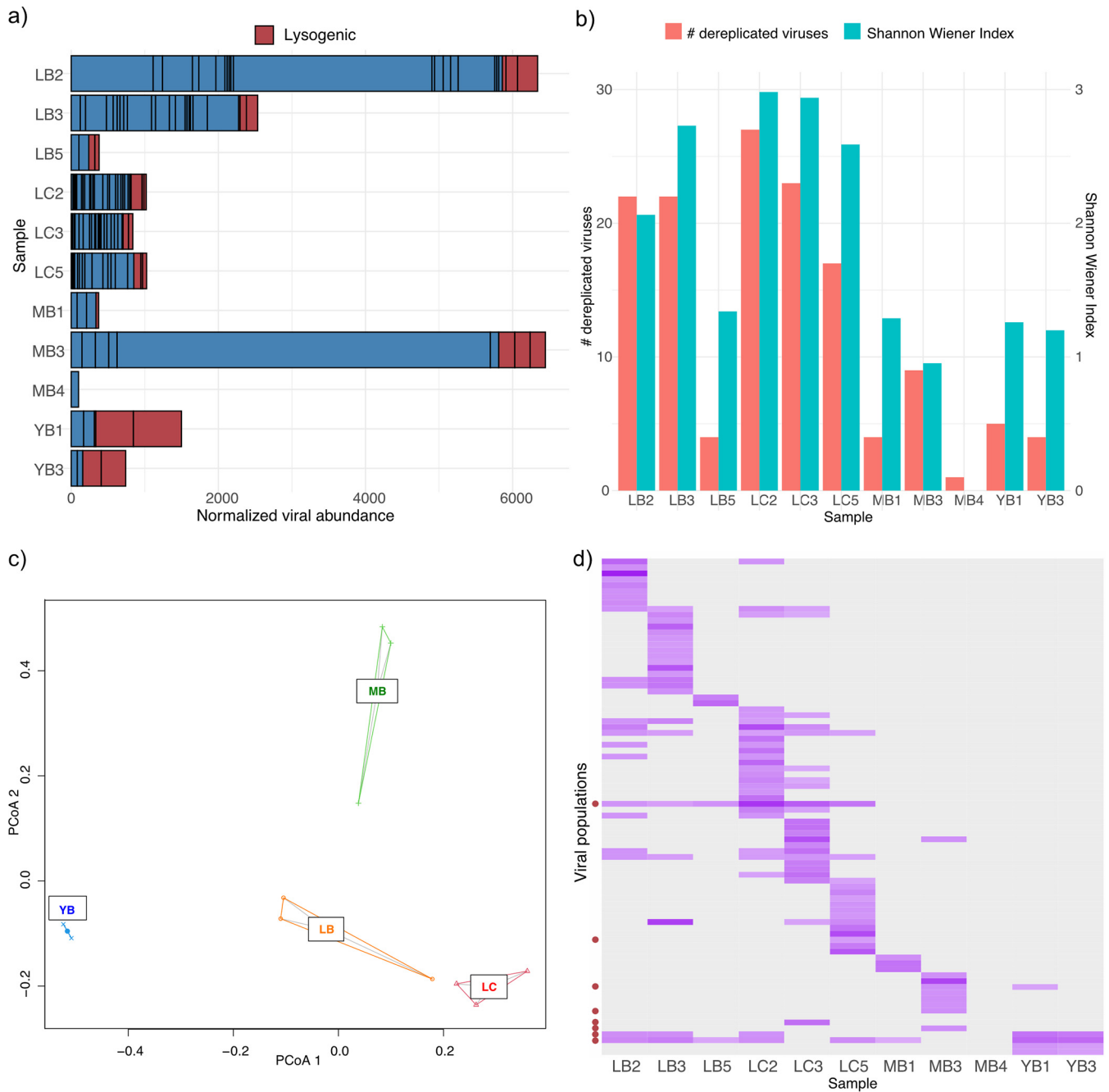
FIG 2 Abundance and diversity analyses of viromes. (a) Abundance of viral contigs normalized by sequencing depth of each sample. Each bar denotes a single viral population, and predicted lysogenic viral populations are marked in red. (b) Alpha diversity and species evenness of viromes. (c) PCoA plot of viromes using a Bray-Curtis distance matrix. (d) Relative abundance profile of each viral population. Darker purple denotes a viral population with higher relative abundance in a sample (gray indicates none detected). Lysogenic populations are marked with red circles on the y axis. Graphs were visualized in R v4.0.2 (81) using ggplot2 (83).

alpha diversity and species evenness (Fig. 2b) than samples from M and Y ($n = 5$) (Welch's t test, $P = 0.0022$ [alpha diversity] and $P = 0.0056$ [species evenness]). Principal-coordinate analysis (PCoA) (Fig. 2c) based on Bray-Curtis distances showed clustering by sampling site, while the beta diversities varied between sites, with YB viromes being particularly conserved and LB and MB viromes exhibiting higher variability than LC viromes. Permutational multivariate analysis of variance (PERMANOVA) (38) confirmed statistically significant differentiation of viral communities based on the sampling site ($R^2 = 0.384$; $P = 0.011$). Biota-environment (BioENV) analysis (39) using geochemical and

a)

Legend

Node fill color
Phage type

- *Arthrobacterphage*
- *Gordoniaphage*
- *Lactococcusphage*
- *Streptomycesphage*
- *Mycobacteriumphage*
- LC phages
- LB phages
- MB phages
- YB phages
- Other phage

Node shape
Phage family

- Myoviridae
- Siphoviridae
- Podoviridae
- Other
- Unassigned

b)

Taxonomy

- AVC1
- AVC10
- AVC12
- AVC14
- AVC15
- AVC16
- AVC17
- AVC19
- AVC2
- AVC20
- AVC3
- AVC4
- AVC6
- AVC7
- AVC8
- AVC9
- Outlier
- Singleton
- Siphoviridae1_Arthrobacterphage
- Siphoviridae2_Streptomycesphage
- Siphoviridae3_Lactococcusphage
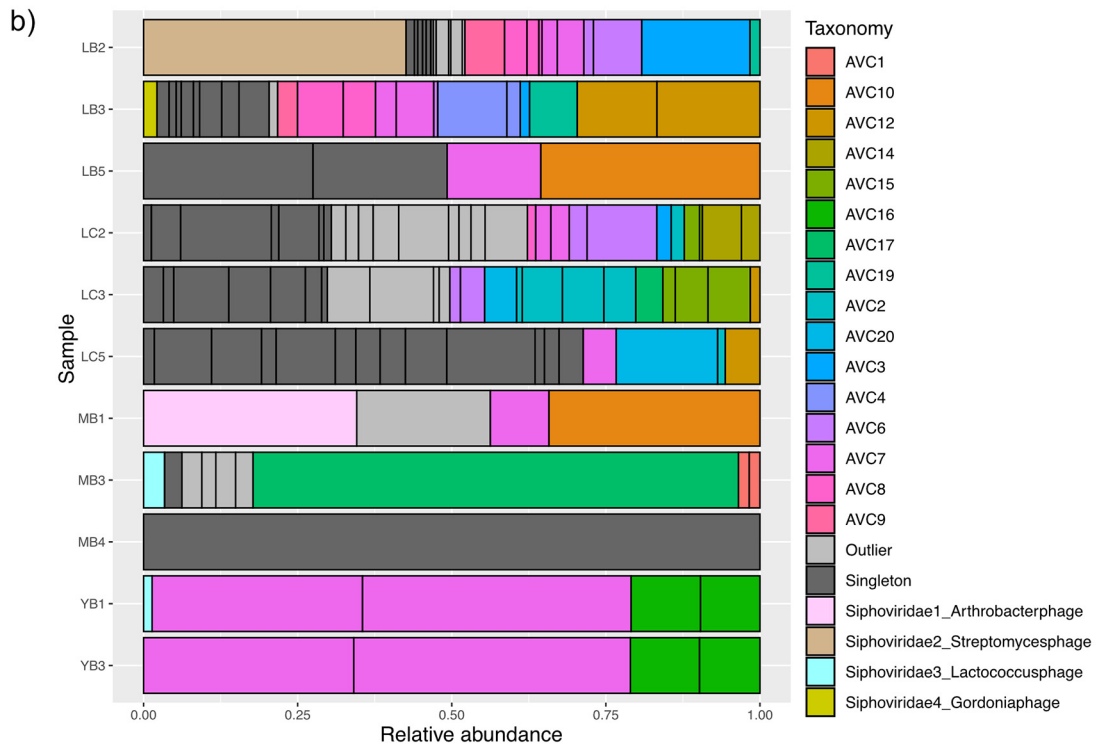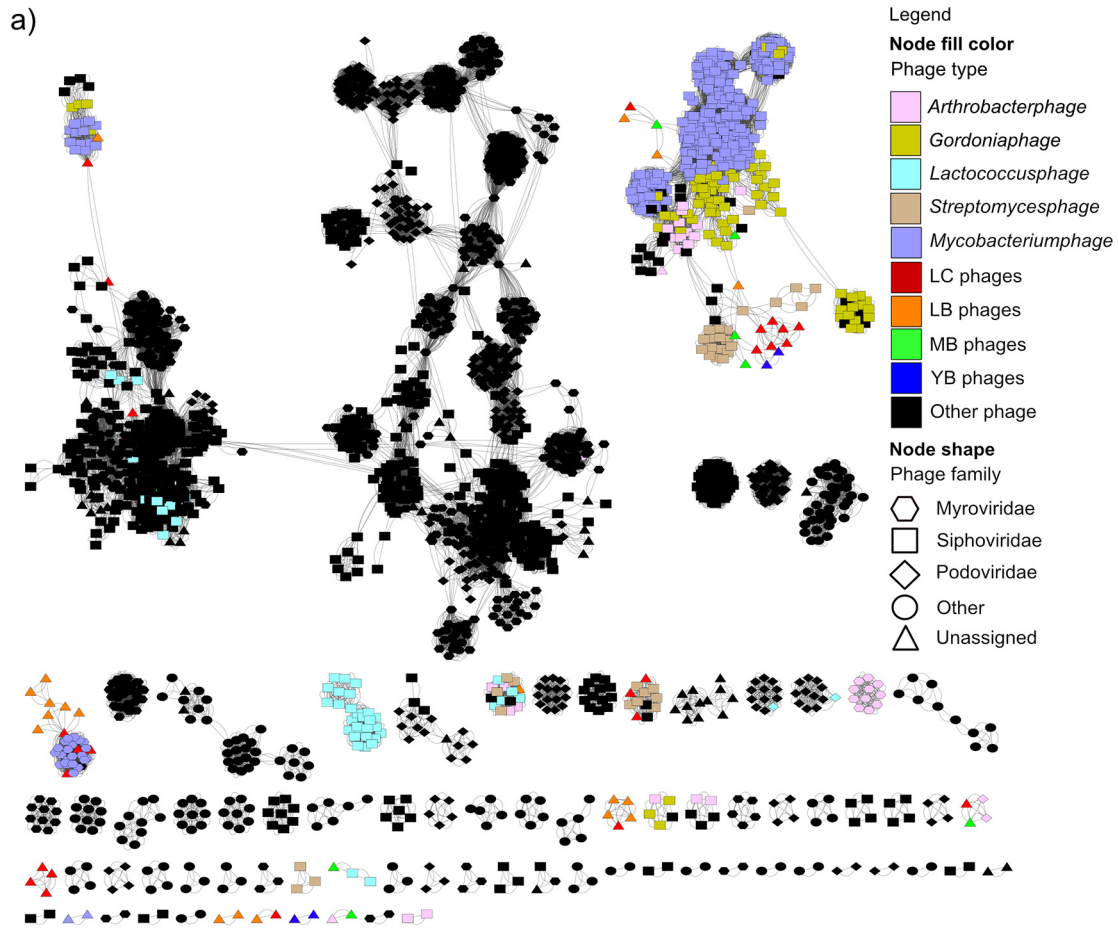- Siphoviridae4_Gordoniaphage

FIG 3 Clustering of Atacama viral genomes with reference viral genomes based on shared genes. (a) vConTACT2 output network with a significance greater than 1. The network was visualized using Cytoscape v3.8.0 (88). Queried viruses are color coded

(Continued on next page)

environmental metadata (Table S1b) revealed temperature and Na$^+$ concentration to be most correlated with the viral community composition (rho = 0.5642; P = 0.001). The majority (66.7%) of the viral populations were unique to the sample; only one lysogenic virus with a representative genome of 12.6 kbp was detected across all three sites in eight samples, and three additional viral populations were observed between two different sites (Fig. 2d).

By comparing the distributions of viral and microbial genome abundances in each sample (Fig. S2), we found that only LC samples featured statistically significantly lower abundances (P < 0.05; Welch's t test) of the viral genomes than microbial genomes, while no statistically significant differences could be observed in other samples. However, the ranges of viral genome abundances were greater than those of microbes in all samples, with samples LB2 and MB3 featuring viruses with abundances up to two orders of magnitude greater than those of microbes. We conducted symmetric co-correspondence analysis (sCoCA) to test whether the viral and microbial community compositions covary. The best sCoCA model using the first three axes determined the common variance between the viral and microbial communities to explain 37.2% and 56.8% of the total variances of viral and microbial communities, respectively (P = 0.006). The first three axes computed by sCoCA accounted for the 62.3% (CoCA1 = 26.5%; CoCA2 = 20.4%; CoCA3 = 15.3%) of the common variance. The first three axes of the ordinations in viral and microbial communities were highly correlated with each other (Pearson product-moment correlation coefficient > 0.99). The relative abundances of microbial taxa across samples used to conduct sCoCA are visualized in Fig. S3a, and the ordination biplots (Fig. S3b and c) illustrate highly similar positioning of the samples along the first two axes identified by sCoCA between microbial and viral communities. Higher species richness and species evenness (Shannon indices) could be observed in LB and LC samples for both microbial and viral communities; however, little pattern in the microbial population could be observed at the phylum level, suggesting that the covariance patterns are rooted in the abundance profiles of individual populations. Interestingly, neither viral nor microbial communities were predictive of each other when CoCA was conducted in predictive mode (P > 0.05) (40).

Atacama viruses are phylogenetically novel and diverse. We clustered 84 dereplicated viral population genomes using intergenomic similarities (41). Eighty-two clusters were formed at the genus level (intergenomic similarity threshold at 70%), indicating that all except two viral population genomes recovered in this study are of different genera. The OPTSIL clustering of viruses (42) yielded 84, 64, and 14 clusters at the species, genus, and family levels (Fig. S4). vConTACT2 (43) was used to cluster the Atacama viral genomes with 2,616 known prokaryotic viruses (Fig. 3a). Twenty-two Atacama viral genomes were related to phages infecting *Gordonia*, *Mycobacterium*, *Streptomyces*, and *Arthrobacter* at a taxonomic level higher than genus. Only four viral contigs were predicted to be in the same genus as a *Gordonia* phage, a *Streptomyces* phage, a *Lactococcus* phage, and an *Arthrobacter* phage from the reference database, all belonging to the order *Caudovirales*, family *Siphoviridae* (tailed double-stranded-DNA [dsDNA] phage). vConTACT2 also predicted 20 genus-level clusters consisting of two to five Atacama viruses (Atacama viral clusters [AVC]). Of the remaining viral populations that could not be clustered, 27 were classified as "singleton" viruses, which share no or very few genes with the database and each other, and 19 "outlier" viruses, which could be associated with existing sequences but could not be clustered due to a low confidence level. Atacama viral genomes tended to cluster with each other in the gene sharing network rather than with the viral sequences in the database. Notably, the majority of the Atacama viruses that are related to reference *Streptomyces* phages and *Mycobacterium* phage were recovered from the LC samples.

FIG 3 Legend (Continued)
based on the sampling site they were recovered from, and reference viruses are color coded based on the host they were annotated to infect. Node shape denotes the phage family of reference viruses. (b) Relative abundance of identified taxonomic groups per sample. AVC denotes genus-level Atacama viral clusters identified using vConTACT2. Each bar represents a distinct viral population.

Across the three different tools, we estimated between 64 and 80 genus-level clusters among the 84 viral populations.

The relative abundance profiles of different taxonomic groups in each sample (Fig. 3b) illustrated a high level of heterogeneity between the three sites as well as among the samples collected from sites L and M. In contrast, the two Y samples were almost identical in the taxonomic composition of their viromes. For all the samples, singletons and outlier clusters of viral genomes that are biologically novel (due to very little protein homology to the existing database) constituted the majority.

Sequence-informed putative host-virus interactions indicate dispersal of hosts and/or viruses. Previously, we recovered 73 medium- to high-quality (>75% completeness; <15% contamination) metagenome-assembled genomes (MAGs) across 11 metagenomes from three sampling locations (29). The MAGs were classified as 34 *Actinobacteria*, 30 *Chloroflexi*, eight *Thaumarchaeota*, and one *Firmicutes*. We resolved 74 unique interactions between 30 MAGs and 15 viruses using the following four sequence-based methods: (i) protospacer-to-spacer match for MAGs containing CRISPR arrays, (ii) oligonucleotide frequency similarity (VirHostMatcher) (44), (iii) tRNA matching, and (iv) nucleotide sequence homology. Figure 4a illustrates the putative host-virus interactions, specifying the method by which the interactions were resolved.

We identified 14 unique interactions between six actinobacterial MAGs with CRISPRs and seven viruses. Of 73 MAGs surveyed, nine actinobacterial and two *Chloroflexi*-derived MAGs contained CRISPR arrays from which direct repeat (DR) sequences were extracted. All MAGs carried unique sets of DR sequences, although three identical DRs appeared across six MAGs (see Table S1c for details). In total, 18 identified DR sequences were used to recover 3,438 unique spacers directly from the reads in their respective metagenomes. Five actinobacterial MAGs classified as *Rubrobacter* recovered from site L clustered based on their shared infection histories with three phylogenetically distinct viruses also detected from site L, assuming that the CRISPR arrays were not horizontally transferred (45). Interestingly, two additional actinobacterial MAGs (one *Rubrobacter* and one *Acidimicrobiia*) had acquired spacers that results in resistance against viruses detected from sites between 87 and 205 km away (Fig. 2), which indicates viral and/or host dispersal across the desert. In sum, spacer-to-protospacer-based identification of host-virus interactions revealed potentially widely dispersed viruses preying on *Actinobacteria*, particularly *Rubrobacter*.

Although the spacer-to-protospacer matches between CRISPR-containing MAGs and viral genomes provide high-confidence evidence of historical infections between a host population and viruses, many bacteria and archaea do not have CRISPR-Cas defense systems (46) or the respective CRISPR arrays do not get assembled or binned into MAGs. In the metagenomes studied, only 17% of the medium- to high-quality MAGs contained CRISPR arrays. To predict possible host-virus interactions for hosts that lack CRISPR systems, VirHostMatcher (44) identified 54 putative interactions between 23 MAGs (17 *Actinobacteria* and six *Chloroflexi*) and six viral genomes based on shared k-mer frequency patterns. Most linkages were established between 16 actinobacterial MAGs belonging to class *Acidimicrobiia* and three viruses belonging to two genus-level clusters. Interestingly, VirHostMatcher matched some viruses to hosts that are taxonomically distant, some differing at the order level and a few even in different phyla. For instance, three viruses were matched to both *Chloroflexi* and *Actinobacteria* (Fig. 4b), suggesting the possibility that these viruses have broad host ranges. No overlaps between interactions based on spacer-to-protospacer matches and VirHostMatcher host-virus linkage were identified.

Additionally, tRNA matching and a sequence homology search were conducted. We detected between 1 and 35 tRNAs across 20 viral population genomes. Only complete identity between viral and host tRNAs was used as an indication of potential host-virus interaction. We identified four interactions between four actinobacterial MAGs and one virus detected from the Y site. These interactions overlapped interactions inferred from oligonucleotide frequency similarity. Finally, nucleotide sequence homology was used to identify putative host-virus interactions between two lysogenic viruses and a firmicutal MAG. Figure 4b summarizes the putative host-virus linkages for each viral
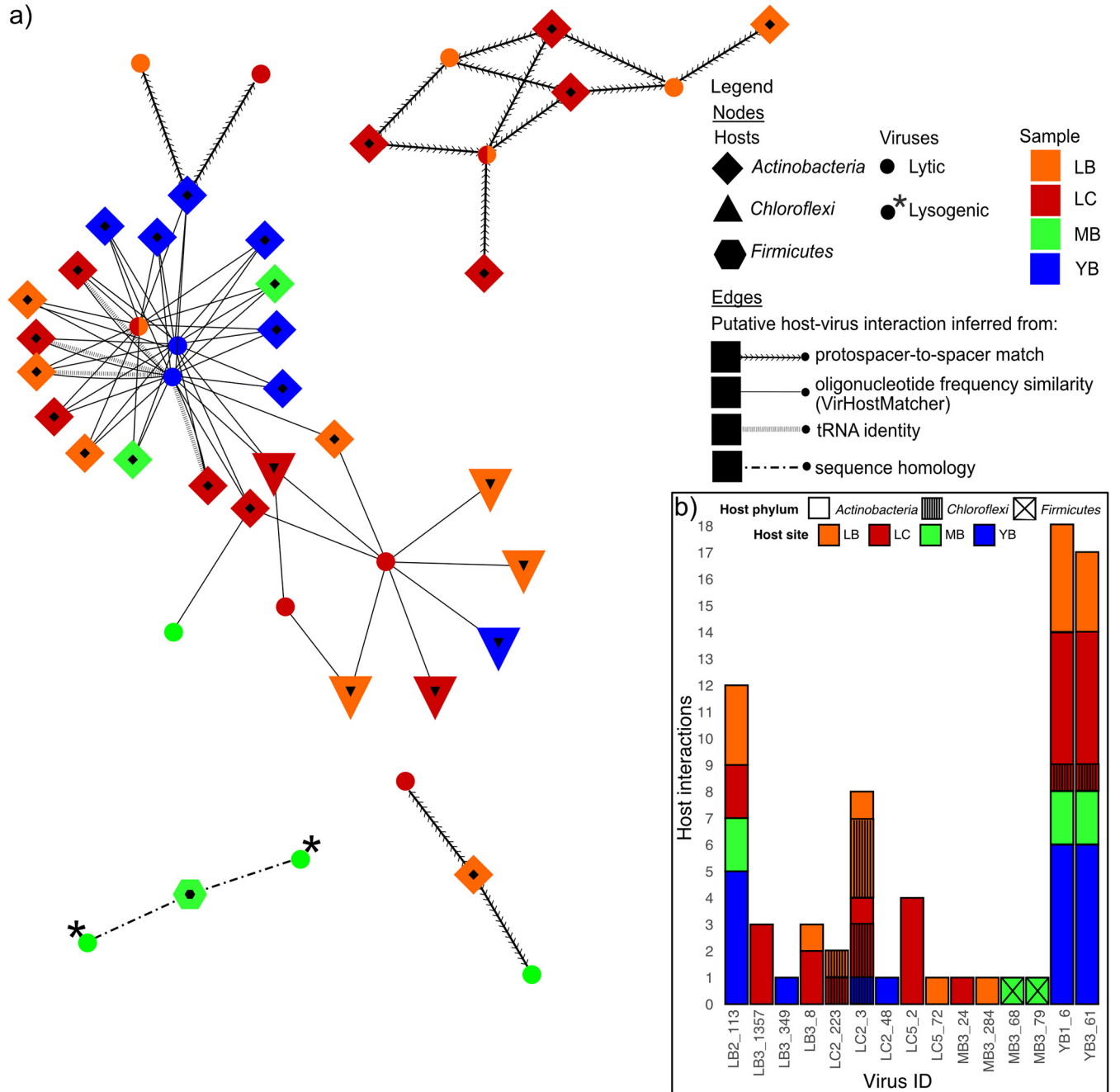
FIG 4 Host-virus interactions. (a) Putative host-virus interaction network displaying host taxonomy, host location, virus location (if a virus was detected in multiple samples, multiple colors are used), and virus lifestyle. Edges denote the method used to predict the host-virus interaction. Visualization was done using Cytoscape v3.8.0 (88). (b) Number of host interactions per virus, where the bar color denotes the site from which the matched MAGs were assembled and bar texture denotes host taxonomy.

genome, visualizing a high degree of variance in the number of identified interactions per virus, as well as the evidence of putative cross-site host-virus interactions and the potential for broad host ranges for some viruses.

Atacama viruses carry genes against environmental stress. Across 84 viral populations, we identified and annotated 4,288 proteins using DRAM-v (47). Thirty-nine percent of these proteins could be associated with the sequences in the queried database with high confidence, of which approximately half were "uncharacterized" and "hypothetical" proteins. We found 58 genes likely involved in extremotolerance and/or metabolism of microorganisms (for locus information, see Table S1d) across 30 viral populations. For
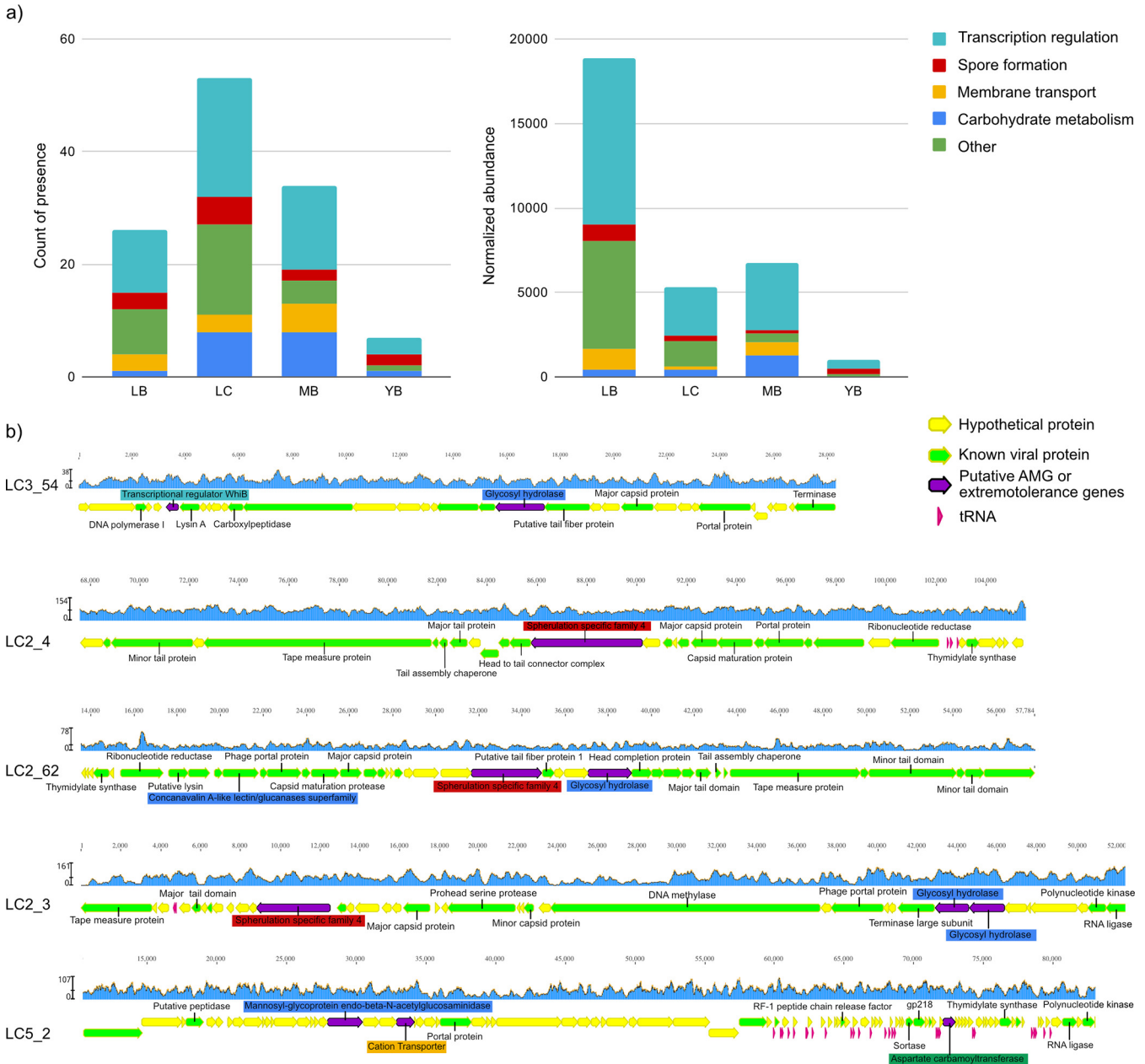
a)

b)

FIG 5 Extremotolerance and auxiliary metabolic genes. (a) Sums of counts of presence (left) and normalized abundance (right) of putative extremotolerance genes and AMGs across sampling sites. Abundances were calculated using the coverage of scaffolds where the gene was found and normalized across samples by sequencing depth. (b) Genomic neighborhoods of putative extremotolerance genes and AMGs. Genes of interest are colored purple, and labels are color coded by the functional categories according to the color scheme used in panel a. Identified viral genes are labeled and colored green, while gene homologs to a viral protein in the DRAM-v (47) database without a known function are colored green without a label. Genes of uncharacterized function and tRNAs are in yellow and pink, respectively. Coverages calculated based on mapped reads are shown in the blue graph above each genomic region visualized with Geneious v11.1.5.

instance, we found a sporulation protein (spherulation-specific family 4 [SSF] protein) in three viral genomes assembled from the LC2 sample. Some viruses also encoded membrane transport proteins for cation transporters and potassium channels, while others encoded transcriptional factors, including WhiB (48), which is also involved in bacterial sporulation initiation (49). Other putative AMGs included PhoD-like phosphatase, esterase, glucanases, glycosyl hydrolases, and endo-beta-*N*-acetylglucosaminidase. Interestingly, we also identified LuxR among the AMGs, the response regulator involved in quorum sensing of bacteria, which binds homoserine lactones and activates genes in the respective operon (50). Putative extremotolerance genes and AMGs were present across all sampling sites (Fig. 5a), with

the LB site exhibiting the highest abundance of viruses carrying these genes, while the LC site harboring the highest diversity of these genes. Figure 5b visualizes genomic regions of example scaffolds carrying genes of interest, displaying relatively uniform coverage across all gene loci and a close vicinity of viral hallmark genes, providing direct evidence that these genes are likely bona fide features of these viral genomes.

We further investigated the evolutionary relationship of viral genes relative to host homologs for two genes of interest: SSF genes and *whiB*. These two genes are generally known to be associated with bacterial spore formation and therefore could provide key extremotolerance for microbes in harsh desert environments. We compared the three genomic regions containing the SSF genes (Fig. S5a) and found that the genes upstream and downstream of the SSF genes in the viral scaffolds LC2_4 and LC2_62 showed synteny despite the lack of sequence homology between the SSF proteins themselves. On the other hand, SSF proteins found in LC2_62 and LC2_3 contained homologous regions, despite the lack of synteny and the disparity in the GC content throughout the genome. We examined the evolutionary history of these SSF proteins by placing them on a phylogenetic tree with similar bacterial proteins deposited in the NCBI nr database (Fig. S5b) and 10 SSF protein homologs found across nine MAGs from this environment. Interestingly, only actinobacterial MAGs binned from the L site metagenomes contained this gene and formed two distinct clusters. Compared to publicly available SSF proteins, the viral SSF proteins identified herein were not closely related to their homologs found in reconstructed MAGs. The closest homologs to the viral SSF proteins were found in genomes of *Actinobacteria* (e.g., *Streptomyces*), which contains many spore-forming bacteria. Notably, the viral genome of LC2_3 was predicted to have interactions with *Actinobacteria* and *Chloroflexi* in our samples (based on oligonucleotide frequency similarity) (Fig. 4), and two actinobacterial MAGs also encoded distantly related SSF proteins (predicted virus-MAG interactions are shown with red lines in Fig. S5b). Based on the evidence of unique SSF proteins found in three distinct viral genomes and the phylogenetic divergence from host proteins, the likelihood that they are randomly packaged host genes is low. Although the function of the viral SSF proteins requires experimental confirmation, we suspect that this relatively large gene (mean protein length = 1,265 amino acids) is beneficial for the viral populations, as it was identified in multiple viruses despite the presumed high cost of maintenance of AMGs in viral genomes (51, 52).

A WhiB-like transcription factor has previously been shown to control the spherulation septation in *Actinobacteria* (49). In addition, WhiB-like proteins previously identified in several *Mycobacterium* phages and *Streptomyces* phages (53) have been shown to regulate host cell wall component alteration in mycobacteria (54). Twelve WhiB-like transcription factor genes were identified across 11 viral populations as well as in 49 actinobacterial MAGs across all three sites. We phylogenetically placed the 12 viral WhiB-like proteins with bacterial WhiB proteins found in medium- to high-quality MAGs (Fig. S6a). Viral WhiB-like proteins were generally more related to each other than to the bacterial proteins. Additionally, WhiB-like proteins found in interacting virus-microbe pairs (Fig. S6a) were not closely related compared to other homologs. Notably, viral WhiB-like proteins from the same site tend to cluster together phylogenetically, with the exception of the LC samples. When comparing these proteins with related sequences in the NCBI nr database, we found four clusters with phage proteins, while the rest (eight proteins) clustered with homologs found in bacteria (Fig. S6b).

DISCUSSION

The hyperarid core of the Atacama Desert harbors an abundant and diverse soil virome that has been mostly understudied due to the scarcity of microbial biomass available. Recent improvements in soil DNA isolation methods and deeper sequencing of metagenomes (28, 29) not only allowed the discovery of microbes actively replicating *in situ* but also shed light upon the viral fraction of the hyperarid soil ecosystem that coexists with their microbial hosts. Our investigation of the Atacama viromes

reveals taxonomically diverse viruses and complex interactions between viruses and their hosts across the desert. Notably, the viruses contained key extremotolerance genes, and we propose a mutualistic model of host-virus interaction, where viruses seek protection in microbes as lysogens and pseudolysogens and, in return, aid host extremotolerance and survival.

Diversity of Atacama viruses contradicts the scarcity of microbial hosts. The Atacama Desert soil virome analyzed in this study consists of 84 viral populations belonging to at least 60 novel genera. The diversity of the viruses in the Atacama Desert soils is astonishing considering that the microorganisms that inhabit these soils are low in both biodiversity and biomass. Typically, groups of viruses infecting the same host often exchange genetic materials and form genotypic clusters (55). Therefore, in a low-diversity ecosystem, where many viruses infect the same hosts, one may expect a stronger genotypic clustering of viruses. Additionally, a large diversity of viral predators coupled with low diversity of prokaryotic prey seems to go against the competitive exclusion principle (56). High abundance and diversity of viruses in an environment with reduced encounters with viable microbial hosts suggest that some of the Atacama soil viruses may be dormant virions waiting for the appropriate host population to thrive, while others remain protected by residing in the host cells as lysogens (integrated into host genomes or plasmids) or pseudolysogens (as virus particles in the host cytoplasm) (24, 57). In particular, viruses have been shown to seek protection in their host cells, and this mode of viral survival has been observed in hot hyperarid desert soils, where lysogenic and pseudolysogenic [referred to together here as (pseudo)lysogenic] phages were found to be more prevalent (24–26). In our study, we predicted eight viral "populations" to be lysogenic based on their representative genomes being proviral. However, computational prediction tools tend to underestimate lysogenic viruses (35) and cannot distinguish pseudolysogens from lytic viruses. Therefore, isolation, cultivation, and visualization of the viruses and their hosts would shed light on the lifestyle of the viruses in the Atacama Desert hyperarid soils once sufficient biomass can be harvested from the ecosystem.

Viral dispersal is likely host mediated across the Atacama Desert. Our study demonstrated high co-correspondence between the viral and microbial communities, possibly due to the sample- and site-specific environmental stressors controlling both microbial and virus populations. Despite the high heterogeneity and specificity between samples, we identified four viral populations (two of which were lysogenic) detected in two or more sites (L, M, and Y). None of these genomes could be linked to a host, while 35 of 74 predicted host-virus interactions were between microbes and viruses detected at different sites. In particular, four host-virus interactions based on protospacer-to-spacer matches provide evidence for past dispersal events of either microbial and/or viral populations across distances up to 205 km. We hypothesize that one specific mechanism of dispersal in the Atacama Desert is the frequent sandstorms and powerful winds (58, 59) transporting infected microbes and virions in organic aerosols (60, 61). Notably, the most frequently detected viral population across samples (in 8 of 11 samples) was predicted to be lysogenic, supporting the scenario of viral dispersal through the transport of infected hosts. Additionally, we observed statistically significantly lower abundance of viral entities relative to microbes (Fig. 5S) in C samples, suggesting that viruses are vulnerable to irradiation in desert environments, perhaps more so than microbes. This result may also indicate that viral entities sheltered below boulders undergo lytic cycles and therefore exhibit higher abundance than microbes, while those beside boulders are primarily (pseudo)lysogenic. Further work is required to reveal the predominant lifestyle and viability of viruses in the Atacama hyperarid core.

A mutualistic model of host virus interactions in the Atacama Desert hyperarid core. A closer look at the genes carried by the Atacama viruses suggested an intriguing interaction between viruses and hosts, where a fine balance between viral predation and host extremotolerance sustains the continuum of the ecosystem. We posit that the viruses may serve as vectors delivering extremotolerance genes to their microbial hosts, increasing the chance of microbial survival under the harsh conditions of the
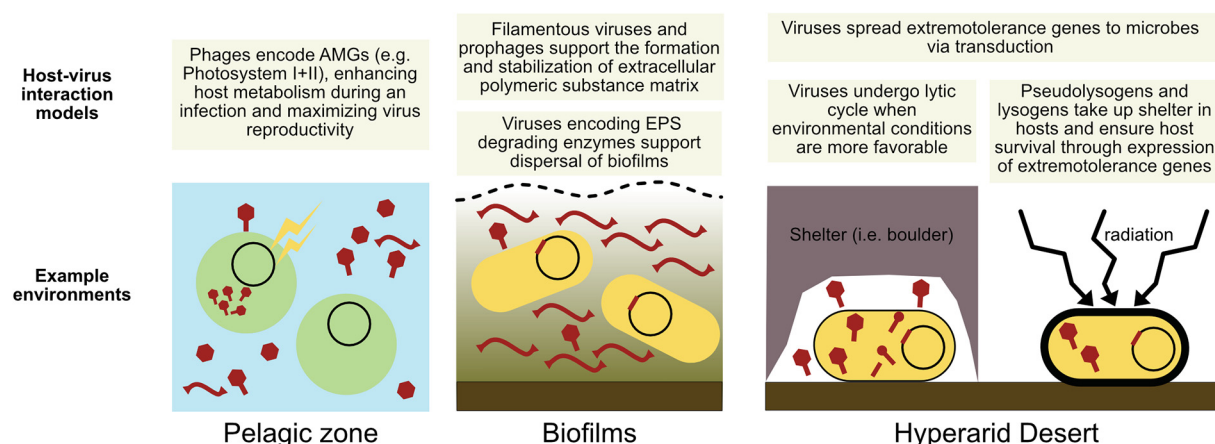
FIG 6 Visual schematic of the proposed host-virus interaction models in the hyperarid core of the Atacama Desert and comparison with models in other environments. Viruses (virions, pseudolysogens, and prophages) are shown in dark red, and microbial cells are in yellow and green.

Atacama Desert hyperarid core. In particular, we propose a model specific to extreme deserts, where (pseudo)lysogenic viruses encoding extremotolerance genes could support microbial survival in exchange for taking up shelter inside the microbial cytoplasm or genome. This mutualistic model closely parallels viral AMGs found in temperate environments (e.g., photosystem I and II genes in marine cyanophages [62–64] and CAZymes in mangrove soil viruses [65]), where AMGs are selected to maximize viral production by enhancing host metabolism during an infection. In the hyperarid core of the Atacama, viral extremotolerance genes likely increase the chance of viral production by aiding host survival, even if they result in temporary dormancy of the host through sporulation (in the case of SSF protein and WhiB).

A similar model of virus-host interactions (66) has been described in biofilms, where lysogenic viruses support formation, stabilization, and dispersal of biofilms, and biofilms in return provide protection for viruses against environmental stress (67–69). For instance, in hot desert soils, where microbes are known to form biofilms to protect themselves from desiccation, UV radiation, and low nutrient availability (70), Zablocki et al. (21) hypothesized a positive selection for temperate viruses in desert biofilms.

In even more extreme desert environments, such as the hyperarid core of the Atacama, where we did not detect any formation of biofilms, viruses may instead seek protection in the cytoplasm of the microbes as (pseudo)lysogens, as previously identified in other hot deserts (24–26). In this case, viral genes encoding extremotolerance may be selected for two reasons: (i) to ensure the survival of both the microbe and the (pseudo)lysogenic virus in the short term and (ii) to spread extremotolerance genes among microbes via transduction or lysogeny, resulting in the long-term increased fitness of the hosts against environmental stress. A visual schematic of the proposed host-virus interactions in the Atacama hyperarid core and a comparison with proposed host-virus interactions in other environments are found in Fig. 6. This mutualistic model does not exclude the antagonistic interactions between viruses and their hosts, as evidenced by the diverse innate and adaptive antiphage systems encoded in host MAGs (for an analysis of innate and adaptive immune systems, see Text S1).

We posit that viruses undergo lytic cycles when more favorable environmental conditions are met (i.e., rain events) or even in sheltered environments, resulting in the observed higher relative viral abundances under boulders compared to beside boulders. Phylogenetic analyses of the viral homologs to SSF proteins and the WhiB-like transcription factors suggest that these genes are indeed distinct from bacterial homologs, including those found in putative hosts. We also observed phylogenetic clustering of viral homologs to WhiB-like transcription factors by sampling site, suggesting the possibility of individual adaptation to the host communities and/or the environmental

conditions of each sampling site. Model host-virus systems will be required to confirm our proposed model (Fig. 6) and identify to what extent extremotolerance genes provide an increase in fitness for microbes and viruses in hyperarid environments.

Conclusion. The hyperarid core of the Atacama Desert is a much more biologically complex ecosystem than previously thought. We investigated hyperarid soil metagenomes to uncover a diverse virome interacting with a wide range of microbial hosts. Viruses in the Atacama Desert not only endure long periods of desiccation and extreme oxidative stress themselves but may also deliver extremotolerance genes to their hosts and aid their survival. This study expands the ecological significance of viruses in terrestrial systems, particularly in deserts. Life seems to persist even in the most hostile environments on Earth, and so do viruses. The Atacama Desert virome and its complex interplay with extremotolerant host populations highlight the role viruses play in microbial evolution and dynamics and illustrate a new dimension to host-virus interactions in extreme environments.

MATERIALS AND METHODS

Sampling location, procedure, and metagenomic library preparation. Briefly, sampling was conducted in March 2019. Three sampling sites, Yungay (Y), Maria Elena (M), and Lomas Bayas (L), were chosen from the hyperarid core of the Atacama Desert (a map of the sampling locations is shown in Fig. 1). Samples were collected from below boulders (B) and in the exposed surface soil (control [C]) beside boulders. Three B samples (LB2, LB3, and LB5) were collected at the Lomas Bayas boulder field, three B samples (MB1, MB3, and MB4) were from the Maria Elena boulder field, and two B samples (YB1 and YB3) were from the Yungay boulder field. Each B sample was collected from soil below one unique boulder, where the number in the sample name corresponds to the specific boulder. Three C samples (LC2, LC3, and LC5) were taken from the Lomas Bayas boulder field, from the exposed surface soil beside corresponding sampled boulders. Eleven metagenomic libraries of DNA extracted from eight B samples and three C samples were sequenced on Illumina HiSeq 2500 (Illumina, California, USA). Details of the sampling procedure, site coordinates, DNA extraction, and Illumina library preparation and sequencing can be found in our previous study (29).

Metagenomic analysis, host genome binning, and taxonomic classification. Descriptions of the assembly of metagenomic reads, contig binning, and bin analyses can be found in our previous study (29). Only medium- to high-quality bins (>75% completeness and <15% contamination calculated using CheckM v1.0.13 [71]) were considered host genomes. Host taxonomy was predicted using GTDB-Tk classify_wf (72). MAGs underwent gene prediction using Prodigal v2.6.3 (73) in meta mode and were annotated using Diamond v0.9.9 (74) against the UniRef100 database (75). The annotations were subsequently screened for identification of host innate defense marker genes identified by Bezuidt et al. (76).

Prediction and analysis of viral scaffolds. A schematic illustration of the analyses conducted can be found in Fig. S1. VirSorter v1 (35) with default settings and the –diamond flag, as well as VIBRANT v1.2.1 with default settings (36), was used for viral signal prediction across all assembled metagenomes and scaffolds length of ≥1,000 bp. VirSorter-predicted viral scaffolds in categories 1 and 2 were combined with VIBRANT-predicted viral scaffolds with qualities of medium, high, and complete. A viral contig was considered lysogenic if it was predicted as such by at least one of the following tools: VIBRANT, VirSorter, and CheckV v0.6.0 (37). Similarly, a virus was considered complete if at least one of the three previously mentioned tools predicted it to be circular or complete. Viral contigs were dereplicated using CD-HIT v4.6 (77) at 99% identity to identify viral populations, which were used for all subsequent analyses. CheckV v0.6.0 was used for completeness and quality estimation. Abundances of viral genomes were estimated using coverage calculated across all samples using a method described by Roux et al. (78). In short, reads were mapped using Bowtie2 (79) at ≥90% identity using the options –ignore-quals –mp = 1,1 –np = 1 –rdg = 0,1 –rfg = 0,1 –score-min = L,0,-0.1 as suggested by Nilsson et al. (80). Coverages were calculated only for scaffolds with mapped reads across ≥75% of the scaffold length with ≥1× coverage, for which average per-base coverage was calculated. Abundances of MAGs were estimated using the coverage of the ribosomal protein S3 (rpS3)-containing scaffold as described in our previous study (29). Calculated coverages were then subsequently normalized across samples by the total number of reads per sequenced library, to control for the sequencing depth of each sample. Statistically significant differences between viral and microbial abundance per sample were determined using Welch's t test with a P value threshold of <0.05. Viral genome annotation was performed using DRAM-v (47) using the UniRef90 database (75) and all other default databases in DRAM without the KEGG annotations.

Statistical characterization of viromes. Following community statistical analyses based on the normalized coverages of viral populations across samples were conducted and visualized in R version 4.0.2 (81) using the vegan (82) and ggplot2 (83) packages, respectively: Shannon-Wiener indices, Bray-Curtis distance matrix (84) calculations, and subsequent PCoA, PERMANOVA (using the adonis package) (38), CoCA (using the cocorresp package [40]), and BioENV analysis (39). Previously reported (29) environmental and geochemical variables (Table S1b) were used as input for BioENV. CoCA was performed on microbial and viral relative abundance profiles using the cocorresp package in both symmetric and predictive modes. Significance and degree of covariance in symmetric mode were computed, and the ordination

biplots were visualized using a method described previously by Alric et al. (85). All permutation-based tests were conducted with 999 iterations. Microbial relative abundance profiles were calculated using read-mapping-based abundances of *rpS3* genes from our previous study (29).

Intergenomic distance clustering and phylogenetic analysis of putative viruses. Intergenomic distances of viruses were calculated to identify genus- and species-level clusters using VIRIDIC with default settings (41). Phylogenetic trees were constructed using nucleic acid sequence-based VICTOR (42). vConTACT2 v0.9.19 (43, 86) was used to cluster and classify selected viral scaffolds against the ProkaryoticViralRefseq v94 database (87); resulting clusters were subsequently visualized using Cytoscape v3.8.0 (88). Per-sample relative abundances of viral clusters were calculated by summing up the calculated coverages for each viral population genome in the cluster and then dividing the total by the total coverage of all viral population genomes in the respective sample.

CRISPR-Cas analysis and spacer extraction of medium- to high-quality host genomes. For each high-quality host genome, direct repeats and Cas genes associated with CRISPR systems were extracted by combining the tools PILER-CR v1.06 (89) in default settings and CRISPRCasFinder (90), with results filtered for evidence level 4 for the latter. Filtered direct repeats were subsequently used for spacer extraction using MetaCRAST (91) with the flags -d 3 -l 60 -c 0.9 -a 0.9 -r from the raw reads of the respective metagenome from which the MAG was binned. Spacers were dereplicated using CD-HIT v4.6 (77) at 100% identity to identify the number of unique spacers across all metagenomes and within each sample.

Host-virus matching. (i) Protospacer-to-spacer matching. Extracted CRISPR spacers from all metagenomes were subjected to BLAST searching (92) with the blastn –short algorithm against the predicted viral sequences across all metagenomes and filtered with an 80% similarity threshold (similarity = alignment length × identity/query length).

(ii) Oligonucleotide frequency-based matching. VirHostMatcher (44) was used to determine putative interactions between medium- to high-quality MAGs from all metagenomes and predicted viral genomes based on shared oligonucleotide frequency pattern ($k = 6$). A d_2^* dissimilarity threshold of <0.2 was used to filter all potential host-virus interactions based on the benchmarking performed by Ahlgren et al. (44), where the lowest dissimilarity score threshold of 0.2 yielded above 90% accuracy in host prediction at the class level and approximately 60% accuracy at the order level.

(iii) tRNA identity-based matching. tRNAs in viral genomes and MAGs were predicted using DRAM-v (47) and tRNAscan-SE (93), respectively. Viral tRNAs were subjected to BLAST searching (92) against microbial tRNAs, and only complete (100% identity) matches were considered to imply host-virus interactions.

(iv) Nucleotide sequence homology-based matching. Viral genomes were subjected to BLAST searching against medium- to high-quality MAGs with a cutoff of ≥75% coverage over the length of the viral contig, ≥70% minimum nucleotide identity, ≥50 -bit score, and E value of ≤0.001. Exact matches resulting from a viral genome being binned in a MAG were excluded as potential binning errors, except when the viral genome was identified by VirSorter to be a defined prophage region inside a longer scaffold. Identified interactions were combined, and the host-virus interaction network was visualized using Cytoscape v3.8.0 (88).

Genomic neighborhood visualization and phylogenetic analyses of viral genes. Genomic neighborhoods were compared and visualized using Geneious v11.1.5. software (https://www.geneious.com) and Easyfig v2.2.2 (94) with a BLASTn E-value threshold of 1E−5. Three viral spherulation-specific proteins and 12 viral WhiB-like protein amino acid sequences were queried against the NCBI nr database using BLASTP. Twenty highest-identity matches per spherulation-specific sequence and five highest-identity matches per WhiB-like sequence were selected for subsequent phylogenetic analyses. Spherulation-specific proteins were searched for in the medium- to high-quality MAGs using hmmsearch (HMMER v3.2; www.hmmer.org) with spherulin4.hmm with an E-value threshold of 1E−10. Bacterial WhiB-like protein sequences were selected from the medium- to high-quality MAGs based on their annotation against the UniRef100 database (75). Duplicate sequences were removed prior to alignment using MUSCLE v3.8.31 (95). Spherulation protein alignments were trimmed using BMGE v1.12 (96) due to the presence of larger gaps. Trees were constructed using iqtree v1.5.5 (97) with the flags -m MFP -alrt 1000 -bb 1000 and visualized using iToL (98). Branches with bootstrap (99) values at least 95 and SH-aLRT (100) test values of at least 80 were marked as strongly supported. Phage promoters were predicted using PromoterHunter from phiSITE (101). Rho-dependent and Rho-independent terminators were predicted using RhoTermPredict (102) and ARNold (103), respectively.

Data availability. MAGs and viral genomes used in the analyses have been deposited in NCBI under BioProject no. PRJNA665391. NCBI accession information for the viral genomes and MAGs are found in Table S1a and e, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
TEXT S1, DOCX file, 0.02 MB.
FIG S1, TIF file, 0.8 MB.
FIG S2, TIF file, 0.2 MB.
FIG S3, TIF file, 1.9 MB.
FIG S4, TIF file, 1.2 MB.
FIG S5, TIF file, 1.7 MB.
FIG S6, TIF file, 1.4 MB.
TABLE S1, XLSX file, 0.4 MB.

mSystems®

ACKNOWLEDGMENTS

REFERENCES

1. Suttle CA. 2005. Viruses in the sea. Nature 437:356–361. https://doi.org/10.1038/nature04160.
2. Weinbauer MG, Rassoulzadegan F. 2003. Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11. https://doi.org/10.1046/j.1462-2920.2003.00539.x.
3. Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548. https://doi.org/10.1038/21119.
4. Suttle CA. 1994. The significance of viruses to mortality in aquatic microbial communities. Microb Ecol 28:237–243. https://doi.org/10.1007/BF00166813.
5. Gobler CJ, Hutchins DA, Fisher NS, Cosper EM, Saňudo-Wilhelmy SA. 1997. Release and bioavailability of C, N, P Se, and Fe following viral lysis of a marine chrysophyte. Limnol Oceanogr 42:1492–1504. https://doi.org/10.4319/lo.1997.42.7.1492.
6. Middelboe M, Jørgensen NOG. 2006. Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. J Mar Biol Assoc 86:605–612. https://doi.org/10.1017/S0025315406013518.
7. Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732. https://doi.org/10.1038/nrmicro1235.
8. Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasić L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7:828–836. https://doi.org/10.1038/nrmicro2235.
9. Suttle CA. 2007. Marine viruses–major players in the global ecosystem. Nat Rev Microbiol 5:801–812. https://doi.org/10.1038/nrmicro1750.
10. Bettarel Y, Sime-Ngando T, Amblard C, Dolan J. 2004. Viral activity in two contrasting lake ecosystems. Appl Environ Microbiol 70:2941–2951. https://doi.org/10.1128/AEM.70.5.2941-2951.2004.
11. Pratama AA, van Elsas JD. 2018. The "neglected" soil virome—potential role and impact. Trends Microbiol 26:649–662. https://doi.org/10.1016/j.tim.2017.12.004.
12. Emerson JB. 2019. Soil viruses: a new hope. mSystems 4:e00120-19. https://doi.org/10.1128/mSystems.00120-19.
13. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. 2017. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. Annu Rev Virol 4:201–219. https://doi.org/10.1146/annurev-virology-101416-041639.
14. Trubl G, Hyman P, Roux S, Abedon ST. 2020. Coming-of-age characterization of soil viruses: a user's guide to virus isolation, detection within metagenomes, and viromics. Soil Syst 4:23. https://doi.org/10.3390/soilsystems4020023.
15. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. Nature 536:425–430. https://doi.org/10.1038/nature19094.
16. Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, Solden L, Ellenbogen J, Runyon AT, Bolduc B, Woodcroft BJ, Saleska SR, Tyson GW, Wrighton KC, Sullivan MB, Rich VI. 2018. Soil viruses are underexplored players in ecosystem carbon processing. mSystems 3:e00076-18. https://doi.org/10.1128/mSystems.00076-18.
17. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li C, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol 3:870–880. https://doi.org/10.1038/s41564-018-0190-y.
18. Roy K, Ghosh D, DeBruyn JM, Dasgupta T, Wommack KE, Liang X, Wagner RE, Radosevich M. 2020. Temporal dynamics of soil virus and bacterial populations in agricultural and early plant successional soils. Front Microbiol 11:1494. https://doi.org/10.3389/fmicb.2020.01494.
19. Kuzyakov Y, Mason-Jones K. 2018. Viruses in soil: nano-scale undead drivers of microbial life, biogeochemical turnover and ecosystem functions. Soil Biol Biochem 127:305–317. https://doi.org/10.1016/j.soilbio.2018.09.032.
20. Li Y, Sun H, Yang W, Chen G, Xu H. 2019. Dynamics of bacterial and viral communities in paddy soil with irrigation and urea application. Viruses 11:347. https://doi.org/10.3390/v11040347.
21. 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.
22. 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.
23. Dávila-Ramos S, Castelán-Sánchez HG, Martínez-Ávila L, Sánchez-Carbente MDR, Peralta R, Hernández-Mendoza A, Dobson ADW, Gonzalez RA, Pastor N, Batista-García RA. 2019. A review on viral metagenomics in extreme environments. Front Microbiol 10:2403. https://doi.org/10.3389/fmicb.2019.02403.
24. Prigent M, Leroy M, Confalonieri F, Dutertre M, DuBow MS. 2005. A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. Extremophiles 9:289–296. https://doi.org/10.1007/s00792-005-0444-5.
25. Prestel E, Salamitou S, DuBow MS. 2008. An examination of the bacteriophages and bacteria of the Namib desert. J Microbiol 46:364–372. https://doi.org/10.1007/s12275-008-0007-4.
26. Prestel E, Regeard C, Salamitou S, Neveu J, Dubow MS. 2013. The bacteria and bacteriophages from a Mesquite Flats site of the Death Valley desert. Antonie Van Leeuwenhoek 103:1329–1341. https://doi.org/10.1007/s10482-013-9914-4.
27. Clarke JDA. 2006. Antiquity of aridity in the Chilean Atacama Desert. Geomorphology 73:101–114. https://doi.org/10.1016/j.geomorph.2005.06.008.
28. Schulze-Makuch D, Wagner D, Kounaves SP, Mangelsdorf K, Devine KG, de Vera J-P, Schmitt-Kopplin P, Grossart H-P, Parro V, Kaupenjohann M, Galy A, Schneider B, Airo A, Frösler J, Davila AF, Arens FL, Cáceres L, Cornejo FS, Carrizo D, Dartnell L, DiRuggiero J, Flury M, Ganzert L, Gessner MO, Grathwohl P, Guan L, Heinz J, Hess M, Keppler F, Maus D, McKay CP, Meckenstock RU, Montgomery W, Oberlin EA, Probst AJ, Sáenz JS, Sattler T, Schirmack J, Sephton MA, Schloter M, Uhl J, Valenzuela B, Vestergaard G, Wörmer L, Zamorano P. 2018. Transitory microbial habitat in the hyperarid Atacama Desert. Proc Natl Acad Sci U S A 115:2670–2675. https://doi.org/10.1073/pnas.1714341115.
29. Hwang Y, Schulze-Makuch D, Arens FL, Saenz JS, Adam PS, Bornemann TLV, Airo A, Schloter M, Probst AJ. 2020. Leave no stone unturned: the hidden potential of carbon and nitrogen cycling by novel, highly adapted Thaumarchaeota in the Atacama Desert hyperarid core. bioRxiv https://doi.org/10.1101/2020.07.17.208546.
30. Azua-Bustos A, Caro-Lara L, Vicuña R. 2015. Discovery and microbial content of the driest site of the hyperarid Atacama Desert, Chile. Environ Microbiol Rep 7:388–394. https://doi.org/10.1111/1758-2229.12261.
31. Navarro-González R, Rainey FA, Molina P, Bagaley DR, Hollen BJ, de la Rosa J, Small AM, Quinn RC, Grunthaner FJ, Cáceres L, Gomez-Silva B, McKay CP. 2003. Mars-like soils in the Atacama Desert, Chile, and the dry

limit of microbial life. Science 302:1018–1021. https://doi.org/10.1126/science.1089143.

32. Crits-Christoph A, Gelsinger DR, Ma B, Wierzchos J, Ravel J, Davila A, Cristina Casero M, DiRuggiero J. 2016. Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. Environ Microbiol 18:2064–2077. https://doi.org/10.1111/1462-2920.13259.

33. Uritskiy G, Tisza MJ, Gelsinger DR, Munn A, Taylor J, DiRuggiero J. 2020. Cellular life from the three domains and viruses are transcriptionally active in a hypersaline desert community. Environ Microbiol https://doi.org/10.1111/1462-2920.15023. Epub ahead of print.

34. Tchakerian V, Pease P. 2015. The critical zone in desert environments. Dev Earth Surface Processes 19:449–472. https://doi.org/10.1016/B978-0-444-63369-9.00014-8.

35. 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.

36. Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8:90. https://doi.org/10.1186/s40168-020-00867-0.

37. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2020. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol https://doi.org/10.1038/s41587-020-00774-7. Epub ahead of print.

38. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance: non-parametric MANOVA for ecology. Austral Ecol 26:32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.

39. Clarke KR, Ainsworth M. 1993. A method of linking multivariate community structure to environmental variables. Mar Ecol Prog Ser 92:205–205. https://doi.org/10.3354/meps092205.

40. ter Braak CJF, Schaffers AP. 2004. Co-correspondence analysis: a new ordination method to relate two community compositions. Ecology 85:834–846. https://doi.org/10.1890/03-0021.

41. Moraru C, Varsani A, Kropinski AM. 2020. VIRIDIC—a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. Viruses 12:1268. https://doi.org/10.3390/v12111268.

42. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396–3404. https://doi.org/10.1093/bioinformatics/btx440.

43. Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Rodney Brister J, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol 37:632–639. https://doi.org/10.1038/s41587-019-0100-8.

44. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. 2017. Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res 45:39–53. https://doi.org/10.1093/nar/gkw1002.

45. Godde JS, Bickerton A. 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol 62:718–729. https://doi.org/10.1007/s00239-005-0223-z.

46. 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.

47. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48:8883–8900. https://doi.org/10.1093/nar/gkaa621.

48. Ghai R, Mehrshad M, Mizuno CM, Rodriguez-Valera F. 2017. Metagenomic recovery of phage genomes of uncultured freshwater actinobacteria. ISME J 11:304–308. https://doi.org/10.1038/ismej.2016.110.

49. Bush MJ. 2018. The actinobacterial WhiB-like (Wbl) family of transcription factors. Mol Microbiol 110:663–676. https://doi.org/10.1111/mmi.14117.

50. Li Z, Nair SK. 2012. Quorum sensing: how bacteria can coordinate activity and synchronize their response to external signals? Protein Sci 21:1403–1417. https://doi.org/10.1002/pro.2132.

51. Mara P, Vik D, Pachiadaki MG, Suter EA, Poulos B, Taylor GT, Sullivan MB, Edgcomb VP. 2020. Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. ISME J 14:3079–3092. https://doi.org/10.1038/s41396-020-00739-3.

52. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. Nat Microbiol 3:754–766. https://doi.org/10.1038/s41564-018-0166-y.

53. Van Dessel W, Van Mellaert L, Liesegang H, Raasch C, De Keersmaeker S, Geukens N, Lammertyn E, Streit W, Anné J. 2005. Complete genomic nucleotide sequence and analysis of the temperate bacteriophage VWB. Virology 331:325–337. https://doi.org/10.1016/j.virol.2004.10.028.

54. Rybniker J, Nowag A, van Gumpel E, Nissen N, Robinson N, Plum G, Hartmann P. 2010. Insights into the function of the WhiB-like protein of mycobacteriophage TM4—a transcriptional inhibitor of WhiB2. Mol Microbiol 77:642–657. https://doi.org/10.1111/j.1365-2958.2010.07235.x.

55. Szymczak P, Rau MH, Monteiro JM, Pinho MG, Filipe SR, Vogensen FK, Zeidan AA, Janzen T. 2019. A comparative genomics approach for identifying host-range determinants in Streptococcus thermophilus bacteriophages. Sci Rep 9:7991. https://doi.org/10.1038/s41598-019-44481-z.

56. Haerter JO, Mitarai N, Sneppen K. 2014. Phage and bacteria support mutual diversity in a narrowing staircase of coexistence. ISME J 8:2317–2326. https://doi.org/10.1038/ismej.2014.80.

57. Ripp S, Miller RV. 1997. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. Microbiology 143:2065–2070. https://doi.org/10.1099/00221287-143-6-2065.

58. Azua-Bustos A, González-Silva C, Fernández-Martínez MÁ, Arenas-Fajardo C, Fonseca R, Martín-Torres FJ, Fernández-Sampedro M, Fairén AG, Zorzano M-P. 2019. Aeolian transport of viable microbial life across the Atacama Desert, Chile: implications for Mars Sci Rep 9:11024. https://doi.org/10.1038/s41598-020-57444-6.

59. Maki T, Lee KC, Kawai K, Onishi K, Hong CS, Kurosaki Y, Shinoda M, Kai K, Iwasaka Y, Archer SDJ, Lacap-Bugler DC, Hasegawa H, Pointing SB. 2019. Aeolian dispersal of bacteria associated with desert dust and anthropogenic particles over continental and oceanic surfaces. J Geophys Res Atmos 124:5579–5588. https://doi.org/10.1029/2018JD029597.

60. Reche I, D'Orta G, Mladenov N, Winget DM, Suttle CA. 2018. Deposition rates of viruses and bacteria above the atmospheric boundary layer. ISME J 12:1154–1162. https://doi.org/10.1038/s41396-017-0042-4.

61. Prospero JM, Blades E, Mathison G, Naidu R. 2005. Interhemispheric transport of viable fungi and bacteria from Africa to the Caribbean with soil dust. Aerobiologia 21:1–19. https://doi.org/10.1007/s10453-004-5872-7.

62. Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, Pinter RY, Partensky F, Koonin EV, Wolf YI, Nelson N, Béjà O. 2009. Photosystem I gene cassettes are present in marine virus genomes. Nature 461:258–262. https://doi.org/10.1038/nature08284.

63. Fridman S, Flores-Uribe J, Larom S, Alalouf O, Liran O, Yacoby I, Salama F, Bailleul B, Rappaport F, Ziv T, Sharon I, Cornejo-Castillo FM, Philosof A, Dupont CL, Sánchez P, Acinas SG, Rohwer FL, Lindell D, Béjà O. 2017. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected Prochlorococcus cells. Nat Microbiol 2:1350–1357. https://doi.org/10.1038/s41564-017-0002-9.

64. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol 4:e234. https://doi.org/10.1371/journal.pbio.0040234.

65. Jin M, Guo X, Zhang R, Qu W, Gao B, Zeng R. 2019. Diversities and potential biogeochemical impacts of mangrove soil viruses. Microbiome 7:58. https://doi.org/10.1186/s40168-019-0675-9.

66. Obeng N, Pratama AA, van Elsas JD. 2016. The significance of mutualistic phages for bacterial ecology and evolution. Trends Microbiol 24:440–449. https://doi.org/10.1016/j.tim.2015.12.009.

67. Secor PR, Sweere JM, Michaels LA, Malkovskiy AV, Lazzareschi D, Katznelson E, Rajadas J, Birnbaum ME, Arrigoni A, Braun KR, Evanko SP, Stevens DA, Kaminsky W, Singh PK, Parks WC, Bollyky PL. 2015. Filamentous bacteriophage promote biofilm assembly and function. Cell Host Microbe 18:549–559. https://doi.org/10.1016/j.chom.2015.10.013.

68. Rice SA, Tan CH, Mikkelsen PJ, Kung V, Woo J, Tay M, Hauser A, McDougald D, Webb JS, Kjelleberg S. 2009. The biofilm life cycle and virulence of Pseudomonas aeruginosa are dependent on a filamentous prophage. ISME J 3:271–282. https://doi.org/10.1038/ismej.2008.109.

69. McDougald D, Rice SA, Barraud N, Steinberg PD, Kjelleberg S. 2012. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. Nat Rev Microbiol 10:39–50. https://doi.org/10.1038/nrmicro2695.

70. Lebre PH, De Maayer P, Cowan DA. 2017. Xerotolerant bacteria: surviving through a dry spell. Nat Rev Microbiol 15:285–296. https://doi.org/10.1038/nrmicro2017.16.

71. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

mSystems®

72. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

73. 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.

74. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

75. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23:1282–1288. https://doi.org/10.1093/bioinformatics/btm098.

76. Bezuidt OKI, Lebre PH, Pierneef R, León-Sobrino C, Adriaenssens EM, Cowan DA, Van de Peer Y, Makhalanyane TP. 2020. Phages actively challenge niche communities in Antarctic soils. mSystems 5:e00234-20. https://doi.org/10.1128/mSystems.00234-20.

77. Li W, Jaroszewski L, Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics 17:282–283. https://doi.org/10.1093/bioinformatics/17.3.282.

78. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ 5:e3817. https://doi.org/10.7717/peerj.3817.

79. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

80. Nilsson E, Bayfield OW, Lundin D, Antson AA, Holmfeldt K. 2020. Diversity and host interactions among virulent and temperate Baltic Sea Flavobacterium phages. Viruses 12:158. https://doi.org/10.3390/v12020158.

81. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

82. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P. 2011. The vegan package: community ecology package. R package version 2.0–2.

83. Wickham H, Chang W. 2012. ggplot2: an implementation of the grammar of graphics (0.9.3 ed). https://ggplot2.tidyverse.org.

84. Bray JR, Curtis JT. 1957. An ordination of upland forest communities of southern Wisconsin. In Ecological Monographs, vol 27, p 325–349. https://doi.org/10.2307/1942268

85. Alric B, Braak CJF, Desdevises Y, Lebredonchel H, Dray S. 2020. Investigating microbial associations from sequencing survey data with co-correspondence analysis. Mol Ecol Resour 20:468–480. https://doi.org/10.1111/1755-0998.13126.

86. Bolduc B, Jang HB, Doulcier G, You ZQ, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ 5:e3243. https://doi.org/10.7717/peerj.3243.

87. 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.

88. 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.

89. Edgar RC. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics 8:18. https://doi.org/10.1186/1471-2105-8-18.

90. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246–W251. https://doi.org/10.1093/nar/gky425.

91. Moller AG, Liang C. 2017. MetaCRAST: reference-guided extraction of CRISPR spacers from unassembled metagenomes. PeerJ 5:e3788. https://doi.org/10.7717/peerj.3788.

92. 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.

93. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.

94. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010. https://doi.org/10.1093/bioinformatics/btr039.

95. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.

96. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. https://doi.org/10.1186/1471-2148-10-210.

97. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.

98. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. https://doi.org/10.1093/nar/gkw290.

99. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.

100. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.

101. Klucar L, Stano M, Hajduk M. 2010. phiSITE: database of gene regulation in bacteriophages. Nucleic Acids Res 38:D366–D370. https://doi.org/10.1093/nar/gkp911.

102. Di Salvo M, Puccio S, Peano C, Lacour S, Alifano P. 2019. RhoTermPredict: an algorithm for predicting Rho-dependent transcription terminators based on Escherichia coli, Bacillus subtilis and Salmonella enterica databases. BMC Bioinformatics 20:117. https://doi.org/10.1186/s12859-019-2704-x.

103. Lesnik EA, Sampath R, Levene HB, Henderson TJ, McNeil JA, Ecker DJ. 2001. Prediction of rho-independent transcriptional terminators in Escherichia coli. Nucleic Acids Res 29:3583–3594. https://doi.org/10.1093/nar/29.17.3583.

104. NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. 2019. Global Digital Elevation Model V003 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/ASTER/ASTGTM.003.