Viruses carrying genes for microbial extremotolerance are abundant and widely dispersed in the Atacama Desert hyperarid soil

Yunha Hwang[1,2*], Janina Rahlff[2], Dirk Schulze-Makuch[1,3,4], Michael Schloter[5], Alexander J. Probst[2*]

Running title: Atacama viruses with genes for microbial resilience

Affiliations:

[1] Center for Astronomy and Astrophysics, Technische Universität Berlin, 10623, Berlin, Germany

[2] Group for Aquatic Microbial Ecology, Environmental Microbiology and Biotechnology, Department of Chemistry, University of Duisburg-Essen, 45141, Essen, Germany

[3] German Research Centre for Geosciences (GFZ), Section Geomicrobiology, 14473, Potsdam, Germany

[4] Department of Experimental Limnology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), 12587, Stechlin, Germany

[5] Research Unit for Comparative Microbiome Analysis, Helmholtz Zentrum München, 85758, Oberschleißheim, Germany.

*corresponding authors

To whom the correspondence should be addressed:

Yunha Hwang (yhwang@g.harvard.edu)

Alexander J. Probst (alexander.probst@uni-due.de)

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 dispersal of diverse and abundant viruses that infect a wide range of hosts over 205 km across the desert. Host genomes encoded both adaptive and innate immune systems, providing evidence of viral predation being a key selective pressure along with abiotic stresses. Viral genomes carried extremotolerance features (i.e. DNA repair proteins, enzymes against oxidative damage) and other auxiliary metabolic genes, indicating that viruses could mediate the spread of microbial resilience against environmental stress across the desert. 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 indicate that viruses are abundant, diverse and 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 evolutionary dynamics that shape the unique and poorly understood host-virus relationships in desert soils.

Introduction

Viruses are considered to be the most abundant entities on Earth (1), with high genomic diversity (2) and an expanding ecological and biogeochemical importance. Viruses, particularly bacteriophages (thereafter: phages), have been shown to shape microbial community turnover and composition (3, 4), nutrient cycling (5, 6) as well as microbial evolution (7, 8) in marine (9) and freshwater (10) environments. For soil viruses, slower research progress has been made in the past, mainly due to difficulties in isolating viruses from heterogeneous and complex soil environments (11). However, recent metagenomic and metatranscriptomic approaches revealed diverse viruses in high abundance (12), which play significant roles in carbon processing (13–16) and other nutrient turnover (17, 18). Even less explored are viruses in extreme soil environments, where life itself becomes scarce in biomass and low in biodiversity (19). Understanding the abundance and diversity of viruses as well as their interactions with extremotolerant microbes in respective environments can highlight unique roles viruses may play in driving the adaptation of their hosts, and reveal 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. Biota that permanently inhabits these environments are often limited to a few bacterial and archaeal phyla. Recent studies of hot (i.e. Namib Desert, Sahara Desert) and cold (i.e. Antarctic soil) hyperarid desert viromes have revealed abundant viruses of diverse lineages and sizes, with temperate viruses being more prevalent than lytic ones in hot deserts (19–22). 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 aquatic environments, in which both viruses and hosts freely diffuse (11). Added to the viral selection pressure are the extremely slow or halted turnover rates of host populations and highly localized microbial communities (i.e. biofilms) with difficult-to-penetrate matrices consisting of extracellular polymeric substance (EPS) (23).

The Atacama Desert is one of the harshest environments on Earth with its hyperarid core experiencing extreme desiccation with mean annual precipitation < 2 mm (24). The surface soil of

the Atacama hyperarid core generally contains less than 1 wt% of water, and experiences high daily ultraviolet (UV) radiation (30 $J \cdot m^{-2}$), extreme diurnal temperature fluctuation (~ 60°C) (25) and additional osmotic pressure from the accumulation of salts (26, 27). Scarce populations of highly adapted microbial communities consisting of *Actinobacteria*, *Firmicutes*, *Chloroflexi* (26–28) and more recently, *Thaumarchaeota* (25) were found to inhabit soils of the Atacama hyperarid core. However, very little is known about viruses from these desert soil microbiomes. Schulze-Makuch *et al*. (2018) detected viruses in the Atacama Desert soils and showed a positive correlation of sequence abundances of viruses with their potential hosts after a rain event. Viruses have also been studied in hypersaline Atacama environments; Crits-Christoph *et al*. (29) identified viral sequences and their potential hosts in halite endoliths. Additionally, Uritskiy *et al.* (30) 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 (~35% of the Earth's land surface (31)).

To understand the diversity and ecological impact of viruses inhabiting the 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 stress resilience genes and auxiliary metabolic genes (AMGs) found in the predicted viral sequences, providing evidence that widely dispersed viruses in the Atacama Desert may play a key role in spreading microbial tolerance to extreme environmental conditions.

Results

Soil metagenomes of the Atacama hyperarid core feature abundant viral signatures. We predicted viral genomes in eleven assembled metagenomes (4.09 Gbp in total) from three different boulder fields (Lomas Bayas, L; Maria Elena, M; Yungay; Y) and two different soil compartments (Below boulder, B; Control - exposed soil adjacent to Boulder, C) (for analysis workflow see **Figure 1**, for the map of sampling locations see **Figure 2**). The metagenomes 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 (25). In total, 6809 out of 707,509 examined scaffolds were predicted to be viral. In detail, Virsorter (32) predicted 735 viral scaffolds, while VIBRANT (33) predicted 6437, including 363 overlapping scaffolds between both tools. The average length of the predicted viral scaffolds (hereafter referred to as "viral genomes") was 3857 bp (± 5275 bp), with the longest being 177 kbp and smallest being 1 kbp (only scaffolds with length >= 1 kbp were considered for viral prediction). The average G+C content of viral genomes was 63.9 % (± 6.3 %) and the average coding density was 86.6 % (± 10.7 %). The viral genomes were of varying quality: 13 "high quality", 16 "medium quality", 3163 "low quality" and 3617 "not-determined" according to CheckV (34). Virsorter and VIBRANT predicted 16 of these viral genomes to be circular and 424 (6.2% of all predicted viral scaffolds) to be lysogenic. Approximately twice as many unique viral scaffolds (dereplicated at 99% identity) per sample were predicted from exposed control soil (C) metagenomes in comparison to below boulder (B) metagenomes (Welch's t-test: $p = 0.0317$, $n = 11$; **Figure S3, Table S3**). By comparing the distributions of viral and microbial genome abundances (based on rpS3-carrying scaffold coverage) in each sample (**Figure S1**), we found the median of viral genome abundances to be approximately half of the median of microbial genome abundances in the same sample. However, the ranges of viral genome abundances were approximately two orders of magnitude greater than those of microbes in the respective samples. All samples featured highly abundant viruses, with abundances up to two orders of magnitude greater than those of microbes detected in the same sample (see **Figure S2** for an example of high coverage mapping of raw reads).

Spacer to protospacer matches reveal infection histories of bacteria and viruses across the Atacama hyperarid core. Previously, we recovered 73 high quality (>75% completeness; <15% contamination) MAGs across eleven metagenomes from three sampling locations (25). The MAGs were classified as 34 *Actinobacteria*, 30 *Chloroflexi*, eight *Thaumarchaeota* and one *Firmicutes*. Searches for CRISPR arrays revealed that nine actinobacterial and two Chloroflexi-derived MAGs carried repetitive elements 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 S4** for detail). In total, 18 identified DRs were used to recover 3438 unique spacers directly from the reads in their respective metagenomes (for distribution across samples see **Table S3**). 706 unique spacer to protospacer matches were predicted, resulting in 268 unique interactions between ten MAGs (nine *Actinobacteria*, and one *Chloroflexi*) and 175 viral genomes (**Figure 3a**) assuming that the CRISPR arrays were not horizontally transferred (35). Six actinobacterial MAGs classified as *Rubrobacters* recovered from site L shared infection histories with a set of viruses. Interestingly, seven out of nine analyzed MAGs had acquired resistances against viruses recovered from sites between 87 and 205 km away (**Figure 2**) also indicating viral dispersal across the desert.

An actinobacterial MAG belonging to the *Acidimicrobiia* class recovered from below a boulder at site Y had acquired resistances against a unique set of diverse viruses detected across three different locations. Interestingly, 15 out of 66 matched viruses were predicted to be lysogenic by VirSorter (32) and/or VIBRANT (32, 33). The likelihood for lysogenic infections in this ecosystem appears to be specific to the host taxa. Another *Rubrobacter* MAG recovered below a boulder from site L, had acquired resistance against a unique set of 16 viruses from sites L and M. Considering that this MAG was found in the same site as other rubrobacterial MAGs, it is surprising that none of the spacer-protospacer matches overlap with viruses infecting other *Rubrobacteria*, despite the close phylogenetic relationships between them based on the recovered ribosomal protein S3 (rpS3) sequences (90.1 - 97.3% identity). This suggests a narrow host-range based on CRISPR-*Cas* infection histories for this bacterial genus. In summary, spacer to protospacer-based identification of host-virus interactions revealed potentially widely dispersed viruses preying on

Actinobacteria, particularly *Rubrobacters*. Evidence of both lytic and lysogenic viral interactions with host populations were identified, with one particular *Acidimicrobiia* host population being more susceptible to lysogenic infections.

Oligonucleotide frequency-based host-virus matches suggest a broad host range of the Atacama viruses. The spacer to protospacer matches between CRISPR containing MAGs and viral genomes are high confidence evidence of historical infections between a host population and viruses. However, many bacteria and archaea do not have CRISPR-*Cas* defense systems (36) or the respective CRISPR arrays do not get assembled or binned in metagenomics. In the studied metagenomes, only 17% of the high quality MAGs contained CRISPR arrays. In extreme environments such as the Atacama Desert, where only slow microbial replication is supported (25, 26), it is plausible that CRISPR-*Cas* may not yield fast enough immune responses (36) for it to be prevalent. To predict possible host-virus interactions for hosts that lack CRISPR systems, VirHostMatcher (37) identified 3897 putative interactions between 73 MAGs and 132 viral scaffolds using d2* threshold of 0.3 (**Figure 3b** shows highest confidence interactions with d2* threshold of 0.25 for visualisation, **Table S3** contains sample specific overview). Most putative interactions were established with actinobacterial MAGs, which were infected by a largely shared set of multiple viruses, while *Thaumarchaeota* and *Firmicutes* only interacted with three and four viruses specific to the hosts' taxonomic group respectively. The four viral scaffolds belonging to the virus putatively infecting *Thaumarchaeota* were between 10kb and 40kb in length. They belonged to an uncharacterized genus according to vConTACT2 (38) and only 24 out of the predicted 163 genes could be functionally predicted by VIBRANT using Hidden Markov Models (HMM) (**Table S5**), highlighting the lack of archaeal virus protein homolog entries in public databases (39). 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, 91 viruses matched to both *Chloroflexi* and *Actinobacteria* using d2* threshold of 0.3.

Atacama viruses are phylogenetically novel and diverse. We clustered 299 matched (spacer-protospacer match and/or VirHostMatcher; **Figure 1**) viral genomes using intergenomic similarities (40). 269 clusters were formed, with 249 singleton clades, and the largest clade consisting of seven viral genomes. Intergenomic similarities between non-singleton clusters depicted as a heatmap in **Figure S4a** and a phylogenomic tree of the clustered sequences (n = 50) are shown in **Figure S4b**. Interestingly, 15 clusters of viruses contained members that were recovered from sampling locations separated by distances between 87 and 205 km (**Figure S4**). vConTACT2 (38) was used to cluster all 299 viral scaffolds with 2616 known prokaryotic viruses. Many Atacama viral genomes were related to phages infecting *Gordonia*, *Mycobacterium*, and *Streptomyces* at a taxonomic level higher than the genus level. Only one viral contig was classified at the genus level and was affiliated to the *Nyceiraevirus* (a *Gordonia* phage) in the order *Caudovirales*, family *Siphoviridae*, a tailed dsDNA phage (highlighted with an arrow in **Figure 4**). vConTACT2 also predicted 55 genus-level clusters of size between two and seven Atacama viruses and 174 singleton clusters. In total, different tools estimated between 228 and 267 genera amongst the 299 host-matched viruses (**Table S6**). **Figure 4** illustrates the network of reference viral genomes and viral genomes in this study based on shared genes. Notably, the majority of the Atacama viruses that are related to reference *Streptomycesphages* and *Mycobacteriumphage* were recovered from the LC samples collected from the control soil adjacent to boulders.

Atacama viruses encode genes against environmental stress. We predicted 5,817 proteins in 299 host-matched viruses and only 36% of the proteins were annotated based on sequence similarity with a protein in the UniRef100 (41) database. Only 38% of the annotations were with known function and the rest were annotated as “hypothetical proteins” or “uncharacterized proteins”. Similarly, VIBRANT predicted the function of 19% of the proteins based on sequence homology using HMM search. We found a number of putative stress-resistance genes in host-matched viruses (for loci information see **Table S7**). For instance, we found numerous DNA repair proteins (*YkoV*, very short patch repair endonuclease, *RecA*, *RecB*, *RecE*, resolvase, *UvdE*) against radiation stress, enzymes against oxidative damage (NTP pyrophosphohydrolase,

Glyoxalase, Alpha/beta hydrolases, oxidoreductase, multi-copper oxidase, *MutT* family protein, Aspartyl/Asparaginyl beta-hydroxylase). Coat protein for spore protection, genes involved in biofilm formation (MSHA pili biosynthesis protein *MshQ* and glycosyltransferase) and Gas vesicle protein *GvpU* that could provide resistance against osmotic or temperature shock (42, 43). Additionally, VIBRANT predicted 41 auxiliary metabolic genes (AMGs). Notable AMGs identified using either sequence similarity and/or sequence homology include glycogen synthase (involved in glycogenesis), phosphoglycerate mutase (involved in glycolysis), cellulose 1,4-beta-cellobiosidase (involved in cellulose degradation), phosphoadenosine phosphosulfate reductase and methanethiol S-methyltransferase (both involved in sulfur metabolism), and Poly(3-hydroxyalkanoate) synthetase. Some viruses also encoded membrane transport proteins for Magnesium (*CorA*), Ribosyl nicotinamide (*PnuC*-like), Chloride (Ca-activated channel) and potassium (Voltage-gated channel). .

Interestingly, we also found a gene of the *HigA* family, coding for the antitoxin part of the Type II toxin-antitoxin Abi antiphage system. Phages have been shown to counter the host Abortive Infection (Abi) system by encoding a mimicked antitoxin gene (44) and this may be evidence of viruses adapting to host defense systems (for detection of Abi systems in host genomes, please see below). Additionally, many viruses encoded *RelG* toxin protein, a toxin component of the host Abi system; the evolutionary advantage of *RelG* toxin in viruses has not yet been studied. We also found a *RpfB* gene resuscitation-promoting factor in a viral contig, which may be transcribed and secreted to promote resuscitation of dormant microorganisms. Finally, many viruses harbored restriction endonucleases (REases) which was previously shown to aid host DNA degradation (45) and thereby add to the “fitness” of viruses.

Atacama Desert bacterial genomes feature diverse antiphage defense systems. Many of the potential hosts identified through VirHostMatcher (37) did not have CRISPR-Cas adaptive defense systems. To understand the non-CRISPR-Cas defense mechanisms these extremotolerant microbes use to protect themselves against diverse and abundant viruses, we examined the matched host MAGs for genes involved in innate antiphage defense systems. Of the nine (Restriction-Modification, DISARM, BREX, Bruantia, Abortive

Infection, Zorya, Septu, Gabija, Theoris) innate antiphage defense systems (46–50) surveyed we found Restriction-Modification (RM) and Abi to be the most widespread mechanism across the 73 MAGs (**Table S8**). We found 264 genes involved in the RM system across 53 MAGs belonging to *Actinobacteria*, *Chloroflexi*, *Firmicutes* and *Thaumarchaeota*. Additionally, we found 158 genes involved in the Abi system across 50 MAGs. In total 64 MAGs contained at least one gene associated with either RM or Abi system with 39 MAGs putatively possessing both systems. Although we identified 12 genes carrying a *pglZ*-domain (one of the genes in the BREX locus (48)) in 10 MAGs, no other key BREX or DISARM genes could be identified. This is in contrast to a recent cold analog study (46) that reported a much more diverse set of innate immune systems. However, *pglZ-domains* have been found to be enriched in genomic regions called “defense islands” (51) harboring phage resistance genes, many of which are yet to be characterized. Thus, these 10 MAGs with *pgIZ*-domain containing genes may possess currently unknown defense mechanisms. Interestingly, MAGs with CRISPR arrays contained more innate anti-phage genes compared to those without any binned CRISPRs (Welch's t-test; $p = 0.00946$, $n = 73$).

Discussion

The Atacama Desert hyperarid core harbors an abundant and diverse soil virome that was previously overlooked due to the scarcity of microbial biomass detected from these extreme environments. Recent improvements in soil DNA isolation methods and deeper sequencing of the metagenomes (25) 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 coexist with their microbial hosts. Our investigation of the Atacama virome reveals a broad dispersal of viruses across the desert and complex interactions between viruses and their hosts. Notably, the viruses contain key extremotolerance genes, and we propose a unique host-virus dynamics characterized by the trade-off between viral predation and viral delivery of extremotolerance genes.

Viral diversity and abundance. The Atacama Desert soil virome analyzed in this study consists of 299 host-matched viruses belonging to at least 228 novel genera, as well as additional 6510 viruses without an identified host. The majority of the matched hosts belonged to *Actinobacteria* and *Chloroflexi.* Notably, we also identified a putative thaumarchaeal virus matched to eight thaumarchaeal MAGs found across all three sampling sites. While recent studies have isolated a novel family of viruses infecting marine *Thaumarchaeota* (52), and a provirus in order *Caudovirales* in mesophilic soil *Nitrososphaera viennensis* (53, 54), very little is known about lytic viruses infecting soil *Thaumarchaeota*, despite their ubiquity and abundance in many terrestrial ecosystems (55, 56).

The abundance and diversity of the viruses in the Atacama Desert soils are surprising considering that the microorganisms that inhabit these soils are low in both biodiversity and biomass. Typically, groups of viruses infecting the same host exchange genetic materials, forming genotypic clusters. Therefore in a low diversity ecosystem, where many viruses infect the same hosts, we would expect stronger genotypic clustering of viruses. Additionally, large diversity of predators (viruses) in low diversity of prey (microorganisms) seems to go against the ‘competitive exclusion principle’ (57). 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 extracellular viruses (virions) waiting for the right host population to bloom, while others remain protected by residing in the host cells as proviruses (integrated into host genomes or plasmids) or pseudolysogens (as inactive virus particles in the host cytoplasm) (20, 58). Virions have been shown to endure harsh environments using a variety of mechanisms such as highly stable capsid structures (59), mineralization in silica (60) and formation of virion aggregates with lipid vesicles (61). Alternatively, viruses have been shown to seek protection in their host cells and this mode of viral survival have been observed in hot hyperarid desert soils, where lysogenic and pseudolysogenic bacteriophages were found to be more prevalent (20–22). We observed relatively low proportions of lysogenic viruses in contrast to previous studies, possibly due to underestimation by prediction tools (32). Therefore, isolation, cultivation and visualization approaches of the viruses and their hosts would shed light on the lifestyle of the viruses in the Atacama Desert hyperarid soils.

Viral dispersal by wind. The genetic similarity between viruses from distant locations along with evidence of past infections in hosts by viruses recovered from distant locations suggest that the Atacama Desert soils viruses and microbes experience dynamic and broad dispersal and mixing. Wide distributions of similar phage genomes were also observed in other environments (62–64). One specific mechanism of dispersal in the Atacama Desert may be frequent sandstorms and powerful winds (65, 66) transporting infected microbes and viral entities in organic aerosols (67, 68). This mode of dispersal is consistent with the higher counts of unique predicted viral genomes in the control soils (adjacent to boulders) compared to the soils below the boulders that remain stable for years except during occasional seismic activities (69). Additionally, nearly all viral genomes that clustered closely with the reference database viruses were recovered from the control soil samples. This indicates that the control soils may harbor aerially deposited non-indigenous viruses from less extreme and thereby better studied ecosystems, while viruses found below the boulders are probably specific and endemic to the previously uncharacterized Atacama hyperarid soil virome. Viral predation, particularly by new viruses introduced via winds, is an important addition to a wide range of selection pressures previously thought to be limited to harsh abiotic stressors. However,

future work is required to reveal viral dispersal patterns in exposed and sheltered sites in desert ecosystems as well as the actual origin and travel distances of these viruses. Furthemore, the viability of the transported viruses must also be verified, as metagenomic studies cannot distinguish degrading and inactivated viral DNA from those that retain infectivity.

Host-virus relationship in the Atacama Desert. Identified host-virus relationships reveal a variety of host-ranges of Atacama viruses; some spacer to protospacer matches indicated a narrow host-range at an intraspecies-level, while VirHostMatcher predicted potential interactions with host-ranges extended to the inter-phylum level. Our results suggest a high degree of flexibility in host-ranges that could be beneficial in environments with low possibilities for host encounters. Viruses infecting multiple phyla of bacteria have been observed and are hypothesized to be competitive in oligotrophic environments (70). Similarly, broad host-ranges of many detected viruses may be positively selected in extremely harsh environments like the Atacama Desert, where the microbial density and diversity are very low and many microbes remain vegetative until a rain event takes place (71). If these viruses are truly "omnivorous" and capable of infecting bacteria belonging to two separate phyla, they could mediate inter-phyla horizontal transfers of genes.

A closer look at the genes carried by the host-matched viruses revealed an intriguing interaction between viruses and hosts where a fine balance between viral predation and host extremotolerance sustains the continuum of the ecosystem. We conclude that the viruses may serve as the vessels delivering extremotolerance genes to their microbial hosts, increasing the chance of microbial survival in harsh conditions of the Atacama Desert hyperarid core. This mutualistic model of virus-host relationships (72) have also been described in biofilms where lysogenic phages support stabilization of biofilms (23) and biofilms in return provide protection for viruses against environmental stress. For instance, in hot desert soils, where microbes are known to form biofilms to protect themselves from desiccation, UV radiation, and poor nutrients (73), Zablocki *et al* (19) hypothesized a positive selection of temperate viruses in biofilms. In extreme environments, mutualistic host-virus relationships may be crucial for the survival of

both the microbiome and the virome. The soil samples considered in this study did not exhibit visibly recognizable biofilms (personal observation), and the majority of the detected viruses were predicted to be lytic. Here, we propose another mutualistic mechanism of host-virus interactions in desert soils, in which widely dispersed viruses spread extremotolerance genes via transduction or lysogeny, and increased fitness of the hosts ensure the reproduction of viruses. This mutualistic model does not exclude the prevalence of antagonistic interactions between viruses and their hosts. For instance, in marine lytic cyanophages have been shown to carry genes for photosystem I and II that augment host photosynthesis during an infection, which in turn maximizes the viral reproductivity (74–76). Other AMGs (77) previously hypothesized to enhance host metabolism include *dsr*, *sox* and *amo* genes (78, 79) in marine viruses and CAZYmes (80) in mangrove soil viruses. In temperate environments, AMGs are selected to maximize viral reproductivity by increasing host fitness during an infection. In extreme environments such as the Atacama, viruses encode extremotolerance genes that increase the chance of viral reproduction by aiding host survival. Our results also provide evidence of host-virus arms races, where the majority of host genomes contain multiple host immunity elements and viral genomes show potential signs of adaptation against the host immune responses. Additionally, we observed an uneven distribution of both innate and adaptive antiphage systems, with microbes with CRISPRs also carrying more innate antiphage genes than those without CRISPRs, suggesting that the viral predation stress varies for each host and elicits different immune responses. In summary, previously overlooked desert viromes may play a key role in driving the evolution and adaptation of extremotolerant microbes through a combination of mutualistic and antagonistic interactions specific to desert environments.

Conclusion. The hyperarid core of the Atacama Desert is a much more biologically complex ecosystem than previously thought. In this study, we investigated hyperarid soil metagenomes to uncover an abundant and diverse virome interacting with a wide range of microbial hosts. Viruses in the Atacama Desert are widely dispersed (likely through winds and sandstorms), endure long periods of desiccation and irradiation, and potentially deliver extremotolerance genes to their hosts. 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. Further investigations on viruses in extreme environments will provide key insights to viral ecology, physiology and evolution on Earth and may have significant astrobiological implications as we look for extraterrestrial lifeforms. The Atacama Desert virome and its complex interplay with the extremophilic host populations highlight the role viruses play in microbial evolution and dynamics, even in the most extreme ecosystems previously thought to be shaped solely by abiotic stressors.

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 (Coordinates and other sampling information available in **Table S1,** map of the sampling locations shown in **Figure 2**). Samples were collected from below boulders (B) and in the exposed surface soil (control; C) beside boulders. Three B samples (LB2, LB3, LB5) were collected at the Lomas Bayas boulder field, three B (MB1, MB3, MB4) samples from the Maria Elena boulder field and two B (YB1, YB3) samples 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 (LC2, LC3, LC5) samples were taken from the Lomas Bayas boulder field, from the exposed topsoil 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, CA, USA). Detailed sampling procedure, DNA extraction and Illumina library preparation and sequencing can be found in Hwang *et al. (25)*.

Metagenomic analysis, host genome binning and taxonomic classification. Assembly of metagenomic reads, contig binning, and bin analyses can be found in Hwang *et al. (25)*. Only high quality bins (>75% completeness and <15% contamination calculated using CheckM (81)) were considered as host genomes. Host taxonomy was estimated using GTDB-Tk classify_wf (82).

Prediction and analysis of viral scaffolds. A schematic illustration of analyses conducted can be found in **Figure 1**. VirSorter (32) with default settings and –diamond flag, as well as VIBRANT v3 in default settings (33) were used for viral signal prediction across all assembled metagenomes and scaffolds length >=1000 bp. Virsorter predicted viral scaffolds in categories 1, 2, 4 and 5 were combined with VIBRANT predicted viral scaffolds. VIBRANT and VirSorter predicted lysogeny and determined circularity of viral scaffolds. CheckV (34) was used for completeness and quality estimation. Abundances of viral genomes and

microbial hosts were estimated using coverage calculated by mapping raw reads with Bowtie2 in sensitive mode (83) to each viral scaffold or ribosomal protein S3 (rpS3) containing scaffold respectively. Calculated coverages were then subsequently normalized across samples by the total number of reads per sequenced library. Putative viral genes in the matched viral sequences were extracted using Prodigal (84) and functionally annotated by using DIAMOND (85) against UniRef100 (41). HMM-based protein annotations including putative AMGs were conducted using VIBRANT (33) in agreement with the concept of class 1 AMGs according to (86). Viral scaffolds were dereplicated using CD-HIT v 4.6 (87) at 99% identity to identify the number of unique sequences per sample.

CRISPR-Cas analysis and spacer extraction of high quality host genomes. For each high quality host genome, direct repeats and Cas genes associated with clustered-regularly interspaced short palindromic repeats (CRISPR) systems were extracted using PILER-CR v1.06 (88) in default settings (89), which were then verified using CRISPRCasFinder (89) and filtered for evidence level 4. Filtered direct repeats were subsequently used for spacer extraction using MetaCRAST (90) with flags -d 3 -l 60 -c 0.9 -a 0.9 -r from the raw reads of the respective metagenome the MAG was binned from. Spacers were dereplicated using CD-HIT v 4.6 (87) at 100% identity to identify the number of unique spacers across all metagenomes and within each sample.

Host-virus matching. Extracted CRISPR spacers from all metagenomes were BLAST-ed (91) with blastn --short algorithm against the predicted viral sequences across all metagenomes and filtered with an 80 % similarity threshold, (similarity=alignmentLength*Identity)/QueryLength), to determine host-virus interactions based on spacer-protospacer matches. The interaction network was visualized using Cytoscape version 3.8.0 (92). VirHostMatcher (93) was used to determine putative interactions between high quality host genomes from all metagenomes and predicted viral sequences with length >= 10 kb including all circular ones across all metagenomes based on shared oligonucleotide frequency pattern (k=6). A d_2* dissimilarity threshold of < 0.3 was used to filter all potential host-virus interactions based on the

benchmarking performed by Algren *et al.*, (2017) where the lowest dissimilarity score threshold of 0.3 yielded above 80 % accuracy in host prediction at class level and approximately 60% accuracy at class and order levels.

Integenomic distance clustering and phylogenetic analysis of putative viruses. 299 host-matched (via spacer-protospacer matching and/or VirHostMatcher) viral genomes were compiled for further clustering and phylogenetic analysis (see **Figure 1** for visual schematics). Intergenomic distance was calculated to identify genus and species level clusters using VIRIDIC with default settings (40). Viral scaffolds that were clustered with at least one other contig using VIRIDIC at 70 % intergenomic similarity (40) were selected for further phylogenetic analysis and tree construction using both nucleic acid and amino acid-based VICTOR (94). vConTACT2 v.0.9.19 (38, 95) was used to cluster and classify selected viral scaffolds against the ProkaryoticViralRefseq v94 database (96), resulting clusters were subsequently visualized using cytoscape (92)

Statistical tests. For detecting differences in the number of unique viral scaffolds between sample types (below boulder n=8, open n=3) and between the number of innate antiphage genes between MAGs with (n = 11) and without (n = 64) CRISPRs, unpaired, two-sided Welch's t-tests at the 95% significance level were conducted using R (97) version 4.0.2 (2020-06-22).

Acknowledgements

This work was funded by ERC Advanced Grant HOME (# 339231) to DSM. AJP was supported by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein- Westfalen ("Nachwuchsgruppe Dr. Alexander Probst"). We acknowledge support by the German Aerospace Center (DLR) under contract DISPERS (50WB1922).

Competing Interests

All authors declare that they have no competing interests.

Author contributions

YH, AJP and DSM conceived the project; YH conducted sampling; MS generated the raw sequence data; YH assembled, curated and analyzed sequence data with contribution from JR and AJP; AJP provided computational resources; YH wrote the manuscript with contribution from JR; all authors discussed and revised the manuscript.

References

1. Suttle CA. 2005. Viruses in the sea. Nature.

2. Weinbauer MG, Rassoulzadegan F. 2004. Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11.

3. Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.

4. Suttle CA. 1994. The significance of viruses to mortality in aquatic microbial communities. Microb Ecol 28:237–243.

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. Limnology and Oceanography.

6. Middelboe M, Jørgensen NOG. 2006. Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. Journal of the Marine Biological Association of the United Kingdom.

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.

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.

9. Suttle CA. 2007. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol 5:801–812.

10. Bettarel Y, Sime-Ngando T, Amblard C, Dolan J. 2004. Viral activity in two contrasting lake ecosystems. Appl Environ Microbiol 70:2941–2951.

11. 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 Systems 4:23.

12. Graham EB, Paez-Espino D, Brislawn C, Neches RY, Hofmockel KS, Wu R, Kyrpides NC, Jansson JK, McDermott JE. 2019. Untapped viral diversity in global soil metagenomes.

13. 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.

14. 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.

15. 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.

16. Starr EP, Nuccio EE, Pett-Ridge J, Banfield JF, Firestone MK. 2019. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil.

17. 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.

18. 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.

19. Zablocki O, Adriaenssens EM, Cowan D. 2016. Diversity and Ecology of Viruses in Hyperarid Desert Soils. Appl Environ Microbiol 82:770–777.

20. 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.

21. Prestel E, Salamitou S, DuBow MS. 2008. An examination of the bacteriophages and bacteria of the Namib desert. J Microbiol 46:364–372.

22. 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.

23. 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. Nature Reviews Microbiology.

24. Clarke JDA. 2006. Antiquity of aridity in the Chilean Atacama Desert. Geomorphology.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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. Environmental Microbiology.

30. 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.

31. Tchakerian V, Pease P. 2015. The Critical Zone in Desert Environments. Developments in Earth Surface Processes.

32. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.

33. 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.

34. Nayfach S, Camargo AP, Eloe-Fadrosh E, Roux S. 2020. CheckV: assessing the quality of metagenome-assembled viral genomes. BioRxiv.

35. 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.

36. 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.

37. Wang W, Ren J, Tang K, Dart E, Ignacio-Espinoza JC, Fuhrman JA, Braun J, Sun F, Ahlgren NA. 2020. A network-based integrated framework for predicting virus-prokaryote interactions. NAR Genom Bioinform 2:lqaa044.

38. 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. Nature Biotechnology.

39. Dellas N, Snyder JC, Bolduc B, Young MJ. 2014. Archaeal Viruses: Diversity, Replication, and Structure. Annu Rev Virol 1:399–426.

40. Moraru C, Varsani A, Kropinski AM. 2020. VIRIDIC – a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. BioRxiv.

41. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23:1282–1288.

42. van Keulen G, Hopwood DA, Dijkhuizen L, Sawers RG. 2005. Gas vesicles in actinomycetes: old buoys in novel habitats? Trends Microbiol 13:350–354.

43. Li Y, Pinto-Tomás AA, Rong X, Cheng K, Liu M, Huang Y. 2019. Population Genomics Insights into Adaptive Evolution and Ecological Differentiation in Streptomycetes. Appl Environ Microbiol 85.

44. Blower TR, Short FL, Fineran PC, Salmond GPC. 2012. Viral molecular mimicry circumvents abortive infection and suppresses bacterial suicide to make hosts permissive for replication. Bacteriophage 2:234–238.

45. Agarkova IV, Dunigan DD, Van Etten JL. 2006. Virion-associated restriction endonucleases of chloroviruses. J Virol 80:8114–8123.

46. 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.

47. 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 microbial pangenome. Science.

48. 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.

49. 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 anti-phage activities. Nat Microbiol 3:90–98.

50. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317–327.

51. Makarova KS, Wolf YI, Snir S, Koonin EV. 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J Bacteriol 193:6039–6056.

52. Kim J-G, Kim S-J, Cvirkaite-Krupovic V, Yu W-J, Gwak J-H, López-Pérez M, Rodriguez-Valera F, Krupovic M, Cho J-C, Rhee S-K. 2019. Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. Proc Natl Acad Sci U S A 116:15645–15650.

53. Krupovic M, Spang A, Gribaldo S, Forterre P, Schleper C. 2011. A thaumarchaeal provirus testifies for an ancient association of tailed viruses with archaea. Biochem Soc Trans 39:82–88.

54. Krupovic M, Makarova KS, Wolf YI, Medvedeva S, Prangishvili D, Forterre P, Koonin EV. 2019. Integrated mobile genetic elements in Thaumarchaeota. Environ Microbiol 21:2056–2078.

55. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. 2011. Examining the global distribution of dominant archaeal populations in soil. ISME J 5:908–917.

56. Ochsenreiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. 2003. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. Environ Microbiol 5:787–797.

57. Haerter JO, Mitarai N, Sneppen K. 2014. Phage and bacteria support mutual diversity in a narrowing staircase of coexistence. The ISME Journal.

58. Ripp S, Miller RV. 1997. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. Microbiology.

59. Stone NP, Demo G, Agnello E, Kelch BA. 2019. Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure. Nat Commun 10:4471.

60. Laidler JR, Shugart JA, Cady SL, Bahjat KS, Stedman KM. 2013. Reversible inactivation and desiccation tolerance of silicified viruses. J Virol 87:13927–13929.

61. Sanjuán R, Thoulouze M-I. 2019. Why viruses sometimes disperse in groups? Virus Evol 5:vez014.

62. Short CM, Suttle CA. 2005. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Appl Environ Microbiol 71:480–486.

63. Breitbart M, Miyake JH, Rohwer F. 2004. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol Lett 236:249–256.

64. Bellas CM, Schroeder DC, Edwards A, Barker G, Anesio AM. 2020. Flexible genes establish

widespread bacteriophage pan-genomes in cryoconite hole ecosystems. Nature Communications.

65. 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.

66. 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. Journal of Geophysical Research: Atmospheres.

67. 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.

68. 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.

69. Quade J, Reiners P, Placzek C, Matmon A, Pepper M, Ojha L, Murray K. 2012. Seismicity and the strange rubbing boulders of the Atacama Desert, northern Chile. Geology.

70. Malki K, Kula A, Bruder K, Sible E, Hatzopoulos T, Steidel S, Watkins SC, Putonti C. 2015. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. Virol J 12:164.

71. Crits-Christoph A, Robinson CK, Barnum T, Fricke WF, Davila AF, Jedynak B, McKay CP, Diruggiero J. 2013. Colonization patterns of soil microbial communities in the Atacama Desert. Microbiome 1:28.

72. Obeng N, Pratama AA, van Elsas JD. 2016. The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. Trends Microbiol 24:440–449.

73. Lebre PH, De Maayer P, Cowan DA. 2017. Xerotolerant bacteria: surviving through a dry spell. Nat Rev Microbiol 15:285–296.

74. 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.

75. 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. Nature Microbiology 2:1350–1357.

76. 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.

77. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. Virol J 16:15.

78. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537:689–693.

79. Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, Woyke T, Hallam SJ, Sullivan MB. 2014. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. Elife 3:e03125.

80. 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.

81. 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.

82. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics.

83. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

84. 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.

85. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nature Methods.

86. Hurwitz BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and taxonomic niche specialization in the “core” and “flexible” Pacific Ocean Virome. ISME J 9:472–484.

87. Li W, Jaroszewski L, Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics.

88. Edgar RC. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics 8:18.

89. 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.

90. Moller AG, Liang C. 2017. MetaCRAST: reference-guided extraction of CRISPR spacers from unassembled metagenomes. PeerJ 5:e3788.

91. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

92. 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.

93. 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.

94. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396–3404.

95. 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.

96. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. Nucleic Acids Res 43:D571–7.

97. Core Team R, Others. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available.

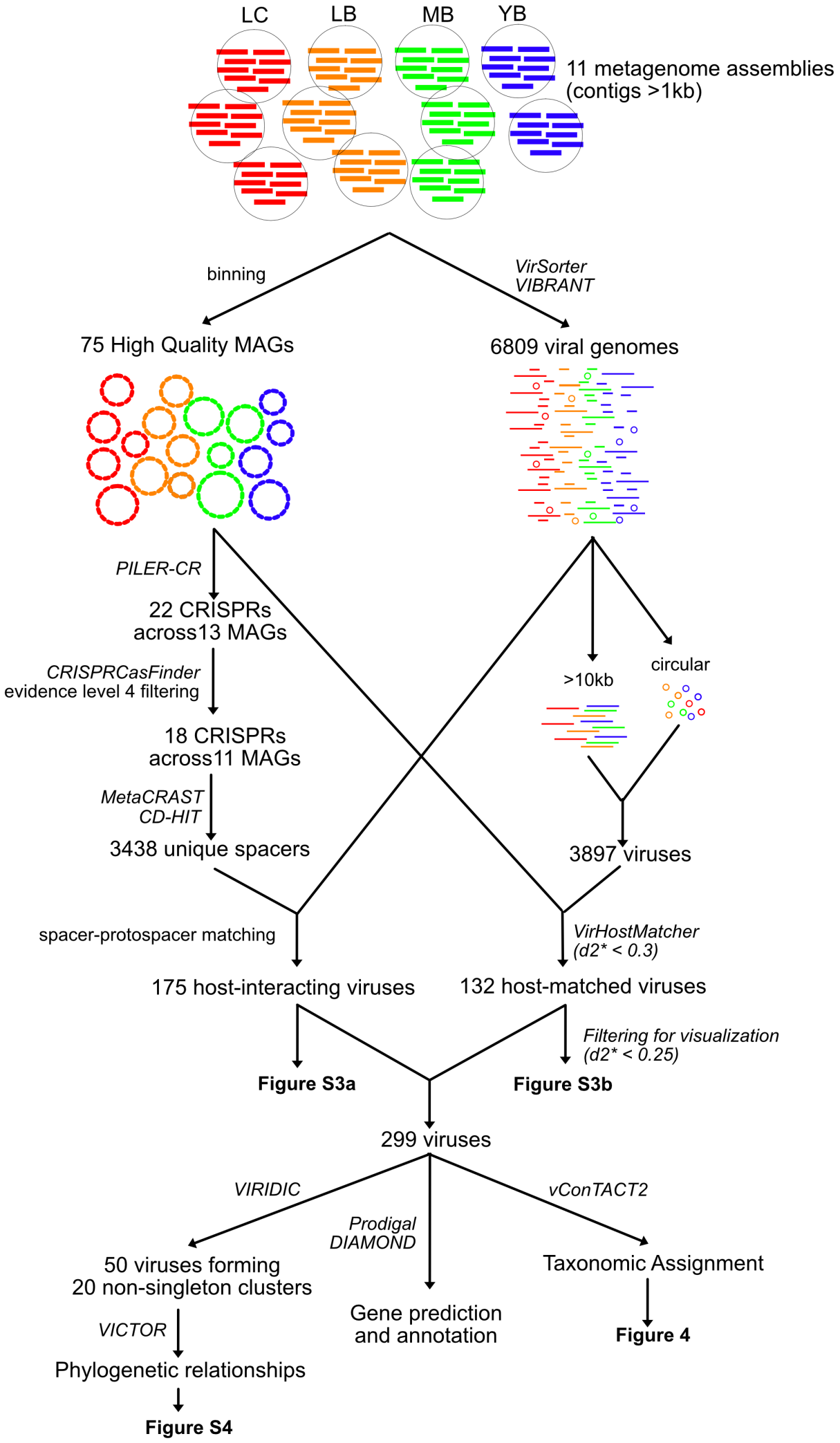
Figures legends

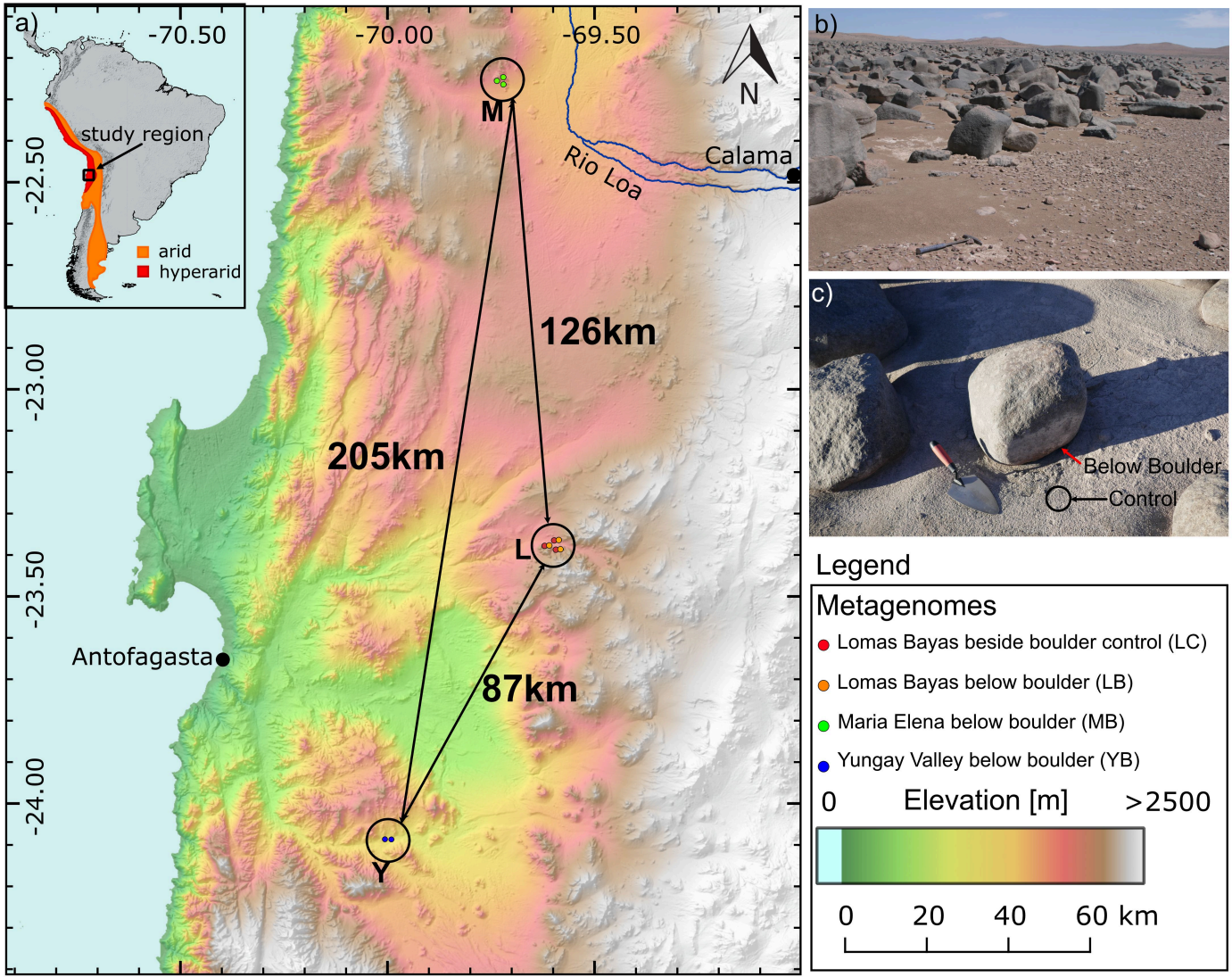
Figure 1. Flowchart outlining analyses carried out in this study. Software tools used are in italic, resulting figures are in bold. Graphics are for illustrative purposes and not quantitative.

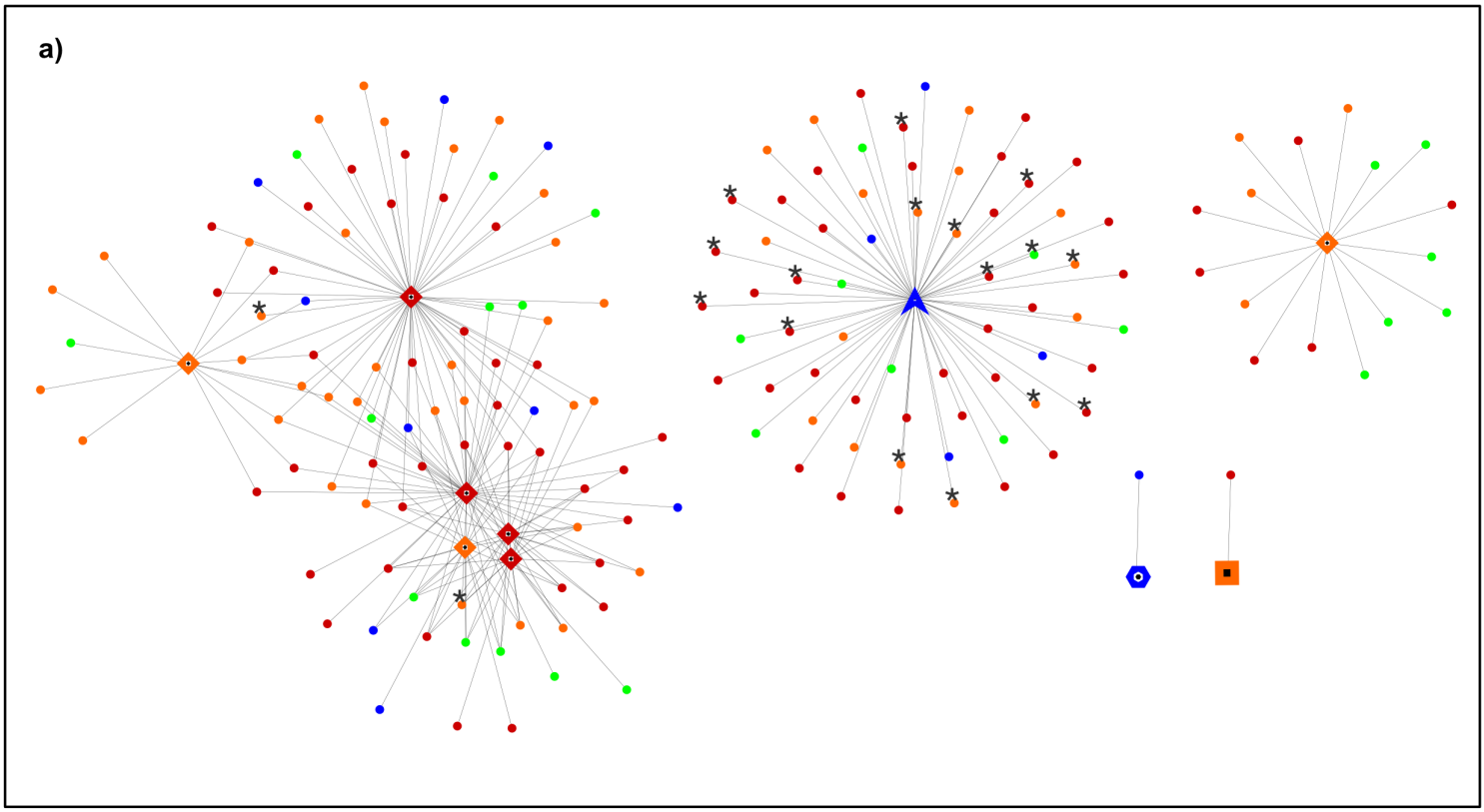
Figure 2. Sampling location information. a) Map of the sampling locations M: Maria Elena, L: Lomas Bayas, Y: Yungay Valley. Distances between sampling locations are shown. Approximate metagenome sample locations are shown in smaller circles. b) Yungay boulder field. c) A sampled boulder in Maria Elena boulder field. Red and black arrows distinguish sample types B and C respectively.

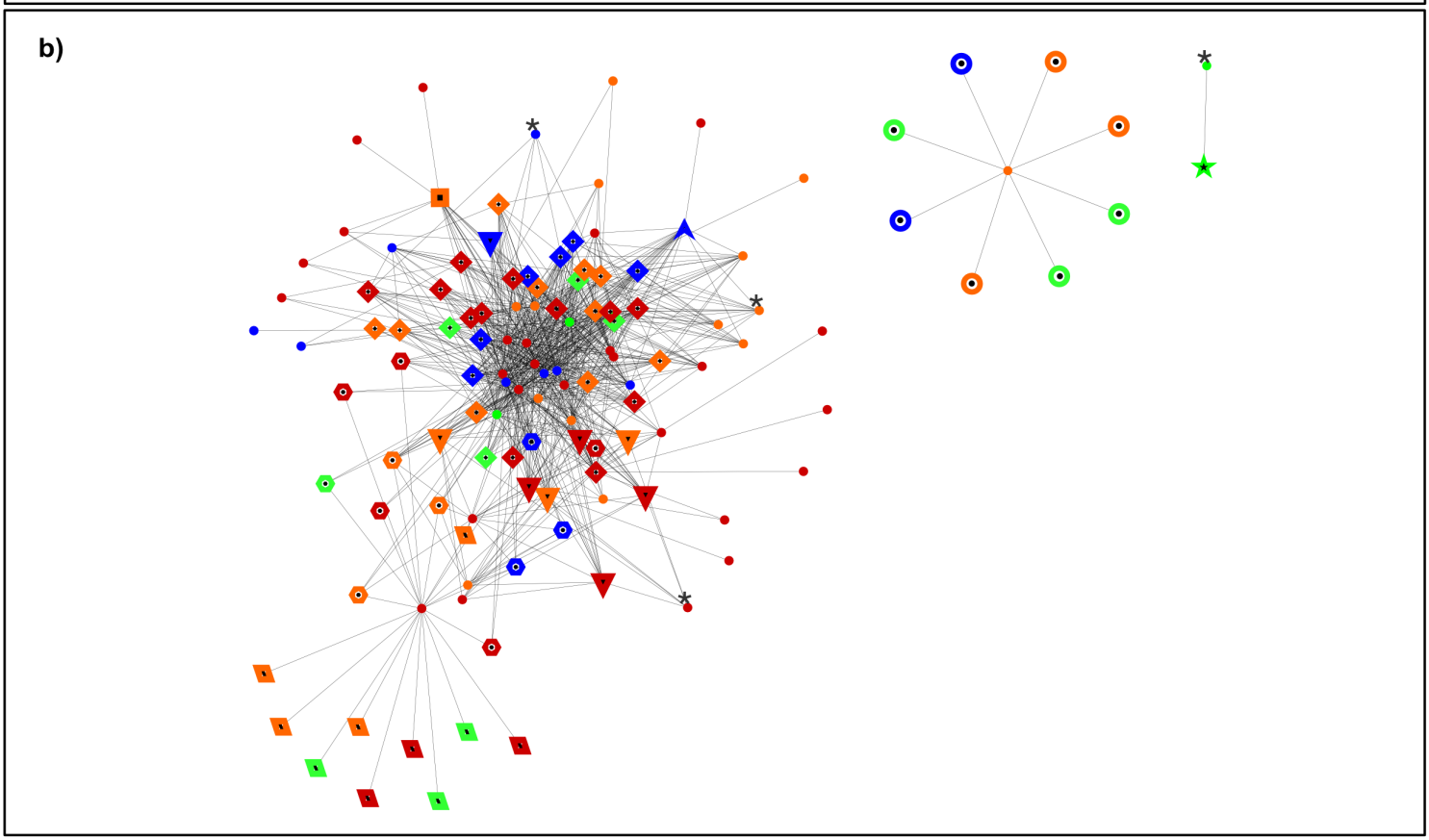
Figure 3. Host-virus interactions based on a) spacer-protospacer matches, and b) oligonucleotide frequencies (VirHostMatcher). Dissimilarity d2* threshold of < 0.25 was used for simplified visualization via Cytoscape (92). Both hosts and viruses are color coded based on the sampling site. Host node shape denotes taxonomy and lysogenic viruses are identified using asterisks.

Figure 4. Clustering of Atacama viral genomes with reference viral genomes based on shared genes. vConTACT2 output network with significance greater than 1. Arrow indicates the Atacama viral genome classified as a *Nyceiraevirus* (a *Gordonia* phage) in the order *Caudovirales*, family *Siphoviridae*, a tailed dsDNA phage. Visualization was done with Cytoscape (92). Viruses are color coded based on the sampling site they were recovered from.


LC
LB
MB
YB
11 metagenome assemblies
(contigs >1kb)
binning
VirSorter
VIBRANT
75 High Quality MAGs
6809 viral genomes
PILER-CR
22 CRISPRs
across13 MAGs
CRISPRCasFinder
evidence level 4 filtering
18 CRISPRs
across11 MAGs
MetaCRAST
CD-HIT
3438 unique spacers
>10kb
circular
3897 viruses
spacer-protospacer matching
VirHostMatcher
(d2* < 0.3)
175 host-interacting viruses
132 host-matched viruses
Filtering for visualization
(d2* < 0.25)
Figure S3a
Figure S3b
299 viruses
VIRIDIC
Prodigal
DIAMOND
vConTACT2
50 viruses forming
20 non-singleton clusters
Gene prediction
and annotation
Taxonomic Assignment
VICTOR
Figure 4
Phylogenetic relationships
Figure S4


a)
-70.50
-22.50
study region
arid
hyperarid
-70.00
-69.50
N
M
Calama
Rio Loa
126km
205km
-23.00
L
-23.50
Antofagasta
87km
-24.00
Y
b)
c)
Below Boulder
Control
Legend
Metagenomes
Lomas Bayas beside boulder control (LC)
Lomas Bayas below boulder (LB)
Maria Elena below boulder (MB)
Yungay Valley below boulder (YB)
0
Elevation [m]
>2500
0
20
40
60 km





Legend

Hosts

Actinobacteria

- *Rubrobacterales*
- *Acidimicrobiia*
- *Solirubrobacterales*

Chloroflexi

- *Chloroflexales*
- *Thermomicrobiales*
- *Thermobaculales*

Other bacteria

- *Firmicutes*

Archaea

- *Thaumarchaea*

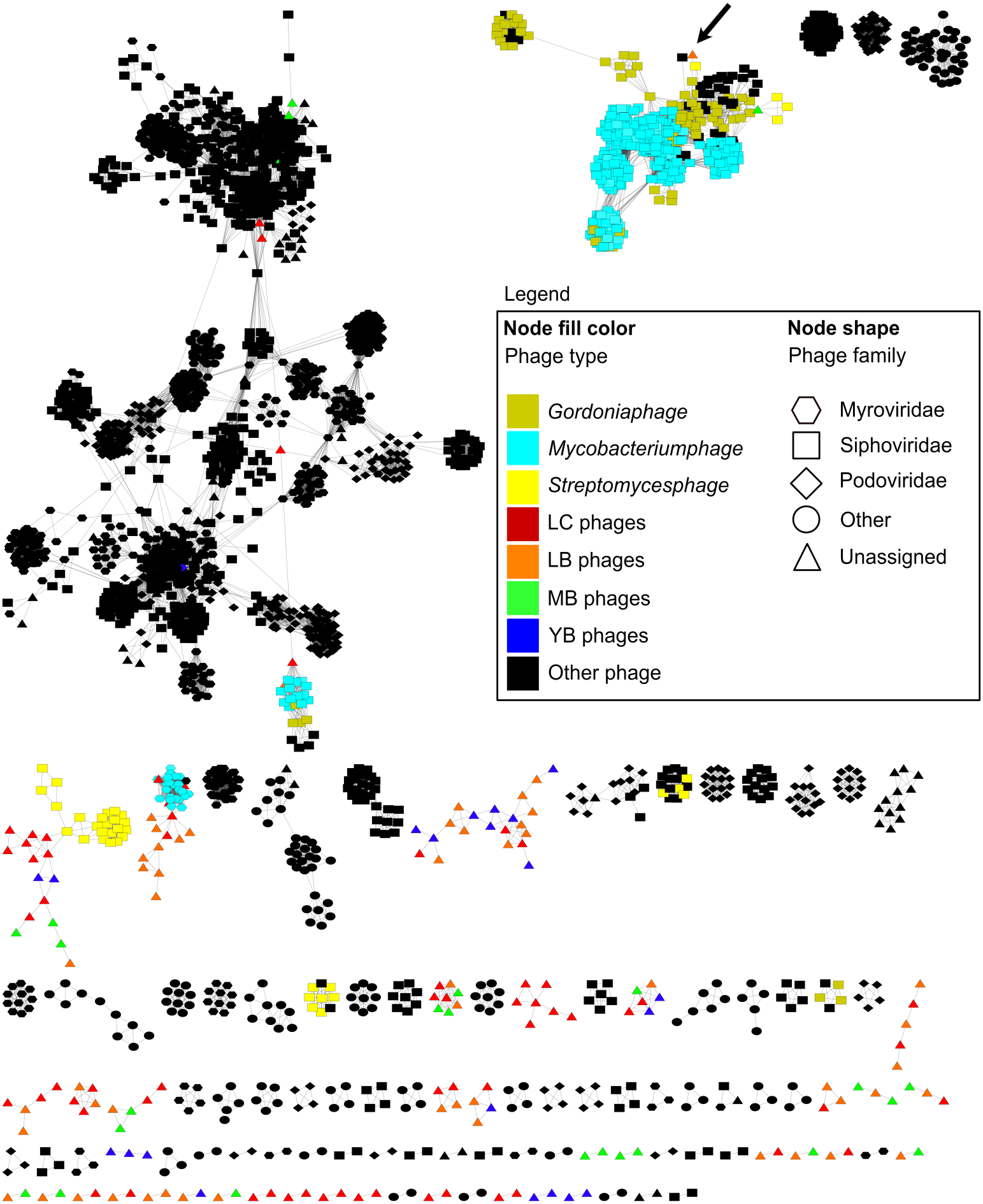
Viruses

- Lytic
- * Lysogenic

Putative host-virus interaction

Sample

- LC
- LB
- MB
- YB


Legend
Node fill color
Phage type
Gordoniaphage
Mycobacteriumphage
Streptomycesphage
LC phages
LB phages
MB phages
YB phages
Other phage
Node shape
Phage family
Myroviridae
Siphoviridae
Podoviridae
Other
Unassigned