Characteristics of Wetting-Induced Bacteriophage Blooms in Biological Soil Crust

Marc W. Van Goethem,[a] Tami L. Swenson,[a] Gareth Trubl,[b] Simon Roux,[c] Trent R. Northen[a,c]

[a]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[b]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA
[c]DOE Joint Genome Institute, Walnut Creek, California, USA

ABSTRACT Biological soil crusts (biocrusts) are photosynthetic "hot spots" in deserts and cover ~12% of the Earth's terrestrial surface, and yet they face an uncertain future given expected shifts in rainfall events. Laboratory wetting of biocrust communities is known to cause a bloom of *Firmicutes* which rapidly become dominant community members within 2 days after emerging from a sporulated state. We hypothesized that their bacteriophages (phages) would respond to such a dramatic increase in their host's abundance. In our experiment, wetting caused *Firmicutes* to bloom and triggered a significant depletion of cyanobacterial diversity. We used genome-resolved metagenomics to link phage to their hosts and found that the bloom of the genus *Bacillus* correlated with a dramatic increase in the number of *Caudovirales* phages targeting these diverse spore-formers ($r = 0.762$). After 2 days, we observed dramatic reductions in the relative abundances of *Bacillus*, while the number of *Bacillus* phages continued to increase, suggestive of a predator-prey relationship. We found predicted auxiliary metabolic genes (AMGs) associated with sporulation in several *Caudovirales* genomes, suggesting that phages may influence and even benefit from sporulation dynamics in biocrusts. Prophage elements and CRISPR-Cas repeats in *Firmicutes* metagenome-assembled genomes (MAGs) provide evidence of recent infection events by phages, which were corroborated by mapping viral contigs to their host MAGs. Combined, these findings suggest that the blooming *Firmicutes* become primary targets for biocrust *Caudovirales* phages, consistent with the classical "kill-the-winner" hypothesis.

IMPORTANCE This work forms part of an overarching research theme studying the effects of a changing climate on biological soil crust (biocrust) in the Southwestern United States. To our knowledge, this study was the first to characterize bacteriophages in biocrust and offers a view into the ecology of phages in response to a laboratory wetting experiment. The phages identified here represent lineages of *Caudovirales,* and we found that the dynamics of their interactions with their *Firmicutes* hosts explain the collapse of a bacterial bloom that was induced by wetting. Moreover, we show that phages carried host-altering metabolic genes and found evidence of proviral infection and CRISPR-Cas repeats within host genomes. Our results suggest that phages exert controls on population density by lysing dominant bacterial hosts and that they further impact biocrust by acquiring host genes for sporulation. Future research should explore how dominant these phages are in other biocrust communities and quantify how much the control and lysis of blooming populations contributes to nutrient cycling in biocrusts.

KEYWORDS *Firmicutes*, bacteriophages, biocrust, soil microbiology, virus-host interactions, wetting

The Southwestern United States is predicted to experience alterations in the timing of rainfall events with concomitant increases in soil surface temperatures (1). In arid lands, moisture pulses are infrequent, and yet their timing, intensity, and duration dictate biological activity (2). A significant avenue of recent microbial ecological research has been elucidation of the responses of microbial communities to various manifestations of climate change, such as increased rainfall, and of how these responses in turn affect ecosystem processes (3–5).

Biological soil crusts (referred to as biocrusts) are dominant topsoil microbial communities in deserts and serve as the basis for soil surface colonization, stabilization, and activity (6). Biocrusts can be comprised of photosynthetic cyanobacteria, bryophytes (mosses and liverworts), lichens, fungi, a myriad of heterotrophic lineages that cooperate to cycle both soil carbon and nitrogen, and other elements (7). Our previous research in biocrusts dominated by cyanobacteria has shown that biocrust wetting results in significant shifts in microbial community structure (8) and dramatically alters microbial metabolic outputs within the first day of wetting (9). The most dramatic of these observations is that multiple *Firmicutes* lineages bloom and then collapse following wetting, which we hypothesize is due to bacteriophage (phage) predation.

Recent studies using cultivation-independent metagenomic sequencing have revealed the roles that viruses play across terrestrial biomes, including predation by giant viruses in forests (10) and viral predation of dominant and ecologically relevant microbial lineages along a permafrost thaw gradient (11, 12) and of dominant tailed phages in Antarctic soils (13, 14). Viral metagenomics offers insight into viral diversity within complex samples (15) and unlocks the potential for novel viral diversity to be uncovered without the need for cultivation (16).

We have previously shown that *Firmicutes* spp. respond positively to a biocrust wetting event (8), which is thought to be due to their emergence from dormancy (17). However, we noted previously that many *Bacillus* strains declined in abundance soon after the initial wetting event whereas others continued to increase in abundance until a sudden decline occurred (9). We hypothesized that the increase in phage activity upon wetting leads to the predation of blooming *Firmicutes* populations in biocrust. Several lines of evidence motivate this view: First, phages that infect *Bacillus* in hot desert soils in both the Mojave Desert and the Namib Desert have been described previously (18, 19). Second, phages have been shown to target blooming bacterial populations in aquatic ecosystems (20), including *Prochlorococcus* spp. in the ocean (21, 22) and blooms of the cyanobacterium *Microcystis aeruginosa* in freshwater lakes (23). Finally, the "kill-the-winner" hypothesis posits that viruses act as moderating agents within a community by attacking the most active population of prokaryotes, thus making the most of the limited resources available to them (24, 25). Accordingly, we expected biocrust phages to target the most active hosts following a latent period. To test this concept, we mined microbial and phage sequences from metagenomes obtained from biocrust exposed to a controlled laboratory wetting event and tracked their responses over 2 days. The samples originated from the Moab Desert, a site of long-term research aimed at exploring the effects of climate change in arid regions in Utah in the Southwestern United States.

RESULTS

Our 24 biocrust communities were subjected to a simulated wetting experiment, which we leveraged to explore microbial and phage responses along the 2-day time series. The assembled sequence data comprised 28.4 million contigs (~20 Gbp) (see Table S1A and B in the supplemental material) from which we reconstructed 40 nearly complete bacterial metagenome-assembled genomes (MAGs). According to the Genomic Standards Consortium standards (26), all of the MAGs were of medium quality, as most lacked an identifiable 16S rRNA gene, and were >75% complete with <10% contamination (Table S2). These genomes encompassed bacterial members of *Firmicutes* (n = 16), *Bacteroidetes* (n = 7), *Proteobacteria* (n = 5), *Acidobacteria* (n = 4), *Chloroflexi* (n = 2), and *Actinobacteria* (n = 2) and included a single representative of

TABLE 1 Statistics describing the taxonomic diversity within the biocrust metagenomes grouped according to time after wetting

Sample	α-diversity ($P < 1.89e^{-5}$)	γ-diversity	β-diversity ($P < 0.001$)	β-dispersion	Shannon (H) ($P < 0.05$)	Simpson (D) ($P < 0.005$)	Fisher's alpha test value	Pielou's evenness (J) ($P < 0.05$)
Dry	117	219	1.496	0.1724	3.993	0.9689	8.579	0.8913
3 min	126.5	240	1.526	0.1752	4.011	0.9714	8.737	0.8922
9 h	111	211	1.516	0.1907	3.996	0.9673	9.179	0.8816
18 h	94.5	205	1.683	0.2266	3.953	0.9632	9.219	0.8756
42 h	76.50	174	1.699	0.2341	3.859	0.9584	8.499	0.8699
49.5 h	56.75	138	1.827	0.2851	3.782	0.9545	8.195	0.8678

Deinococcus-Thermus (see Fig. S1A and B in the supplemental material). The assembled genomes averaged 4.4 Mbp in size, with a range of relative abundances from high abundance (1.33% of assembled metagenome contigs) to rare (0.0067%) community members. We interrogated our MAGs for genes of interest relating to nutrient cycling, sporulation, osmoregulation, carbon starvation, and heat shock tolerances (see Results in Text S1 in the supplemental material).

The most phylogenetically diverse biocrust communities were the prewetting ("dry") metagenomes and those collected 3 min and 9 h after wetting ($\sim$58% of the community was sequenced on average as estimated using Nonpareil [27]). Biocrust metagenomes from later time points, 18 and 42 h after wetting, captured $\sim$64% of the community, while the final time point, 49.5 h after wetting, produced communities that were less complex, whereby 71% of the community was sequenced (Fig. S2A). Diversity estimates corroborated these rarefaction estimates and showed significant decreases in α-diversity following wetting (analysis of variance [ANOVA], $P < 0.05$) (Table 1). The 49.5-h group had significantly lower α-diversity than any of the other groups (Tukey *post hoc* test, $P < 0.05$). In contrast, the number of single-copy ribosomal genes (large and small subunits) increased significantly with wetting ($P < 0.05$). We also observed a significant increase in β-diversity over the course of our wetting experiment (within treatment variation). Note that the decrease in diversity may also have been due to our limited sampling and sequencing of rare community members as their relative abundances decreased dramatically after wetting.

We observed significant differences in community composition according to time after wetting (permutational multivariate analysis of variance [PERMANOVA], $P < 0.05$) (Fig. 1, left panel). The effect of wetting induces changes in composition that are strong enough to supersede those of the maturity stage of the biocrust, which developed over decades ($P > 0.05$). The increase in β-diversity over time was also observable as individual communities separated within ordination space (nonmetric multidimensional scaling [NMDS] ordinations based on Bray-Curtis dissimilarities of 16S rRNA gene content) (Fig. 1, left panel), which was the result of taxonomic divergence over time and of overall increases in 16S rRNA gene content (Fig. 1, right panel).

Overall, the biocrust communities encompassed 23 bacterial phyla and 1 archaeal phylum (*Thaumarchaeota*). The metagenomes also provided contigs belonging to *Ascomycota* (filamentous, saprobic *Fusarium* and *Purpureocillium* lineages), metazoa (*Tardigrada*), and mosses (*Physcomitrella patens*). The low diversity of fungi might have been due to database bias, and using a dedicated fungal database such as UNITE might have revealed more fungal diversity in these samples. Next, we explored the taxonomic shifts that occurred during biocrust wetting by tracking changes in 16S rRNA gene content in the unassembled reads. On the basis of our previous 16S rRNA gene survey, we anticipated a significant enrichment of *Bacillales* and other *Firmicutes* taxa $\sim$18 h after wetting (8, 9). Notably, these blooms should have been accompanied by substantial reductions in the proportion of filamentous cyanobacterial community members across all samples as nonbloomers are outcompeted by faster growers. However, *Cyanobacteria* are known to be sensitive to wetting (28, 29), and excessive cellular swelling and lysis are thought to cause their negative response. The dry and 3-min
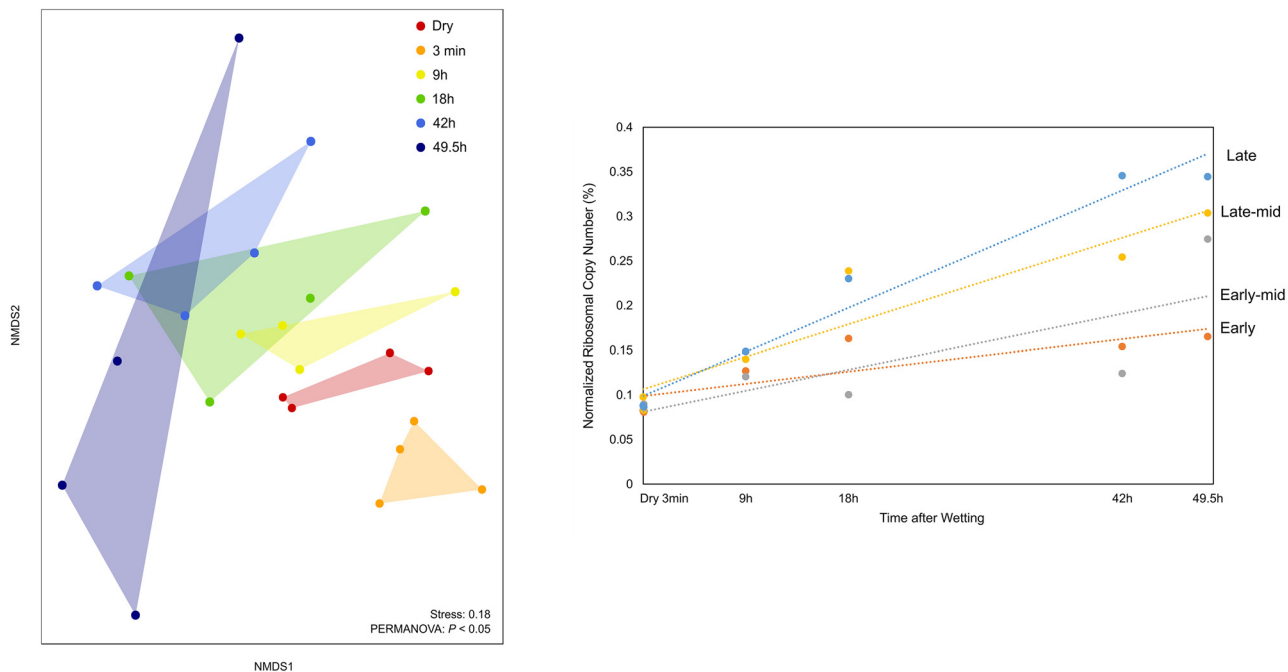
FIG 1 (Left) The changes in taxonomy over the course of the wetting experiment are observable in NMDS ordination space. Biocrust communities grouped according to time after wetting, which was sufficient to explain the differences in community composition (PERMANOVA; $P < 0.05$). (Right) The communities grew more dissimilar as the copy numbers of 16S rRNA genes increased over time.

biocrust metagenomes were dominated by photosynthetic *Cyanobacteria* and by heterotrophic *Actinobacteria*, *Bacteroidetes*, and *Proteobacteria* (Fig. 2, left panel). Overall, the biocrust communities encompassed 23 bacterial phyla and 1 archaeal phylum (*Thaumarchaeota*). The metagenomes also provided contigs belonging to *Ascomycota*
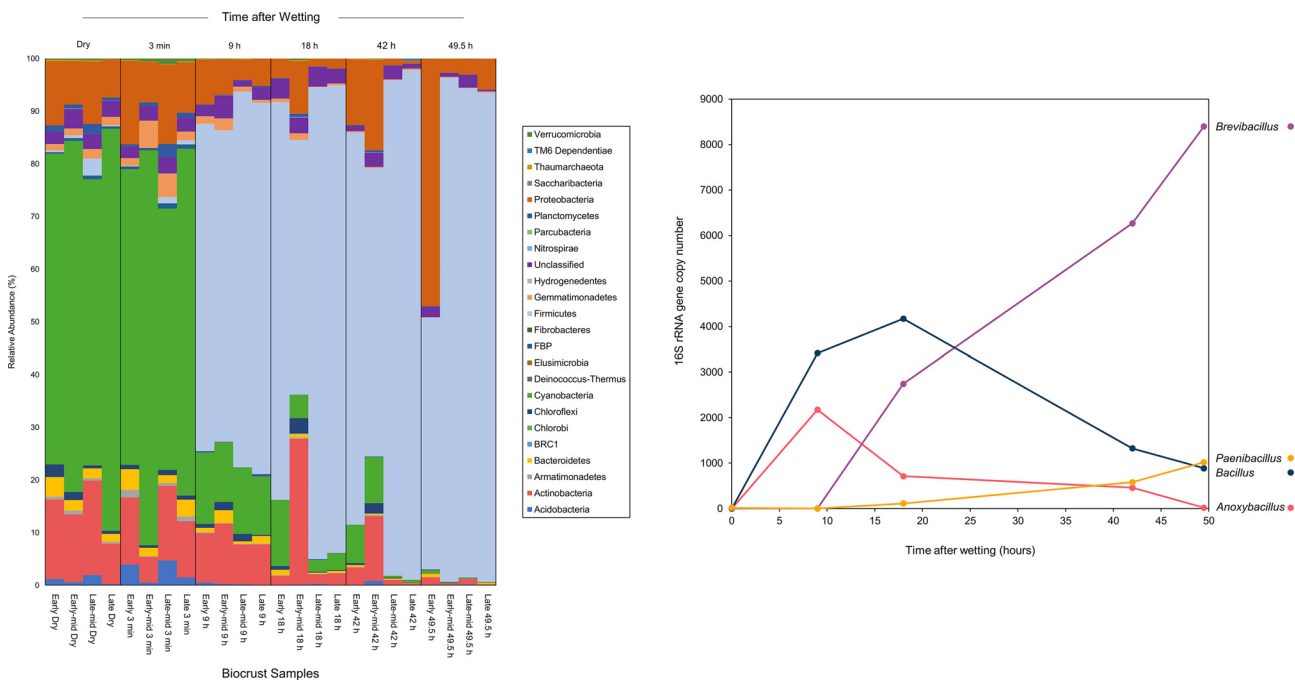
FIG 2 Exposure to a sudden wetting event had strong effects on the taxonomy of the biocrust communities across all successional stages. (Left) Relative abundances of microbial phyla in the biocrust communities. Archaea and eukaryotes are classified at the domain level, while *Proteobacteria* are distinguished at the class level. (Right) The responses of *Firmicutes* genera are shown as the proportion of 16S rRNA gene copies per metagenome.

(filamentous, saprobic *Fusarium* and *Purpureocillium* lineages), metazoa (*Tardigrada*), and mosses (*Physcomitrella patens*).

The wetting event drove decreases in the relative abundances of all nonblooming phyla. Notably, the relative abundance of *Cyanobacteria* decreased significantly from ~64% of the community in the dry samples to just 0.23% 49.5 h after wetting ($P < 2.30e^{-10}$). We observed striking, significant decreases in the relative abundances of most cyanobacterial lineages over the course of the experiment, culminating in a nearly complete loss of the genera *Microcoleus* ($P < 2.15e^{-7}$), unclassified *Cyanobacteria* ($P < 4.78e^{-5}$), *Tychonema* ($P < 2.36e^{-3}$), and Subsection III, Family I members ($P < 0.025$) by 49.5 h (Fig. 2, left panel). *Cyanobacteria* are known to be sensitive to wetting (28, 29) and excessive cellular swelling, and lysis may have caused their negative response here. The reduction in the abundances of cyanobacterial members also had dramatic effects on network topology across the experiment on the basis of co-occurrence analyses of taxonomic information (see Text S1, Fig. S3A to C, and Table S5).

Only members of the *Firmicutes* significantly increased in relative abundance after wetting; their levels increased from 1.07% of 16S rRNA genes in dry samples to 82.3% of 16S rRNA genes at 49.5 h (relative abundance of 16S rRNA genes, ANOVA, Bonferroni corrected, $P < 1.77e^{-6}$) (Table S3A). The presence of the *Firmicutes* bloom was pronounced across all biocrust successional stages, but the responses of individual genera were more nuanced. Notably, *Brevibacillus* showed a steady increase in abundance after wetting (Fig. 2, right panel), whereas *Paenibacillus* showed slower growth. The population of *Bacillus* became dominant 9 h after wetting but had collapsed by 49.5 h. We speculated that these diverse responses to wetting would be reflected by increases in the proportions of functional genes affiliated with the changes in the proportions of these taxa.

We tracked how changes in taxonomic composition affected the relative contributions of functional genes. Most of the upregulated genes in our analysis were involved in sporulation (Fig. 3), suggesting that previously dormant *Firmicutes* had replicated within 9 h after wetting. Specifically, we observed significant increases in the abundance of sporulation cluster stage IIIA genes (*AB*, *AC*, *AD*, and *D*), stage V genes (*spoVS*), spore maturation protein A, and genes encoding the spore cortex-lytic enzyme *CwlJ* (Fig. 3). Consistently, the dramatic reduction in the populations of cyanobacterial taxa over the course of the wetting experiment led to significant reductions in the relative abundances of genes involved in the assembly of photosystem I (*ycf4* and *psaLK2*) and photosystem II (*psbADMPVY*) and of genes involved in the production of bilin (*pcyA*) and the phycobilisome (*nblB* and *cpeS*). The relative abundances of genes associated with the cyanobacterial circadian clock (input kinase A) and myxoxanthophyll biosynthesis (*cruF*) and genes encoding products associated with photorespiration (D-glycerate 3-kinase and glycolate oxidase) were also significantly reduced (Table S3B). Strikingly, numerous genes involved in carbon dioxide (CO_2) uptake (carboxysome) were also significantly reduced in abundance. These included RuBisCO chaperonin genes (*rbcX*) as well as CO_2-concentrating genes (*ccmO*), high-affinity carbon uptake genes (*hatR*), and low-affinity CO_2 hydration genes (*cphX*). Overall, 95 genes decreased in relative abundance over the wetting experiment, while only 22 genes became significantly more abundant following wetting (Table S3B).

Next, we used VirSorter to mine viral signals from our 28 metagenomes (including the 4 bundle metagenomes), and the mined signals were coassembled into 5 data sets to assist viral sequence recovery. This procedure produced a total of 1,849 unambiguous viral sequences (see Materials and Methods). The removal of short (<10-kb) contigs, together with manual curation (see Materials and Methods), refined our data set to 1,210 viral genomes (ViGs) or genome fragments.

To place our ViGs within the known, relevant diversity of publicly available viral genomes, we created a virus genome-based network based on shared protein content (30, 31) by combining our contigs with the viral genomes and genomic fragments present in the RefSeq viral database (v85) (32). The analysis produced 40,796 virus
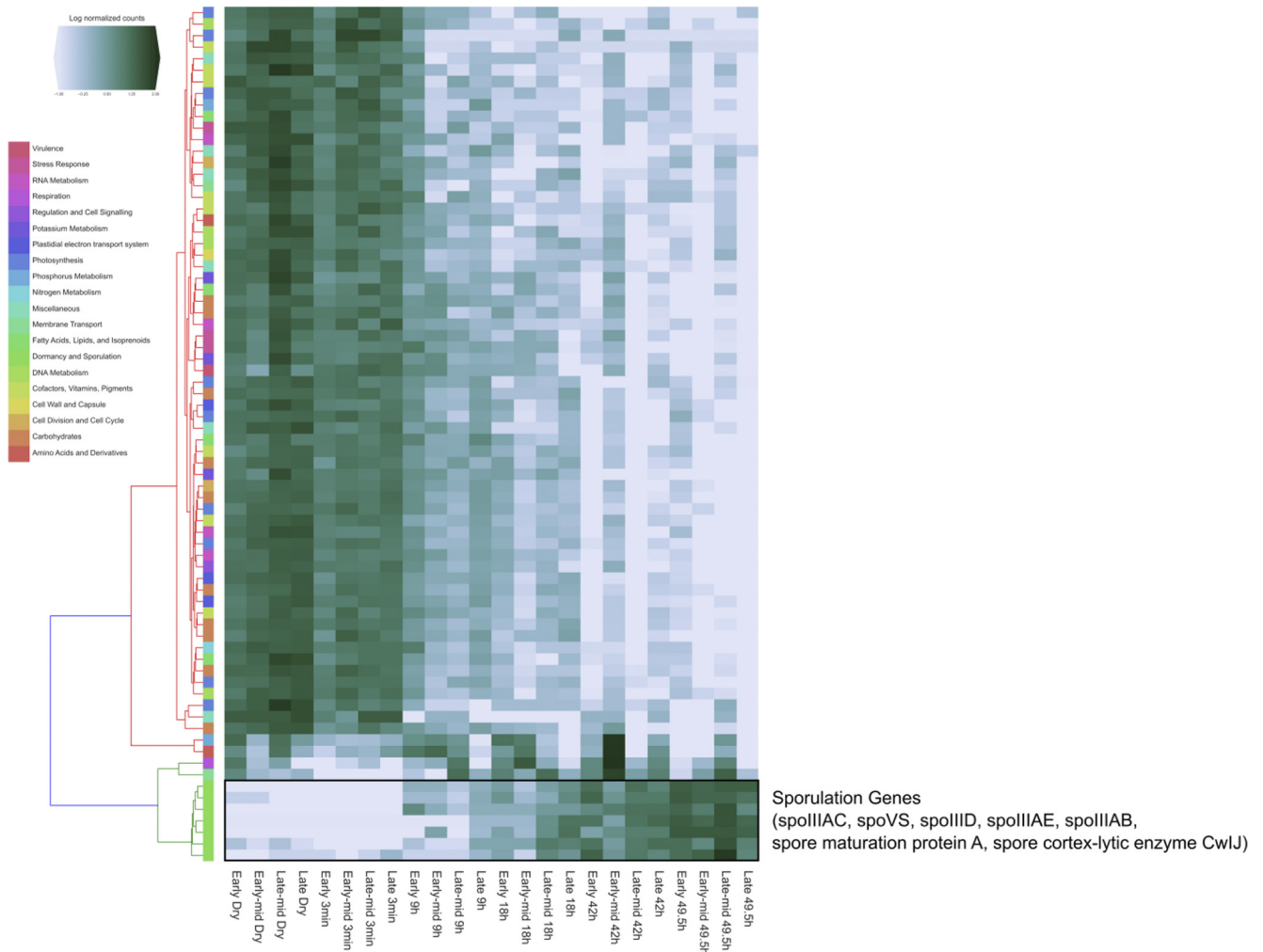
FIG 3 Heat map indicating the distribution of major functional genes in biocrust metagenomes following soil wetting. Count data are log normalized to enable comparisons across different sequencing efforts. Genes that are overrepresented are represented by darker green colors; white data represent genes that are underrepresented (i.e., the genes with the lowest relative abundance levels). The x axis indicates biocrust communities following wetting, i.e., ranging from the dry state to 49.5 h after wetting. The y axis data include all functional genes that differed significantly in abundance across the wetting event.

clusters (VCs) representing approximately genus-level taxonomy, of which 2,294 (6%) contained biocrust viral contigs, representing 504 uncultivated viral genomes (UViGs). Overall, 148 (29%) of the 504 biocrust UViGs that we identified were found to be taxonomically affiliated using these VCs. Clustering the 504 UViGs at 95% average nucleotide identity (ANI) and 85% alignment fraction (AF) (20) produced 433 viral operational taxonomic units (vOTUs), of which 49 could be assigned at the genus level. All classified vOTUs belonged to *Caudovirales* families (double-stranded DNA [dsDNA] tailed phages), although six vOTUs were broadly categorized as representing dsDNA viruses (Fig. 4, left and top right panels; see also Table S4A).

The longest circular UViGs belonged to three identical *Spbetavirus* phages (163,801 bp), which are *Siphoviridae* phages known to infect soil *Bacillus* (33). These UViGs were reconstructed independently biocrust metagenomes seen in the early, late-middle (late-mid), and late successional stages, while the early-middle (early-mid) metagenome also contained a smaller *Spbetavirus* contig of 137,161 bp. the *Siphoviridae* phages were the most abundant, with 89 vOTUs spanning 17 genera. *Podoviridae* were represented by 19 vOTUs, including three T7 viruses, while 21 vOTUs assigned to the *Myoviridae* family were found to include three genera (*Abouovirus*, *Bxz1virus*, and *P2virus*). A further 10 genomes were coarsely assigned to the *Caudovirales*. Mapping
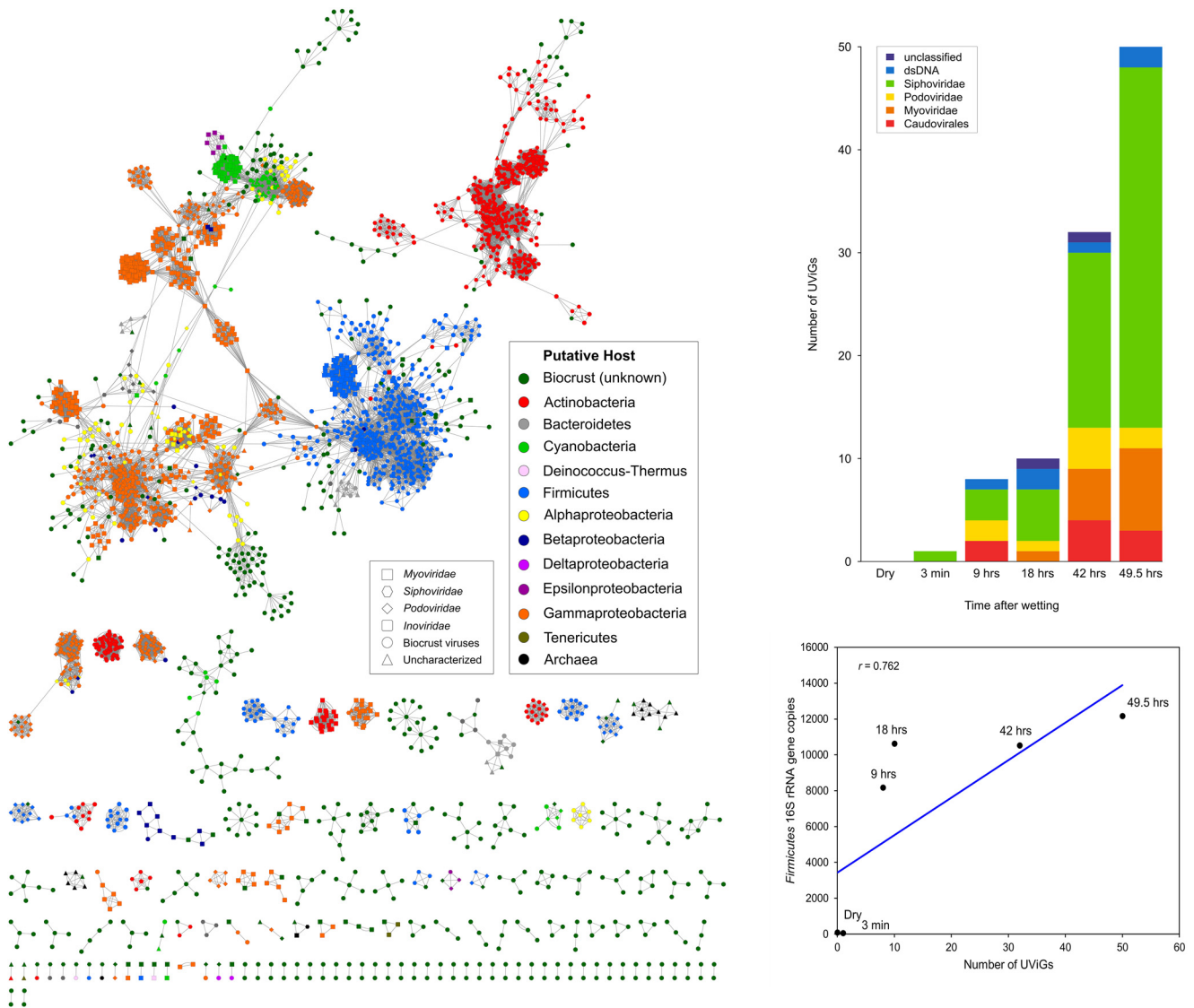
FIG 4 (Left) Network clustering of recovered virus genomes and genome fragments with RefSeq viral genomes (v85) based on shared predicted protein content. Major viral families (nodes) are indicated by unique colors and shapes, with vOTUs found in this study indicated by green circles. Edges (lines between nodes) indicate significant shared protein content. (Top right) Stacked bar chart indicating the number of UViGs that were recovered across the wetting experiment. (Bottom right) Correlation between the relative proportions of *Firmicutes* and UViGs per time point.

UViGs back to individual metagenomes revealed that all of the phages were detected only after wetting; i.e., no phages were recovered from the bundles or the dry samples (Fig. 4, top right panel), which culminated in the greatest number of UViGs ($n = 50$) being recovered more than 2 days after wetting. For example, the full-length spbetaviruses emerged at 42 h and 49.5 h after wetting, which correlates with the decline in the abundances of the putative *Bacillaceae* hosts at those time points (Fig. 2, right panel). Andromedaviruses (length range, 47,888 to 48,100 bp) were found in all successional stages at 18, 42, and 49.5 h after wetting and have previously been shown to infect *Bacillus* hosts (33). Combined, these UViG data indicated that *Firmicutes* were predominant targets, particularly at the end of the wetting experiment, which led us to further explore this putative link. The classification of our viral genomes provided primary evidence of phages targeting *Firmicutes* such as *Bacillus*, *Staphylococcus*, and *Brevibacillus*.

Putative auxiliary metabolic genes (AMGs; genes that originate from microbes and can increase fitness) that code for sporulation proteins (*spoIIIE*) were identified in 9
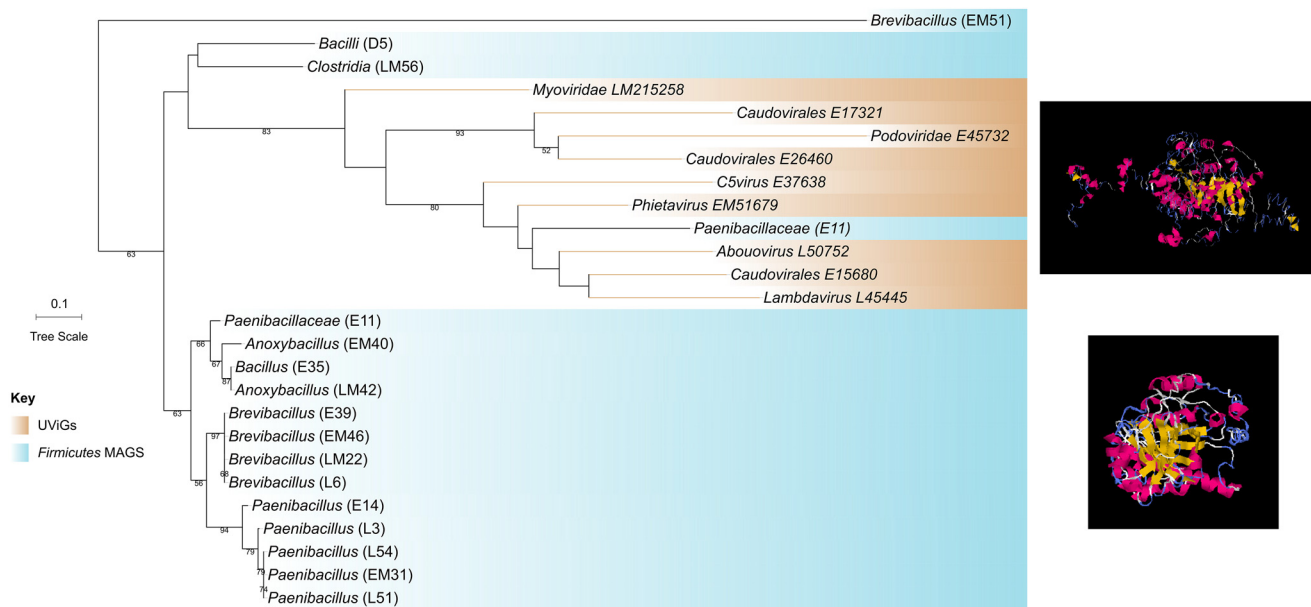
FIG 5 Maximum likelihood tree indicating the phylogenetic placement of *spoIIIE* genes retrieved from UViGs with their nearest bacterial neighbors (from *Firmicutes* MAGs). Protein structures were predicted using I-TASSER, and the first predicted models are presented. Tree lineages are colored blue for bacterial sequences and orange for viral sequences.

UViGs. To confirm that the AMGs were of viral origin, we checked that flanking regions present both upstream and downstream contained common virus genes (integrases, structural genes, or terminases), had significant matches to the gene of interest (E value $< 1 \times 10^{-3}$; bit score > 60), and represented the only AMG per contig. BLASTP queries indicated that all 9 sporulation genes (*spoIIIE*) were assigned to *Bacillus* strains. We used the sporulation AMGs to explore the proportion of phages infecting the blooming *Firmicutes*. As an initial proof of concept, we found that the numbers of UViGs correlated strongly with the relative abundances of *Firmicutes* across the wetting experiment ($r = 0.762$). That is, as the proportion of *Firmicutes* increased throughout the wet-up, so too did the number of recoverable phage genomes (Fig. 4, bottom right panel; see also Fig. S4B).

Genes for sporulation were detected in 15 of the 16 *Firmicutes* MAGs. Genes involved in spore DNA protection (dipicolinate synthase subunits A and B, *dpaAB*) and small acid-soluble spore proteins (*SASP1*, *SASPS6*, and *gpr*) were detected exclusively in the *Firmicutes* genomes and are well-known components of endospore production (34, 35). *DpaAB* genes code for dipicolinic acid, which allows endospores to maintain dormancy under harsh conditions. Sporulation-related genes, including *spoIIIE* and *gerE*, were among the most highly overrepresented genes in the metagenomes examined in our study by the end of our wetting experiment (Tables S3B and S4B), reiterating the increase in the relative abundances of spore-formers following wetting. A maximum likelihood tree indicated that the *spoIIIE* AMGs found in our UViGs showed high similarity to those found in the *Firmicutes* genomes (Fig. 5). We confirmed the clustering result by predicting protein structures of each *spoIIIE* gene using I-TASSER and comparing models using TM-align, which showed a similarity score of 0.488.

Additionally, we searched for CRISPR-Cas repeats and proviral genes within the reconstructed MAGs to link viruses to potential hosts. We found evidence of CRISPR-Cas repeats in 16 MAGs, which links them to previous or ongoing infection events by biocrust viruses (11, 36). Six of the 16 *Firmicutes* MAGs contained CRISPR elements (Table S2). Aligning the CRISPR-Cas repeats and spacers from the MAGs with our virus contigs revealed perfect homology between a *Phietavirus* (EM_51679) and the *Anoxybacillus* MAG EM_40. Both the repeat sequence (5′-CTTCACATCACACATAGTTTCCTCAG

Bacillus-Phage Relationship
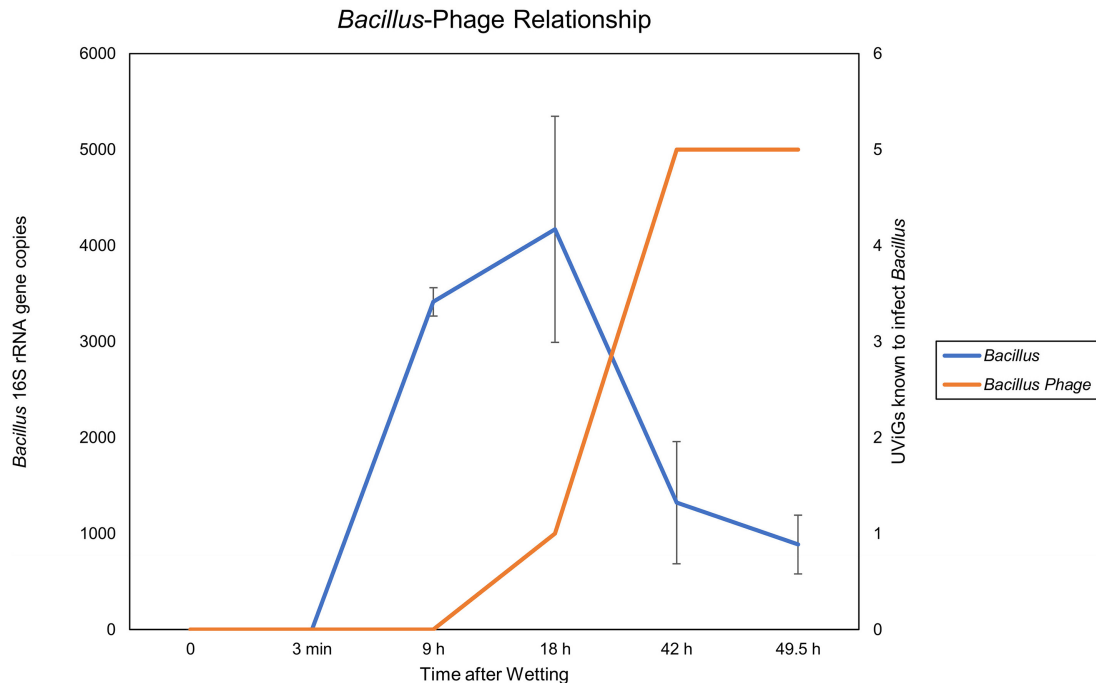
FIG 6 Predator-prey relationship between *Bacillus* organisms and their phages. *Bacillus* rRNA copy number is provided on the left-hand *y* axis with error bars calculated across four metagenomes from each time point. The numbers of UViGs linked to *Bacillus* infection are indicated on the right-hand *y* axis.

AAAC-3′) and three spacers (data not shown) were identical between the virus and its cognate host, which offers a direct link to the microbial host *in silico*. Next, we searched for prophage elements within each MAG to uncover temperate phages that utilized a lysogenic infection cycle. We found 12 MAGs with viral elements within their genomes, of which 7 were *Firmicutes* (Table S2).

DISCUSSION

The results presented here provide an initial assessment of the phage diversity in biocrust communities—dominant photosynthetic topsoil communities found around the globe. This description of dominant phage lineages in biocrust further extends our knowledge of viruses in important soil ecosystems (37), which lags considerably behind what is known about marine viruses (20, 38). Our work sought to resolve the contradicting fates of taxa within the *Firmicutes* after biocrust communities were exposed to a simulated wetting event (8), which we predicted might have been due to phage predation. Here, we provide evidence of a predator-prey relationship between *Bacillus* and their phages. Comparing the 16S rRNA gene copy numbers of *Bacillus* throughout wet-up to the numbers of their known UViGs showed that phage abundance peaked as that of their hosts began to decline (Fig. 6).

The presence of proviral sequences within *Firmicutes* genomes suggests that lysogenic replication strategies are common among biocrust viruses. Temperate viruses shift their replication strategy from lysogenic to lytic replication as bacterial host production increases (39). We found prophages in 12 of the 40 MAGs. Seven *Paenibacillus* MAGs contained proviral sequences, and each of these contained multiple prophages. While the genus *Bacillus* abundance declined, both *Paenibacillus* and *Brevibacillus* continued to increase in relative abundance throughout the wet-up. According to the kill-the-winner model, we would expect phages to attack the fastest-growing host population first (i.e., *Bacillus*). We anticipate that phages would target members of the *Paenibacillus* and *Brevibacillus* after the *Bacillus* populations have been depleted on the basis of the trajectory of these genera over 2 days (Fig. 2, right panel).

For *Bacillus* hosts, phage-encoded sporulation genes could play an additional role in

these dynamics by providing a mechanism for lysis with concomitant phage and spore release where uninfected host spores would represent a reservoir of future host cells. Note that we cannot differentiate between population reduction mediated by lysis from sporulation and that mediated by some combination of factors with these data. In addition, spores that include integrated phage genomes would also provide additional protection, which may be important in the harsh biocrust environment. Regardless, the lysis of *Firmicutes* cells probably stimulates the biocrust community as metabolites, such as nitrate, are released into the soil. Conceivably, this viral shunt could provide essential nutrients to mosses or other higher eukaryotes, as predicted in marine ecosystems (40), which has the potential to influence ecological services of the crust.

Our metagenome-assembled genomes also revealed that *Firmicutes*-infecting phages acquired genes horizontally from their hosts' genomes. Specifically, nine *spoIIIE* genes in phage genomes link them to *Firmicutes* infections. The *spoIIIE* gene codes for a DNA translocase that plays a critical role during sporulation by assisting the translocation of chromosomal DNA from mother cells into the forespore during the final stages of engulfment (41). It has been shown that marine cyanophages contain photosynthetic genes which maintain essential host functions (22). These *spoIIIE* genes may have a similar role in these biocrusts if sporulation is an important component of the life history of these *Bacillus* phages, as endospores can be carriers of phage genomes, thus offering protection from various environmental stresses (42, 43). Alternatively, it is possible that these phage-encoded proteins were originally sporulation related but now perform another function related to DNA export and/or ATP-dependent motor activity. AMGs encoding sporulation machinery (*spoVS*) in vOTUs from permafrost have been detected previously (12), and it would be intriguing to investigate how widespread these genes are in other soil systems as they may influence host sporulation dynamics (42).

As *Firmicutes* bloom during precipitation events, they are probably preyed upon by phages, exerting control on their abundance throughout the year. Biocrust saturation after rainfall releases adsorbed phages, which may initiate repeated cycles of lytic replication (37). The drying soil likely draws phages into small pores by shrinking water films (44). Soil microheterogeneity may further limit lytic virus-host interactions as cells become heterogeneously distributed within the dry soil matrix. This paradigm is especially true of desert biocrusts, which experience prolonged periods of dormancy for much of the year and reawaken after moisture events (45, 46). Future research should investigate these dynamics in the field, particularly as our long-term storage of samples (~11 months in the dark) may have influenced the stress levels that the microbes experienced. Moreover, it is unclear whether blooms of *Firmicutes* occur as rapidly in naturally occurring biocrust communities or if this response is mitigated by other factors.

Previous surveys of viral populations from desert biotopes, such as cyanobacterium-dominated hypoliths (18) and exposed desert soils (47), revealed a high proportion of *Caudovirales* (dsDNA tailed phages), which appear to be dominant dsDNA phages in biocrust as well. We found that the family-level hierarchical abundance of *Siphoviridae* > *Myoviridae* > *Podoviridae* was remarkably similar to that described in Antarctic hypoliths (14), potentially indicating that these tailed phages are core constituents of "hot spots" of microbial activity in desert ecosystems. Moreover, the average genome sizes of phages identified here (average length, ~43 kb) are like those described in Namib soils identified from pulsed-field gel electrophoresis profiles (55 kb). Additionally, those studies provided striking evidence that cyanophages are a potentially low-abundance component of desert viral communities despite the abundance of putative cyanobacterial hosts in hypoliths (47) and biocrust, including the genera *Microcoleus*, *Scytonema*, *Schizothrix* and *Nostoc* (7, 48). These results should be interpreted with caution, however, as our long-term storage of biocrust samples in the dark for 11 months may have diminished the responsiveness of certain taxa to wetting. Moreover, most knowledge regarding cyanophage diversity stems from oceanic surveys, which may have introduced biases into the RefSeq database that was used here.

The number of studies on soil viruses continues to increase rapidly, which certainly offers researchers the potential to detect novel cyanophages in these underexplored soil ecosystems.

Summary. Biocrust wetting induced a bloom of *Firmicutes* and their phages. Our data suggest that *Firmicutes* emerge from spores, whereby fewer individuals remain in sporulated states after 2 days, which is reflected in significant enrichment for genes relating to sporulation. In contrast, the reduction in the relative abundance of *Cyanobacteria* resulted in significant depletion of functions relating to both carbon and nitrogen fixation, and this was observed in collapses in co-occurrence network structures (see Fig. S3C in the supplemental material). Our investigation of phage diversity in biocrust, performed here for the first time to our knowledge, did not link phage predation to cyanobacterial decline but instead revealed that biocrust phages primarily infect blooming *Firmicutes*, particularly members of the *Bacillus* genus. The phage-bacterium dynamics observed here appear to follow the kill-the-winner model, in which phages expand on the fastest-growing host population in biocrust. The collapses of population blooms are often mediated by phage infection and lysis and can occur on very short time scales such as in marine systems where phages target blooming *Synechococcus* species within 30 min (40). Here, we showed that the increase in *Firmicutes* abundance over 2 days is closely followed by peaks in the frequency of recoverable *Caudovirales* phage genomes targeting different members of the phylum, followed by a decline of specific *Firmicutes* lineages as their phages replicate. The presence of putative auxiliary metabolic genes associated with sporulation in these phages suggests that the sporulation process may be an important component of the life history of these phages. Future research could further investigate the contributions of sporulation to phage dispersal and maintenance within endospores. Additional analyses could consolidate the fate of the biocrust communities by investigating how biocrusts return to their initial steady state following drying.

MATERIALS AND METHODS

Sample collection, experimental setup, and DNA extraction. Details of the sample collection procedure, the biocrust wetting experiment, and DNA extraction protocols have been described previously (9) and are illustrated in Fig. S4A in the supplemental material. Briefly, 28 petri dishes (6 cm² by 1 cm in depth) were used to core biocrust samples from the Green Butte Site near Canyonlands National Park (38°42′54.1′′N, 109°41′27.0′W, Moab, UT, USA) in 2014. The site is an area of long-term ecological research interest aimed at exploring climatic changes in arid regions. Biocrust samples (28 petri dishes) were collected along a maturity gradient of *Cyanobacteria*-dominated biocrusts ranging from lightly colored, young crust to dark mature crust, which encompassed four maturity stages overall. Dry samples collected in the field were maintained in a dark desiccation chamber in the laboratory for 11 months until required. Biocrust samples underwent a laboratory (12 h light/12 h dark cycle) wetting event by wetting of each soil sample (0.5 g) with liquid chromatography-mass spectrometry (LC-MS)-grade water (1 ml). Soil was removed at five time points following wetting (in addition to dry and bundled samples) for each successional stage (3 min, 9 h, 18 h, 42 h, and 49.5 h after wetting). After pelleting of soil by centrifugation (5,000 × *g* for 5 min) and removal of porewater, DNA was extracted from soils (0.25 g) using a DNeasy PowerSoil kit (Qiagen, Hilden, Germany), resulting in 100 μl of eluted DNA. Library preparation and sequencing were performed at the QB3 facility at the University of California, Berkeley (UC Berkeley), using an Illumina HiSeq 4000 system.

Metagenomic sequencing and analysis. Metagenomic sequencing provided ~413 Gbp of Illumina HiSeq data in 2,754,282,840 raw sequences across 28 biocrust samples. The data included four separate metagenomes for microbial communities collected before wetting and those collected 3 min, 9 h, 18 h, 42 h, and 49.5 h after wetting. Sequencing was performed on an Illumina HiSeq 4000 instrument using 2 × 150-bp chemistry. Raw sequences were filtered using Prinseq v0.20.4 (49), which eliminated reads with phred scores of <20 for six consecutive bases (–*min_qual_mean*), singletons, and those with ambiguous bases (*ns_max_n*). Estimates of sequence coverage were obtained using Nonpareil v3.30 (27). High-quality sequences were assembled under default conditions using MEGAHIT v1.1.3 (50) as previously suggested for highly complex soil samples (51). These 24 metagenomes were also coassembled according to their maturity stage, i.e., dry, early, early-mid, late-mid, and late successional stage biocrust, as physically identified *a priori* during the field sampling expedition. Statistics about each assembled metagenome were obtained using MetaQUAST v4.6.3 (52) and are available in Table S1B in the supplemental material. Open reading frames (ORFs) were extracted from each assembled metagenomes with Prodigal v2.6.3 (53). Annotation of contigs was performed using DIAMOND v0.9.14.155 (54) with queries made against the entire NCBI nonredundant nucleotide database. Here, annotations were provided at E values of <0.00001 (-*e*), while the five best hits were retained (-*k*). MEGAN software v6.12.5 was used to visualize functional and taxonomic attributes of communities (55). We extracted all 16S rRNA

genes from the curated metagenomes using SortMeRNA (56). Finally, we resolved exact sequence variants (ESVs) using DADA2 (Divisive Amplicon Denoising Algorithm [57]), with default parameters except for *truncLen* (120, 150) and *maxEE* (2, 5). We assigned taxonomy against the entire SILVA reference database (58) after random rarefaction of each community to the lowest sequencing effort (8,805).

Network construction. We constructed co-occurrence networks at the genus level using our log-transformed taxonomic data. Networks were constructed using Sparse Correlations for Compositional (SparCC) data (59). We chose to divide our 24 metagenomes according to time after wetting. Individual networks were created for the beginning (dry and 3 min after wetting), middle (9 h and 18 h after wetting), and end (42 h and 49.5 h after wetting) of the experiment. We used significant SparCC values ($P < 0.05$) that were greater than 0.6 or less than -0.6 and included only genera present in at least half of the samples within our OTU table. All networks were visualized using Cytoscape v3.6.1 (60), and highly connected modules were automatically identified using MCODE (61).

Reconstructing genomes from metagenomes. Due to our ultradeep sequencing effort, we were able to reconstruct metagenome-assembled genomes (MAGs) from our assembled biocrust communities using MetaBAT v2.12.1 (62). We increased the sensitivity of the binning process by increasing the minimum contig length to 5,000 bp (–m), the percentage of good contigs to 97 (–maxP), the minimum edge binning score to 85 (–minS), and the maximum number of edges per node to 250 (–maxEdges). We also included the –noAdd option, which prevented small contigs from being binned and prevented the creation of bins that were <1 Mbp in size. MAGs were checked for completeness and contamination using CheckM v1.0.8 (63). Medium- to high-quality MAGs were retained according to the conventional values for minimum information about a MAG (>75% completeness, <10% contamination) (26). MAGs were then functionally annotated online using Prokka v1.12 (64) and the RAST v2.0 server (65), with ORFs validated with UGENE (66). Quality statistics about the MAGs are presented in Table S2.

Phylogenetic tree building. Phylogenetic trees were built using universal bacterial protein gyrase subunit B (*gyrB*) extracted from MAGs ($n = 19$) and 2,878 reference sequences. The *gyrB* sequences were collated and aligned using MAFFT v7 (67). Neighbor-joining trees were built with bootstrap values reported, and visualization was performed in iTOL v3 (68). The phylogenetic placement of the MAGs was also evaluated using the Microbial Genome Atlas (MiGA) online (69).

Statistical analyses. We used STAMP v2.1.3 (70) and R v3.5.0 to make statistical comparisons between the communities according to both time after wetting and successional stage. Within the R statistical environment, we used the *vegan* package (71) to explore ecologically meaningful differences between communities according to time after wetting and successional stage. Bray-Curtis dissimilarity matrices were used to explore biocrust community structure after Hellinger transformations of species relative abundance data. PERMANOVA was used to test for differences in community structure on the basis of time after wetting (dry biocrust and biocrust at 3 min, 9 h, 18 h, 42 h, and 49.5 h after wetting) and successional maturity (early, early-mid, late-mid, and late) using the *adonis* function. We used STAMP to test for significant differences in taxonomic and functional composition between time points and successional stages and used ANOVA with Tukey-Kramer *post hoc* tests and Bonferroni corrections of the *P* value statistic.

To obtain viral contigs from biocrust metagenomes, we coassembled samples according to their characteristics and community complexity. The coassemblies were produced by combining the 4 filamentous bundles, the 4 dry biocrust samples, the 5 early successional stage biocrust samples, the 5 early-mid succession biocrust samples, the 5 late-mid succession biocrust samples, and the 5 late succession biocrust metagenomes into single, concatenated metagenomes. This produced 6 coassemblies to complement the 28 individual metagenomes.

Recovering and annotating viral contigs. Viral contigs were recovered from the 6 coassembled biocrust metagenomes using VirSorter 1.0.2, which relies on the detection of "hallmark" viral genes, such as those genes involved in the formation of the viral capsid (72). The Virome db database, without additional viral sequences as references, was used to recover viral contigs from all coassembled biocrust metagenomes. All contigs classified into category 1 or category 2 (complete viral contigs) were considered to be of unambiguous viral origin, while contigs in categories 4 and 5 (proviruses within microbial contigs) were modified to include only the predicted proviral regions, thus excluding host genomic signatures (see host analysis below). Only the viral contigs that scored in the highest confidence categories were retained and analyzed. The viral contigs within our metagenomes are likely sequences derived from double-stranded DNA phages and include prophage elements (temperate phages integrated within microbial chromosomes) and plasmids (extrachromosomal DNA) as well as lytic phages. All predicted viral contigs were retained if they were ≥10 kb in length. The most common class was composed of category 2 viral signatures which were "likely" predictions ($n = 1,665$), followed by category 1 ("most confident" predictions; $n = 184$). These viral signals were retrieved from our 28 metagenomes (including the 4 bundle metagenomes), which were coassembled into 5 data sets to assist viral sequence recovery. The removal of short (<10-kb) contigs refined our data set to 1,210 viral genomes or genome fragments. The identified viral protein sequences were clustered and taxonomically classified using vConTACT2 (30), which is supported by the International Committee on Taxonomy of Viruses (ICTV). We built a network on the basis of significant gene-sharing similarities (represented as edges) between viral genomes with contigs represented as nodes. We performed *all-vs-all* BLASTp queries for comparisons between our 2,140 VirSorted sequences and all predicted viral proteins in the NCBI RefSeq database (v85) with an E value of 1×10^{-3}. Protein clusters were defined using the Markov Clustering Algorithm (MCL) and then ClusterONE as described previously (73). These viral genomes represent both high-quality draft genomes (single genome fragment greater than 90% of the expected genome size) and viral genome fragments (single genome fragments less than 90% of the expected genome size or of unknown genome size), defined within

the Genomics Standards Consortium framework as the minimum information about an uncultivated virus genome (MIUViG) (16). The UViGs identified by this procedure were visualized within the Cytoscape v3.6.1 program. We also mapped our MAGs and UViGs back to our metagenomes to calculate mapping rates and to determine which genomes were recovered from each time point using Bowtie 2 (74). UViGs were clustered into viral operational taxonomic units (vOTUs) at 95% ANI and 85% AF.

Viral host prediction and CRISPR-Cas analyses. To link viruses with their cognate hosts, we explored the distribution of CRISPR-Cas elements within each annotated MAG and metagenome. Clustered regularly interspaced short palindromic repeats (CRISPR) represent phage infection legacy and thus provide a palpable link between phages and their cognate hosts (36). CRISPRs identified with MinCED 0.3.2 were confirmed using BLASTn queries by allowing for 1 mismatch. We also searched for prophage elements within each MAG using VirSorter, within the CyVerse Discovery Environment.

Auxiliary metabolic gene (AMG) analysis. We considered the 148 viral genomes and genomic fragments for exploration of auxiliary metabolic genes (AMGs) on the basis of VirSorter-predicted contigs (only VirSorter categories 1 and 2, thus excluding proviral contigs). This reduced the possibility of including host genes in our analysis. The identified sporulation related AMGs were reannotated to confirm their classifications through BLASTP comparisons whereby those with E values of 1×10^{-3} and bit scores of >60 were retained. Next, the genomic regions both upstream and downstream of the retained genes were explored for common viral genes, including those encoding terminases, integrases, and viral structural genes. Finally, maximum likelihood trees were built using the AMG sequences and those extracted from the host genomes, as described in the "Phylogenetic tree building" section. The protein structures of our AMGs of interest were predicted with I-TASSER (75), after which we tested for correct predictions using TM-align (76).

Data availability. The sequencing data used in this study are available at the NCBI Sequence Read Archive (project accession number PRJNA395099; sample accession numbers SRX3024638 to SRX3024657; https://www.ncbi.nlm.nih.gov/bioproject/395099).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .02287-19.

TEXT S1, DOCX file, 0.02 MB.
FIG S1, TIF file, 1.6 MB.
FIG S2, TIF file, 2.8 MB.
FIG S3, TIF file, 2.3 MB.
FIG S4, TIF file, 1.6 MB.
TABLE S1, XLSX file, 0.02 MB.
TABLE S2, XLSX file, 0.02 MB.
TABLE S3, XLSX file, 0.03 MB.
TABLE S4, XLSX file, 0.04 MB.
TABLE S5, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Schlesinger WH, Reynolds JF, Cunningham GL, Huenneke LF, Jarrell WM, Virginia RA, Whitford WG. 1990. Biological feedbacks in global desertification. Science 247:1043–1048. https://doi.org/10.1126/science.247 .4946.1043.

2. Noy-Meir I. 1973. Desert ecosystems: environment and producers. Annu Rev Ecol Syst 4:25–51. https://doi.org/10.1146/annurev.es.04.110173 .000325.

3. Reed SC, Coe KK, Sparks JP, Housman DC, Zelikova TJ, Belnap J. 2012. Changes to dryland rainfall result in rapid moss mortality and altered soil fertility. Nat Clim Chang 2:752–755. https://doi.org/10.1038/nclimate1596.

4. Rutherford WA, Painter TH, Ferrenberg S, Belnap J, Okin GS, Flagg C, Reed SC. 2017. Albedo feedbacks to future climate via climate change impacts on dryland biocrusts. Sci Rep 7:44188. https://doi.org/10.1038/ srep44188.

5. Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, Blazewicz SJ, Harden J, Turetsky MR, McGuire AD, Shah MB, VerBerkmoes NC, Lee LH, Mavrommatis K, Jansson JK. 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature 521:208. https://doi.org/10.1038/nature14238.

6. Belnap J, Gillette DA. 1998. Vulnerability of desert biological soil crusts to wind erosion: the influences of crust development, soil texture, and disturbance. J Arid Environ 39:133–142. https://doi.org/10.1006/jare.1998.0388.

7. Elbert W, Weber B, Burrows S, Steinkamp J, Büdel B, Andreae MO, Pöschl U. 2012. Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. Nat Geosci 5:459. https://doi.org/10.1038/ngeo1486.

8. Karaoz U, Couradeau E, Da Rocha UN, Lim H-C, Northen T, Garcia-Pichel F, Brodie EL. 2018. Large blooms of Bacillales (Firmicutes) underlie the response to wetting of cyanobacterial biocrusts at various stages of maturity. mBio 9:e01366-16. https://doi.org/10.1128/mBio.01366-16.

9. Swenson TL, Karaoz U, Swenson JM, Bowen BP, Northen TR. 2018. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. Nat Commun 9:19. https://doi.org/10.1038/s41467-017-02356-9.

10. Schulz F, Alteio L, Goudeau D, Ryan EM, Feiqiao BY, Malmstrom RR, Blanchard J, Woyke T. 2018. Hidden diversity of soil giant viruses. Nat Commun 9:4881. https://doi.org/10.1038/s41467-018-07335-2.

11. 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. https://doi.org/10.1038/s41564-018-0190-y.

12. 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.

13. Adriaenssens EM, Kramer R, Van Goethem MW, Makhalanyane TP, Hogg I, Cowan DA. 2017. Environmental drivers of viral community composition in Antarctic soils identified by viromics. Microbiome 5:83. https://doi.org/10.1186/s40168-017-0301-7.

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

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. https://doi.org/10.1038/nature19094.

16. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, et al. 2019. Minimum information about an uncultivated virus genome (MIUVIG). Nat Biotechnol 37:29. https://doi.org/10.1038/nbt.4306.

17. Placella SA, Brodie EL, Firestone MK. 2012. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. Proc Natl Acad Sci U S A 109:10931–10936. https://doi.org/10.1073/pnas.1204306109.

18. Adriaenssens EM, Van Zyl L, De Maayer P, Rubagotti E, Rybicki E, Tuffin M, Cowan DA. 2015. Metagenomic analysis of the viral community in Namib Desert hypoliths. Environ Microbiol 17:480–495. https://doi.org/10.1111/1462-2920.12528.

19. 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.

20. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801. https://doi.org/10.1038/nrmicro1750.

21. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. 2004. Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci U S A 101:11013–11018. https://doi.org/10.1073/pnas.0401526101.

22. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol 3:e144. https://doi.org/10.1371/journal.pbio.0030144.

23. Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki K, Hiroishi S. 2008. Ecological dynamics of the toxic bloom-forming cyanobacterium Microcystis aeruginosa and its cyanophages in freshwater. Appl Environ Microbiol 74:3269–3273. https://doi.org/10.1128/AEM.02240-07.

24. Winter C, Bouvier T, Weinbauer MG, Thingstad TF. 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. Microbiol Mol Biol Rev 74:42–57. https://doi.org/10.1128/MMBR.00034-09.

25. Thingstad T, Lignell R. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat Microb Ecol 13:19–27. https://doi.org/10.3354/ame013019.

26. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725. https://doi.org/10.1038/nbt.3893.

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

28. Angel R, Conrad R. 2013. Elucidating the microbial resuscitation cascade in biological soil crusts following a simulated rain event. Environ Microbiol 15:2799–2815. https://doi.org/10.1111/1462-2920.12140.

29. Belnap J. 2006. The potential roles of biological soil crusts in dryland hydrologic cycles. Hydrol Process 20:3159–3178. https://doi.org/10.1002/hyp.6325.

30. Jang HB, Bolduc B, Zablocki O, Kuhn J, Roux S, Adriaenssens E, Brister JR, Kropinski A, Krupovic M, Turner D. 2019. Gene sharing networks to automate genome-based prokaryotic viral taxonomy. BioRxiv https://doi.org/10.1101/533240.

31. Bolduc B, Jang HB, Doulcier G, You Z-Q, 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.

32. 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.

33. Mahmud RS, Garifulina K, Ulyanova V, Evtugyn V, Mindubaeva L, Khazieva L, Dudkina E, Vershinina V, Kolpakov A, Ilinskaya O. 2017. Bacteriophages of soil bacilli: a new multivalent phage of Bacillus altitudinis. Mol Genet Microbiol Virol 32:87–93. https://doi.org/10.3103/S0891416817020082.

34. Onyenwoke RU, Brill JA, Farahi K, Wiegel J. 2004. Sporulation genes in members of the low G+ C Gram-type-positive phylogenetic branch (Firmicutes). Arch Microbiol 182:182–192. https://doi.org/10.1007/s00203-004-0696-y.

35. Galperin MY. 2013. Genome diversity of spore-forming Firmicutes. Microbiol Spectr 1:TBS-0015-2012. https://doi.org/10.1128/microbiolspectrum.TBS-0015-2012.

36. Skennerton CT, Imelfort M, Tyson GW. 2013. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. Nucleic Acids Res 41:e105. https://doi.org/10.1093/nar/gkt183.

37. 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.

38. Rohwer F, Thurber RV. 2009. Viruses manipulate the marine environment. Nature 459:207. https://doi.org/10.1038/nature08060.

39. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. 2016.

Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. ISME J 10:437. https://doi.org/10.1038/ismej.2015.125.

40. Ortmann A, Lawrence J, Suttle C. 2002. Lysogeny and lytic viral production during a bloom of the cyanobacterium Synechococcus spp. Microb Ecol 43:225–231. https://doi.org/10.1007/s00248-001-1058-9.

41. Sharp MD, Pogliano K. 1999. An in vivo membrane fusion assay implicates SpoIIIE in the final stages of engulfment during Bacillus subtilis sporulation. Proc Natl Acad Sci U S A 96:14553–14558. https://doi.org/10.1073/pnas.96.25.14553.

42. Gabiatti N, Yu P, Mathieu J, Lu GW, Wang X, Zhang H, Soares HM, Alvarez PJ. 2018. Bacterial endospores as phage genome carriers and protective shells. Appl Environ Microbiol 84:e01186-18. https://doi.org/10.1128/AEM.01186-18.

43. Felde V, Peth S, Uteau-Puschmann D, Drahorad S, Felix-Henningsen P. 2014. Soil microstructure as an under-explored feature of biological soil crust hydrological properties: case study from the NW Negev Desert. Biodivers Conserv 23:1687–1708. https://doi.org/10.1007/s10531-014-0693-7.

44. Lebre PH, De Maayer P, Cowan DA. 2017. Xerotolerant bacteria: surviving through a dry spell. Nat Rev Microbiol 15:285. https://doi.org/10.1038/nrmicro.2017.16.

45. Pointing SB, Belnap J. 2012. Microbial colonization and controls in dryland systems. Nat Rev Microbiol 10:551. https://doi.org/10.1038/nrmicro2831.

46. Griffin JS, Wells GF. 2017. Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly. ISME J 11:500. https://doi.org/10.1038/ismej.2016.121.

47. 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.

48. Garcia-Pichel F, Loza V, Marusenko Y, Mateo P, Potrafka RM. 2013. Temperature drives the continental-scale distribution of key microbes in topsoil communities. Science 340:1574–1577. https://doi.org/10.1126/science.1236404.

49. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.

50. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

51. Van der Walt AJ, Van Goethem MW, Ramond J-B, Makhalanyane TP, Reva O, Cowan DA. 2017. Assembling metagenomes, one community at a time. BMC Genomics 18:521. https://doi.org/10.1186/s12864-017-3918-9.

52. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32:1088–1090. https://doi.org/10.1093/bioinformatics/btv697.

53. 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.

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

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

56. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217. https://doi.org/10.1093/bioinformatics/bts611.

57. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581. https://doi.org/10.1038/nmeth.3869.

58. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J,

Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and Web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.

59. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687.

60. 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.

61. Bader GD, Hogue CW. 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4:2. https://doi.org/10.1186/1471-2105-4-2.

62. Kang D, Li F, Kirton ES, Thomas A, Egan RS, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ Preprints 7:e27522v1. https://doi.org/10.7717/peerj.7359.

63. 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.

64. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

65. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

66. Okonechnikov K, Golosova O, Fursov M, UGENE team. 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166–1167. https://doi.org/10.1093/bioinformatics/bts091.

67. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.

68. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128. https://doi.org/10.1093/bioinformatics/btl529.

69. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282–W288. https://doi.org/10.1093/nar/gky467.

70. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30:3123–3124. https://doi.org/10.1093/bioinformatics/btu494.

71. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ. 2007. The vegan package: community ecology package. https://cran.r-project.org/package=vegan.

72. 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.

73. 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, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537:689. https://doi.org/10.1038/nature19366.

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

75. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40. https://doi.org/10.1186/1471-2105-9-40.

76. Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309. https://doi.org/10.1093/nar/gki524.