Title: Ecology of active viruses and their bacterial hosts in frozen Arctic peat soil revealed with $H_2{}^{18}O$ stable isotope probing metagenomics

Running title: SIP-metagenomics of arctic peat viral populations

Gareth Trubl[1], Jeffrey A. Kimbrel[1], Jose Liquet-Gonzalez[1], Erin E. Nuccio[1], Peter K. Weber[1], Jennifer Pett-Ridge[1], Janet K. Jansson[2], Mark P. Waldrop[3], Steven J. Blazewicz[1]

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

[2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

[3]U.S. Geological Survey, Geology, Minerals, Energy, and Geophysics Science Center, Menlo Park, CA, USA

Corresponding authors:

Gareth Trubl (Trubl1@llnl.gov) and Steve Blazewicz (blazewicz1@llnl.gov)

Key Words: soil viruses, bacteriophage, stable isotope probing, permafrost, peat, ^{18}O-water, metagenomics

Abstract

Winter carbon loss in northern ecosystems is estimated to be greater than the average growing season carbon uptake. However, most ecosystem carbon measurements neglect winter months since carbon losses (primarily driven by microbial decomposers) are assumed to be negligible at low temperatures. We used stable isotope probing (SIP) targeted metagenomics to reveal the genomic potential of active soil microbial populations under winter conditions, with an emphasis on viruses and virus-host dynamics. Peat soils from the Bonanza Creek LTER site in Alaska were incubated under subzero anoxic conditions with $H_2{}^{18}O$ for 184 and 370 days. We identified 46 bacterial populations (MAGs; spanning 9 bacterial phyla) and 243 viral populations (vOTUs) that actively took up ^{18}O and produced significant CO_2 throughout the incubation. Active hosts, predicted for 33% of the active vOTUs, were some of the most abundant MAGs and capable of fermentation and organic matter degradation. Approximately three-quarters of the active vOTUs carried auxiliary metabolic genes that spanned five functional categories, including carbon utilization, highlighting the potential impact of viruses in this peat soil's microbial biogeochemistry. These results illustrate significant bacterial and viral activity and interactions occur in frozen soils, revealing viruses are a major community-structuring agent throughout winter months.

Introduction

Northern peatlands are important terrestrial ecosystems for carbon (C) storage, estimated to contain one-third of soil C (~1,000 gigatons) [1], yet the fate of this C is unknown as these soils are experiencing dramatic changes from anthropogenic climate change [2-4]. While soil-warming experiments indicate increased carbon dioxide (CO_2) and methane emissions under global climate change, it is likely that these C losses from northern peatlands are underestimated because virtually all measurements neglect winter processes [5-8]. Recent estimates of winter C loss alone are estimated to be greater than the average growing season C uptake [9]. While the winter months include large temperature fluctuations and extreme temperature minimums [10], the temperatures found in much of the soil profile of permafrost or seasonally frozen bogs can remain stable and just below the freezing point (□1 to □5°C) [5, 11, 12]. Bacteria have been shown to remain active below the freezing point [13] with both catabolic and anabolic activities observed in frozen soils [14, 15]. Therefore, it is critical to understand the taxonomy, functional capacity, and activities of bacteria and viruses that cause microbial turnover in frozen soils to better predict their biogeochemical implications.

In soils, viruses may play a major role in microbial population dynamics and the fate of soil C, similar to the role they play in marine systems, where viruses kill ~40% of bacteria daily and sustain up to 55% of bacterial production via C liberation [16, 17]. Soils hold an enormous viral reservoir, but we know remarkably little about their diversity, activity, host interactions, lysis-induced C-cycling, and persistence as compared to other environments [18, 19]. Viruses can affect soil C cycling by stopping microbial metabolism during lytic infections and releasing cell-derived nutrients into the environment [20]. These nutrients can either fuel other organisms' metabolism or be stabilized via entombing effects [21]. This process can be prolonged by

temperate viruses, which remain latent in their host for long periods, before eventually switching into the lytic infection cycle. Viruses can also carry auxiliary metabolic genes (AMGs), host-derived genes that can be expressed during infection and typically aid the infection process by overcoming energetic limitations [22, 23]. While viruses and their impacts have been well-characterized in marine environments via meta-omic approaches [24] and development of tractable virus-host model systems, these linkages remain enigmatic in soil environments. Growing season studies of viruses in northern peatlands indicate they are largely unrelated to other known viruses, highly diverse, endemic to their habitat, infect dominant microbial populations, and carry AMGs that could impact C cycling [25, 26]. The question of whether viruses are active during the 7–9 months of the year when Arctic peatlands are frozen remains unanswered.

Stable isotope probing (SIP) combined with metagenomics is an effective tool for tracking the active microbial taxa in complex communities, linking individuals to specific functions [27-29], or characterizing specific microbes and the viruses that infect them [30]. While many isotope tracing studies use ^{13}C-enriched tracers (e.g., $^{13}CO_2$, ^{13}C-plant biomass, or ^{13}C-glucose) to identify specific substrate degraders [28] or assess microbial C use efficiency [31], heavy water ($H_2{}^{18}O$) SIP-metagenomics has the unique benefit of labelling all actively growing microbes because water is a universal substrate for nucleic acid synthesis [32-34]. Over adequate time scales, this technique is sensitive enough to label slow-growing or less abundant microbes and identify taxon-specific microbial growth and mortality patterns [35-38], but previously, has not been used to label active viruses.

Here we used heavy-water SIP-metagenomics to label active microbial taxa and viruses and characterize whether viruses play a role in controlling active microbial population dynamics

under subzero and anoxic soil conditions. To our knowledge, viral activities have never been confirmed under such conditions, and understanding their impact on microbial activity and soil C transformation over winter months may be pivotal for understanding the future fate of the vast organic C stocks in Arctic peatlands.

Methods

Experimental setup

We collected soil samples from the Alaska Peatland Experiment (APEX), part of the Bonanza Creek Long Term Ecological Research (LTER) site southwest of Fairbanks, AK (64.70 °N, 148.3 °W), in a zone of discontinuous permafrost (for overview see Fig. S1). Three peat soil cores were collected from the active layer of a *Sphagnum*-dominated thermokarst bog on June 16, 2013 from the beta site approximately 20 m south of the flux tower. During the week of sampling, midday air temperatures were 27°C, midday surface 10 cm peat temperature was 10°C, active layer depth in the bog was 35 cm, and the water table was 3–8 cm below the surface. The pH was approximately 4.80 (1:1 soil:water ratio). Vegetation and other soil and flux data are given in Waldrop et al., [39]. To collect the cores, moss and peat were removed down to the top of the water table using scissors. The top 20 cm below the water table were collected using a 6.3 cm diameter sharpened steel barrel corer attached to an electric drill. Three cores were collected 1 meter apart along a transect. Cores were stored in mason jars filled with porewater collected from the core hole. Sealed jars were immediately shipped on ice to the U.S. Geological Survey (Menlo Park, CA, USA) where they were homogenized in an anaerobic glovebox maintained at 4°C. Anoxic synthetic porewater was created by freeze-drying filtered (0.45 µm PTFE then 0.2 µm nylon filter) soil porewater collected in tandem with the cores, and then rehydrating the remaining particulates with either heavy water (97 atom% $H_2{}^{18}O$,

Cambridge Isotope Laboratories) or natural abundance water (control), and then sparging with N_2 to remove O_2. Soil subsamples (2 g of soil wet weight) were collected from each core, pressed to remove porewater (using a 5 ml syringe fitted with a nylon screen and a glass fiber filter), and transferred to Wheaton serum vials (10 ml), creating 12 incubation vessels. Vials were sealed with blue butyl rubber stoppers and the headspace was purged (to remove H_2 from the glovebox atmosphere) by vacuuming and filling with N_2 (10 inches Hg/5 psi) 10 times in a cold room. Anoxic synthetic ^{18}O-enriched porewater was added (2.5 ml) to half of the incubation vials and anoxic synthetic natural abundance porewater was added to the other half using a 5 ml syringe with a 23 G needle that was purged with N_2. Six parallel samples were set up in a similar manner for headspace gas analysis to quantify CO_2 production, except these used proportionally larger amounts of soil (18.15 g wet soil) and larger incubation vials (100 ml). All the sample vials were submerged in a glycerol bath at 4°C and the temperature was slowly and steadily reduced to □1.5°C, over 48 h. Samples were then continuously maintained at □1.5°C for 184 d (midyear) and 370 d (full year). At each timepoint, biological replicates (n=3) were destructively harvested and snap frozen in liquid N_2 and stored at □80°C. Three of the gas production vials were incubated at □20°C as a control.

CO_2 production quantification

Gas samples were collected from gas production vials at 10 timepoints over the 370 d incubation. To prevent oxygen from contaminating the incubation vials, a 5 ml syringe with a stopcock and 23 G needle was cleared 3 times with O_2-free N_2. The syringe was then inserted into the vial septa and plunged 3 times to mix the headspace; 2 ml headspace was collected, and the stopcock locked. The 2 ml samples were transferred to 10 ml serum bottles that had been

purged and then filled with N_2 (1 atm). Gas samples were analyzed via gas chromatography (SRI 8610GC, SRI instruments, Torrance, CA) to quantify headspace CO_2.

DNA extraction and density gradient SIP

DNA was extracted from each replicate using a modified phenol/chloroform protocol [40]. In summary, 0.5 g (+/- 0.01 g) wet soil was added to lysing matrix E tubes, and 100 µl 1x TE (pH 7.5), 150 µl PO_4 buffer (0.2 M in 1 M NaCl), and 100 µl 10% SDS were added and vortexed. Tubes were bead beat for 30 s at 30 1/s and briefly centrifuged. 0.6 ml phenol:chloroform:isoamyl alcohol (25:24:1) was added, vortexed, and incubated at 65°C for 10 min. Tubes were spun for 5 min at 10K x g, and the supernatant was transferred to a new tube. The bead beat tubes were then re-extracted using 220 µl 1x TE and 80 µl PO_4 buffers. The supernatant from the first and second extracts were combined in a new 2 ml tube. 550 µl phenol/chloroform/isoamyl alcohol was added, mixed, and centrifuged (10K x g, 5 min). The aqueous top layer was transferred to a new 2 ml tube. 900 µl chloroform:isoamyl alcohol (24:1) was added, mixed, centrifuged (10K x g, 5 min), and the supernatant transferred to a new 2 ml tube. Then, 850 µl chloroform:isoamyl alcohol (24:1) was added, mixed, centrifuged (10K x g, 5 min), and the supernatant transferred to a new 1.7 ml tube. RNAase was added (6.44 µl, 10 mg/ml), mixed, and incubated at 50°C for 10 min. 244 µl 10 M NH_4^+ acetate was added, mixed, and incubated at 4°C for 2 h. Tubes were centrifuged at 16.1K x g for 15 min, and the supernatant transferred to a new tube. Isopropanol (670 µl) was added, mixed, and centrifuged (16.1K x g, 20 min). The supernatant was removed, and the DNA pellet dried in a PCR hood for 15 min. 30 µl 1xTE was added, mixed, and the DNA was then stored at □80°C.

Extracted DNA was fractionated via CsCl density gradient ultracentrifugation to separate ^{18}O-enriched DNA as described previously [35]. DNA was binned into 5 fractions based on

density, and the binned DNA from the two heaviest fractions (medium-heavy [MH; 1.717–1.725 g/ml] and heavy [H; 1.725–1.750 g/ml]) were sequenced.

Sequencing and metagenome generation

DNA from the SIP fractions was sent to the Keck Sequencing Facility at Yale University. For each sample, 100 ng of DNA was sheared to 500bp using the Covaris E210 (Covaris, Inc., Woburn, MA, USA), followed by a SPRI bead cleanup using Ampure XP (Beckman Coulter Life Sciences, Brea, CA, USA); DNA quality was checked using a Bioanalyzer chip. The sheared gDNA from the 24 samples was then end-repaired, A-tailed, adapters ligated, and sequenced on an Illumina HiSeq 2500 to generate metagenomes (Table S1). Paired-end 151 nt reads were processed in three steps with bbduk v38 (Bushnell, B.): 1) adapter-trimming (ftl=10 ktrim=r k=23 mink=11 hdist=1 tpe tbo minlen=50), 2) PhiX and Illumina adapter/barcode removal (k=31 hdist=1 minlen=50), and 3) quality-trimming (qtrim=r trimq=10 minlen=50). Metagenomes were successfully generated for 23 of the samples (one did not have enough DNA recovered), with 302 Gbp of sequencing data.

Recovery of vOTUs from metagenomes

Virus-specific informatics were used to increase the number of viral sequences detected in these soil datasets [19]. Processed reads were assembled into contigs using SPAdes v3.11.1 (--only-assembler --phred-offset 33 --meta -k 25,55,95 –12)[41]. From the 23 SIP-fractionated metagenomes, we assembled 51 487 619 contigs, with 63% of the total reads mapped to the contigs. Contigs were processed with VirSorter (virome decontamination mode) [42] and DeepVirFinder (DVF) [43] to detect viral contigs. We retained contigs that were ≥10kb, sorted into VirSorter categories 1 and 2, and had a DVF score ≥0.9 and p value <0.05. Viral contigs were clustered at 95% average nucleotide identity (ANI) across 85% of the shorter contig [44]

using nucmer [45] to generate a nonredundant set of viral populations (vOTUs). vOTU quality was assessed with CheckV (default parameters) [46]. Coverage of the vOTUs was estimated based on post quality-controlled read mapping at ≥90% ANI and covering ≥75% of the contig [44] using Bowtie2 [47]. Coverage was then normalized per gigabase-pair of metagenome and by length of the contig [48]. Activity of vOTUs was determined by a vOTU being present in the $H_2{}^{18}O$ samples and not present in the paired natural abundance water samples. The vOTUs were annotated using the virus-centric multiPhATE pipeline (default parameters) [49] and the AMG-centric DRAM-v pipeline (with --skip_uniref) [50]. We note that DRAM-v provides AMG scores only for vOTUs detected via VirSorter; AMGs predicted from the 208 vOTUs recovered from DVF were manually curated. For this manual inspection, we removed putative AMGs that were at contig ends or those with annotations from multiple functional categories. To determine the proportion of temperate vOTUs, we searched for genes associated with proviruses, such as integrase and parA [51], leveraged classification from VirSorter (categories 4 or 5) and our genome-similarity host linkages (≤90% similarity), and used two tools — CheckV [46] and PHASTER [52].

MAG curation and host linking

To generate metagenome assembled genomes (MAGs), read-pairs from the biological replicates were grouped for 8 separate co-assemblies (2 timepoints x 2 SIP fractions x 2 O isotopes) with MEGAHIT v1.1.4 (--k-min 27 --k-max 127 --k-step 10) [53]. Contigs greater than 1 KB were separately binned with Concoct v1.0.0 [54], MaxBin v2.2.6 [55] and MetaBAT v2.12.1 [56]. Genome bins from these three binning tools were refined using the bin_refinement

module of MetaWRAP v1.2.1 (-c 50 -x 10) [57] with CheckM v1.0.12 [58]. Only genome bins with at least 'medium quality' according to MIMAG standards [59] were retained.

Two methods were used to define MAGs as active: 1) a read-subtraction approach, and 2) a log-fold-change approach. For the read-subtraction approach, contigs from the ^{16}O assemblies were used as a reference to subtract the ^{18}O reads that aligned with the unlabeled dataset using bbsplit v38 (maxindel=1) [60]. The isotopically labeled reads that did not align with the unlabeled dataset were considered 'active' reads and were then processed through the same MAG assembly workflow described above (starting with the MEGAHIT assembly through MetaWRAP refinement); this generated genome bins reconstructed from distinct ^{18}O reads from the two SIP fractions at two timepoints.

Refined bins from all 12 groups (8 total + 4 active) were dereplicated into a final set of metagenome assembled genomes (MAGs) using dRep v2.2.3 (-comp 50 -con 10 -p 6 -nc 0.6) [61]. MAG taxonomy was determined using GTDB-tk v0.3.2 [62] with the version r89 GTDB database. Structural annotation was done using Patric [63], and general functional annotation with RASTtk [64] for subsystems within Patric v3 (retaining subsystem variants predicted to be active or likely), and KofamScan v1.1.0 [65] with the KEGG [66] 93 database. Per-sample MAG abundances were determined by aligning each sample's filtered reads against the MAG genomes using bbmap v38 [67] to obtain total assigned reads per contig and average fold coverage per contig.

For the log-fold-change approach to define active MAGs, we assessed significant (5% false-discovery rate) and positive log2-fold change in ^{18}O versus ^{16}O read abundances within a time point and SIP fraction. Fold changes were determined using wrench-normalized [68] total

assigned reads per MAG with a zero-inflated log-normal model implemented in metagenomeSeq [69].

The vOTUs and MAGs were linked via clustered regularly interspaced short palindromic repeats (CRISPR) spacers and shared genomic content as previously described [70]. Briefly, CRISPR regions were detected in the MAGs using MinCED (options -minNR 2 -spacers) [71] and linked to the vOTUs with blastn (options -max_target_seqs 10000000 -dust no -word_size 7) [72]. In addition, BLAST (options -dust no -perc_identity 70) was used to link vOTUs and MAGs based on shared genomic content [73]. Virus-host abundance estimates were made by summing microbial host abundances at the phylum level in each timepoint and linked vOTU abundances.

Data availability

The 23 metagenomes were deposited to NCBI under BioProject identifier (ID) PRJNA634918 with BioSample information included in Table S1. Figures were generated with Microsoft Excel and R, using packages Vegan for diversity and pheatmap for heat maps. T-tests and linear regressions were performed using the data analysis package in Excel.

Results

Characterization of viruses

To characterize viral activity in Arctic peat soil under winter conditions, we analyzed viral sequences from heavy-water SIP-targeted metagenomic reads. Using two viral detection methods, we identified 5 737 putative viruses (≥5 kb) that clustered into a nonredundant set of 332 vOTUs ≥10 kb (Table S2) that span 66 viral genera (Fig. S2; Table S3). The size range of these vOTUs was 10 039 – 437 858 bp (average 32 954 bp) with 15 vOTUs ≥100 kb, including 5

so-called ‘huge’ viruses (i.e., ≥200 kb) [74]. The vOTUs were well-covered with an average of 17x normalized coverage per metagenome, but with a large range, 3–147x (Fig. 1A). After quality checks, we identified 58 medium–high quality vOTUs, of which four were considered ‘complete’ according to community standards [44]. Genome quality could not be assessed for 93 vOTUs because they did not possess a known viral gene and were detected via a reference-free machine learning method [43]. Annotation of the vOTU genomes with the virus-centric Multiphate pipeline predicted 15 772 genes, of which 61% were novel (Table S4). With the AMG-centric pipeline DRAM-v, we predicted 86 putative AMGs (score 1–3) distilled into five functional categories (carbon utilization, energy, organic nitrogen, transporters, and miscellaneous) from 31 vOTUs (Table S5); 21 of the vOTUs were active and carried 63 AMGs (Fig. 2). To identify temperate viruses, we searched the annotations for genes used in the lysogenic infection cycle and predicted nearly half (43%) of our vOTUs were temperate viruses. More than half (59%) of these temperate viruses were active, and the majority (80%) had at least one member of their population integrated at the time of sampling (provirus; Table S6).

Significantly more vOTUs were observed in the SIP fractions from heavy-water incubations relative to control (natural abundance water) incubations, confirming that many viruses incorporated ^{18}O into their DNA and were actively replicating (Fig. S3). A majority (73%) of the vOTUs found in the dense ^{18}O SIP fractions were active at some point over the 370-d incubation, with about half active the full year and the other half only active in 184 d or 370 d samples (Fig. 1B). We measured active vOTU diversity to assess changes in the viral community structure from midyear to a full year of incubation. The Shannon’s H metric indicates a significantly ($p \leq 0.05$) more diverse viral community at 370 d compared to 184 d (Fig. 1C). Shannon’s H diversity, which includes both richness and evenness, was driven by an average

increase of more than 60% for vOTU richness from 184 to 370 d. Of these, 64 vOTUs became more abundant from 184 to 370 d, 110 became newly active, and 18 were no longer detected as labeled at 370 d.

Host characterization based on SIP-metagenomics

To identify potential viral hosts, we used a suite of metagenome assembly and binning methods which yielded 153 medium to high quality MAGs, spanning 16 bacterial phyla and 1 archaeal phylum (Table S7; GTDB taxonomy). The dominant phyla detected were Chloroflexota, Acidobacteria, and Myxococcota (formally Proteobacteria). Incubation with heavy water indicated 30% (46) of these MAGs were actively growing, spanning 9 bacterial phyla, with the most abundant active populations from the Acidobacteria, Bacteroidetes, and Firmicutes. By sequencing both unlabeled and isotopically labeled samples, we gained insight into genetic repertoire from both active and inactive bacteria (Table S8) but focused our efforts on characterizing the active bacterial lineages and metabolisms that viruses may be altering over winter months. We used CO_2 production measurements to confirm microbial activity and positive fluxes occurred continuously throughout the □1.5°C incubation, but not from the □20°C control incubation (Fig. 3).

Linking vOTUs to MAG hosts

To understand the influence of the viruses on microbial dynamics during the winter season, we predicted potential microbial hosts via two different *in silico* approaches based on sequence similarity. First, we identified 11 879 CRISPR spacers in 136 of the MAGs and used these to link 10 vOTUs to 6 MAGs from 4 bacterial phyla (Fig. 4, Table S9). Leveraging the SIP activity patterns, we found most of these identified linkages connected active vOTUs and active MAGs (12 linkages between 8 active vOTUs and 5 active MAGs, 1 linkage between an

unlabeled vOTU and an active MAG, and 3 linkages between unlabeled vOTUs and unlabeled MAGs). In a second approach, we used vOTU-MAG genome sequence similarity and recovered 798 similarity matches that linked 141 vOTUs to 65 MAGs from 10 bacterial phyla (Fig. 4; Table S10). Combined, the two virus-host linkage approaches indicated 318 unique connections between 145 vOTUs and 65 MAGs spanning 10 bacterial phyla. A significantly higher proportion of these vOTU-host matches were between active populations (Table 1). Notably, four vOTUs (#s 153, 161, 162, and 270) had a broad-host range, and were linked to bacteria from several bacterial phyla (three of these vOTUs infected two phyla and one infecting four phyla; Fig. 4). Two of these broad-host-range vOTUs (153 and 270) were active and linked to only active MAGs from bacterial phyla Bacteriodota and Firmicutes.

All 145 of the vOTUs we identified as host-linked may therefore be classified as dsDNA bacteriophages (since the vOTUs were linked to MAGs with a bacterium taxonomic assignment). These represented the majority (88%) of the host-virus matches, and almost all (92%) of the unlabeled vOTUs that were linked to unlabeled MAGs. In our soil incubations, Actinobacteriota (56%), Chloroflexota (24%), and Firmicutes (12%) were the most ‘infected’ bacterial phyla (i.e., with the most vOTU-MAG linkages). Firmicutes (55%), Bacteroidota (34%), Patescibacteria (9%), and Proteobacteria (2%) were the only phyla that had active MAGs linked to active vOTUs. Of the active populations, 81 (33%) vOTUs and 33 (51%) MAGs were linked, with the top 15% most abundant active vOTUs predicted to infect Firmicutes and Bacteroidota, and the abundances of both these vOTUs and their host populations increased over the year incubation (Fig. 5).

Discussion

Quantifying and predicting the drivers of C loss in northern latitude peatlands underlain with permafrost is a complex and critical issue, especially as these environments continue to be disproportionately impacted by climate change. Microbes largely govern C release from Arctic soils, and recent work has highlighted the important role viruses could play in microbial C processing and helping their hosts adapt to subzero temperatures [25, 26]. What is not well understood is what proportion of these viruses are active, how the viral community dynamics change over time, and how long viruses persist in frozen anoxic peat soils. To our knowledge, this is the first study to address these concerns, using heavy water SIP-metagenomics to directly identify active microbes and their associated viruses and characterize their dynamics from a half of a year to a full year under subzero and anoxic conditions.

Viral activity in cryoecosystems

Over the course of a year, we found hundreds of active vOTUs, in stark contrast to the general paradigm that frozen soils have little to no activity [75]. A third of these active vOTUs were linked to active MAGs, highlighting that not only are microbes active in these anoxic subzero temperature soils, but their viruses are also active and likely impacting soil microbial community assembly and ecosystem biogeochemistry, as previously proposed [25, 26, 76, 77]. We observed an increase in vOTU richness and abundance from midyear to a full year of incubation, although vOTU evenness remained the same, likely due to a combination of proviruses replicating with active microbes and many generations of virions actively infecting newly active host(s). The few vOTUs that decreased in abundance or were present midyear and not detected after a full year, could have been isotopically enriched then some persisted in the soil as environmental DNA (eDNA), or as inert virions from internal nucleic acid errors (e.g., lethal mutations) or structural damage, or as competent virions that were no longer infective due

to their host evolving resistance (e.g., the ‘host-virus arms race’ that has been observed in other systems)[78].

While most viruses are host-specific, having a broad host range can influence a virus’ ecological significance. Generalist viruses that can infect more than one species of host are thought to have different abundance patterns, infection efficiency, movement within an ecosystem, and other unique attributes [79]. We identified fifteen generalists and four highly promiscuous generalist vOTUs that were predicted to infect hosts from different phyla. In addition, these four generalists were some of the most abundant vOTUs, suggesting that having a very broad host range may offer an advantage in these subzero and anoxic soils. Recently, Malki et al. [80] showed that four viruses from an oligotrophic lake in Michigan could infect bacteria across several phyla including Proteobacteria, Actinobacteria, and Bacteroidetes. Interestingly, the host phyla identified by Malki et al. are the same bacterial phyla that our generalists vOTUs infected, with the addition of Firmicutes and Chloroflexota (also predicted by our network analyses, see Table S11; Fig. S4).

Putative viral influence on winter biogeochemistry

Soil viruses shape the abundance, diversity, and metabolic outputs of their microbial hosts. As recent literature has made clear, viruses may modulate ecosystem biogeochemical processes via an intriguing mechanism—reprogramming host metabolism with AMGs during infection [24]. The metabolic functions and ubiquity of AMGs can vary by environment, with marine viruses carrying AMGs for photosynthesis, central carbon metabolism, sulfur cycling, and nitrogen cycling [24]. Soil viruses, although understudied, appear to carry AMGs for polysaccharide degradation and sporulation [25, 26, 81]. In our cryoecosystem, we identified carbon degradation AMGs (e.g., galactose degradation and Xylan 1,4-beta-xylosidase) that could

provide fitness advantages for utilizing much-needed carbon sources, as well as genes for scavenging (e.g., ABC-2 type transport system) that may provide their host an avenue to acquire essential nutrients [82]. Additionally, AMGs such as methionine degradation [83] and Glycosyl transferases [84] were found in high abundance which can help during infection.

To better understand the roles viruses play in controlling central biogeochemical processes in Arctic peat under winter conditions, we assessed the active bacterial phyla that were virus-infected. Firmicutes had the largest increase in abundance from midyear to a full year and had the most viral infections (both the number of linked active vOTUs and increase in vOTU abundance), supporting the 'kill-the-winner' theory described in aquatic systems (Fig. 3; Fig. 4) [85]. In this Lotka–Volterra predator-prey framework, the most abundant host would become infected by viruses, leading to a decline in host abundance and an increase in the prevalence of their viruses, at least until microbial hosts with genetic resistance to the currently dominant viruses had time to respond. The dominance of Firmicutes hosts makes sense, since this diverse bacterial phylum is known for adaptations to anoxic conditions, including fermentation and creating endospores [86]. In our samples, all active infected Firmicutes contained fermentation genes for the capacity to produce ethanol, lactate, or both (Table S8). Bacteroidota were also an active and frequently infected bacterial phylum that increased in abundance through time; of this group, all the infected MAGs shared taxonomic affiliation to the order Bacteroidales. Most of these bacteria are obligate anaerobes and known for their diverse arrays of carbohydrate-active enzymes arranged into polysaccharide utilization loci and fermentation [87, 88]. Many active MAGs within this group had the genomic capacity for polysaccharide degradation (e.g., ICE 1 encoded pectin degradation protein KdgF) and all encoded genes for lactate fermentation (Table S8). After one year of subzero and anoxic incubation, Bacteroidales had become the most

abundant active lineage in our soils and were infected by the most abundant active vOTUs (#s 23, 37, 41, 124, and 190), further supporting 'kill-the-winner' theory and highlighting viruses influencing host winter activities. While the abundance of active Bacteroidota and Firmicutes populations were increasing, C mineralization to CO_2 occurred steadily throughout the incubation and vOTUs linked to these populations carried several AMGs that would support central carbon metabolism (Table S5). Together, these results suggest that these two bacterial lineages play an important role in over-winter C loss from these Artic peat soils, and the viruses that infect them likely shape both their population dynamics and functional impact.

Another way viruses can impact soil biogeochemistry is by infecting hosts that occupy different metabolic niche spaces. The Patescibacteria we identified (part of the Candidate phyla radiation; CPR) are known for their ultrasmall cell size with reduced genomes and most have a symbiotic relationship with Bacteroidota [89], suggesting direct interactions between these two lineages in our cryoecosystem. Patescibacteria are thought to resist phage infection by lacking or reducing the number of potential phage receptors on their cell membrane [90], but in our study, it was notably one of the most infected lineages and had none of the previously reported phage receptors (Table S8). The prevalence of these infections may be the result of their interactions with Bacteroidota. One member of the Patescibacteria was linked to an unlabeled vOTU with a CRISPR spacer match, suggesting this adapted immune system element was successful at protecting the host from infection. Typical CRISPRs are rarely found in CPR bacteria and recent work suggests this may be due to this group using a compact CRISPR-CasY system that is highly divergent to typical CRISPR systems [91]. Another infected and active bacterial phylum we observed was Proteobacteria, with only one active MAG from the class Micavibrionales. Little is known about this clade, and even less about their viruses, because they have an epibiotic

lifestyle where they feed on other organisms to survive, making them difficult to culture. These predatory bacteria may be active under winter conditions by either consuming non-active or dead cells, or they may benefit from attaching to a host that can utilize alternative energy sources and recalcitrant organic matter [92]. We identified multiple active vOTUs infecting this lineage, revealing another indirect way that viruses may impact nutrient cycling and microbial diversity, via preying upon bacterial predators.

Challenges associated with identifying virus 'activity'

Heavy-water SIP has proven to be a robust method for identifying metabolically active microbes in soils [32, 33, 36]. In many ways, the approach is superior to other techniques that track activity such as RNA-to-DNA ratios, rRNA levels, bioorthogonal non-nanonical amino acid tagging (BONCAT) or other stable isotopes (e.g., ^{13}C) because active microbes synthesize DNA when their cells divide, incorporating ^{18}O, and therefore the DNA of all active microbes is labeled. Data from RNA studies can be hard to interpret as RNA levels often do not correlate with growth and often have weak or no correlation with proteins levels [93-95]. Compared to other isotopes as tracers, ^{18}O labeling via heavy-water SIP does not rely on the microbe's carbon use efficiency or prior knowledge of the microbe's substrate preference [37].

The application of isotope tracers for direct assessment of activity is not as straightforward for viruses as it is for other microorganisms, and worthy of reasoned consideration. Characterizing activity for viruses is quite different compared to 'free-living' organisms because of their different infection cycles, their lack of metabolism, and the many states in which they can be present [19]. One of the main reasons to identify active entities is to help quantify their impact on their hosts and their environment, with the assumption that inactive entities have less of an impact but are still important [96]. For a microbe, this may be true, but

for a virus there is a range in magnitude of impact for different host metabolic states that depends heavily on the infection cycle. Assessing the prevalence of an infection cycle, however, is challenging due to difficulties in quantifying lysogeny or virion abundance, burst size, diversity, and ecology [51].

Arctic soils, that predominantly exist under anoxic and subzero temperature conditions, might be generally considered a harsh environment, and would limit microbial growth. For this reason, we hypothesized most of our viruses would be temperate [51, 97] and they would be detected as proviruses. We did also see a high rate of lysogeny (43%), and in support of our hypotheses, the majority (80%) of temperate viruses were present as proviruses. Further, half of the active vOTUs linked to active MAGs (i.e., bacterial phyla Bacteroidota, Firmicutes, Patescibacteria, and Proteobacteria) were temperate viruses. The number of temperate viruses present in this study was higher than previously reported for desert Antarctic soils (4–20%) and almost identical (44%) to whole soil assays from temperate wetlands soils [98]. We hypothesize the increased incidence of temperate viruses is linked to low host abundances and environmental conditions, therefore increasing the potential for sporadic viral infection.

Temperate viruses can undergo lytic infection, where activity is identified by progeny viruses, or lysogenic infection, where activity is difficult to assess. A temperate virus that is undergoing lysogenic infection (present as a provirus) during a heavy-water SIP incubation would become isotopically enriched and depend on its host's division rate for abundance. Active proviruses undergoing lysogenic infection need to be distinguished from viruses undergoing lytic infection because the effect of proviruses on the host metabolism (and therefore ecosystem) will not be as pronounced. This is because proviruses do not shut down and redirect host metabolism for progeny production during the lysogenic cycle as compared to the lytic cycle. During

lysogeny, viruses still impact host metabolism via host gene regulation and acquisition of new virulence factors, but it is primarily for maintaining the provirus in the host genome [99]. Proviruses may also be labeled, but not currently active if they are proviral remnants of a past infection [100]. These remnants have no negative impact on host metabolism (beyond occupying genome space) but may confer some advantage as a gene transfer agent [51] or by contributing virulence factors which can still impact host physiology and metabolism [99]. SIP-enabled metagenomics alone cannot unequivocally identify a virus' state (e.g., in maintenance mode or remnant), making it currently difficult to fully assess viral activity. Combining SIP-metagenomics with other approaches—such as a SIP-virome or induction assays—may identify vOTUs that have undergone lytic infection, and therefore provide a more holistic view of vOTU dynamics.

Conclusions

Winter carbon losses in Arctic peat soils are estimated to be significant and growing, but the mechanisms that drive these losses are poorly understood. Using an ^{18}O-heavy-water incubation under subzero, anoxic conditions, we found that bacteria and their viruses are active over the long winter months in northern peatlands, and these active populations drive significant CO_2 fluxes. Our approach, SIP-targeted metagenomics, allowed us to move beyond a general catalogue of the genetic repertoire of these soil communities, and expose the specific population-level dynamics and functional capacities of the activity viral community. Given the high abundance of unlabeled bacteria and their viruses, these bacterial and viral populations would have been traditionally described first and foremost, thus occluding the characterization of the active bacteria and viruses in these soils. Despite the subzero temperatures and lack of oxygen in

these peat soils, many active hosts and active viruses (both temperate and lytic) appear to be engaged in a surprisingly high level of biotic interactions and biogeochemical processing.

Acknowledgements

We would like to thank the Millard lab for curating a list of Genbank bacteriophage genomes used for taxonomic identification of our vOTUs. Thanks to Jack McFarland and Monica Haw (USGS) for help and support with the experimental setup and maintenance. Sample collection and processing was supported by a US Geological Survey Mendenhall Fellowship to S.J.B., the Bonanza Creek LTER Program, jointly funded by NSF (DEB 1026415) and the USDA Forest Service Pacific Northwest Research Station (PNW01-JV112619320-16), and support from the USGS Climate R&D Program and AK Climate Science Center. Analysis of results was supported by a Lawrence Livermore National Laboratory, Laboratory Directed Research & Development grant (18-ERD-041) to S.J.B. and by LLNL's U.S. Department of Energy, Office of Biological and Environmental Research, Genomic Science Program LLNL 'Microbes Persist' Scientific Focus Area (#SCW1632). Work conducted at PNNL, operated for the DOE by Battelle Memorial Institute, was conducted under Contract DE-AC05-76RLO1830. Work conducted at LLNL was conducted under the auspices of the US Department of Energy under Contract DE-AC52-07NA27344.

1. Nichols, J.E. and D.M. Peteet, *Rapid expansion of northern peatlands and doubled estimate of carbon storage.* Nature Geoscience, 2019. 12(11): p. 917-921.
2. Koven, C.D., et al., *A simplified, data-constrained approach to estimate the permafrost carbon-climate feedback.* Philos Trans A Math Phys Eng Sci, 2015. 373(2054).
3. McGuire, A.D., et al., *Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change.* Proc Natl Acad Sci U S A, 2018. 115(15): p. 3882-3887.
4. Schuur, E.A., et al., *Climate change and the permafrost carbon feedback.* Nature, 2015. 520(7546): p. 171-9.
5. Hodgkins, S.B., et al., *Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production.* Proc Natl Acad Sci U S A, 2014. 111(16): p. 5819-24.
6. Hopple, A.M., et al., *Massive peatland carbon banks vulnerable to rising temperatures.* Nat Commun, 2020. 11(1): p. 2373.

7. Parazoo, N.C., et al., *Detecting regional patterns of changing CO2 flux in Alaska.* Proc Natl Acad Sci U S A, 2016. 113(28): p. 7733-8.

8. Wilson, R.M., et al., *Stability of peatland carbon to rising temperatures.* Nat Commun, 2016. 7: p. 13723.

9. Natali, S.M., et al., *Large loss of CO2 in winter observed across the northern permafrost region.* Nature Climate Change, 2019. 9(11): p. 852-857.

10. Przybylak, R., *Diurnal temperature range in the Arctic and its relation to hemispheric and Arctic circulation patterns.* International Journal of Climatology, 2000. 20(3 %@ 0899-8418): p. 231-253.

11. Bolduc, B., et al., *The IsoGenie database: an interdisciplinary data management solution for ecosystems biology and environmental research.* PeerJ, 2020. 8: p. e9467.

12. Männistö, M.K., et al., *Acidobacteriadominate the active bacterial communities of Arctic tundra with widely divergent winter-time snow accumulation and soil temperatures.* FEMS Microbiology Ecology, 2013. 84(1): p. 47-59.

13. Murray, A.E., et al., *Microbial life at -13 C in the brine of an ice-sealed Antarctic lake.* Proceedings of the National Academy of Sciences, 2012. 109(50): p. 20626-20631.

14. Drotz, S.H., et al., *Both catabolic and anabolic heterotrophic microbial activity proceed in frozen soils.* Proceedings of the National Academy of Sciences, 2010. 107(49): p. 21046-21051.

15. Margesin, R. and T. Collins, *Microbial ecology of the cryosphere (glacial and permafrost habitats): current knowledge.* Applied Microbiology and Biotechnology, 2019. 103(6): p. 2537-2549.

16. Brum, J.R. and M.B. Sullivan, *Rising to the challenge: accelerated pace of discovery transforms marine virology.* Nature Reviews Microbiology, 2015. 13(3): p. 147-159.

17. Wilhelm, S.W. and C.A. Suttle, *Viruses and Nutrient Cycles in the Sea.* BioScience, 1999. 49(10): p. 781-788.

18. Suttle, C.A., *Marine viruses — major players in the global ecosystem.* Nature Reviews Microbiology, 2007. 5(10): p. 801-812.

19. Trubl, G., et al., *Coming-of-Age Characterization of Soil Viruses: A User's Guide to Virus Isolation, Detection within Metagenomes, and Viromics.* Soil Systems, 2020. 4(2): p. 23.

20. Braga, L.P.P., et al., *Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios.* Microbiome, 2020. 8(1).

21. Liang, C., J.P. Schimel, and J.D. Jastrow, *The importance of anabolism in microbial control over soil carbon storage.* Nature Microbiology, 2017. 2(8): p. 17105.

22. Rosenwasser, S., et al., *Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean.* Trends in Microbiology, 2016. 24(10): p. 821-832.

23. Thompson, L.R., et al., *Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism.* Proceedings of the National Academy of Sciences, 2011. 108(39): p. E757-E764.

24. Zimmerman, A.E., et al., *Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems.* Nature Reviews Microbiology, 2020. 18(1): p. 21-34.

25. Emerson, J.B., et al., *Host-linked soil viral ecology along a permafrost thaw gradient.* Nature Microbiology, 2018. 3(8): p. 870-880.

26. Trubl, G., et al., *Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing.* mSystems, 2018. 3(5).

27. Barnett, S.E. and D.H. Buckley, *Simulating metagenomic stable isotope probing datasets with MetaSIPSim.* BMC Bioinformatics, 2020. 21(1).

28. Dumont, M.G. and M. Hernández García, *Stable Isotope Probing.* Methods and Protocols Totowa, NJ, US: Humana Press, 2019.

29. Starr, E.P., et al., *Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon.* Microbiome, 2018. 6(1).

30. Haig, S.-J., et al., *Stable-isotope probing and metagenomics reveal predation by protozoa drives E. coli removal in slow sand filters.* The ISME Journal, 2015. 9(4): p. 797-808.

31. Gross, A., et al., *The role of soil redox conditions in microbial phosphorus cycling in humid tropical forests.* Ecology, 2020. 101(2).

32. Aanderud, Z.T. and J.T. Lennon, *Validation of Heavy-Water Stable Isotope Probing for the Characterization of Rapidly Responding Soil Bacteria.* Applied and Environmental Microbiology, 2011. 77(13): p. 4589-4596.

33. Blazewicz, S.J. and E. Schwartz, *Dynamics of 18O Incorporation from H 2 18O into Soil Microbial DNA.* Microbial Ecology, 2011. 61(4): p. 911-916.

34. Schwartz, E., *Characterization of Growing Microorganisms in Soil by Stable Isotope Probing with H218O.* Applied and Environmental Microbiology, 2007. 73(8): p. 2541-2546.

585 35. Blazewicz, S.J., et al., *Taxon-specific microbial growth and mortality patterns reveal*
586 *distinct temporal population responses to rewetting in a California grassland soil.* The
587 ISME Journal, 2020. 14(6): p. 1520-1532.
588 36. Koch, B.J., et al., *Estimating taxon-specific population dynamics in diverse microbial*
589 *communities.* Ecosphere, 2018. 9(1): p. e02090.
590 37. Papp, K., et al., *Quantitative stable isotope probing with H218O reveals that most*
591 *bacterial taxa in soil synthesize new ribosomal RNA.* The ISME Journal, 2018. 12(12): p.
592 3043-3045.
593 38. Schwartz, E., et al., *Characterization of growing bacterial populations in McMurdo Dry*
594 *Valley soils through stable isotope probing with18O-water.* FEMS Microbiology Ecology,
595 2014. 89(2): p. 415-425.
596 39. Waldrop, M.P., et al., *Carbon Fluxes and Microbial Activities from Boreal Peatlands*
597 *Experiencing Permafrost Thaw.* Journal of Geophysical Research: Biogeosciences, 2021.
598 40. Russell, D.W. and J. Sambrook, *Molecular cloning: a laboratory manual*. Vol. 1. 2001:
599 Cold Spring Harbor Laboratory Cold Spring Harbor, NY.
600 41. Bankevich, A., et al., *SPAdes: A New Genome Assembly Algorithm and Its Applications to*
601 *Single-Cell Sequencing.* Journal of Computational Biology, 2012. 19(5): p. 455-477.
602 42. Roux, S., et al., *VirSorter: mining viral signal from microbial genomic data.* PeerJ, 2015.
603 3: p. e985.
604 43. Ren, J., et al., *Identifying viruses from metagenomic data using deep learning.*
605 Quantitative Biology, 2020. 8(1): p. 64-77.

606 44. Roux, S., et al., *Minimum Information about an Uncultivated Virus Genome (MIUViG).*
607 Nature Biotechnology, 2019. 37(1): p. 29-37.
608 45. Delcher, A.L., S.L. Salzberg, and A.M. Phillippy, *Using MUMmer to identify similar regions*
609 *in large sequence sets.* Current protocols in bioinformatics, 2003(1): p. 10.3. 1-10.3. 18.
610 46. Nayfach, S., et al., *CheckV assesses the quality and completeness of metagenome-*
611 *assembled viral genomes.* Nature Biotechnology, 2020.
612 47. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature
613 Methods, 2012. 9(4): p. 357-359.
614 48. Roux, S., et al., *Benchmarking viromics: an in silico evaluation of metagenome-enabled*
615 *estimates of viral community composition and diversity.* PeerJ, 2017. 5: p. e3817.
616 49. Ecale Zhou, C.L., et al., *multiPhATE: bioinformatics pipeline for functional annotation of*
617 *phage isolates.* Bioinformatics, 2019. 35(21): p. 4402-4404.
618 50. Shaffer, M., et al., *DRAM for distilling microbial metabolism to automate the curation of*
619 *microbiome function.* Nucleic Acids Research, 2020. 48(16): p. 8883-8900.
620 51. Howard-Varona, C., et al., *Lysogeny in nature: mechanisms, impact and ecology of*
621 *temperate phages.* The ISME Journal, 2017. 11(7): p. 1511-1520.
622 52. Arndt, D., et al., *PHASTER: a better, faster version of the PHAST phage search tool.*
623 Nucleic Acids Research, 2016. 44(W1): p. W16-W21.
624 53. Li, D., et al., *MEGAHIT: an ultra-fast single-node solution for large and complex*
625 *metagenomics assembly via succinct de Bruijn graph.* Bioinformatics, 2015. 31(10): p.
626 1674-1676.

54. Alneberg, J., et al., *Binning metagenomic contigs by coverage and composition.* Nature Methods, 2014. 11(11): p. 1144-1146.

55. Wu, Y.-W., B.A. Simmons, and S.W. Singer, *MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets.* Bioinformatics, 2016. 32(4): p. 605-607.

56. Kang, D.D., et al., *MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities.* PeerJ, 2015. 3: p. e1165.

57. Uritskiy, G.V., J. Diruggiero, and J. Taylor, *MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis.* Microbiome, 2018. 6(1).

58. Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.* Genome Research, 2015. 25(7): p. 1043-1055.

59. Bowers, R.M., et al., *Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.* Nature Biotechnology, 2017. 35(8): p. 725-731.

60. Bushnell, B., *BBDuk: Adapter.* Quality Trimming and Filtering. httpssourceforgenetprojectsbbmap.

61. Olm, M.R., et al., *dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication.* The ISME Journal, 2017. 11(12): p. 2864-2868.

62. Chaumeil, P.-A., et al., *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.* Bioinformatics, 2019.

648 63. Wattam, A.R., et al., *Improvements to PATRIC, the all-bacterial Bioinformatics Database*
649 *and Analysis Resource Center.* Nucleic Acids Research, 2017. 45(D1): p. D535-D542.
650 64. Brettin, T., et al., *RASTtk: A modular and extensible implementation of the RAST*
651 *algorithm for building custom annotation pipelines and annotating batches of genomes.*
652 Scientific Reports, 2015. 5(1): p. 8365.
653 65. Aramaki, T., et al., *KofamKOALA: KEGG Ortholog assignment based on profile HMM and*
654 *adaptive score threshold.* Bioinformatics, 2020. 36(7): p. 2251-2252.
655 66. Kanehisa, M., et al., *KEGG as a reference resource for gene and protein annotation.*
656 Nucleic Acids Research, 2016. 44(D1): p. D457-D462.
657 67. Bushnell, B., *BBMap: a fast, accurate, splice-aware aligner*. 2014, Lawrence Berkeley
658 National Lab.(LBNL), Berkeley, CA (United States).
659 68. Kumar, M.S., et al., *Analysis and correction of compositional bias in sparse sequencing*
660 *count data.* BMC Genomics, 2018. 19(1).
661 69. Paulson, J.N., et al., *Differential abundance analysis for microbial marker-gene surveys.*
662 Nature Methods, 2013. 10(12): p. 1200-1202.
663 70. Andersson, A.F. and J.F. Banfield, *Virus Population Dynamics and Acquired Virus*
664 *Resistance in Natural Microbial Communities.* Science, 2008. 320(5879): p. 1047-1050.
665 71. Bland, C., et al., *CRISPR Recognition Tool (CRT): a tool for automatic detection of*
666 *clustered regularly interspaced palindromic repeats.* BMC Bioinformatics, 2007. 8(1): p.
667 209.
668 72. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Biology,
669 1990. 215(3): p. 403-410.

670 73. Edwards, R.A., et al., *Computational approaches to predict bacteriophage–host*
671 *relationships.* FEMS Microbiology Reviews, 2016. 40(2): p. 258-272.
672 74. Al-Shayeb, B., et al., *Clades of huge phages from across Earth's ecosystems.* Nature,
673 2020. 578(7795): p. 425-431.
674 75. Nikrad, M.P., L.J. Kerkhof, and M.M. Häggblom, *The subzero microbiome: microbial*
675 *activity in frozen and thawing soils.* FEMS Microbiology Ecology, 2016. 92(6): p. fiw081.
676 76. Adriaenssens, E.M., et al., *Environmental drivers of viral community composition in*
677 *Antarctic soils identified by viromics.* Microbiome, 2017. 5(1).
678 77. Zablocki, O., E.M. Adriaenssens, and D. Cowan, *Diversity and Ecology of Viruses in*
679 *Hyperarid Desert Soils.* Applied and Environmental Microbiology, 2016. 82(3): p. 770-
680 777.
681 78. Breitbart, M., et al., *Phage puppet masters of the marine microbial realm.* Nature
682 Microbiology, 2018. 3(7): p. 754-766.
683 79. Fermin, G., *Host Range, Host–Virus Interactions, and Virus Transmission.* Viruses, 2018:
684 p. 101.
685 80. Malki, K., et al., *Bacteriophages isolated from Lake Michigan demonstrate broad host-*
686 *range across several bacterial phyla.* Virology Journal, 2015. 12(1).
687 81. Van Goethem, M.W., et al., *Characteristics of Wetting-Induced Bacteriophage Blooms in*
688 *Biological Soil Crust.* mBio, 2019. 10(6).
689 82. Greiner, T., et al., *Genes for membrane transport proteins: Not so rare in viruses.*
690 Viruses, 2018. 10(9): p. 456.

691 83. Ankrah, N.Y.D., et al., *Phage infection of an environmentally relevant marine bacterium*
692 *alters host metabolism and lysate composition.* The ISME Journal, 2014. 8(5): p. 1089-
693 1100.
694 84. Markine-Goriaynoff, N., et al., *Glycosyltransferases encoded by viruses.* Journal of
695 General Virology, 2004. 85(10): p. 2741-2754.
696 85. Thingstad, T.F., *Elements of a theory for the mechanisms controlling abundance,*
697 *diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems.* Limnology
698 and Oceanography, 2000. 45(6): p. 1320-1328.
699 86. Zeibich, L., O. Schmidt, and H.L. Drake, *Dietary polysaccharides: fermentation potentials*
700 *of a primitive gut ecosystem.* Environmental Microbiology, 2019. 21(4): p. 1436-1451.
701 87. Lapébie, P., et al., *Bacteroidetes use thousands of enzyme combinations to break down*
702 *glycans.* Nature Communications, 2019. 10(1).
703 88. Larsbrink, J. and L.S. McKee, *Bacteroidetes bacteria in the soil: Glycan acquisition,*
704 *enzyme secretion, and gliding motility*. Advances in applied microbiology, 2020. 110: p.
705 63-98.
706 89. Lemos, L.N., et al., *Genomic signatures and co-occurrence patterns of the ultra-small*
707 *Saccharimonadia (phylum CPR/Patescibacteria) suggest a symbiotic lifestyle.* Molecular
708 Ecology, 2019. 28(18): p. 4259-4271.
709 90. Tian, R., et al., *Small and mighty: adaptation of superphylum Patescibacteria to*
710 *groundwater environment drives their genome simplicity.* Microbiome, 2020. 8(1).

91. Chen, L.-X., et al., *Candidate Phyla Radiation Roizmanbacteria From Hot Springs Have Novel and Unexpectedly Abundant CRISPR-Cas Systems.* Frontiers in Microbiology, 2019. 10.

92. Coutinho, F.H., et al., *Ecogenomics and metabolic potential of the South Atlantic Ocean microbiome.* Science of The Total Environment, 2021. 765: p. 142758.

93. Bashiardes, S., G. Zilberman-Schapira, and E. Elinav, *Use of Metatranscriptomics in Microbiome Research.* Bioinformatics and Biology Insights, 2016. 10: p. BBI.S34610.

94. Blazewicz, S.J., et al., *Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses.* The ISME Journal, 2013. 7(11): p. 2061-2068.

95. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nature Reviews Genetics, 2012. 13(4): p. 227-232.

96. Joergensen, R.G. and F. Wichern, *Alive and kicking: Why dormant soil microorganisms matter.* Soil Biology and Biochemistry, 2018. 116: p. 419-430.

97. Stewart, F.M. and B.R. Levin, *The population biology of bacterial viruses: Why be temperate.* Theoretical Population Biology, 1984. 26(1): p. 93-117.

98. Williamson, K.E., et al., *Incidence of lysogeny within temperate and extreme soil environments.* Environmental Microbiology, 2007. 9(10): p. 2563-2574.

99. Hernández, S. and M.J. Vives, *Phages in Anaerobic Systems.* Viruses, 2020. 12(10): p. 1091.

100. Casjens, S., *Prophages and bacterial genomics: what have we learned so far?* Molecular Microbiology, 2003. 49(2): p. 277-300.

Table 1. Linkages between ^{18}O labeled and unlabeled viruses and bacteria in a SIP incubation of sub-zero, anoxic Arctic peat soils. Viral operational taxonomic units (vOTUs) were linked to metagenome assembled genomes (MAGs) via nucleotide similarity and using CRISPR spacers. 'Active' vOTUs and MAGs were defined based on assimilation of ^{18}O enriched water (heavy water) into DNA. In total, there were 814 linkages from active vOTUs to active MAGs (group 1), active vOTUs to unlabeled MAGs (group 2), unlabeled vOTUs to active MAGs (group 3) and unlabeled vOTUs to unlabeled MAGs (group 4).

Group	Description	No. of vOTU-MAG matches	Unique vOTU-MAG matches	vOTUs	MAGs
1	Active vOTU-active MAG	164	107	74	37
2	Active vOTU-unlabeled MAG	2	2	2	2
3	Unlabeled vOTU-active MAG	226	77	44	11
4	Unlabeled vOTU-unlabeled MAG	422	132	58	30
	Total	**814**	**318**	**178**	**80**

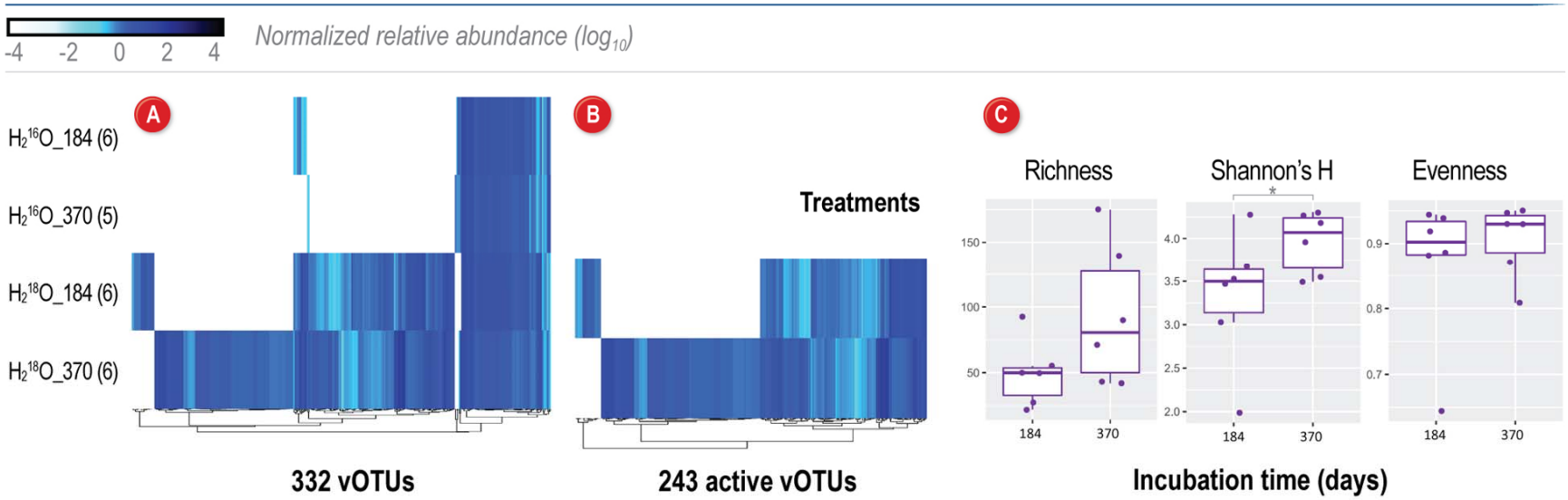

Figure 1. Assessment of viral community structure and activity after a ^{18}O water incubation in Arctic peat soils. (A) Viral sequences were identified from 23 samples grouped by two treatments (natural abundance water "$H_2{}^{16}O$", and heavy water "$H_2{}^{18}O$") and two time points: 184 days and 370 days. Number of replicates is indicated in parentheses. Relative abundances of all 332 vOTUs identified in the peat soils, clustered by abundances in each treatment/timepoint. (B) Relative abundances of 243 vOTUs considered 'active' due to DNA ^{18}O enrichment patterns. Relative abundance for each vOTU (illustrated by blue gradient) was normalized by metagenome size (total base pairs) and contig length, and reads were mapped to the contig if they shared ≥90% average nucleotide identity and had ≥85% alignment fraction. (C) Diversity metrics for the 243 active vOTUs. Box plots show the median, upper and lower quartile range, and the variance among the samples. Asterix denotes significance ($p < 0.05$).

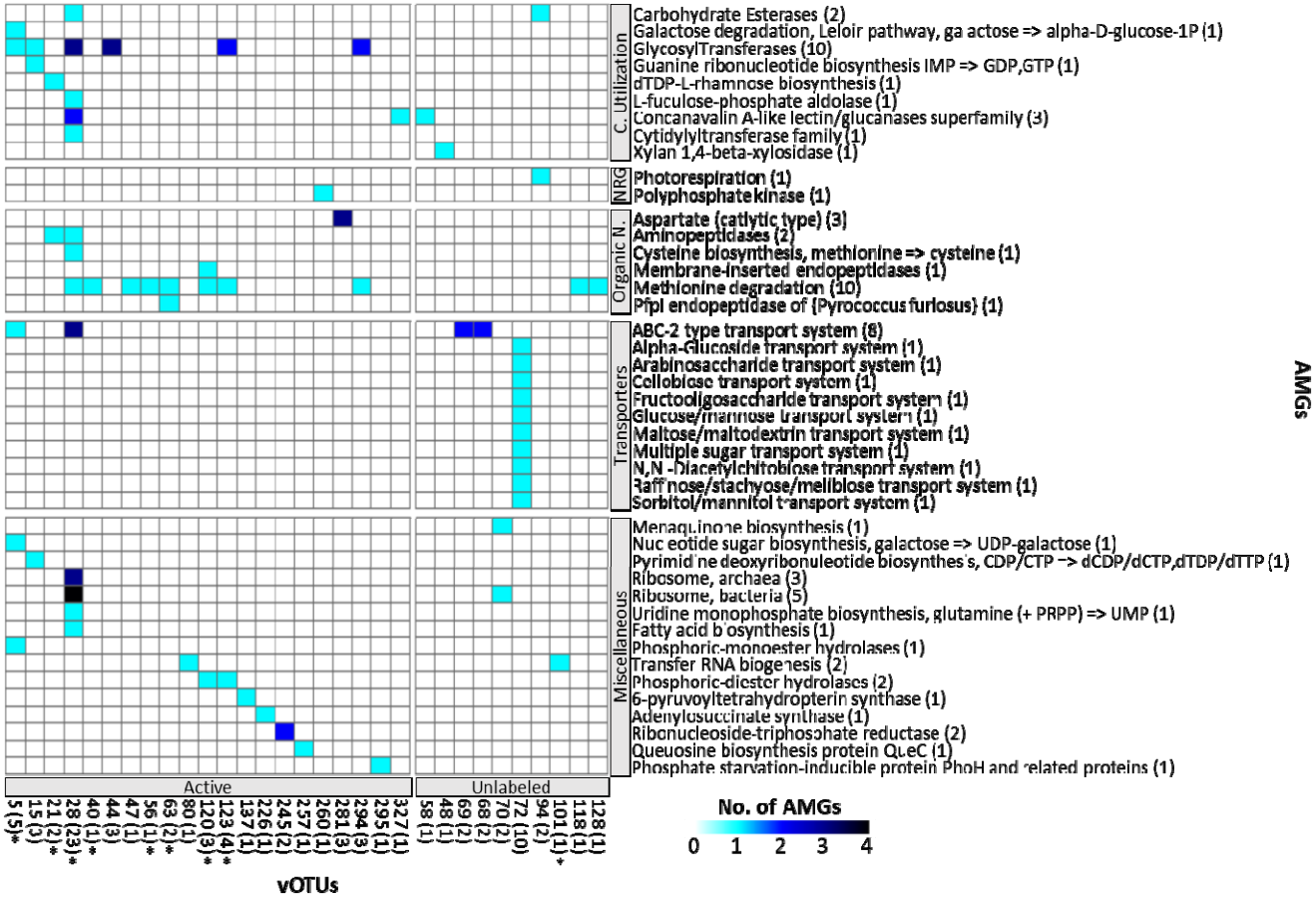

Figure 2. Putative auxiliary metabolic genes associated with peat soil viral genomes. Heatmap of 31 vOTUs carrying confidently predicted auxiliary metabolic genes (AMGs) (Dram-v score 1–3) and their annotation. The vOTUs are grouped by active or unlabeled (see Table S3 for additional detail) with the sum of AMGs per vOTU indicated in parenthesis and those with an asterisk are linked to an active MAG. AMGs are grouped by functional category — carbon utilization, energy generation, organic nitrogen use, transporters, and miscellaneous.

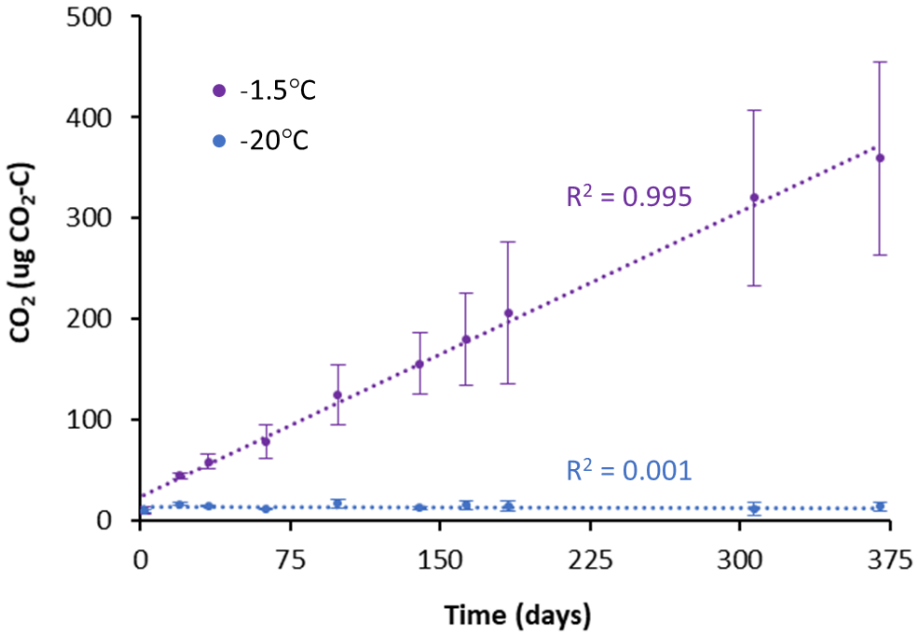

Figure 3. CO_2 production. Cumulative CO_2 production in soil incubated at □1.5°C (experimental conditions) and □20°C (control). Error bars show standard error (n=3) and R^2 is shown for each linear regression.

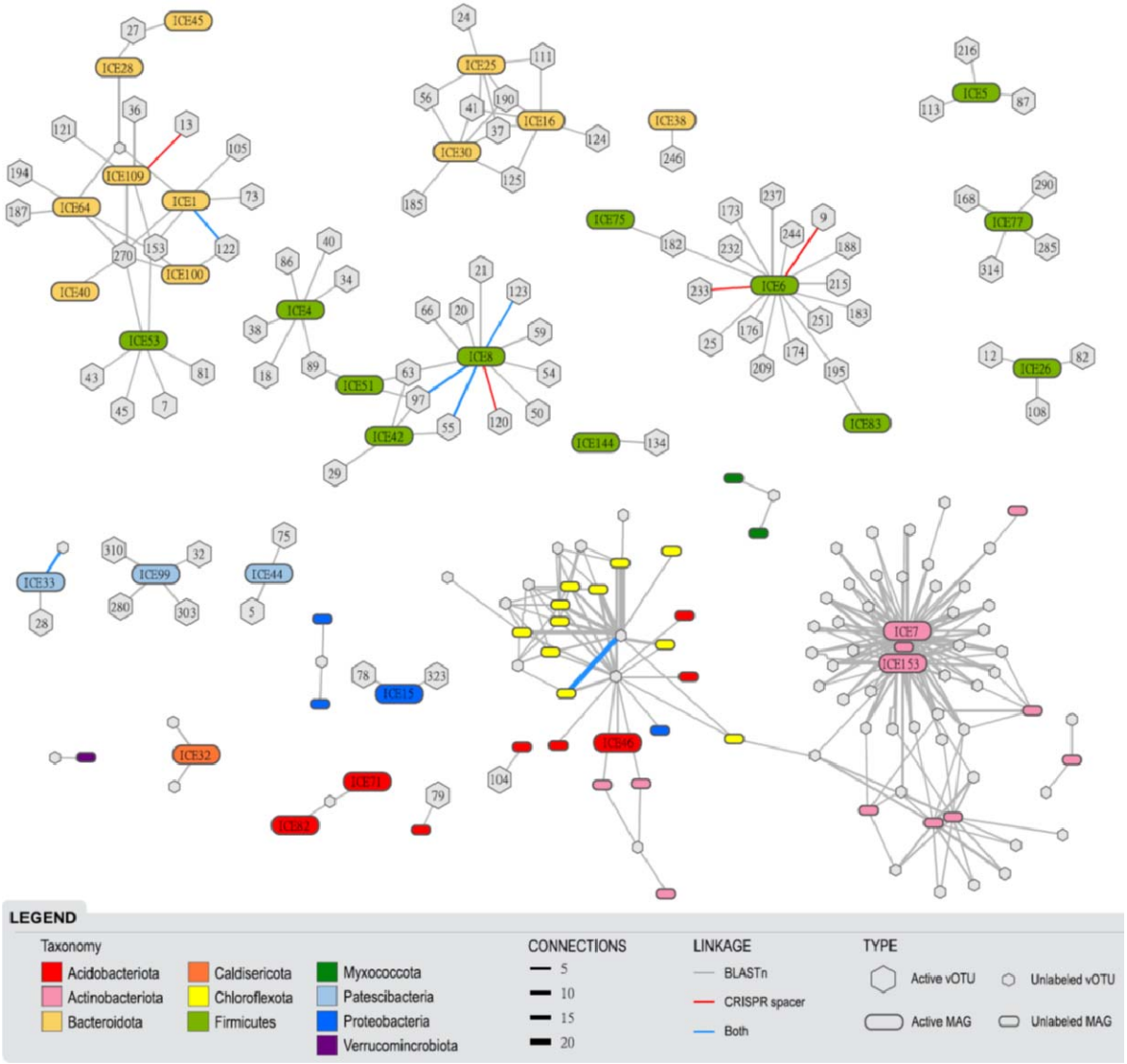

Figure 4. Virus-host linkages in Arctic peat soils incubated with ^{18}O enriched water. Network diagram illustrating vOTUs and their predicted bacterial hosts, colored by bacterial phyla. Active versus inactive vOTUs and MAGs (based on ^{18}O enrichment) are indicated by large named or small hexagon/rounded rectangles, respectively. Lines represent linkages between a vOTU and a bacterial MAG, thickness denotes the number of connections, and are colored by the identification approach used: similarity in genomic content (gray), CRISPR spacer match (red), or both (blue).

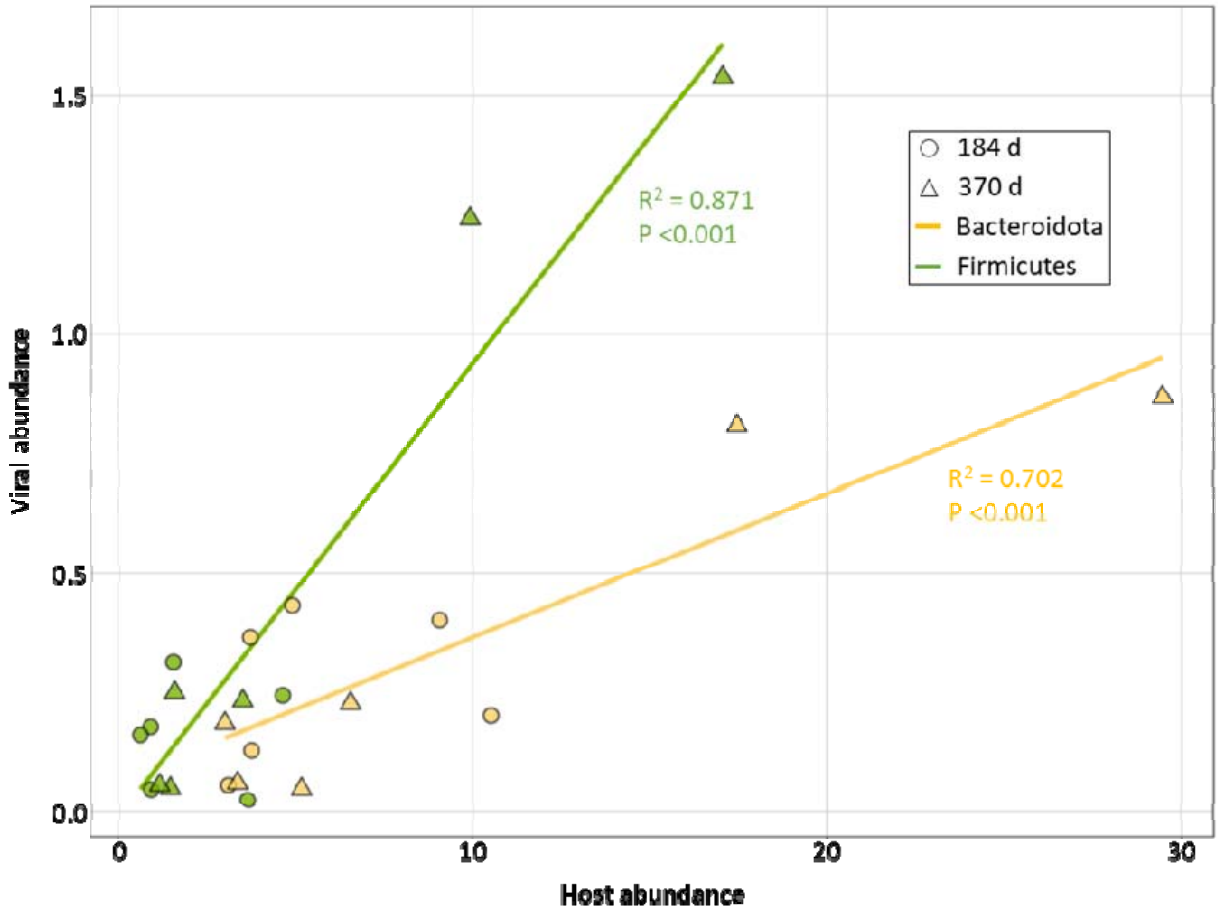

Figure 5. Abundances of active viruses and their predicted bacterial hosts. Average virus:host abundance ratios for bacterial phyla Bacteroidota (yellow) and Firmicutes (green) from heavy-water treatment samples at 184 (n=6) and 370 days (n=6). Host abundance and the abundance of viruses for that host were calculated as the mean coverage depth from metagenomic read mapping, normalized by the number of reads in the sample.