Title: Active viral population dynamics in frozen Arctic peat soil revealed with $H_2{}^{18}O$ stable isotope probing metagenomics

Running title: SIP-metagenomics of arctic peat virus 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, Menlo Park, CA, USA

Corresponding authors:

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

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

Abstract

Winter soil processes are critical to the carbon balance of northern ecosystems, yet the microbial ecology governing biogeochemical cycling in frozen soils is largely unknown. We used stable isotope probing targeted metagenomics to reveal the genomic potential of active microbial populations, with an emphasis on viruses, in soils. Peat soils were incubated under simulated winter conditions (subzero and anoxic) with $H_2{}^{18}O$ or natural abundance water for 184 and 370 days. Isotope incorporation revealed 46 active bacterial populations (MAGs; spanning 9 bacterial phyla) and 243 active viral populations (vOTUs). Active hosts were predicted for 33% of the active vOTUs and were some of the most abundant MAGs, having capacity for fermentation and carbohydrate utilization. Additionally, almost one-third of vOTUs carried auxiliary metabolic genes spanning five functional categories, highlighting the potential impact of viruses in microbial biogeochemistry. CO_2 production throughout the incubation supports our evidence of microbial activities under winter conditions. Our results revealed a multi-trophic and changing microbial community in tandem with a changing viral community targeting dominant active bacteria consistent with the "kill-the-winner" hypothesis. These data have important implications for low-temperature soil processes in northern peatlands and reveal active host-linked soil viral ecology, with potential multifaceted biogeochemical impacts.

Introduction

Northern peatlands are important terrestrial ecosystems for carbon (C) storage, estimated to contain one-third of soil C (~1,000 gigatons; Nichols and Peteet 2019), yet the fate of this C is unknown as these soils experience changes in growing seasons and precipitation. Soil-warming experiments indicate increased carbon dioxide and methane emissions, and it is likely that C loss from northern peatlands is underestimated as virtually all measurements come from the summer months when these soils are not frozen (Hodgkins et al. 2014; Parazoo et al. 2016; Wilson et al. 2016; Hopple et al. 2020). Recent estimates of winter C loss alone are estimated to be greater than the average growing season C uptake (Natali et al. 2019). While the winter months include large temperature fluctuations and extreme temperature minimums (Przybylak 2000), the temperatures found in much of the soil profile for permafrost or seasonally frozen bogs can remain stable and just below the freezing point (-1 to -5°C) (Männistö et al. 2013; Hodgkins et al. 2014; Bolduc et al. 2020). Bacteria have been shown to remain active below the freezing point (Murray et al. 2012) with both catabolic and anabolic activities observed in frozen soils (Drotz et al. 2010; Margesin and Collins 2019). Still, very little is known about the taxonomy or functional capacity of bacteria in frozen soils and their global biogeochemical impacts.

A major unknown in seasonally frozen soils is what roles viruses play in controlling active microbial population dynamics and dictating the fate of soil C. Viruses can affect soil C by stopping microbial metabolism via cell lysis, which effectively maintains microbial diversity by killing abundant lineages and releasing cell nutrients into the environment which can either fuel other organisms' metabolism or be stabilized via entombing effects (Liang et al. 2017). Viruses can also carry auxiliary metabolic genes, which are host-derived genes that can be expressed during

infection and typically aid the infection process by overcoming energetic limitations (Thompson et al. 2011; Rosenwasser et al. 2016). While viruses and their impacts have been well-characterized in marine environments via meta-omics approaches (Zimmerman et al. 2019) and development of tractable virus-host model systems, they remain enigmatic in soil environments. The initial studies on viruses in northern peatlands during the growing season have shown they are largely unrelated to other viruses, highly diverse, endemic to their habitat, infect dominant microbial populations, and carry AMGs impacting C cycling (Trubl et al. 2018; Emerson et al. 2018). These discoveries have drawn substantial attention to viruses in C-rich peatland soils, but the question of whether viruses are active during the majority of the year when Arctic peatlands are frozen (~7–9 months) remains unanswered.

Stable isotope probing (SIP) combined with metagenomics is an effective tool for tracking microbial activities in complex communities, linking individuals to specific functions (Starr et al. 2018; Dumont and Hernández García 2019; Barnett and Buckley 2020), or characterizing specific microbes and the viruses that infect them (Haig et al. 2015). While many isotope tracing studies use ^{13}C-enriched tracers (in $^{13}CO_2$, ^{13}C-plant biomass, or ^{13}C-glucose) to identify specific substrate degraders (Dumont and Hernández García 2019) or assess microbial C use efficiency (Gross et al. 2020), 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 (Schwartz 2007; Aanderud and Lennon 2011; Blazewicz and Schwartz 2011). 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 (Schwartz et al. 2014; Blazewicz et al. 2014; Papp et al. 2018; Blazewicz et al. 2020), but previously, has not been used to label active

viruses. We used heavy-water SIP-metagenomics to label active viruses and characterize their roles in controlling active microbial population dynamics in frozen northern peat soils. To our knowledge, viral activities have never been confirmed under such conditions, and a better understanding of their impact on microbial activity and C transformation over winter months may be pivotal for understanding the fate of Soil C in these C-rich peat soils.

Methods

Experimental setup

Samples were collected 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. 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, that was 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 in 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, except for proportionally larger amounts of soil (18.15 g wet soil) and larger incubation vials (100 ml), for headspace gas analysis to quantify CO_2 production. All of the sample vials, except 3 of the gas production 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 maintained at -1.5°C and half of the samples for each treatment (n=3) were destructively sampled at 184 d (mid-year) and 370 d (whole year), whereupon they were 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). Samples were then analyzed via gas chromatography to quantify headspace CO_2.

DNA extraction and density gradient SIP

DNA was extracted from each replicate using a modified phenol/chloroform protocol (Sambrook and Russell 2001). 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^{3-} 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 new tube. The bead beat tubes were then re-extracted using 220 µl 1xTE and 80 µl PO_4^{3-} buffer. The supernatant from the first and second extractions 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 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 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 (Blazewicz et al. 2014). DNA was binned into 5 fractions

based on density, and the gDNA from the two heaviest fractions (medium-heavy [MH; 1.717–1.725 g/ml] and heavy [H; 1.725–1.750 g/ml]).

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) and DNA quality was checked using a Bioanalyzer chip. The sheared gDNA was then end-repaired, A-tailed, adapters ligated, and sequenced to generate metagenomes. DNA was sequenced on an Illumina HiSeq 2500 for 24 samples (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 (Roux et al. 2019; Trubl et al. 2020). Processed reads were assembled into contigs using SPAdes v3.11.1 (--only-assembler --phred-offset 33 --meta -k 25,55,95 –12; Bankevich et al. 2012). 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 VirSorter (virome decontamination mode; Roux et al. 2015) and DeepVirFinder (DVF; Ren et al. 2020) to

detect viral contigs. Only contigs ≥10kb, those sorted into VirSorter categories 1 and 2, and those with a score ≥0.9 and p value < 0.05 from DVF were retained. Viral contigs were clustered at 95% average nucleotide identify (ANI) across 85% of the shorter contig (Roux et al., 2019) using nucmer (Delcher, Salzberg & Phillippy, 2003) to generate a nonredundant set of viral populations (vOTUs). vOTU quality was assessed with CheckV (default parameters; Nayfach et al. 2020). Coverage of the vOTUs was estimated based on post quality-controlled reads mapping at ≥90% ANI and covering ≥75% of the contig (Roux et al., 2019) using Bowtie2 (Langmead & Salzberg, 2012). Activity of vOTUs was determined by a vOTU being present in the $H_2{}^{18}O$ samples and not present in the natural abundance water samples. The vOTUs were annotated using the virus-centric multiPhATE pipeline (default parameters; Zhou et al. 2019) and the AMG-centric DRAM-v pipeline (with --skip_uniref; Shaffer et al. 2020). 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 (Howard-Verona et al. 2017), leveraged classification from VirSorter (categories 4 or 5) and our genome-similarity host linkages (≤90% similarity), and used two tools — Checkv (Nayfach et al. 2020) and PHASTER (Arndt et al. 2016).

MAG curation and host linking

Read-pairs from the biological replicates were grouped for 8 separate co-assemblies with MEGAHIT v1.1.4 (--k-min 27 --k-max 127 --k-step 10; Li et al. 2015). Contigs greater than 1 KB were separately binned with Concoct v1.0.0 (Alneberg et al. 2014), MaxBin v2.2.6 (Wu, Simmons,

and Singer 2016) and MetaBAT v2.12.1 (Kang et al. 2015). Genome bins from these three binning tools were refined using the bin_refinement module of MetaWRAP v1.2.1 (-c 50 -x 10; Uritskiy, DiRuggiero and Taylor 2018) with CheckM v1.0.12 (Parks et al. 2015). Only genome bins with at least medium quality according to MIMAG standards (Bowers et al. 2017) were retained.

Two methods were used to define MAGs as active; 1) a read-subtraction approach, and 2) a log fold change approach (described below). For the read-subtraction approach, contigs from the ^{16}O assemblies were used as a reference to split the ^{18}O reads into those that aligned (unlabeled) and those that did not align (labeled) using bbsplit v38 (Bushnell, B.; maxindel=1). Mapped reads were considered to be from inactive community members and were discarded, and the 4 sets (2 SIP fractions x 2 timepoints) of active reads were then processed through the same workflow as above, starting with the MEGAHIT assembly through MetaWRAP refinement, obtaining genome bins reconstructed from distinct ^{18}O reads.

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; Olm et al. 2017). MAG taxonomy was determined using GTDB-tk v0.3.2 (Chaumeil et al. 2020) with the version r89 GTDB database. Structural annotation was done using Patric (Wattam et al. 2017), and general functional annotation with RASTtk (Brettin et al. 2015) for subsystems within Patric v3 (retaining subsystem variants predicted to be active or likely) and KofamScan v1.1.0 (Aramaki et al. 2020) against the KEGG (Kanehisa et al. 2016) 93 database. Per-sample MAG abundances were determined by aligning each sample's filtered reads against the MAG genomes using bbmap v38 (https://sourceforge.net/projects/bbmap/) to obtain total assigned reads per contig and average fold coverage per contig.

The second approach to define MAGs as active was by significant (5% false-discovery rate) and positive log fold change in ^{18}O read abundances compared to ^{16}O within a time point and SIP fraction. Fold changes were determined using wrench-normalized (Kumar et al. 2018) total assigned reads per MAG through a zero-inflated log-normal model implemented in metagenomeSeq (Paulson et al., 2013).

Detected vOTUs and MAGs were linked via clustered regularly interspaced short palindromic repeats (CRISPR) spacers and shared genomic content as previously described (Andersson & Banfield 2008; Roux et al. 2016). Briefly, CRISPR regions were detected in the MAGs using MinCED (options -minNR 2 -spacers; Bland et al. 2007) and linked to the vOTUs with blastn (options -max_target_seqs 10000000 -dust no -word_size 7; Altschul et al. 1990). In addition, BLAST (options -dust no -perc_identity 70) was used to link vOTUs and MAGs based on shared genomic content (Edwards et al. 2016).

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 R, using packages Vegan for diversity (Oksanen et al., 2016) and pheatmap for heat maps (Kolde and Kolde 2015). Statistics were performed using the data analysis package in excel.

Results

Characterization of viruses

To gain an understanding of the viruses and their activities from the Bonanza Creek LTER soil incubated under winter-like conditions we analyzed virus sequences from heavy water SIP

targeted metagenomic reads. Using two virus detection methods, we identified 5 737 putative viruses (≥3 kb) that clustered into a nonredundant set of 332 vOTUs ≥10 kb (Table S2) spanning 66 genera (Table S3). The size range of these vOTUs was 10 039–437 858 bp with an average of 32 954 bp and 15 ≥100 kb, including 5 huge viruses (i.e., ≥200 kb; Al-Shayeb et al. 2020). The vOTUs were well-covered with an average of 17x normalized coverage per metagenome but with a large range, 3–147x (Fig. 1A). Quality checks for these vOTUs identified 58 medium–high quality vOTUs and four that were considered 'complete' according to community standards (Roux et al. 2019). Notably, 93 vOTUs could not be assessed because they either did not possess a known viral gene and were detected via a reference-free machine learning method (Ren et al. 2020) or due to a discrepancy between databases. Annotation of the vOTU genomes with the virus-centric Multiphate pipeline predicted 15 772 genes, in 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, miscellaneous, organic nitrogen, and transporters) from 31 vOTUs (Table S5), and 21 of these were from active vOTUs that carried 63 AMGs (Fig. 2). After searching the annotations for genes used in lysogenic replication, we predicted nearly half (43%) of our vOTUs were temperate viruses, with the majority (80%) having at least one member undergoing lysogenic infection at the time of sampling (Table S6). Further, 59% of the predicted temperate viruses were active, implying that most of the active vOTUs in these samples may be strictly lytic viruses.

The vOTU richness was an order of magnitude higher in samples incubated with heavy water compared to natural abundance water (Fig. S2), confirming that viruses incorporated ^{18}O into their DNA. A majority (73%) of the vOTUs found in the dense ^{18}O SIP fractions were active at

some point over the 370-d incubation (Fig. 1B). We measured active vOTU diversity to assess changes in the viral community from mid-year to a whole year. A Shannon's H metric revealed a significantly (p ≤0.05) more diverse viral community at 370 d compared to 184 d (Fig. 1C). Shannon's H diversity, which includes richness and evenness, was driven by an average increase of more than 60% for vOTU richness from 184 to 370 d. Comparing the same active vOTUs between both time points revealed 64 vOTUs became more abundant over time, 110 became newly active, and 18 were no longer labeled and either degraded or went below detection limit from mid-year.

Host characterization based on SIP-metagenomics

To identify potential virus hosts, we used an array of metagenome assembly and binning methods which yielded 153 medium to high quality MAGs, spanning 16 bacterial phyla and 1 phylum of archaea (Table S7; GTDB taxonomy). The dominant phyla detected were Chloroflexota, Acidobacteria, and Myxococcota (formally Proteobacteria). Incubation with heavy water ($H_2{}^{18}O$) indicated 30% (46) of the MAGs were actively growing (Fig. 3A–C), spanning 9 bacterial phyla and the most abundant active populations were from the Acidobacteria, Bacteroidetes, and Firmicutes. Relative abundance declined between mid-year and full year samples for all bacterial phyla, except Bacteroidota, Caldisericota, and Firmicutes. By sequencing both unlabeled and isotope labeled samples, we gained a consensus of the genetic repertoire from these soils (Table S8) and focused our efforts on characterizing the active microbial lineages and metabolisms that viruses may be altering over winter months. To confirm microbial activity, CO_2 production was measured and found to occur continuously throughout the incubation (Fig. 3D)

Linking vOTUs to MAGs

To understand the influence of native viruses on microbial dynamics over the winter season, we predicted potential microbial hosts via two different *in silico* approaches based on sequence similarity. We identified 11 879 CRISPR spacers found in 136 MAGs, which linked 10 vOTUs to 6 MAGs spanning 4 bacterial phyla (Fig. 4, Table S9). Layering on the activity of vOTUs and MAGs revealed 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. Using genome sequence similarity of the vOTUs to the MAGs beyond just CRISPR regions, we recovered 798 similarity matches that linked 141 vOTUs to 65 MAGs spanning 10 bacterial phyla (Fig 4; Table S10). Both linkage approaches combined showed a significantly higher proportion of vOTU-host matches between active populations (Table 1). Notably, four vOTUs (153, 161, 162, and 270) had a broad-host range across several bacterial phyla, with three vOTUs infecting two phyla and one infecting four phyla (Fig. 4). Two of the broad-host-range vOTUs (153 and 270) were active and linked to only active MAGs from bacterial phyla Bacteriodota and Firmicutes.

Virus-host linking approaches had 315 unique connections between 145 vOTUs (44%) and 65 MAGs (43%) spanning 10 bacterial phyla. As such, all 145 vOTUs are classified as dsDNA bacteriophages because the vOTUs were linked to MAGs with a bacteria taxonomic assignment. More than half of the vOTUs (69%) linked to a host were considered temperate viruses, representing the majority (88%) of the 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 (most

linkages between vOTU and MAGs). For the active vOTUs and MAGs, Firmicutes (55%), Bacteroidota (34%), Patescibacteria (9%), and then Proteobacteria (2%) are the most infected. Of the active populations, 81 (33%) vOTUs and 33 (51%) MAGs were linked, with the top 5% almost exclusively infecting bacteria from the order Bacteroidales.

Discussion

Quantifying and predicting 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 these soils, and recent work has highlighted the important role viruses likely play in microbial C processing and helping their hosts adapt to these subzero temperatures (Emerson et al. 2018; Trubl et al. 2018). What is not well understood, is what proportion of these viruses are active, how the viral community dynamics change over time, and how long the viruses persist. This is the first study, to our knowledge, to use 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.

Heavy-water SIP-metagenomics reveals hidden active viral community

Heavy-water SIP has proven to be a robust method for identifying metabolically active microbes in soils (Blazewicz and Schwartz 2011; Aanderud and Lennon 2011; Koch et al. 2018). Heavy-water SIP is superior to other techniques to track activity such as RNA-to-DNA ratios, rRNA levels, 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 are hard to interpret as RNA levels don't correlate with growth and often have weak

or no correlation with proteins levels (Blazewicz and Schwartz 2011; Vogel and Marcotte 2012; Bashiardes, Zilberman-Schapira, and Elinav 2016). 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 (Schwartz et al. 2016; Papp et al. 2018). Heavy-water SIP has not been used in soils to track virus activity and here we evaluated the utility of heavy-water SIP to label viruses and their microbial hosts in anoxic and subzero temperature conditions to characterize their activity and potential role in the C cycle over winter months.

Over the course of a year, hundreds of vOTUs were found to be active, in stark contrast to the general paradigm that frozen soils have little to no activity (Nikrad et al. 2016), and 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 ecosystems as previously described (Zablocki, Adriaenssens, and Cowan 2016; Adriaenssens et al. 2017; Emerson et al. 2018; Trubl et al. 2018). We are unable to determine if the increase in abundance of the vOTUs, from mid-year to a whole year, is due to these viruses staying active the whole time either as proviruses of active microbes or many generations of virions actively infecting host(s). Additionally, it may be that some viruses were momentarily active enough to become isotopically enriched and persisted in the soil as environmental DNA (eDNA; potentially infectious), as inert virions from internal errors (e.g., lethal mutations) or structural damage, or as competent virions that can no longer infect due to their host evolving resistance as seen in host-virus arms races (Breitbart 2018). Nevertheless, our data supports the Kill-the-winner hypothesis where we see a shift in active dominant microbes from mid-year to a whole year corresponding with a shift in the active viral communities with the majority of vOTUs increasing

their abundance and a few vOTUs becoming active from the mid-year to the whole year incubation. It is likely that different members of the vOTUs were active at different times as proposed by the Royal Family model, where variants in the vOTUs have overcome their host's resistance, allowing these vOTUs to become more abundant (Breitbart et al. 2018).

Most virus-host linkages were between unlabeled vOTUs and unlabeled MAGs from the phyla Acidobacteriota, Actinobacteriota, and Chloroflexota, which are highly abundant in global bog ecosystems and represented some of the most abundant phyla here (Table 1; Table S7). It is likely that these infections took place during the last growing season and a viral signal was detected from eDNA or as an infectious/inert virion. Likewise, these unlabeled MAGs are likely inactive or recently deceased (e.g. eDNA) given the harsh winter conditions simulated during the incubations. Given their high abundance, without the addition of heavy-water SIP, these bacteria and their viruses would have been traditionally described first and foremost, thus occluding the characterization of the actual microbes and viruses active in these soils under winter conditions.

Viruses infect dominant active microbial lineages

To better understand the roles viruses may play in controlling central biogeochemical processes in bogs under winter conditions, we assessed the infected active phyla of bacteria and the most abundant active vOTUs. Firmicutes had the largest increase in abundance from mid- to full-year and had the most viral infections, in terms of the number of linked active vOTUs and increase in the vOTUs' abundance, supporting the kill-the-winner hypothesis. Firmicutes are a diverse bacterial phylum known for adaptations to anoxic conditions, including fermentation and creating endospores (Zeibich, Schmidt, and Drake 2019). In our samples, all the active infected Firmicutes contained fermentation genes indicating the capacity to produce ethanol, lactate, or

both (Table S8). Another active and infected bacterial phylum that increased abundance through time was Bacteroidota, and all the MAGs identified shared taxonomic affiliation down to the order Bacteroidales. Most of these bacteria are obligate anaerobes and are known for their diverse arrays of carbohydrate-active enzymes arranged into polysaccharide utilization loci and fermentation (Lapébie et al. 2019; Larsbrink and McKee 2020). Many active MAGs within this group could degrade polysaccharides (e.g., ICE 1 encoded pectin degradation protein KdgF) and all could perform lactate fermentation (Table S8). Bacteroidales became the most abundantly active lineage in the year incubation and was infected by the most abundant active vOTUs (23, 37, 41, 124, and 190), suggesting an important role for phage dynamics influencing associated winter activities. In parallel with the increase in abundance of active Bacteroidota and Firmicutes, C mineralization to CO_2 occurred steadily throughout the incubation. This suggests that these two hosts likely play an important role in over-winter C loss from our soils, and therefore the viruses that infect them also likely influence this process.

Another highly abundant active virus was vOTU 28, and linked to MAG ICE33, which was from the phylum Patescibacteria (part of the Candidate phyla radiation; CPR). These bacteria are known for their ultrasmall cell size with reduced genomes that most have a symbiotic lifestyle and have been shown to have symbiotic relationship with Bacteroidota (Lemos et al. 2019), suggesting direct interactions between these two clades in our system. Patescibacteria is thought to resist phage infection by lacking or reducing the number of potential phage receptors on their cell membrane (Chen et al. 2019), but notably, was one of the most infected lineages. Additionally, one member was linked to a vOTU with a CRISPR spacer match, which are rarely found in CPR bacteria and the prevalence of these infections may be the result of their

interactions with Bacteroidota as both phyla increased their abundance over time (Lemos et al. 2019). The final infected active phylum of bacteria was Proteobacteria, with only one active MAG from the class Micavibrionales. Not much is known about this clade, even less about their viruses, because they have an epibiotic lifestyle where they feed on another organism to survive making it very difficult to culture. These predatory bacteria may be active under winter conditions by either eating non-active or dead cells, or benefit from attaching to a host that can utilize alternative energy sources and recalcitrant organic matter (Coutinho et al. 2020). We identified multiple active vOTUs infecting this lineage, revealing another indirect way viruses may impact nutrient cycling and microbial diversity by preying upon these predators.

Another way to assess virus ecological significance is to understand host range breadth. Some viruses, commonly referred to as generalists or polyvalent, can infect more than one species of bacteria. We identified four generalist vOTUs, two that were unlabeled and two that were active, and all four were predicted to infect hosts from different phyla. In addition, these four generalists were some of the most abundant vOTUs, suggesting that having a broad host range may offer an advantage in these arduous soils. Recently, four viruses from an oligotrophic lake in Michigan, were found to be able to infect bacteria across several phyla including Proteobacteria, Actinobacteria, and Bacteroidetes (Malki et al. 2015). The identified host phyla in that study are the same bacterial phyla that our generalists infected with the addition of Firmicutes and Chloroflexota. Taking into consideration environmental conditions that limit microbial growth and virus-host interactions such as low temperature, anoxia, and the structuring of soil, we would expect more viruses to be generalists, but this is not the case. It may be beneficial for these viruses to be specialists which would allow them to integrate into a

particular host more rapidly and efficiently during the prolonged winter until the growing season when conditions are more favorable.

Challenges associated with identifying virus 'activity'

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 (Trubl et al. 2020). 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 (Joergensen and Wichern 2018). For a microbe, this may be true, but for a virus there is a range in magnitude of impact for different 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 (Howard-Verona et al. 2017). Even the application of isotope tracers for direct assessment of activity is not as straightforward for viruses as it is for other microorganisms, and therefore should be discussed.

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 viruses undergoing lysogenic infection versus lytic infection need to be distinguished because their effect 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 lysogenic cycle as compared to lytic cycle. During lysogeny, viruses may still impact host metabolism via host gene regulation (and thus phenotype), but it is primarily for maintenance. Additionally, proviruses may be labeled, but not currently active if they are

proviral remnants of a past infection (Casjens 2003). These remnants have no negative impact on host metabolism (beyond occupying genome space) but may confer some advantage as a gene transfer agent (Howard-Verona et al. 2017; Lang, Westbye, and Beatty, 2017). Unfortunately, SIP-enabled metagenomics alone cannot unequivocally identify viral metabolic state (e.g., in maintenance mode or remnant), making it currently difficult to fully assess viral activity. In other words, we can use heavy-water SIP to show virus activity, but with this approach alone we cannot identify which viral replication cycle a virus is undergoing or quantify the impact the virus has on the host's metabolism. However, this is the first application of heavy-water SIP-metagenomics to characterize virus activity – a combination of SIP-metagenomics with other approaches such as a SIP-virome can provide a more holistic view of vOTU dynamics.

Soils are thought to generally be harsh environments, especially our soils due to the anoxic and subzero temperature conditions, and for that reason we expected a majority of our viruses would be temperate (Stewart and Levin 1984; Howard-Verona et al. 2017). A main reason why a temperate virus undergoes lysogenic infection rather than lytic infection can be attributed to a poor host physiological state, and so not only did we hypothesize most of our detected viruses would be temperate, but they would be detected as proviruses. Indeed, in our samples we did see a high rate of lysogeny (43%) and it was higher than previously reported for desert Antarctic soils (4–20%) and almost identical (44%) to whole soil assays from temperate wetlands soils (Williamson et al. 2007), and the majority (80%) of viruses were present as proviruses.

Lytic infection is the iconic role of a virus where progeny viruses are made, but this presents its own challenges for understanding activity. Viruses undergoing lytic infection in the presence of only heavy water, would have all their progeny viruses labeled, but the abundance

of the population would depend on its burst size. Once the ^{18}O-labeled virion is in the environment, it does not necessarily imply it is infectious, because it may have internal errors, structural damage, or their host(s) may no longer be around leaving the virus in 'Bank mode' (Breitbart and Rohwer 2005). Nevertheless, soils are known to harbor environmental DNA, which for viruses could be infectious, and even allow virions to persist thousands of years before becoming infectious again (Legendre et al. 2014).

Conclusions

Soil viral ecology remains an immature field, largely due to the complex nature of soil combined with the many states in which viruses can exist, and the diversity of infection cycles and viral nucleic acid types. This work represents a first inroad into labeling active microbes and viruses via heavy-water SIP-metagenomics, providing a new lens into virus-host dynamics and microbial ecology under winter conditions. To further characterize active virus-host dynamics, future work can utilize quantitative SIP (qSIP) characterization and over a time-series with increased frequency of sampling to better resolve the different levels of activity a virus can have. Our study suggests that bacteria and their viruses are active over the long winter months in northern peatlands and the populations that are active are a subset of the community captured using metagenomics alone. While metagenomics offers a lens into the genetic repertoire of these soils, characterization of the active microbial and viral communities may reduce uncertainty in ecosystem models and a better understanding of the ecosystem ecology.

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 at the USGS for their 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, which is jointly funded by NSF (DEB 1026415) and the USDA Forest Service, Pacific Northwest Research Station (PNW01-JV112619320-16), and from the USGS Climate R&D Program and AK Climate Science Center. Analysis of this work was supported by a Lawrence Livermore National Laboratory, Laboratory Directed Research & Development grant (18-ERD-041) awarded to S.J.B. and by awards from the U.S. Department of Energy, Office of Biological and Environmental Research, Genomic Science Program LLNL 'Microbes Persist' Scientific Focus Area (award #SCW1632) and PNNL 'Phenotypic Response of the Soil Microbiome to Environmental Perturbations' Scientific Focus Area (award #70880). Work conducted at LLNL was contributed under the auspices of the US Department of Energy under Contract DE-AC52-07NA27344.

References

1. Parazoo, N.C., Commane, R., Wofsy, S.C., Koven, C.D., Sweeney, C., Lawrence, D.M., Lindaas, J., Chang, R.Y.W. and Miller, C.E., 2016. Detecting regional patterns of changing CO_2 flux in Alaska. Proceedings of the National Academy of Sciences, 113(28), pp.7733-7738.
2. Ecale Zhou, C.L., Malfatti, S., Kimbrel, J., Philipson, C., McNair, K., Hamilton, T., Edwards, R. and Souza, B., 2019. multiPhATE: bioinformatics pipeline for functional annotation of phage isolates. Bioinformatics, 35(21), pp.4402-4404.
3. Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y. and He, C., 2020. Clades of huge phages from across Earth's ecosystems. Nature, 578(7795), pp.425-431.
4. Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, M.H.H. and Wagner, H., 2015. Vegan: community ecology package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. R package ver, pp.2-3.
5. Nichols, J.E. and Peteet, D.M., 2019. Rapid expansion of northern peatlands and doubled estimate of carbon storage. Nature Geoscience, 12(11), pp.917-921.
6. Männistö, M.K., Kurhela, E., Tiirola, M. and Häggblom, M.M., 2013. Acidobacteria dominate the active bacterial communities of Arctic tundra with widely divergent winter-time snow accumulation and soil temperatures. FEMS microbiology ecology, 84(1), pp.47-59.
7. Murray, A.E., Kenig, F., Fritsen, C.H., McKay, C.P., Cawley, K.M., Edwards, R., Kuhn, E., McKnight, D.M., Ostrom, N.E., Peng, V. and Ponce, A., 2012. Microbial life at− 13 C in the

brine of an ice-sealed Antarctic lake. Proceedings of the National Academy of Sciences, 109(50), pp.20626-20631.

8. Natali, S.M., Watts, J.D., Rogers, B.M., Potter, S., Ludwig, S.M., Selbmann, A.K., Sullivan, P.F., Abbott, B.W., Arndt, K.A., Birch, L. and Björkman, M.P., 2019. Large loss of CO_2 in winter observed across the northern permafrost region. Nature Climate Change, 9(11), pp.852-857.
9. Blazewicz, S.J., Hungate, B.A., Koch, B.J., Nuccio, E.E., Morrissey, E., Brodie, E.L., Schwartz, E., Pett-Ridge, J. and Firestone, M.K., 2020. Taxon-specific microbial growth and mortality patterns reveal distinct temporal population responses to rewetting in a California grassland soil. The ISME Journal, pp.1-13.
10. Schwartz, E., 2007. Characterization of growing microorganisms in soil by stable isotope probing with H218O. Appl. Environ. Microbiol., 73(8), pp.2541-2546.
11. Aanderud, Z.T. and Lennon, J.T., 2011. Validation of heavy-water stable isotope probing for the characterization of rapidly responding soil bacteria. Appl. Environ. Microbiol., 77(13), pp.4589-4596.
12. Blazewicz, S.J. and Schwartz, E., 2011. Dynamics of ^{18}O incorporation from $H_2{}^{18}O$ into soil microbial DNA. Microbial ecology, 61(4), pp.911-916.
13. Papp, K., Mau, R.L., Hayer, M., Koch, B.J., Hungate, B.A. and Schwartz, E., 2018. Quantitative stable isotope probing with $H_2{}^{18}O$ reveals that most bacterial taxa in soil synthesize new ribosomal RNA. The ISME journal, 12(12), pp.3043-3045.
14. Blazewicz, S.J., Schwartz, E. and Firestone, M.K., 2014. Growth and death of bacteria and fungi underlie rainfall-induced carbon dioxide pulses from seasonally dried soil. Ecology, 95(5), pp.1162-1172.

15. Schwartz, E., Van Horn, D.J., Buelow, H.N., Okie, J.G., Gooseff, M.N., Barrett, J.E. and Takacs-Vesbach, C.D., 2014. Characterization of growing bacterial populations in McMurdo Dry Valley soils through stable isotope probing with 18O-water. FEMS microbiology ecology, 89(2), pp.415-425.
16. Barnett, S.E. and Buckley, D.H., 2020. Simulating metagenomic stable isotope probing datasets with MetaSIPSim. BMC bioinformatics, 21(1), p.37.
17. Gross, A., Lin, Y., Weber, P.K., Pett-Ridge, J. and Silver, W.L., 2020. The role of soil redox conditions in microbial phosphorus cycling in humid tropical forests. Ecology, 101(2), p.e02928.
18. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320, 1047–1050 (2008).
19. Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A. and Amid, C., 2019. Minimum information about an uncultivated virus genome (MIUViG). Nature biotechnology, 37(1), pp.29-37.
20. Delcher, A.L., Salzberg, S.L. and Phillippy, A.M., 2003. Using MUMmer to identify similar regions in large sequence sets. Current protocols in bioinformatics, (1), pp.10-3.
21. Vogel, C. and Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature reviews genetics, 13(4), pp.227-232.
22. Bashiardes, S., Zilberman-Schapira, G. and Elinav, E., 2016. Use of metatranscriptomics in microbiome research. Bioinformatics and biology insights, 10, pp.BBI-S34610.
23. Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M., Poirot, O., Bertaux, L., Bruley, C. and Couté, Y., 2014. Thirty-thousand-year-old distant relative of

giant icosahedral DNA viruses with a pandoravirus morphology. Proceedings of the National Academy of Sciences, 111(11), pp.4274-4279.

24. Breitbart, M. and Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus?. Trends in microbiology, 13(6), pp.278-284.
25. Trubl, G., Jang, H.B., Roux, S., Emerson, J.B., Solonenko, N., Vik, D.R., Solden, L., Ellenbogen, J., Runyon, A.T., Bolduc, B. and Woodcroft, B.J., 2018. Soil viruses are underexplored players in ecosystem carbon processing. MSystems, 3(5).
26. Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H.B., Singleton, C.M., Solden, L.M., Naas, A.E., Boyd, J.A. and Hodgkins, S.B., 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nature microbiology, 3(8), pp.870-880.
27. Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., Solonenko, N., Lara, E., Poulain, J. and Pesant, S., 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature, 537(7622), pp.689-693.
28. Roux, S., Enault, F., Hurwitz, B.L. and Sullivan, M.B., 2015. VirSorter: mining viral signal from microbial Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), p.357.genomic data. PeerJ, 3, p.e985.
29. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy T, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature biotechnology 35:725.
30. Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B. and Gazitúa, M.C., 2020. DRAM for distilling

microbial metabolism to automate the curation of microbiome function. Nucleic acids research, 48(16), pp.8883-8900.

31. Howard-Varona, C., Hargreaves, K.R., Abedon, S.T. and Sullivan, M.B., 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. The ISME journal, 11(7), pp.1511-1520.
32. Li, D., Liu, C.M., Luo, R., Sadakane, K. and Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10), pp.1674-1676.
33. Kolde, R. and Kolde, M.R., 2015. Package 'pheatmap'. R Package, 1(7), p.790.
34. Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F. and Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nature methods, 11(11), pp.1144-1146.
35. Wu, Y.W., Simmons, B.A. and Singer, S.W., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics, 32(4), pp.605-607.
36. Kang, D.D., Froula, J., Egan, R. and Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ, 3, p.e1165.
37. Uritskiy, G.V., DiRuggiero, J. and Taylor, J., 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome, 6(1), pp.1-13.
38. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research, 25(7), pp.1043-1055.

39. Olm, M.R., Brown, C.T., Brooks, B. and Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME journal, 11(12), pp.2864-2868.

40. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P. and Parks, D.H., 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.

41. Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L. and Gerdes, S., 2017. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic acids research, 45(D1), pp.D535-D542.

42. Sambrook J, Russell DW. Molecular cloning: A laboratory manual. 3rd ed. New York: Cold Spring Harbor Laboratory Press; 2001.

43. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S. and Ogata, H., 2020. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics, 36(7), pp.2251-2252.

44. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. Nucleic acids research, 44(D1), pp.D457-D462.

45. Ren, J., Song, K., Deng, C., Ahlgren, N.A., Fuhrman, J.A., Li, Y., Xie, X., Poplin, R. and Sun, F., 2020. Identifying viruses from metagenomic data using deep learning. Quantitative Biology, pp.1-14.

46. Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D. and Shukla, M., 2015. RASTtk: a modular and extensible

implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Scientific reports, 5(1), pp.1-6.

47. Koch, B.J., McHugh, T.A., Hayer, M., Schwartz, E., Blazewicz, S.J., Dijkstra, P., van Gestel, N., Marks, J.C., Mau, R.L., Morrissey, E.M. and Pett-Ridge, J., 2018. Estimating taxon-specific population dynamics in diverse microbial communities. Ecosphere, 9(1), p.e02090.
48. Nikrad, M.P., Kerkhof, L.J. and Häggblom, M.M., 2016. The subzero microbiome: microbial activity in frozen and thawing soils. FEMS microbiology ecology, 92(6), p.fiw081.
49. Dumont, M.G. and Hernández García, M., 2019. Stable Isotope Probing. Methods and Protocols Totowa, NJ, US: Humana Press.
50. Schwartz, E., Hayer, M., Hungate, B.A., Koch, B.J., McHugh, T.A., Mercurio, W., Morrissey, E.M. and Soldanova, K., 2016. Stable isotope probing with 18O-water to investigate microbial growth and death in environmental samples. Current opinion in biotechnology, 41, pp.14-18.
51. Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S. and Kyrpides, N.C., 2020. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology, pp.1-8.
52. Kumar, M.S., Slud, E.V., Okrah, K., Hicks, S.C., Hannenhalli, S. and Bravo, H.C., 2018. Analysis and correction of compositional bias in sparse sequencing count data. BMC genomics, 19(1), p.799.
53. Paulson, J.N., Stine, O.C., Bravo, H.C. and Pop, M., 2013. Differential abundance analysis for microbial marker-gene surveys. Nature methods, 10(12), pp.1200-1202.

54. Starr, E.P., Shi, S., Blazewicz, S.J., Probst, A.J., Herman, D.J., Firestone, M.K. and Banfield, J.F., 2018. Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. Microbiome, 6(1), p.122.

55. Haig, S.J., Schirmer, M., D'amore, R., Gibbs, J., Davies, R.L., Collins, G. and Quince, C., 2015. Stable-isotope probing and metagenomics reveal predation by protozoa drives E. coli removal in slow sand filters. The ISME journal, 9(4), pp.797-808.

56. Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J. and Chisholm, S.W., 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proceedings of the National Academy of Sciences, 108(39), pp.E757-E764.

57. Rosenwasser, S., Ziv, C., Van Creveld, S.G. and Vardi, A., 2016. Virocell metabolism: metabolic innovations during host–virus interactions in the ocean. Trends in microbiology, 24(10), pp.821-832.

58. Zimmerman, A.E., Howard-Varona, C., Needham, D.M., John, S.G., Worden, A.Z., Sullivan, M.B., Waldbauer, J.R. and Coleman, M.L., 2019. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. Nature Reviews Microbiology, pp.1-14.

59. Joergensen, R.G. and Wichern, F., 2018. Alive and kicking: why dormant soil microorganisms matter. Soil Biology and Biochemistry, 116, pp.419-430.

60. Casjens, S., 2003. Prophages and bacterial genomics: what have we learned so far?. Molecular microbiology, 49(2), pp.277-300.

61. Lang, A.S., Westbye, A.B. and Beatty, J.T., 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. Annual review of virology, 4, pp.87-104.

62. Williamson, K.E., Radosevich, M., Smith, D.W. and Wommack, K.E., 2007. Incidence of lysogeny within temperate and extreme soil environments. Environmental microbiology, 9(10), pp.2563-2574.

63. Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C. and Hugenholtz, P., 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC bioinformatics, 8(1), pp.1-8.

64. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

65. Edwards, R.A., McNair, K., Faust, K., Raes, J. and Dutilh, B.E., 2016. Computational approaches to predict bacteriophage–host relationships. FEMS microbiology reviews, 40(2), pp.258-272.

66. Breitbart, M., Bonnain, C., Malki, K. and Sawaya, N.A., 2018. Phage puppet masters of the marine microbial realm. Nature Microbiology, 3(7), pp.754-766.

67. Stewart, F.M. and Levin, B.R., 1984. The population biology of bacterial viruses: why be temperate. Theoretical population biology, 26(1), pp.93-117.

68. Malki, K., Kula, A., Bruder, K., Sible, E., Hatzopoulos, T., Steidel, S., Watkins, S.C. and Putonti, C., 2015. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. Virology journal, 12(1), pp.1-5.

69. Lapébie, P., Lombard, V., Drula, E., Terrapon, N. and Henrissat, B., 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. Nature communications, 10(1), pp.1-7.

70. Lemos, L.N., Medeiros, J.D., Dini-Andreote, F., Fernandes, G.R., Varani, A.M., Oliveira, G. and Pylro, V.S., 2019. Genomic signatures and co-occurrence patterns of the ultra-small Saccharimonadia (phylum CPR/Patescibacteria) suggest a symbiotic lifestyle. Molecular ecology, 28(18), pp.4259-4271.

71. Liang, C., Schimel, J.P. and Jastrow, J.D., 2017. The importance of anabolism in microbial control over soil carbon storage. Nature microbiology, 2(8), pp.1-6

72. Coutinho, F.H., von Meijenfeldt, F.A.B., Walter, J.M., Haro-Moreno, J.M., Lopéz-Pérez, M., van Verk, M.C., Thompson, C.C., Cosenza, C.A.N., Appolinario, L., Paranhos, R. and Cabral, A., 2020. Ecogenomics and metabolic potential of the South Atlantic Ocean microbiome. Science of The Total Environment, p.142758.

73. Bolduc, B., Hodgkins, S.B., Varner, R.K., Crill, P.M., McCalley, C.K., Chanton, J.P., Tyson, G.W., Riley, W.J., Palace, M., Duhaime, M.B. and Hough, M.A., 2020. The IsoGenie database: an interdisciplinary data management solution for ecosystems biology and environmental research. PeerJ, 8, p.e9467.

74. Przybylak, R., 2000. Diurnal temperature range in the Arctic and its relation to hemispheric and Arctic circulation patterns. International Journal of Climatology: A Journal of the Royal Meteorological Society, 20(3), pp.231-253.

75. Hopple, A.M., Wilson, R.M., Kolton, M., Zalman, C.A., Chanton, J.P., Kostka, J., Hanson, P.J., Keller, J.K. and Bridgham, S.D., 2020. Massive peatland carbon banks vulnerable to rising temperatures. Nature communications, 11(1), pp.1-7.

76. Hodgkins, S.B., Tfaily, M.M., McCalley, C.K., Logan, T.A., Crill, P.M., Saleska, S.R., Rich, V.I. and Chanton, J.P., 2014. Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. Proceedings of the National Academy of Sciences, 111(16), pp.5819-5824.

77. Wilson, R.M., Hopple, A.M., Tfaily, M.M., Sebestyen, S.D., Schadt, C.W., Pfeifer-Meister, L., Medvedeff, C., McFarlane, K.J., Kostka, J.E., Kolton, M. and Kolka, R.K., 2016. Stability of peatland carbon to rising temperatures. Nature Communications, 7(1), pp.1-10.

78. Margesin, R. and Collins, T., 2019. Microbial ecology of the cryosphere (glacial and permafrost habitats): current knowledge. Applied microbiology and biotechnology, 103(6), pp.2537-2549.

79. Drotz, S.H., Sparrman, T., Nilsson, M.B., Schleucher, J. and Öquist, M.G., 2010. Both catabolic and anabolic heterotrophic microbial activity proceed in frozen soils. Proceedings of the National Academy of Sciences, 107(49), pp.21046-21051.

80. Larsbrink, J. and McKee, L.S., 2020. Bacteroidetes bacteria in the soil: Glycan acquisition, enzyme secretion, and gliding motility. In Advances in Applied Microbiology (Vol. 110, pp. 63-98). Academic Press.

81. Zeibich, L., Schmidt, O. and Drake, H.L., 2019. Dietary polysaccharides: fermentation potentials of a primitive gut ecosystem. Environmental microbiology, 21(4), pp.1436-1451.

82. Zablocki, O., Adriaenssens, E.M. and Cowan, D., 2016. Diversity and ecology of viruses in hyperarid desert soils. Applied and environmental microbiology, 82(3), pp.770-777.

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

84. Adriaenssens, E.M., Kramer, R., Van Goethem, M.W., Makhalanyane, T.P., Hogg, I. and Cowan, D.A., 2017. Environmental drivers of viral community composition in Antarctic soils identified by viromics. Microbiome, 5(1), pp.1-14.

85. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology, 19(5), pp.455-477.

86. Chen, L.X., Al-Shayeb, B., Méheust, R., Li, W.J., Doudna, J.A. and Banfield, J.F., 2019. Candidate Phyla Radiation Roizmanbacteria from hot springs have novel and unexpectedly abundant CRISPR-Cas systems. Frontiers in microbiology, 10, p.928.

87. Trubl, G., Hyman, P., Roux, S. and Abedon, S.T., 2020. Coming-of-age characterization of soil viruses: A user's guide to virus isolation, detection within metagenomes, and viromics. Soil Systems, 4(2), p.23.

Table 1. Linkages between labeled and unlabeled viruses and bacteria. vOTUs were linked to MAGs via nucleotide similarity or using CRISPR spacers. 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).

Description	No. of matches	vOTUs	MAGs
Active vOTU-active MAG	164	74	37
Active vOTU-unlabeled MAG	2	2	2
Unlabeled vOTU-active MAG	226	44	11
Unlabeled vOTU-unlabeled MAG	422	58	30

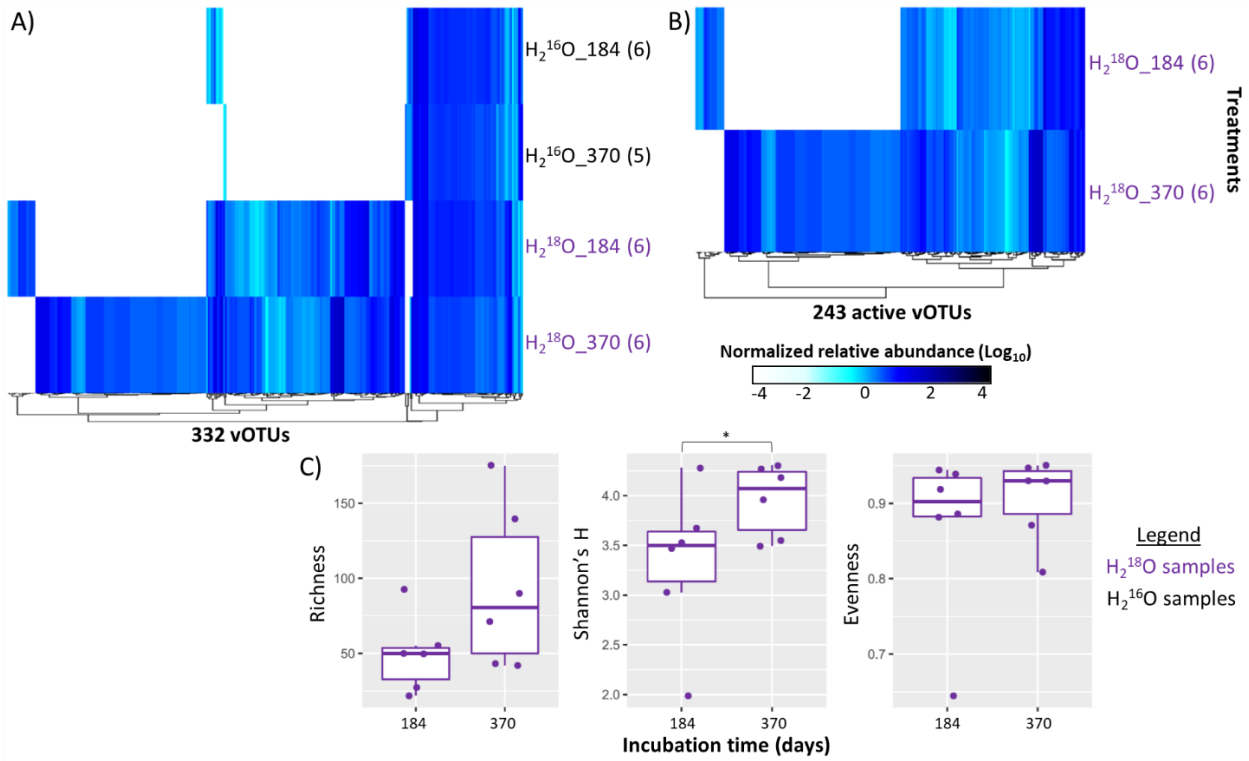

Figure 1. Assessment of viral community structure and activity. The relative abundances of the 332 vOTUs (A) and the 243 active vOTUs (B) are shown with samples grouped by treatment. There were 23 samples grouped into two treatments and spanned two time points: natural abundance water for 184 days ($H_2^{16}O$_184) or 370 days ($H_2^{16}O$_370) and heavy water for 184 days ($H_2^{18}O$_184) or 370 days ($H_2^{18}O$_370). Relative abundance for each vOTU 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. Panel C shows diversity metrics of the 243 active vOTUs. In all panels, $H_2^{18}O$ samples are highlighted with purple text. Significance denoted by asterix (* = $p < 0.05$).

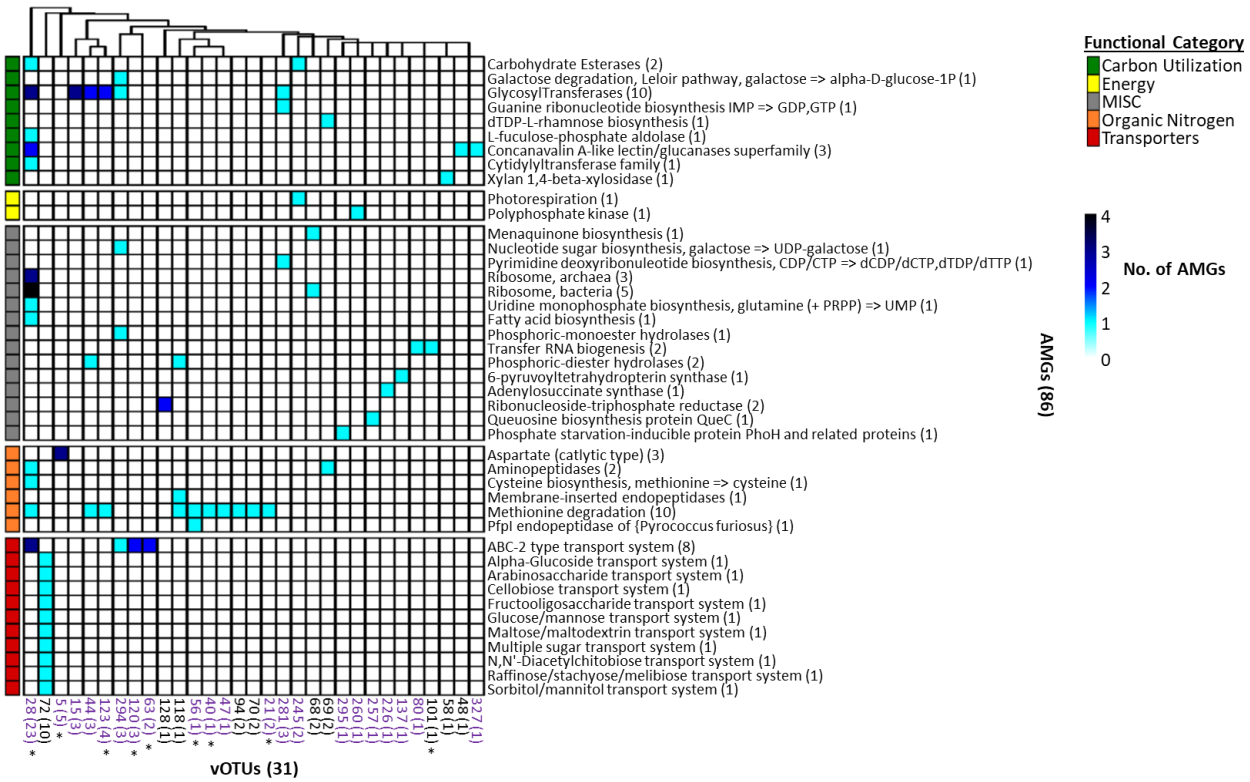

Figure 2. Putative auxiliary metabolic genes. A heatmap of vOTUs (x-axis) carrying confidently predicted AMGs (Dram-v score 1–3) and their annotation (y-axis). AMGs are further grouped by their respective functional category — green is carbon utilization, yellow is energy, orange is misc, pink is organic nitrogen, and red is transporter. The number of AMGs are displayed in parenthesis, active vOTUs are in purple font, and vOTUs with an asterix are linked to active MAGs.

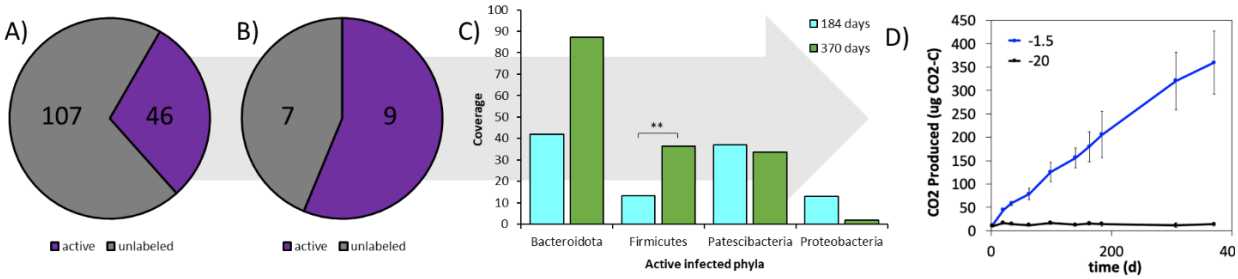

Figure 3. Abundances and activity of MAGs. (A) A pie chart showing the proportion of the 153 MAGs that are active (purple) or unlabeled (gray). (B) A pie chart showing the proportion of the 16 phyla that are active (purple) or unlabeled. (C) A bar chart showing the normalized coverage of the four active infected phyla at 184 d (light blue) and 370 d (green). Significance was determined via a paired t test and a ** denotes $p < .01$. (D) Cumulative CO_2 production in soil incubated at -1.5°C (experimental conditions) and -20°C (control). Error bars show standard error (n=3).

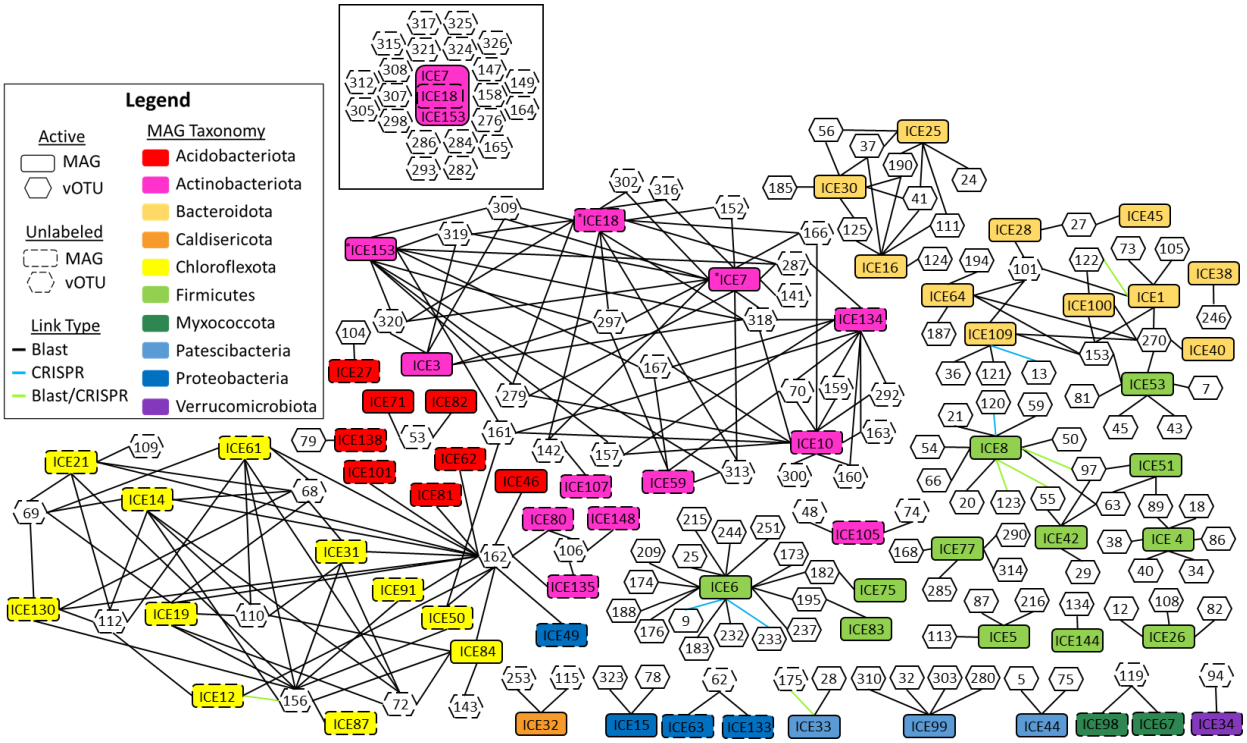

Figure 4. Virus-host linkages. A network was constructed showing vOTUs and the 10 phyla of their predicted bacterial hosts. Edges represent linkages between a vOTU (hexagon) and a bacterial MAG (rounded rectangle) and are colored by the method used: similarity in genomic content (black), CRISPR spacer match (light blue), or both (light green). MAG phylum is color coded and active MAGs and vOTUs are denoted by complete nodes. Linkages shared (via genomic similarity) among MAGs ICE7, ICE18, and ICE153 are shown in a separate box to reduce figure complexity.