

 Open access • Posted Content • DOI:10.1101/2021.01.25.428156

## Ecology of active viruses and their bacterial hosts in frozen Arctic peat soil revealed with H218O stable isotope probing metagenomics — [Source link](#)

Gareth Trubl, Jeffrey A. Kimbrel, Jose Liquet-Gonzalez, Erin E. Nuccio ...+5 more authors

**Institutions:** Lawrence Livermore National Laboratory, Pacific Northwest National Laboratory, United States Geological Survey

**Published on:** 25 Mar 2021 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** Carbon utilization, Stable-isotope probing, Biogeochemistry and Peat

Related papers:

- [Active viral population dynamics in frozen Arctic peat soil revealed with H218O stable isotope probing metagenomics](#)
- [Active virus-host interactions at sub-freezing temperatures in Arctic peat soil.](#)
- [Soil Phage Ecology: Abundance, Distribution, and Interactions with Bacterial Hosts](#)
- [Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations](#)
- [Characterization of Bacterial and Fungal Communities Reveals Novel Consortia in Tropical Oligotrophic Peatlands.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/ecology-of-active-viruses-and-their-bacterial-hosts-in-3miz9q2obr>

1 **Title:** Ecology of active viruses and their bacterial hosts in frozen Arctic peat soil revealed with  
2 H<sub>2</sub><sup>18</sup>O stable isotope probing metagenomics

3

4 **Running title:** SIP-metagenomics of arctic peat viral populations

5

6 Gareth Trubl<sup>1</sup>, Jeffrey A. Kimbrel<sup>1</sup>, Jose Liquez-Gonzalez<sup>1</sup>, Erin E. Nuccio<sup>1</sup>, Peter K. Weber<sup>1</sup>,

7 Jennifer Pett-Ridge<sup>1</sup>, Janet K. Jansson<sup>2</sup>, Mark P. Waldrop<sup>3</sup>, Steven J. Blazewicz<sup>1</sup>

8 <sup>1</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore,

9 CA, USA

10 <sup>2</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

11 <sup>3</sup>U.S. Geological Survey, Geology, Minerals, Energy, and Geophysics Science Center, Menlo

12 Park, CA, USA

13

14 Corresponding authors:

15 Gareth Trubl ([Trubl1@llnl.gov](mailto:Trubl1@llnl.gov)) and Steve Blazewicz ([blazewicz1@llnl.gov](mailto:blazewicz1@llnl.gov))

16

17 Key Words: soil viruses, bacteriophage, stable isotope probing, permafrost, peat, <sup>18</sup>O-water,

18 metagenomics

19

20

21

22

23 **Abstract**

24 Winter carbon loss in northern ecosystems is estimated to be greater than the average growing  
25 season carbon uptake. However, most ecosystem carbon measurements neglect winter months  
26 since carbon losses (primarily driven by microbial decomposers) are assumed to be negligible at  
27 low temperatures. We used stable isotope probing (SIP) targeted metagenomics to reveal the  
28 genomic potential of active soil microbial populations under winter conditions, with an emphasis  
29 on viruses and virus-host dynamics. Peat soils from the Bonanza Creek LTER site in Alaska  
30 were incubated under subzero anoxic conditions with H<sub>2</sub><sup>18</sup>O for 184 and 370 days. We identified  
31 46 bacterial populations (MAGs; spanning 9 bacterial phyla) and 243 viral populations (vOTUs)  
32 that actively took up <sup>18</sup>O and produced significant CO<sub>2</sub> throughout the incubation. Active hosts,  
33 predicted for 33% of the active vOTUs, were some of the most abundant MAGs and capable of  
34 fermentation and organic matter degradation. Approximately three-quarters of the active vOTUs  
35 carried auxiliary metabolic genes that spanned five functional categories, including carbon  
36 utilization, highlighting the potential impact of viruses in this peat soil's microbial  
37 biogeochemistry. These results illustrate significant bacterial and viral activity and interactions  
38 occur in frozen soils, revealing viruses are a major community-structuring agent throughout  
39 winter months.

40

41

42

43

44

## 45 **Introduction**

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

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

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

79 Stable isotope probing (SIP) combined with metagenomics is an effective tool for  
80 tracking the active microbial taxa in complex communities, linking individuals to specific  
81 functions [27-29], or characterizing specific microbes and the viruses that infect them [30].  
82 While many isotope tracing studies use  $^{13}\text{C}$ -enriched tracers (e.g.,  $^{13}\text{CO}_2$ ,  $^{13}\text{C}$ -plant biomass, or  
83  $^{13}\text{C}$ -glucose) to identify specific substrate degraders [28] or assess microbial C use efficiency  
84 [31], heavy water ( $\text{H}_2^{18}\text{O}$ ) SIP-metagenomics has the unique benefit of labelling all actively  
85 growing microbes because water is a universal substrate for nucleic acid synthesis [32-34]. Over  
86 adequate time scales, this technique is sensitive enough to label slow-growing or less abundant  
87 microbes and identify taxon-specific microbial growth and mortality patterns [35-38], but  
88 previously, has not been used to label active viruses.

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

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

## 95 **Methods**

### 96 *Experimental setup*

97 We collected soil samples from the Alaska Peatland Experiment (APEX), part of the  
98 Bonanza Creek Long Term Ecological Research (LTER) site southwest of Fairbanks, AK (64.70  
99 °N, 148.3 °W), in a zone of discontinuous permafrost (for overview see Fig. S1). Three peat  
100 soil cores were collected from the active layer of a *Sphagnum*-dominated thermokarst bog on  
101 June 16, 2013 from the beta site approximately 20 m south of the flux tower. During the week of  
102 sampling, midday air temperatures were 27°C, midday surface 10 cm peat temperature was  
103 10°C, active layer depth in the bog was 35 cm, and the water table was 3–8 cm below the  
104 surface. The pH was approximately 4.80 (1:1 soil:water ratio). Vegetation and other soil and flux  
105 data are given in Waldrop et al., [39]. To collect the cores, moss and peat were removed down to  
106 the top of the water table using scissors. The top 20 cm below the water table were collected  
107 using a 6.3 cm diameter sharpened steel barrel corer attached to an electric drill. Three cores  
108 were collected 1 meter apart along a transect. Cores were stored in mason jars filled with  
109 porewater collected from the core hole. Sealed jars were immediately shipped on ice to the U.S.  
110 Geological Survey (Menlo Park, CA, USA) where they were homogenized in an anaerobic  
111 glovebox maintained at 4°C. Anoxic synthetic porewater was created by freeze-drying filtered  
112 (0.45 µm PTFE then 0.2 µm nylon filter) soil porewater collected in tandem with the cores, and  
113 then rehydrating the remaining particulates with either heavy water (97 atom% H<sub>2</sub><sup>18</sup>O,

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

### 130 *CO<sub>2</sub> production quantification*

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

136 purged and then filled with N<sub>2</sub> (1 atm). Gas samples were analyzed via gas chromatography (SRI  
137 8610GC, SRI instruments, Torrance, CA) to quantify headspace CO<sub>2</sub>.

### 138 *DNA extraction and density gradient SIP*

139 DNA was extracted from each replicate using a modified phenol/chloroform protocol  
140 [40]. In summary, 0.5 g (+/- 0.01 g) wet soil was added to lysing matrix E tubes, and 100 µl 1x  
141 TE (pH 7.5), 150 µl PO<sub>4</sub> buffer (0.2 M in 1 M NaCl), and 100 µl 10% SDS were added and  
142 vortexed. Tubes were bead beat for 30 s at 30 1/s and briefly centrifuged. 0.6 ml  
143 phenol:chloroform:isoamyl alcohol (25:24:1) was added, vortexed, and incubated at 65°C for 10  
144 min. Tubes were spun for 5 min at 10K x g, and the supernatant was transferred to a new tube.  
145 The bead beat tubes were then re-extracted using 220 µl 1x TE and 80 µl PO<sub>4</sub> buffers. The  
146 supernatant from the first and second extracts were combined in a new 2 ml tube. 550 µl  
147 phenol/chloroform/isoamyl alcohol was added, mixed, and centrifuged (10K x g, 5 min). The  
148 aqueous top layer was transferred to a new 2 ml tube. 900 µl chloroform:isoamyl alcohol (24:1)  
149 was added, mixed, centrifuged (10K x g, 5 min), and the supernatant transferred to a new 2 ml  
150 tube. Then, 850 µl chloroform:isoamyl alcohol (24:1) was added, mixed, centrifuged (10K x g, 5  
151 min), and the supernatant transferred to a new 1.7 ml tube. RNAase was added (6.44 µl, 10  
152 mg/ml), mixed, and incubated at 50°C for 10 min. 244 µl 10 M NH<sub>4</sub><sup>+</sup> acetate was added, mixed,  
153 and incubated at 4°C for 2 h. Tubes were centrifuged at 16.1K x g for 15 min, and the  
154 supernatant transferred to a new tube. Isopropanol (670 µl) was added, mixed, and centrifuged  
155 (16.1K x g, 20 min). The supernatant was removed, and the DNA pellet dried in a PCR hood for  
156 15 min. 30 µl 1xTE was added, mixed, and the DNA was then stored at □80°C.

157 Extracted DNA was fractionated via CsCl density gradient ultracentrifugation to separate  
158 <sup>18</sup>O-enriched DNA as described previously [35]. DNA was binned into 5 fractions based on



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

### 161 *Sequencing and metagenome generation*

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

### 173 *Recovery of vOTUs from metagenomes*

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

182 using nucmer [45] to generate a nonredundant set of viral populations (vOTUs). vOTU quality  
183 was assessed with CheckV (default parameters) [46]. Coverage of the vOTUs was estimated  
184 based on post quality-controlled read mapping at  $\geq 90\%$  ANI and covering  $\geq 75\%$  of the contig  
185 [44] using Bowtie2 [47]. Coverage was then normalized per gigabase-pair of metagenome and  
186 by length of the contig [48]. Activity of vOTUs was determined by a vOTU being present in the  
187  $\text{H}_2^{18}\text{O}$  samples and not present in the paired natural abundance water samples. The vOTUs  
188 were annotated using the virus-centric multiPhATE pipeline (default parameters) [49] and the  
189 AMG-centric DRAM-v pipeline (with `--skip_uniref`) [50]. We note that DRAM-v provides AMG  
190 scores only for vOTUs detected via VirSorter; AMGs predicted from the 208 vOTUs recovered  
191 from DVF were manually curated. For this manual inspection, we removed putative AMGs that  
192 were at contig ends or those with annotations from multiple functional categories. To determine  
193 the proportion of temperate vOTUs, we searched for genes associated with proviruses, such as  
194 integrase and *parA* [51], leveraged classification from VirSorter (categories 4 or 5) and our  
195 genome-similarity host linkages ( $\leq 90\%$  similarity), and used two tools — CheckV [46] and  
196 PHASTER [52].

### 197 *MAG curation and host linking*

198 To generate metagenome assembled genomes (MAGs), read-pairs from the biological  
199 replicates were grouped for 8 separate co-assemblies (2 timepoints x 2 SIP fractions x 2 O  
200 isotopes) with MEGAHIT v1.1.4 (`--k-min 27 --k-max 127 --k-step 10`) [53]. Contigs greater than  
201 1 KB were separately binned with Concoct v1.0.0 [54], MaxBin v2.2.6 [55] and MetaBAT  
202 v2.12.1 [56]. Genome bins from these three binning tools were refined using the `bin_refinement`

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

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

213 Refined bins from all 12 groups (8 total + 4 active) were dereplicated into a final set of  
214 metagenome assembled genomes (MAGs) using dRep v2.2.3 (-comp 50 -con 10 -p 6 -nc 0.6)  
215 [61]. MAG taxonomy was determined using GTDB-tk v0.3.2 [62] with the version r89 GTDB  
216 database. Structural annotation was done using Patric [63], and general functional annotation  
217 with RASTtk [64] for subsystems within Patric v3 (retaining subsystem variants predicted to be  
218 active or likely), and KofamScan v1.1.0 [65] with the KEGG [66] 93 database. Per-sample MAG  
219 abundances were determined by aligning each sample’s filtered reads against the MAG genomes  
220 using bbmap v38 [67] to obtain total assigned reads per contig and average fold coverage per  
221 contig.

222 For the log-fold-change approach to define active MAGs, we assessed significant (5%  
223 false-discovery rate) and positive log<sub>2</sub>-fold change in <sup>18</sup>O versus <sup>16</sup>O read abundances within a  
224 time point and SIP fraction. Fold changes were determined using wrench-normalized [68] total

225 assigned reads per MAG with a zero-inflated log-normal model implemented in metagenomeSeq  
226 [69].

227 The vOTUs and MAGs were linked via clustered regularly interspaced short palindromic  
228 repeats (CRISPR) spacers and shared genomic content as previously described [70]. Briefly,  
229 CRISPR regions were detected in the MAGs using MinCED (options -minNR 2 -spacers) [71]  
230 and linked to the vOTUs with blastn (options -max\_target\_seqs 10000000 -dust no -word\_size 7)  
231 [72]. In addition, BLAST (options -dust no -perc\_identity 70) was used to link vOTUs and  
232 MAGs based on shared genomic content [73]. Virus-host abundance estimates were made by  
233 summing microbial host abundances at the phylum level in each timepoint and linked vOTU  
234 abundances.

#### 235 *Data availability*

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

## 240 **Results**

### 241 *Characterization of viruses*

242 To characterize viral activity in Arctic peat soil under winter conditions, we analyzed  
243 viral sequences from heavy-water SIP-targeted metagenomic reads. Using two viral detection  
244 methods, we identified 5 737 putative viruses ( $\geq 5$  kb) that clustered into a nonredundant set of  
245 332 vOTUs  $\geq 10$  kb (Table S2) that span 66 viral genera (Fig. S2; Table S3). The size range of  
246 these vOTUs was 10 039 – 437 858 bp (average 32 954 bp) with 15 vOTUs  $\geq 100$  kb, including 5

247 so-called ‘huge’ viruses (i.e.,  $\geq 200$  kb) [74]. The vOTUs were well-covered with an average of  
248 17x normalized coverage per metagenome, but with a large range, 3–147x (Fig. 1A). After  
249 quality checks, we identified 58 medium–high quality vOTUs, of which four were considered  
250 ‘complete’ according to community standards [44]. Genome quality could not be assessed for 93  
251 vOTUs because they did not possess a known viral gene and were detected via a reference-free  
252 machine learning method [43]. Annotation of the vOTU genomes with the virus-centric  
253 Multiphate pipeline predicted 15 772 genes, of which 61% were novel (Table S4). With the  
254 AMG-centric pipeline DRAM-v, we predicted 86 putative AMGs (score 1–3) distilled into five  
255 functional categories (carbon utilization, energy, organic nitrogen, transporters, and  
256 miscellaneous) from 31 vOTUs (Table S5); 21 of the vOTUs were active and carried 63 AMGs  
257 (Fig. 2). To identify temperate viruses, we searched the annotations for genes used in the  
258 lysogenic infection cycle and predicted nearly half (43%) of our vOTUs were temperate viruses.  
259 More than half (59%) of these temperate viruses were active, and the majority (80%) had at least  
260 one member of their population integrated at the time of sampling (provirus; Table S6).

261       Significantly more vOTUs were observed in the SIP fractions from heavy-water  
262 incubations relative to control (natural abundance water) incubations, confirming that many  
263 viruses incorporated  $^{18}\text{O}$  into their DNA and were actively replicating (Fig. S3). A majority  
264 (73%) of the vOTUs found in the dense  $^{18}\text{O}$  SIP fractions were active at some point over the 370-  
265 d incubation, with about half active the full year and the other half only active in 184 d or 370 d  
266 samples (Fig. 1B). We measured active vOTU diversity to assess changes in the viral community  
267 structure from midyear to a full year of incubation. The Shannon’s H metric indicates a  
268 significantly ( $p \leq 0.05$ ) more diverse viral community at 370 d compared to 184 d (Fig. 1C).  
269 Shannon’s H diversity, which includes both richness and evenness, was driven by an average

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

### 273 *Host characterization based on SIP-metagenomics*

274 To identify potential viral hosts, we used a suite of metagenome assembly and binning  
275 methods which yielded 153 medium to high quality MAGs, spanning 16 bacterial phyla and 1  
276 archaeal phylum (Table S7; GTDB taxonomy). The dominant phyla detected were  
277 Chloroflexota, Acidobacteria, and Myxococcota (formally Proteobacteria). Incubation with  
278 heavy water indicated 30% (46) of these MAGs were actively growing, spanning 9 bacterial  
279 phyla, with the most abundant active populations from the Acidobacteria, Bacteroidetes, and  
280 Firmicutes. By sequencing both unlabeled and isotopically labeled samples, we gained insight  
281 into genetic repertoire from both active and inactive bacteria (Table S8) but focused our efforts  
282 on characterizing the active bacterial lineages and metabolisms that viruses may be altering over  
283 winter months. We used CO<sub>2</sub> production measurements to confirm microbial activity and  
284 positive fluxes occurred continuously throughout the  $\square$ 1.5°C incubation, but not from the  $\square$ 20°C  
285 control incubation (Fig. 3).

### 286 *Linking vOTUs to MAG hosts*

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

293 unlabeled vOTU and an active MAG, and 3 linkages between unlabeled vOTUs and unlabeled  
294 MAGs). In a second approach, we used vOTU-MAG genome sequence similarity and recovered  
295 798 similarity matches that linked 141 vOTUs to 65 MAGs from 10 bacterial phyla (Fig. 4;  
296 Table S10). Combined, the two virus-host linkage approaches indicated 318 unique connections  
297 between 145 vOTUs and 65 MAGs spanning 10 bacterial phyla. A significantly higher  
298 proportion of these vOTU-host matches were between active populations (Table 1). Notably,  
299 four vOTUs (#s 153, 161, 162, and 270) had a broad-host range, and were linked to bacteria  
300 from several bacterial phyla (three of these vOTUs infected two phyla and one infecting four  
301 phyla; Fig. 4). Two of these broad-host-range vOTUs (153 and 270) were active and linked to  
302 only active MAGs from bacterial phyla Bacteroidota and Firmicutes.

303 All 145 of the vOTUs we identified as host-linked may therefore be classified as dsDNA  
304 bacteriophages (since the vOTUs were linked to MAGs with a bacterium taxonomic  
305 assignment). These represented the majority (88%) of the host-virus matches, and almost all  
306 (92%) of the unlabeled vOTUs that were linked to unlabeled MAGs. In our soil incubations,  
307 Actinobacteriota (56%), Chloroflexota (24%), and Firmicutes (12%) were the most ‘infected’  
308 bacterial phyla (i.e., with the most vOTU-MAG linkages). Firmicutes (55%), Bacteroidota  
309 (34%), Patescibacteria (9%), and Proteobacteria (2%) were the only phyla that had active MAGs  
310 linked to active vOTUs. Of the active populations, 81 (33%) vOTUs and 33 (51%) MAGs were  
311 linked, with the top 15% most abundant active vOTUs predicted to infect Firmicutes and  
312 Bacteroidota, and the abundances of both these vOTUs and their host populations increased over  
313 the year incubation (Fig. 5).

## 314 **Discussion**

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

### 325 *Viral activity in cryoecosystems*

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



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

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

### 352 *Putative viral influence on winter biogeochemistry*

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

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

365 To better understand the roles viruses play in controlling central biogeochemical  
366 processes in Arctic peat under winter conditions, we assessed the active bacterial phyla that were  
367 virus-infected. Firmicutes had the largest increase in abundance from midyear to a full year and  
368 had the most viral infections (both the number of linked active vOTUs and increase in vOTU  
369 abundance), supporting the ‘kill-the-winner’ theory described in aquatic systems (Fig. 3; Fig. 4)  
370 [85]. In this Lotka–Volterra predator-prey framework, the most abundant host would become  
371 infected by viruses, leading to a decline in host abundance and an increase in the prevalence of  
372 their viruses, at least until microbial hosts with genetic resistance to the currently dominant  
373 viruses had time to respond. The dominance of Firmicutes hosts makes sense, since this diverse  
374 bacterial phylum is known for adaptations to anoxic conditions, including fermentation and  
375 creating endospores [86]. In our samples, all active infected Firmicutes contained fermentation  
376 genes for the capacity to produce ethanol, lactate, or both (Table S8). Bacteroidota were also an  
377 active and frequently infected bacterial phylum that increased in abundance through time; of this  
378 group, all the infected MAGs shared taxonomic affiliation to the order Bacteroidales. Most of  
379 these bacteria are obligate anaerobes and known for their diverse arrays of carbohydrate-active  
380 enzymes arranged into polysaccharide utilization loci and fermentation [87, 88]. Many active  
381 MAGs within this group had the genomic capacity for polysaccharide degradation (e.g., ICE 1  
382 encoded pectin degradation protein KdgF) and all encoded genes for lactate fermentation (Table  
383 S8). After one year of subzero and anoxic incubation, Bacteroidales had become the most

384 abundant active lineage in our soils and were infected by the most abundant active vOTUs (#s  
385 23, 37, 41, 124, and 190), further supporting ‘kill-the-winner’ theory and highlighting viruses  
386 influencing host winter activities. While the abundance of active Bacteroidota and Firmicutes  
387 populations were increasing, C mineralization to CO<sub>2</sub> occurred steadily throughout the  
388 incubation and vOTUs linked to these populations carried several AMGs that would support  
389 central carbon metabolism (Table S5). Together, these results suggest that these two bacterial  
390 lineages play an important role in over-winter C loss from these Arctic peat soils, and the viruses  
391 that infect them likely shape both their population dynamics and functional impact.

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

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

### 413 *Challenges associated with identifying virus 'activity'*

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

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

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

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

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

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

## 464 **Conclusions**

465 Winter carbon losses in Arctic peat soils are estimated to be significant and growing, but  
466 the mechanisms that drive these losses are poorly understood. Using an  $^{18}\text{O}$ -heavy-water  
467 incubation under subzero, anoxic conditions, we found that bacteria and their viruses are active  
468 over the long winter months in northern peatlands, and these active populations drive significant  
469  $\text{CO}_2$  fluxes. Our approach, SIP-targeted metagenomics, allowed us to move beyond a general  
470 catalogue of the genetic repertoire of these soil communities, and expose the specific population-  
471 level dynamics and functional capacities of the activity viral community. Given the high  
472 abundance of unlabeled bacteria and their viruses, these bacterial and viral populations would  
473 have been traditionally described first and foremost, thus occluding the characterization of the  
474 active bacteria and viruses in these soils. Despite the subzero temperatures and lack of oxygen in

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

#### 477 **Acknowledgements**

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

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

- 508 1. Nichols, J.E. and D.M. Peteet, *Rapid expansion of northern peatlands and doubled*  
509 *estimate of carbon storage*. Nature Geoscience, 2019. 12(11): p. 917-921.
- 510 2. Koven, C.D., et al., *A simplified, data-constrained approach to estimate the permafrost*  
511 *carbon-climate feedback*. Philos Trans A Math Phys Eng Sci, 2015. 373(2054).
- 512 3. McGuire, A.D., et al., *Dependence of the evolution of carbon dynamics in the northern*  
513 *permafrost region on the trajectory of climate change*. Proc Natl Acad Sci U S A, 2018.  
514 115(15): p. 3882-3887.
- 515 4. Schuur, E.A., et al., *Climate change and the permafrost carbon feedback*. Nature, 2015.  
516 520(7546): p. 171-9.
- 517 5. Hodgkins, S.B., et al., *Changes in peat chemistry associated with permafrost thaw*  
518 *increase greenhouse gas production*. Proc Natl Acad Sci U S A, 2014. 111(16): p. 5819-24.
- 519 6. Hopple, A.M., et al., *Massive peatland carbon banks vulnerable to rising temperatures*.  
520 Nat Commun, 2020. 11(1): p. 2373.



- 521 7. Parazoo, N.C., et al., *Detecting regional patterns of changing CO<sub>2</sub> flux in Alaska*. Proc  
522 Natl Acad Sci U S A, 2016. 113(28): p. 7733-8.
- 523 8. Wilson, R.M., et al., *Stability of peatland carbon to rising temperatures*. Nat Commun,  
524 2016. 7: p. 13723.
- 525 9. Natali, S.M., et al., *Large loss of CO<sub>2</sub> in winter observed across the northern permafrost*  
526 *region*. Nature Climate Change, 2019. 9(11): p. 852-857.
- 527 10. Przybylak, R., *Diurnal temperature range in the Arctic and its relation to hemispheric and*  
528 *Arctic circulation patterns*. International Journal of Climatology, 2000. 20(3 %@ 0899-  
529 8418): p. 231-253.
- 530 11. Bolduc, B., et al., *The IsoGenie database: an interdisciplinary data management solution*  
531 *for ecosystems biology and environmental research*. PeerJ, 2020. 8: p. e9467.
- 532 12. Männistö, M.K., et al., *Acidobacteriadominate the active bacterial communities of Arctic*  
533 *tundra with widely divergent winter-time snow accumulation and soil temperatures*.  
534 FEMS Microbiology Ecology, 2013. 84(1): p. 47-59.
- 535 13. Murray, A.E., et al., *Microbial life at -13 °C in the brine of an ice-sealed Antarctic lake*.  
536 Proceedings of the National Academy of Sciences, 2012. 109(50): p. 20626-20631.
- 537 14. Drotz, S.H., et al., *Both catabolic and anabolic heterotrophic microbial activity proceed in*  
538 *frozen soils*. Proceedings of the National Academy of Sciences, 2010. 107(49): p. 21046-  
539 21051.
- 540 15. Margesin, R. and T. Collins, *Microbial ecology of the cryosphere (glacial and permafrost*  
541 *habitats): current knowledge*. Applied Microbiology and Biotechnology, 2019. 103(6): p.  
542 2537-2549.

- 543 16. Brum, J.R. and M.B. Sullivan, *Rising to the challenge: accelerated pace of discovery*  
544 *transforms marine virology*. Nature Reviews Microbiology, 2015. 13(3): p. 147-159.
- 545 17. Wilhelm, S.W. and C.A. Suttle, *Viruses and Nutrient Cycles in the Sea*. BioScience, 1999.  
546 49(10): p. 781-788.
- 547 18. Suttle, C.A., *Marine viruses — major players in the global ecosystem*. Nature Reviews  
548 Microbiology, 2007. 5(10): p. 801-812.
- 549 19. Trubl, G., et al., *Coming-of-Age Characterization of Soil Viruses: A User's Guide to Virus*  
550 *Isolation, Detection within Metagenomes, and Viromics*. Soil Systems, 2020. 4(2): p. 23.
- 551 20. Braga, L.P.P., et al., *Impact of phages on soil bacterial communities and nitrogen*  
552 *availability under different assembly scenarios*. Microbiome, 2020. 8(1).
- 553 21. Liang, C., J.P. Schimel, and J.D. Jastrow, *The importance of anabolism in microbial control*  
554 *over soil carbon storage*. Nature Microbiology, 2017. 2(8): p. 17105.
- 555 22. Rosenwasser, S., et al., *Virocell Metabolism: Metabolic Innovations During Host–Virus*  
556 *Interactions in the Ocean*. Trends in Microbiology, 2016. 24(10): p. 821-832.
- 557 23. Thompson, L.R., et al., *Phage auxiliary metabolic genes and the redirection of*  
558 *cyanobacterial host carbon metabolism*. Proceedings of the National Academy of  
559 Sciences, 2011. 108(39): p. E757-E764.
- 560 24. Zimmerman, A.E., et al., *Metabolic and biogeochemical consequences of viral infection in*  
561 *aquatic ecosystems*. Nature Reviews Microbiology, 2020. 18(1): p. 21-34.
- 562 25. Emerson, J.B., et al., *Host-linked soil viral ecology along a permafrost thaw gradient*.  
563 Nature Microbiology, 2018. 3(8): p. 870-880.

- 564 26. Trubl, G., et al., *Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing*.  
565 mSystems, 2018. 3(5).
- 566 27. Barnett, S.E. and D.H. Buckley, *Simulating metagenomic stable isotope probing datasets*  
567 *with MetaSIPSim*. BMC Bioinformatics, 2020. 21(1).
- 568 28. Dumont, M.G. and M. Hernández García, *Stable Isotope Probing*. Methods and Protocols  
569 Totowa, NJ, US: Humana Press, 2019.
- 570 29. Starr, E.P., et al., *Stable isotope informed genome-resolved metagenomics reveals that*  
571 *Saccharibacteria utilize microbially-processed plant-derived carbon*. Microbiome, 2018.  
572 6(1).
- 573 30. Haig, S.-J., et al., *Stable-isotope probing and metagenomics reveal predation by protozoa*  
574 *drives E. coli removal in slow sand filters*. The ISME Journal, 2015. 9(4): p. 797-808.
- 575 31. Gross, A., et al., *The role of soil redox conditions in microbial phosphorus cycling in*  
576 *humid tropical forests*. Ecology, 2020. 101(2).
- 577 32. Aanderud, Z.T. and J.T. Lennon, *Validation of Heavy-Water Stable Isotope Probing for*  
578 *the Characterization of Rapidly Responding Soil Bacteria*. Applied and Environmental  
579 Microbiology, 2011. 77(13): p. 4589-4596.
- 580 33. Blazewicz, S.J. and E. Schwartz, *Dynamics of 18O Incorporation from H<sub>2</sub> 18O into Soil*  
581 *Microbial DNA*. Microbial Ecology, 2011. 61(4): p. 911-916.
- 582 34. Schwartz, E., *Characterization of Growing Microorganisms in Soil by Stable Isotope*  
583 *Probing with H<sub>2</sub>18O*. Applied and Environmental Microbiology, 2007. 73(8): p. 2541-  
584 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 H<sub>2</sub><sup>18</sup>O 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 with <sup>18</sup>O-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.

- 627 54. Alneberg, J., et al., *Binning metagenomic contigs by coverage and composition*. Nature  
628 Methods, 2014. 11(11): p. 1144-1146.
- 629 55. Wu, Y.-W., B.A. Simmons, and S.W. Singer, *MaxBin 2.0: an automated binning algorithm*  
630 *to recover genomes from multiple metagenomic datasets*. Bioinformatics, 2016. 32(4): p.  
631 605-607.
- 632 56. Kang, D.D., et al., *MetaBAT, an efficient tool for accurately reconstructing single*  
633 *genomes from complex microbial communities*. PeerJ, 2015. 3: p. e1165.
- 634 57. Uritskiy, G.V., J. Diruggiero, and J. Taylor, *MetaWRAP—a flexible pipeline for genome-*  
635 *resolved metagenomic data analysis*. Microbiome, 2018. 6(1).
- 636 58. Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from*  
637 *isolates, single cells, and metagenomes*. Genome Research, 2015. 25(7): p. 1043-1055.
- 638 59. Bowers, R.M., et al., *Minimum information about a single amplified genome (MISAG)*  
639 *and a metagenome-assembled genome (MIMAG) of bacteria and archaea*. Nature  
640 Biotechnology, 2017. 35(8): p. 725-731.
- 641 60. Bushnell, B., *BBDuk: Adapter. Quality Trimming and Filtering*.  
642 <http://sourceforge.net/projects/BBMap>.
- 643 61. Olm, M.R., et al., *dRep: a tool for fast and accurate genomic comparisons that enables*  
644 *improved genome recovery from metagenomes through de-replication*. The ISME  
645 Journal, 2017. 11(12): p. 2864-2868.
- 646 62. Chaumeil, P.-A., et al., *GTDB-Tk: a toolkit to classify genomes with the Genome*  
647 *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).

- 711 91. Chen, L.-X., et al., *Candidate Phyla Radiation Roizmanbacteria From Hot Springs Have*  
712 *Novel and Unexpectedly Abundant CRISPR-Cas Systems*. *Frontiers in Microbiology*, 2019.  
713 10.
- 714 92. Coutinho, F.H., et al., *Ecogenomics and metabolic potential of the South Atlantic Ocean*  
715 *microbiome*. *Science of The Total Environment*, 2021. 765: p. 142758.
- 716 93. Bashiardes, S., G. Zilberman-Schapira, and E. Elinav, *Use of Metatranscriptomics in*  
717 *Microbiome Research*. *Bioinformatics and Biology Insights*, 2016. 10: p. BBI.S34610.
- 718 94. Blazewicz, S.J., et al., *Evaluating rRNA as an indicator of microbial activity in*  
719 *environmental communities: limitations and uses*. *The ISME Journal*, 2013. 7(11): p.  
720 2061-2068.
- 721 95. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from*  
722 *proteomic and transcriptomic analyses*. *Nature Reviews Genetics*, 2012. 13(4): p. 227-  
723 232.
- 724 96. Joergensen, R.G. and F. Wichern, *Alive and kicking: Why dormant soil microorganisms*  
725 *matter*. *Soil Biology and Biochemistry*, 2018. 116: p. 419-430.
- 726 97. Stewart, F.M. and B.R. Levin, *The population biology of bacterial viruses: Why be*  
727 *temperate*. *Theoretical Population Biology*, 1984. 26(1): p. 93-117.
- 728 98. Williamson, K.E., et al., *Incidence of lysogeny within temperate and extreme soil*  
729 *environments*. *Environmental Microbiology*, 2007. 9(10): p. 2563-2574.
- 730 99. Hernández, S. and M.J. Vives, *Phages in Anaerobic Systems*. *Viruses*, 2020. 12(10): p.  
731 1091.

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

734

735

736

737

738

739

740

741

742

743

744 **Table 1. Linkages between <sup>18</sup>O labeled and unlabeled viruses and bacteria in a SIP**

745 **incubation of sub-zero, anoxic Arctic peat soils.** Viral operational taxonomic units (vOTUs)

746 were linked to metagenome assembled genomes (MAGs) via nucleotide similarity and using

747 CRISPR spacers. ‘Active’ vOTUs and MAGs were defined based on assimilation of <sup>18</sup>O

748 enriched water (heavy water) into DNA. In total, there were 814 linkages from active vOTUs to

749 active MAGs (group 1), active vOTUs to unlabeled MAGs (group 2), unlabeled vOTUs to active

750 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
<b>Total</b>		<b>814</b>	<b>318</b>	<b>178</b>	<b>80</b>

751

752

753

754

755

756

757

758

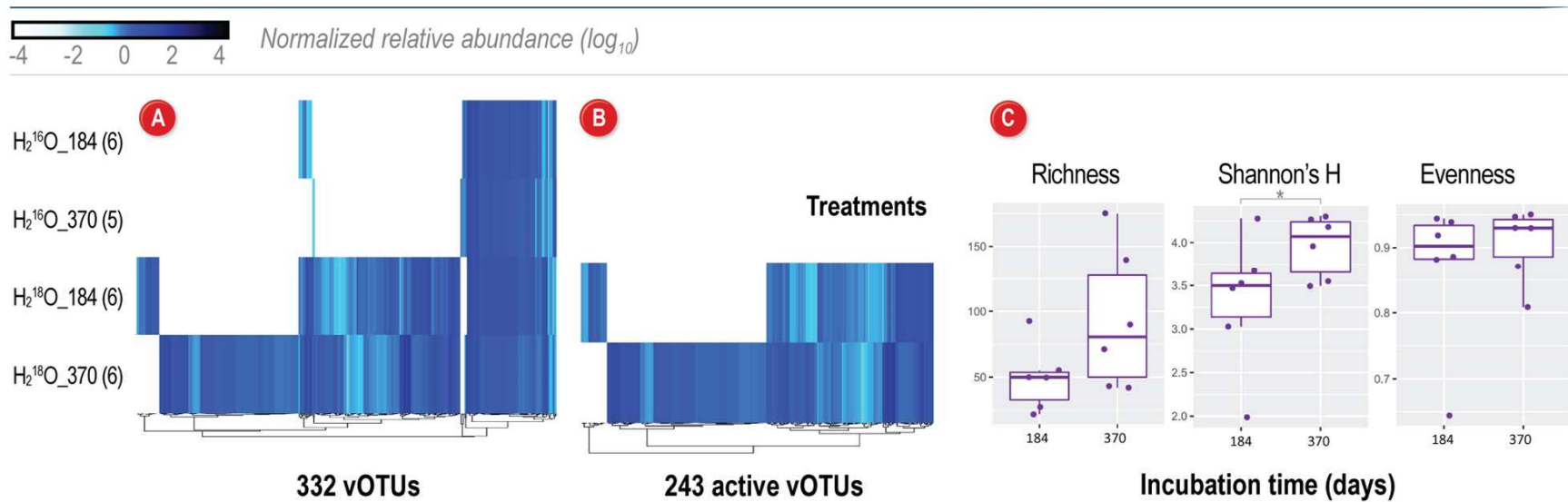
759

760

761

762

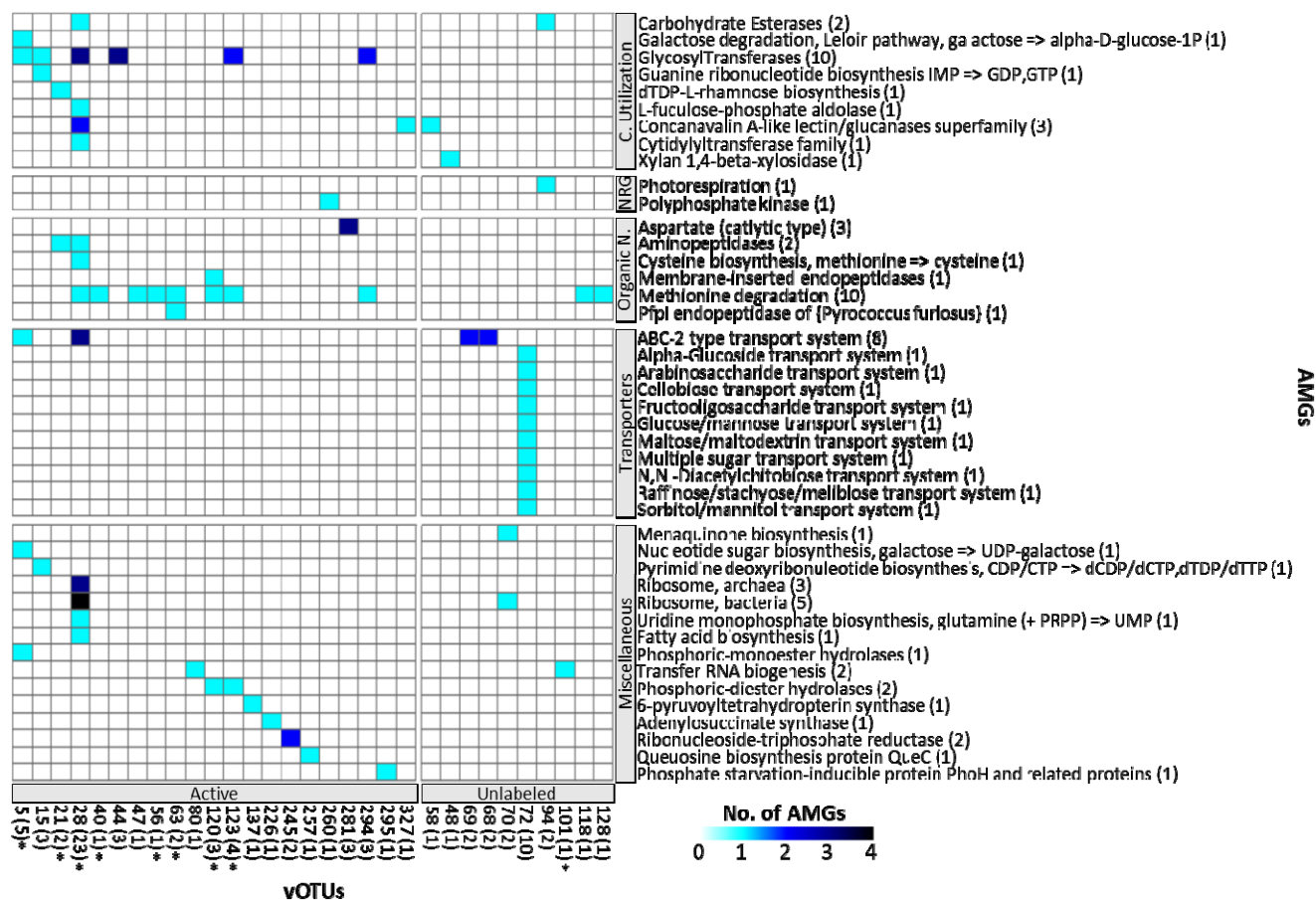
763



764

765 **Figure 1. Assessment of viral community structure and activity after a <sup>18</sup>O water incubation in Arctic peat soils.** (A) Viral  
 766 sequences were identified from 23 samples grouped by two treatments (natural abundance water “H<sub>2</sub><sup>16</sup>O”, and heavy water “H<sub>2</sub><sup>18</sup>O”)  
 767 and two time points: 184 days and 370 days. Number of replicates is indicated in parentheses. Relative abundances of all 332 vOTUs  
 768 identified in the peat soils, clustered by abundances in each treatment/timepoint. (B) Relative abundances of 243 vOTUs considered  
 769 ‘active’ due to DNA <sup>18</sup>O enrichment patterns. Relative abundance for each vOTU (illustrated by blue gradient) was normalized by  
 770 metagenome size (total base pairs) and contig length, and reads were mapped to the contig if they shared ≥90% average nucleotide  
 771 identity and had ≥85% alignment fraction. (C) Diversity metrics for the 243 active vOTUs. Box plots show the median, upper and  
 772 lower quartile range, and the variance among the samples. Asterix denotes significance (p < 0.05).

773



774  
 775 **Figure 2. Putative auxiliary metabolic genes associated with peat soil viral genomes.**  
 776 Heatmap of 31 vOTUs carrying confidently predicted auxiliary metabolic genes (AMGs) (Dram-  
 777 v score 1–3) and their annotation. The vOTUs are grouped by active or unlabeled (see Table S3  
 778 for additional detail) with the sum of AMGs per vOTU indicated in parenthesis and those with an  
 779 asterisk are linked to an active MAG. AMGs are grouped by functional category — carbon  
 780 utilization, energy generation, organic nitrogen use, transporters, and miscellaneous.

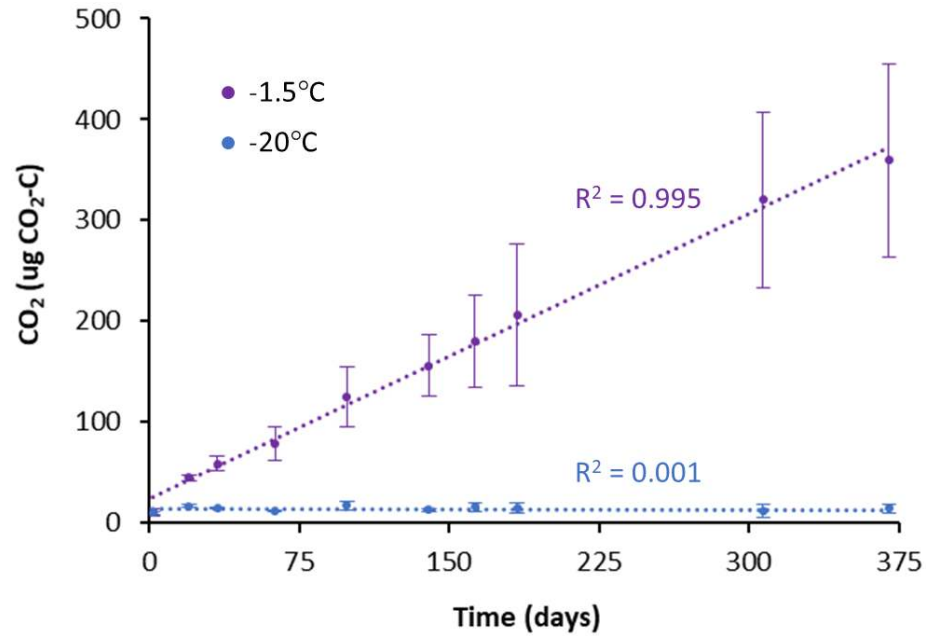
781

782

783

784

785



786

787 **Figure 3. CO<sub>2</sub> production.** Cumulative CO<sub>2</sub> production in soil incubated at □ 1.5°C

788 (experimental conditions) and □ 20°C (control). Error bars show standard error (n=3) and R<sup>2</sup> is

789 shown for each linear regression.

790

791

792

793

794

795

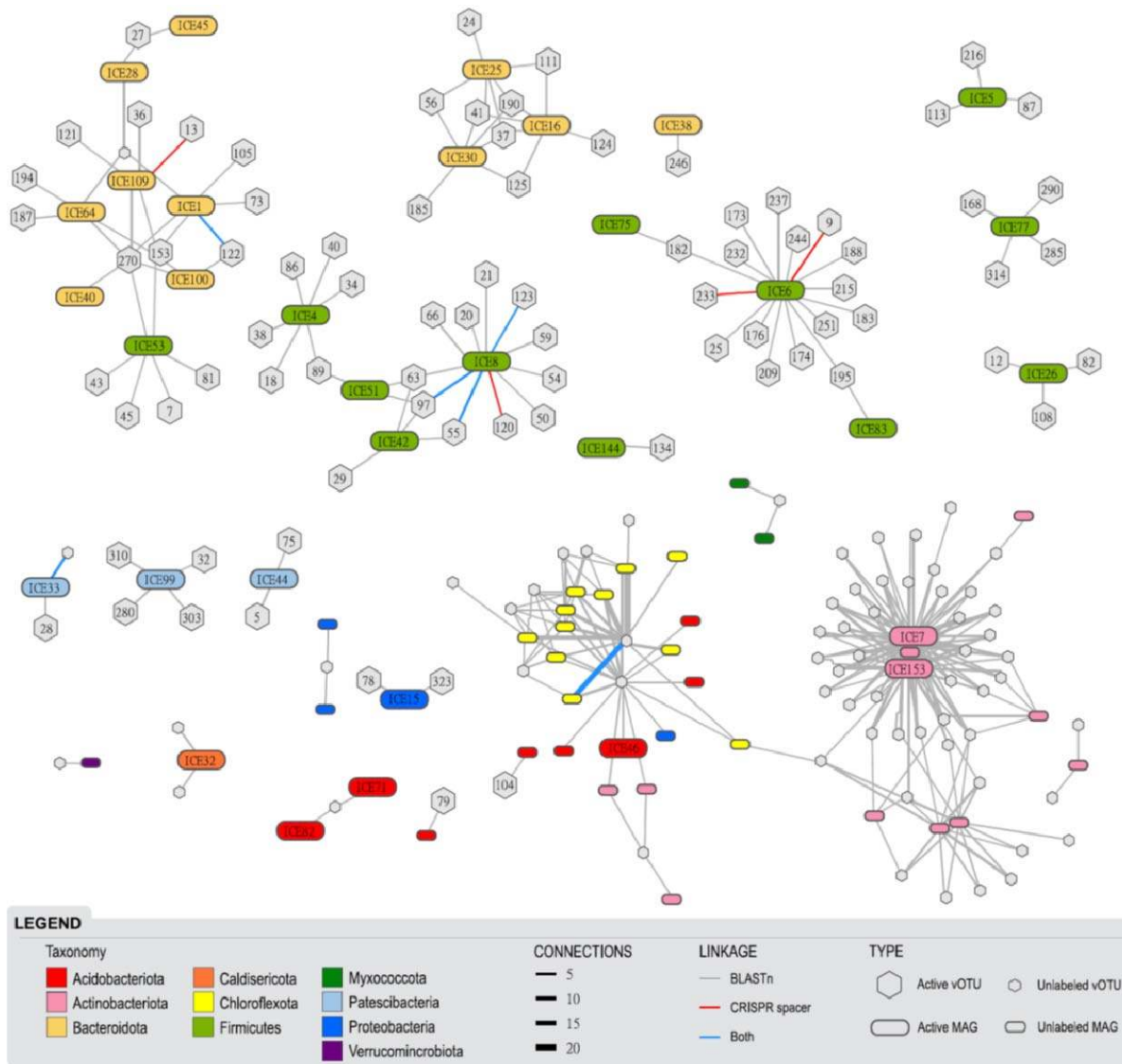
796

797

798

799

800



801

802 **Figure 4. Virus-host linkages in Arctic peat soils incubated with  $^{18}\text{O}$  enriched water.**

803 Network diagram illustrating vOTUs and their predicted bacterial hosts, colored by bacterial

804 phyla. Active versus inactive vOTUs and MAGs (based on  $^{18}\text{O}$  enrichment) are indicated by

805 large named or small hexagon/rounded rectangles, respectively. Lines represent linkages

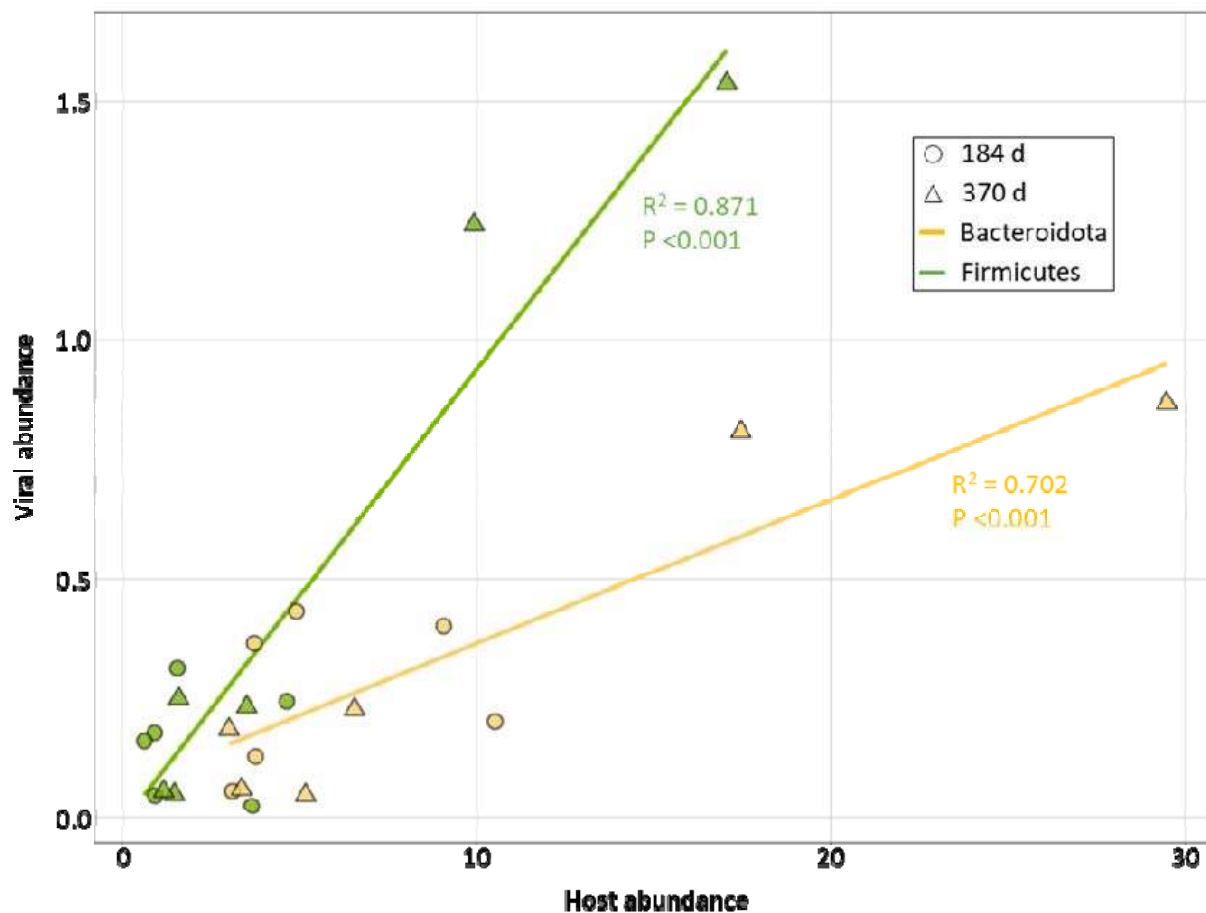
806 between a vOTU and a bacterial MAG, thickness denotes the number of connections, and are

807 colored by the identification approach used: similarity in genomic content (gray), CRISPR

808 spacer match (red), or both (blue).

809





810

811 **Figure 5. Abundances of active viruses and their predicted bacterial hosts.** Average

812 virus:host abundance ratios for bacterial phyla Bacteroidota (yellow) and Firmicutes (green)

813 from heavy-water treatment samples at 184 (n=6) and 370 days (n=6). Host abundance and the

814 abundance of viruses for that host were calculated as the mean coverage depth from

815 metagenomic read mapping, normalized by the number of reads in the sample.

816

817

818