

UC Davis

UC Davis Previously Published Works

Title

Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing

Permalink

<https://escholarship.org/uc/item/6bj6z601>

Journal

mSystems, 3(5)

ISSN

2379-5077

Authors

Trubl, Gareth
Bin Jang, Ho
Roux, Simon
et al.

Publication Date

2018-10-30


DOI

10.1128/msystems.00076-18

Peer reviewed



Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing

 Gareth Trubl,^a Ho Bin Jang,^a Simon Roux,^{a*} Joanne B. Emerson,^{a*} Natalie Solonenko,^a Dean R. Vik,^a Lindsey Solden,^a Jared Ellenbogen,^a Alexander T. Runyon,^a Benjamin Bolduc,^a Ben J. Woodcroft,^b Scott R. Saleska,^c Gene W. Tyson,^b Kelly C. Wrighton,^{a*} Matthew B. Sullivan,^{a,d} Virginia I. Rich^a

^aDepartment of Microbiology, The Ohio State University, Columbus, Ohio, USA

^bAustralian Centre for Ecogenomics, The University of Queensland, St. Lucia, Queensland, Australia

^cDepartment of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA

^dDepartment of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, Ohio, USA

ABSTRACT Rapidly thawing permafrost harbors ~30 to 50% of global soil carbon, and the fate of this carbon remains unknown. Microorganisms will play a central role in its fate, and their viruses could modulate that impact via induced mortality and metabolic controls. Because of the challenges of recovering viruses from soils, little is known about soil viruses or their role(s) in microbial biogeochemical cycling. Here, we describe 53 viral populations (viral operational taxonomic units [vOTUs]) recovered from seven quantitatively derived (i.e., not multiple-displacement-amplified) viral-particle metagenomes (viromes) along a permafrost thaw gradient at the Stordalen Mire field site in northern Sweden. Only 15% of these vOTUs had genetic similarity to publicly available viruses in the RefSeq database, and ~30% of the genes could be annotated, supporting the concept of soils as reservoirs of substantial undescribed viral genetic diversity. The vOTUs exhibited distinct ecology, with different distributions along the thaw gradient habitats, and a shift from soil-virus-like assemblages in the dry palsas to aquatic-virus-like assemblages in the inundated fen. Seventeen vOTUs were linked to microbial hosts (*in silico*), implicating viruses in infecting abundant microbial lineages from *Acidobacteria*, *Verrucomicrobia*, and *Deletaproteobacteria*, including those encoding key biogeochemical functions such as organic matter degradation. Thirty auxiliary metabolic genes (AMGs) were identified and suggested virus-mediated modulation of central carbon metabolism, soil organic matter degradation, polysaccharide binding, and regulation of sporulation. Together, these findings suggest that these soil viruses have distinct ecology, impact host-mediated biogeochemistry, and likely impact ecosystem function in the rapidly changing Arctic.

IMPORTANCE This work is part of a 10-year project to examine thawing permafrost peatlands and is the first virome-particle-based approach to characterize viruses in these systems. This method yielded >2-fold-more viral populations (vOTUs) per gigabase of metagenome than vOTUs derived from bulk-soil metagenomes from the same site (J. B. Emerson, S. Roux, J. R. Brum, B. Bolduc, et al., *Nat Microbiol* 3:870–880, 2018, <https://doi.org/10.1038/s41564-018-0190-y>). We compared the ecology of the recovered vOTUs along a permafrost thaw gradient and found (i) habitat specificity, (ii) a shift in viral community identity from soil-like to aquatic-like viruses, (iii) infection of dominant microbial hosts, and (iv) carriage of host metabolic genes. These vOTUs can impact ecosystem carbon processing via top-down (inferred from lysing dominant microbial hosts) and bottom-up (inferred from carriage of auxiliary metabolic genes) controls. This work serves as a foundation which future studies can build upon to

Received 27 May 2018 Accepted 24 August 2018 Published 2 October 2018


Citation Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, Solden L, Ellenbogen J, Runyon AT, Bolduc B, Woodcroft BJ, Saleska SR, Tyson GW, Wrighton KC, Sullivan MB, Rich VI. 2018. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems* 3:e00076-18. <https://doi.org/10.1128/mSystems.00076-18>.

Editor Seth Bordenstein, Vanderbilt University

Copyright © 2018 Trubl et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Virginia I. Rich, rich.270@osu.edu.

* Present address: Simon Roux, United States Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California, USA; Joanne B. Emerson, Department of Plant Pathology, University of California, Davis, Davis, California, USA; Kelly C. Wrighton, Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA.

 Seven quantitatively-amplified soil viromes reveal viral community structure changes with permafrost thaw and viruses impact ecosystem carbon processing via killing dominant microbial lineages & modulating several carbon metabolic pathways

increase our understanding of the soil virosphere and how viruses affect soil ecosystem services.

KEYWORDS Arctic, carbon cycling, environmental microbiology, microbial ecology, peatlands, permafrost, soil microbiology, soil viromics, viral ecology, viromes

Anthropogenic climate change is elevating global temperatures, most rapidly at the poles (1). High-latitude perennially frozen ground, i.e., permafrost, stores 30 to 50% of global soil carbon (C; $\sim 1,300$ Pg) (2, 3) and is thawing at a rate of ≥ 1 cm of depth yr^{-1} (4, 5). Climate feedbacks from permafrost habitats are poorly constrained in global climate change models (1, 6), due to the uncertainty of the magnitude and nature of carbon dioxide (CO_2) or methane (CH_4) release (7). A model ecosystem for studying the impacts of thaw in a high-C peatland setting is Stordalen Mire, in Arctic Sweden, which is at the southern edge of current permafrost extent (8). The Mire contains a mosaic of thaw stages (9), from intact permafrost palsas, to partially thawed moss-dominated bogs, to fully thawed sedge-dominated fens (10–13). Thaw shifts hydrology (14), altering plant communities (13) and shifting belowground organic matter (OM) toward more labile forms (11, 13), with concomitant shifts in microbiota (15–17) and C gas release (8, 10, 18–20). Of particular note is the thaw-associated increase in emissions of CH_4 , due to its ~ 25 -times-greater climate forcing potential than CO_2 (per kg, at a 100-year time scale [20, 157]), and the associated shifts in key methanogens. These include novel methanogenic lineages (15) with high predictive value for the character of the emitted CH_4 (12). More finely resolving the drivers of C cycling, including microbiota, in these dynamically changing habitats can increase model accuracy (21) to allow a better prediction of greenhouse gas emissions in the future.

Given the central role of microbes to C processing in these systems, it is likely that viruses infecting these microbes impact C cycling, as has been robustly observed in marine systems (22–27). Marine viruses lyse approximately one-third of ocean microorganisms per day, liberating C and nutrients at the global scale (22–24, 28), and viruses have been identified as one of the top predictors of C flux to the deep ocean (29). Viruses can also impact C cycling by metabolically reprogramming their hosts, via the expression of virus-carried auxiliary metabolic genes (AMGs) (28, 30). AMG classification is still in its infancy, with clear definitions still being established (31), but generally these genes are not involved in viral replication and instead allow viruses to directly manipulate host metabolism during infection. This metabolic manipulation potentially affects biogeochemistry, including marine C processing (31–35). In contrast, very little is known about soil virus roles in C processing, or indeed about soil viruses generally. Soils' heterogeneity in texture, mineral composition, and OM content results in significant inconsistency of yields from standard virus capture methods (36–39). While many soils contain large numbers of viral particles (10^7 to 10^9 virus particles per gram of soil [37, 40–42]), knowledge of soil viral ecology has come mainly from the fraction that desorbs easily from soils ($<10\%$ in reference 43) and the much smaller subset that has been isolated (44).

One approach to studying soil viruses has been to bypass the separation of viral particles, by identifying viruses from bulk-soil metagenomes; these are commonly referred to as microbial metagenomes but contain sequences of diverse origin, including proviruses and infecting viruses. Using this approach, several recent studies have powerfully expanded our knowledge of soil viruses and have highlighted the magnitude of genetic novelty that these entities may represent. An analysis of 3,042 publicly available assembled metagenomes spanning 10 ecotypes (19% from soils) increased by 16-fold the total number of known viral genes, doubled the number of microbial phyla with evidence of viral infection, and revealed that the vast majority of viruses appeared to be habitat specific (45). This approach was also applied to 178 metagenomes from the thawing permafrost gradient of Stordalen Mire (46), where viral linkages to potential hosts were appreciably advanced by the parallel recovery of 1,529 microbial metagenome-assembled genomes (MAGs) (17). This effort recovered $\sim 2,000$ thaw-

gradient viruses, more than doubling the known viral genera in RefSeq; identified linkages to abundant microbial hosts encoding important C-processing metabolisms such as methanogenesis; and demonstrated that CH₄ dynamics was best predicted by viruses of methanogens and methanotrophs (46). Viral analyses of bulk-soil metagenomes have, thus, powerfully expanded knowledge of soil viruses and highlighted the large amount of genetic novelty that they represent. However, this approach is by nature inefficient at capturing viral signal, with typically <2% of reads identified as viral (46, 47). The small amount of viral DNA present in bulk-soil extracts can lead to poor or no assembly of viral sequences in the resulting metagenomes and omission from downstream analyses (discussed further in references 37, 39, 48, and 49). In addition, viruses that are captured in bulk-soil metagenomes likely represent a subset of the viral community, since >90% of free viruses adsorb to soil (43), and so, depending on the specific soil, communities, and extraction conditions, bulk-soil metagenomes are likely to be depleted for some free viruses and enriched for actively reproducing and temperate viruses.

Examination of free viruses, while potentially a more efficient and comprehensive approach to soil viral ecology, requires optimized methods to resuspend them (50). Researchers have pursued optimized viral resuspension methods for specific soil types and metagenomically sequenced the recovered viral particles, generating viromes (reviewed in references 40, 42, and 51). In marine systems, viral ecology has relied heavily on viromes, since the leading viral particle capture method is broadly applicable, highly efficient, and relatively inexpensive (52), with now relatively well established downstream pipelines for quantitative sample-to-sequence (53) and sequence-to-ecological-inference (54–56) processing, collectively resulting in great advances in marine viromics (57). Due to the requirement of habitat-specific resuspension optimization, soil viromics is in its early stages. In addition, because particle yields are typically low, most soil virome studies have amplified extracted viral DNA using multiple displacement amplification, which renders the data sets both stochastically and systematically biased and nonquantitative (54, 58–63). The few polar soil viromes have been from Antarctic soils and further demonstrated the genetic novelty of this gene pool while suggesting that resident viral communities were dominated by tailed viruses, had high habitat specificity, and were structured by pH (51, 64, 65).

Having previously optimized viral resuspension methods for the active layer of the permafrost thaw gradient in the Stordalen Mire (41), here we sequenced and analyzed a portion of the viruses recovered from that optimization effort, with no amplification beyond that minor, quantitative form inherent to sequencing library preparation. The seven resulting viromes yielded 378 genuine viral contigs, 53 of which could be classified as viral populations (virus operational taxonomic units [vOTUs]; approximately representing species-level taxonomy) (66–68). The goal of this effort was to efficiently target viral particle genomes via viromes from Stordalen Mire, investigate their ecology and potential impacts on C processing using a variety of approaches, and compare the findings to that of viral analyses of bulk-soil metagenomes from the work of Emerson et al. (46).

RESULTS AND DISCUSSION

Viruses in complex soils. Using recently developed bioinformatics tools to characterize viruses from three different habitats along a permafrost thaw gradient, viral particles were purified from active-layer soil samples (i.e., samples from the upper, unfrozen portion of the soil column) via a previously optimized method tailored for these soils (41) (Fig. 1). DNA from viral particles was extracted and sequenced, to produce seven Stordalen Mire viromes (Table 1), spanning palsas (underlain by intact permafrost), bogs (here underlain by partially thawed permafrost), and fens (where permafrost has thawed entirely). The viromes ranged in size from 2 to 26 million reads, with an average of 19% (range, 5 to 32%) of the reads assembling into 28,025 putative soil viral contigs that clustered into 27,675 unique viral contigs (clustered at 95% average nucleotide identity [ANI] across 80% of the contig length) (2 clustered from 3

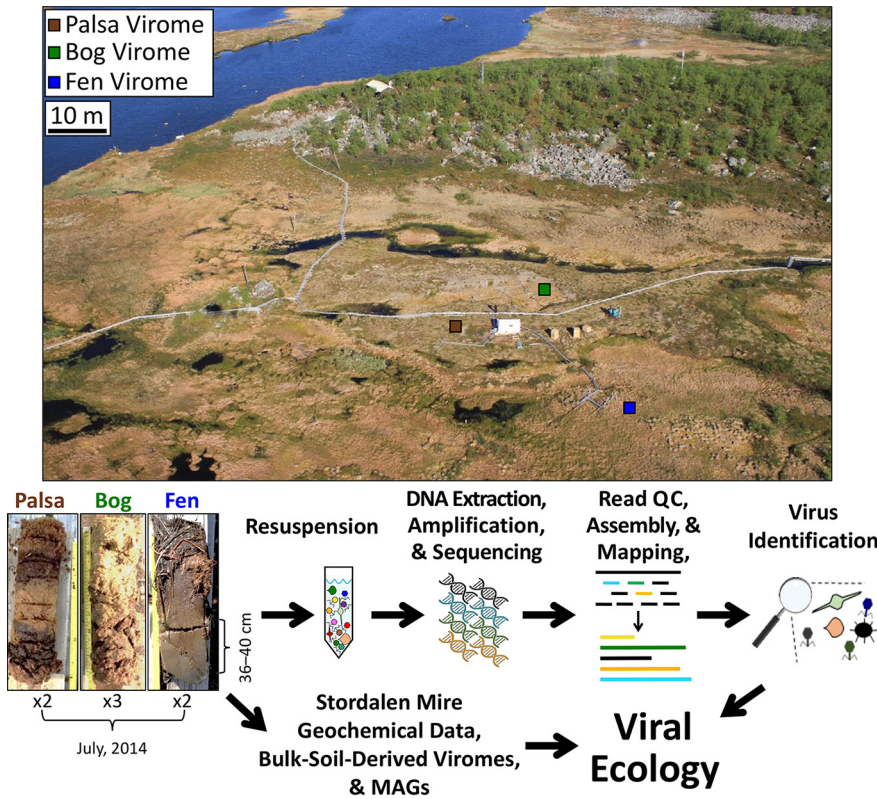


FIG 1 Overview of sample-to-ecology method pipeline. Sampling of the permafrost thaw chronosequence at Stordalen Mire (68°21 N, 19°03 E, 359 m above sea level). Three cores were collected in July 2014; coring locations are indicated on this aerial image of the site (taken in June 2011 by Scott Saleska), and representative photos (taken by Gary Trubl) of cores are shown below. Viruses were resuspended as previously described in the work of Trubl et al. (41), bulk-soil-derived viromes were described in the work of Emerson et al. (46), and metagenome-assembled genomes (MAGs) were described in the work of Woodcroft et al. (17). QC, quality control.

to 10 kb and 348 clustered at <3 kb; see Fig. S1 in the supplemental material). These contigs are putative soil viruses because they passed through a 0.45- μ m filter and remained in the viral fractions for the CsCl densities. Among these, VirSorter predicted that 393 contigs were possible viruses (VirSorter categories 1, 2, and 3 [per reference

TABLE 1 Soil virome read information^a

Sample	BioSample accession no.	DNA quantity (nM)	No. of reads				Avg coverage of 53 vOTUs
			Total	Total assembled	All putative viral	53 vOTU	
Palsa chilled replicate A	SAMN08784142	8.09	10,216,080	540,430	45,972	26,562	15×
Palsa frozen replicate A	SAMN08784143	19.41	26,000,204	1,279,210	70,200	59,800	13×
Bog frozen replicate A	SAMN08784152	9.09	14,499,010	2,611,272	102,944	44,948	14×
Bog frozen replicate B	SAMN08784154	0.8	4,446,734	466,018	53,806	27,126	50×
Bog chilled replicate B	SAMN08784153	2.95	15,578,086	4,210,756	1,190,166	532,770	52×
Fen chilled replicate A	SAMN08784163	0.69	2,108,484	665,226	242,898	182,174	194×
Fen chilled replicate B	SAMN08784165	0.37	2,001,976	649,040	231,428	168,566	160×

^aThe seven viromes are provided along with their DNA quantity, total number of reads, total number of assembled reads, the number of reads that mapped to soil viral contigs, the number of reads that mapped to the 53 vOTUs, and the average adjusted coverage. Adjusted coverage was calculated by mapping reads back to this nonredundant set of contigs to estimate their relative abundance, calculated as number of base pairs mapped to each read normalized by the length of the contig and the total number of base pairs sequenced in the metagenome. For a read to be mapped, it had to have $\geq 90\%$ average nucleotide identity between the read and the contig, and then for a contig to be considered detected, reads had to cover $\geq 75\%$ of the contig.

69]; see Materials and Methods and Table S1). After manual inspection, three putative plasmids were identified and removed (i.e., contigs 5, 394, and 3167 [Table S1]), along with two putatively archaeal viruses (vOTUs 165 and 225; later determined to be bacterial genome fragments; see Text S1 in the supplemental material). Finally, 10 additional contigs that did not meet our threshold for read recruitment of 90% ANI across 75% of contig covered were removed, resulting in 378 probable virus sequences (Table S1). Of these, 53 bacteriophages (phage) were considered well-sampled viral populations (55, 56, 66–68), also known as viral operational taxonomic units (vOTUs), as they had contig lengths of ≥ 10 kb (average, 19.6 kb; range, 10.3 to 129.6 kb), were most robustly viral (VirSorter category 1 or 2 [69]), were clustered at 95%, and were relatively well-covered contigs (average 74 \times coverage [Table 1]). These 53 vOTUs accounted for 10% of the assembled reads and 54% of the reads that recruited to contigs and are the basis for the analyses in this paper due to their genome sizes, which allowed for more reliable taxonomic, functional, and host assignments and fragment recruitment.

There is no universal marker gene (analogous to the 16S rRNA gene in microbes) to provide taxonomic information for viruses. Therefore, we applied a gene-sharing network where nodes were genomes and edges between nodes indicated the gene content similarities and accommodating fragmented genomes of various sizes (70–75). In such networks, viruses sharing a high number of genes localize into viral clusters (VCs) which represent approximately genus-level taxonomy (74, 75). We represented relationships across the 53 vOTUs with 2,010 known bacterial and archaeal viruses (RefSeq, version 75) as a weighted network (Fig. 2A). Only 15% of the Mire vOTUs had similarity to RefSeq viruses (Fig. 2A and B). Four vOTUs fell into 4 VCs comprised of viruses belonging to the *Felixounavirinae* and *Vequintavirinae* (VC10); *Tevenvirinae* and *Eucampyvirinae* (VC4); *Ap22virus* (VC9); and the *Bcep22virus*, *F116virus*, and *Kpp25virus* (VC3) (Fig. 2A). Corroborating its taxonomic assignment by clustering, vOTU_4 contained two marker genes (i.e., major capsid protein and baseplate protein) specific for the *Felixounavirinae* and *Vequintavirinae* viruses (76), phylogenetic analysis of which indicated a close relationship of vOTU_4 to the *Cr3virus* within the *Vequintavirinae* (Fig. S2). The other five populations that clustered with RefSeq viruses were each found in different clusters with taxonomically unclassified viruses (Fig. 2A). Viruses derived from the dry palsa clustered with soil-derived RefSeq viruses, while those from the bog clustered with a mixture of soil and aquatic RefSeq viruses and those from the fen clustered mainly with aquatic viruses (Fig. 2A and C). Though of limited power due to small numbers, this suggests some conservation of habitat preference within genotypic clusters, which has also been observed in marine viruses with only $\sim 4\%$ of VCs being globally ubiquitous (73). Most ($\sim 85\%$) of the Mire vOTUs were unlinked to RefSeq viruses, with 41 vOTUs having no close relatives (i.e., singletons) and the remaining 4 vOTUs clustering in doubletons. This separation between a large fraction of the Mire vOTUs and known viruses is due to a limited number of common genes between them, i.e., $\sim 70\%$ of the total proteins in these viromes are unique (Fig. 2B), reflecting the relative novelty of these viruses and the undersampling of soil viruses (39).

Annotation of the 53 vOTUs resulted in only $\sim 30\%$ of the genes being annotated, which is not atypical; $>60\%$ of genes contained in uncultivated viruses have typically been classified as unknown in other studies (46, 69, 77–81). Of genes with annotations, we first considered those involved in lysogeny, to provide insight into the viruses' replication cycle. Only three viruses carried an integrase gene (other characteristic lysogeny genes were not detected) (82, 83) (Table S2), suggesting that they could be temperate viruses, two of which were from the bog habitat. It had been proposed that since soils are structured and considered harsh environments, a majority of soil viruses would be temperate viruses (84). Although our data set is small, a dominance of temperate viruses is not observed here. We hypothesize that the low encounter rate produced by the highly structured soil environment could, rather than selecting for temperate phage, select for efficient virulent viruses (concept derived from references 85 to 87). Recent analyses of the viral signal mined from bulk-soil metagenomes from this site provide more evidence for our hypothesis of efficient virulent viruses because

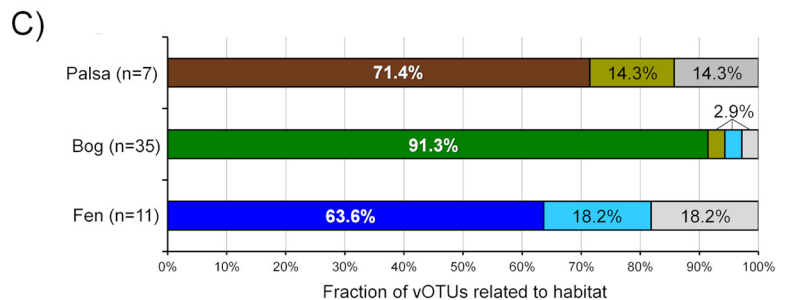
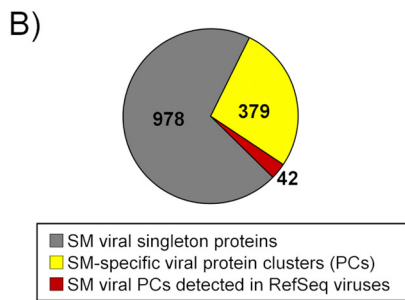
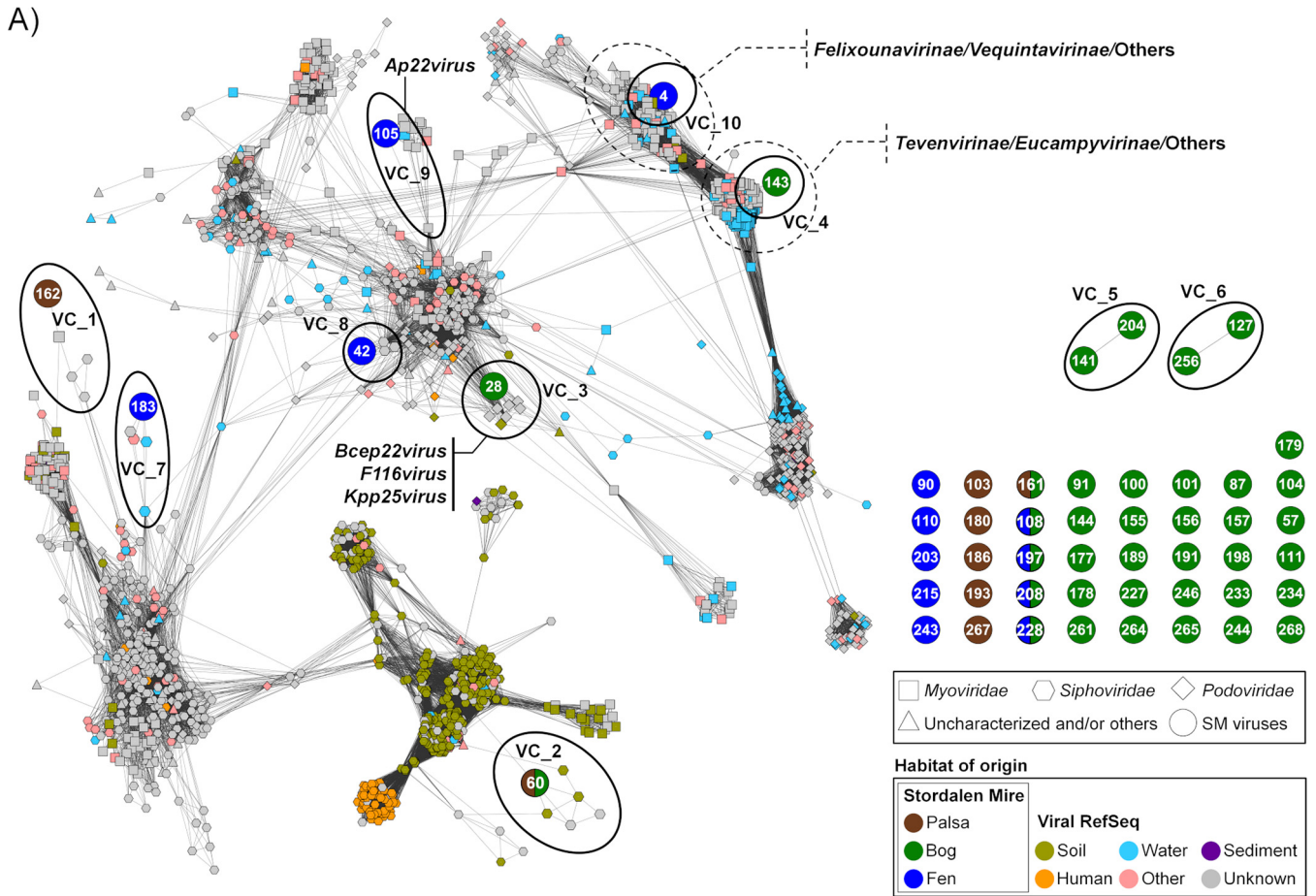
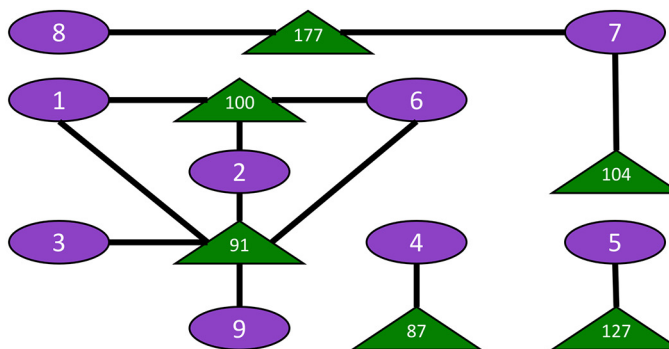


FIG 2 Relating Stordalen Mire viruses to known viral sequence space. (A) Clustering of recovered vOTUs with all RefSeq (v 75) viral genomes or genome fragments with genetic connectivity to these data. Shapes indicate major viral families, and RefSeq sequences only indirectly linked to these data are in gray. The contig numbers are shown within circles. Each node is depicted as a different shape, representing viruses belonging to *Myoviridae* (rectangle), *Podoviridae* (diamond), *Siphoviridae* (hexagon), or uncharacterized viruses (triangle) and viral contigs (circle). Edges (lines) between nodes indicate statistically weighted pairwise similarity scores (see Materials and Methods) of ≥ 1 . Color denotes habitat of origin, with “other” encompassing wastewater, sewage, feces, and plant material. Contig-encompassing viral clusters are encircled by a solid line. (B) The pie chart represents the number of the Stordalen Mire (SM) viral proteins that are recovered by protein clusters (PCs) (yellow and red) and singletons (gray). (C) Percentage of vOTUs that link to those from palsa, bog, and fen as well as RefSeq viruses.

>50% of the identified viruses were likely not temperate (based on the fact that they were not detected as prophage [46]). As a more comprehensive portrait of soil viruses grows, spanning various habitats, this hypothesis can be further tested. Beyond integrase genes, the remaining annotated genes spanned known viral genes and host-like genes. Viral genes included those involved in structure and replication, and their taxonomic affiliations were unknown or highly variable, supporting the quite limited affiliation of these vOTUs with known viruses. Host-like genes included AMGs, which are described in greater detail in the next section.

Verrucomicrobia

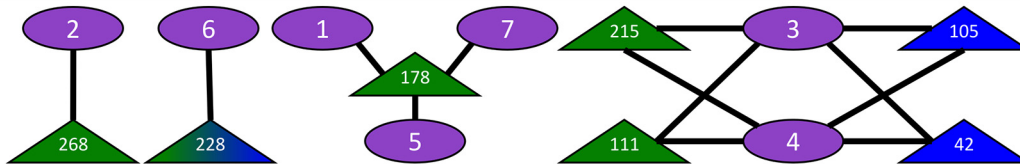
(*Verrucomicrobiae*;
Verrucomicrobiales;
Verrucomicrobia subdivision 3;
Pedosphaera;
Pedosphaera parvula)



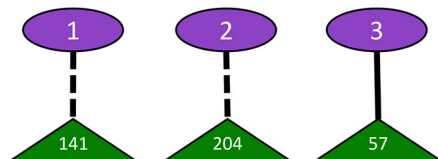
Edge	Evidence
—	CRISPR
- - -	Blast
Node Shape	Organism
△	Virus
○	Microbe
Node Color	Organism
▲	Virus
▲	Virus
▲	Virus
○	Microbe

Acidobacteria

(*Acidobacteriales*;
Acidobacteriaceae;
Acidobacterium)



(*Solibacteres*;
Solibacterales;
Solibacteraceae;
Candidatus Solibacter;
Solibacter usitatus)



Proteobacteria

(*Deltaproteobacteria*;
Syntrophobacteriales;
Syntrophaceae;
Smithella sp. SDB)

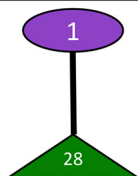


FIG 3 Viral-host linkages between vOTUs and MAGs. Seventeen vOTUs were linked to 4 host lineages by multiple lines of evidence, with 15 linked by CRISPRs (solid line) and 2 linked by BLAST (dashed line). Node shape denotes organism (oval for microbe and triangle for virus), and number references vOTU or bin (see Table S2). Viral nodes are color coded by habitat of origin (green for bog and blue for fen).

Host-linked viruses are predicted to infect key C cycling microbes. In order to examine these viruses' impacts on the Mire's resident microbial communities and processes, we sought to link them to their hosts via emerging standard *in silico* host prediction methods, significantly empowered by the recent recovery of 1,529 MAGs from the site (508 from palsa, 588 from bog, and 433 from fen [17]). Tentative bacterial hosts were identified for 17 of the 53 vOTUs (Fig. 3; Table S3); these hosts spanned four genera among three phyla (*Verrucomicrobia*: *Pedosphaera*, *Acidobacteria*: *Acidobacterium* and "*Candidatus Solibacter*," and *Deltaproteobacteria*: *Smithella*). Eight viruses were linked to more than one host but always within the same species. The four predicted microbial hosts are among the most abundant in the microbial communities and have notable roles in C cycling (16, 17). Three are acidophilic, obligately aerobic chemoorganoheterotrophs and include the Mire's dominant polysaccharide-degrading lineage (*Acidobacteria*), and the fourth is an obligate anaerobe shown to be syntrophic with methanogens (*Smithella*). *Acidobacterium* is a highly abundant, diverse, and ubiquitous soil microbe (88–90) and a member of the most abundant phylum in Stordalen Mire. The relative abundance of this phylum peaked in the bog at 29% but still had a considerably high relative abundance in the other two habitats (5% in palsa and 3% in fen) (17). It is a versatile carbohydrate utilizer, has recently been identified as the primary degrader of large polysaccharides in the palsa and bog habitats in the Mire, and is also an acetogen (17). Seven vOTUs were inferred to infect *Acidobacterium*, implicating these viruses in indirectly modulating a key stage of soil organic matter decomposition. The second identified acidobacterial host was in the newly proposed species "*Candidatus Solibacter usitatus*," another carbohydrate degrader (91). The third

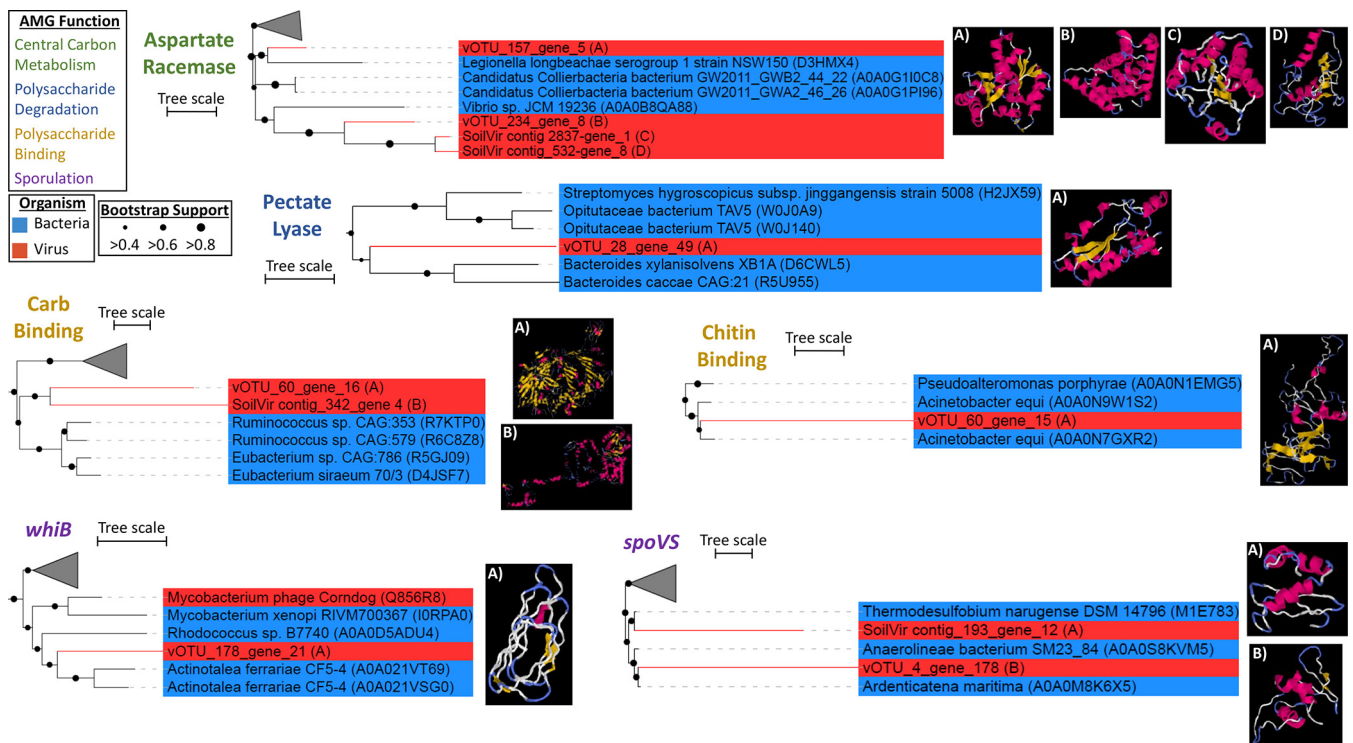


FIG 4 Characterization of select AMGs. FastTree phylogenies were constructed for select AMGs (one from each group), and their structures and those of their nearest neighbors were predicted using I-TASSER (detailed in Table S3). Tree lineages are shaded blue for bacteria and red for viruses. “vOTU” sequences are from the 53-vOTU virome-derived data set, while “SoilVir contig” represents homologs from the 378 probable viral contigs. The predicted protein structure for each AMG is labeled (A, B, C, or D) to match the location in the corresponding phylogenetic tree. The first predicted model for each soil virus is shown and was used for the TM-align comparison. The scale bar indicates the number of substitutions per site.

predicted host was *Pedospaera parvula*, within the phylum *Verrucomicrobia*, which is ubiquitous in soil, is abundant across our soils (~3% in palsa and ~7% in bog and fen habitats, based on metagenomic relative abundance [17]), and utilizes cellulose and sugars (92–96), and in this habitat, this organism could be acetogenic (17). Last, vOTU_28 was linked to the deltaproteobacterium *Smithella* sp. strain SDB, another acidophilic chemoorganoheterotroph but an obligate anaerobe, with a known syntrophic relationship with methanogens (97, 98). Collectively, these virus-host linkages provide evidence for the Mire’s viruses to be impacting the C cycle via population control of relevant C-cycling hosts, consistent with previous results in this system (46) and other wetlands (99).

We next sought to examine viral AMGs for connections to C cycling. To more robustly identify AMGs than by the standard protein family-based search approach, we used a custom-built in-house pipeline previously described in the work of Daly et al. (100) and further tailored to identify putative AMGs based on the metabolisms described in the 1,529 MAGs recently reported from these same soils (17). From this, we identified 30 AMGs from 13 vOTUs (Fig. 4; Tables S2 and S4), encompassing C acquisition and processing (three involved in polysaccharide binding, one involved in polysaccharide degradation, and 23 involved in central C metabolism) and sporulation. Glycoside hydrolases that help break down complex OM are abundant in resident microbiota (17) and may be especially useful in this high-OM environment; notably, they have been found in soil (at our site [46]), rumen (101, 158) and marine systems (albeit scarcely [73]). In addition, central C metabolism genes in viruses may increase nucleotide and energy production during infection and have been increasingly observed as AMGs (25–35). Finally, two different AMGs were found in regulating endospore formation, *spoVS* and *whiB*, which aid in formation of the septum and coat assembly, respectively, improving spores’ heat resistance (102, 103). A WhiB-like protein

has been previously identified in mycobacteriophage TM4 (WhiBTM4) and experimentally shown to not only transcriptionally regulate host septation but also cause superinfection exclusion (i.e., exclusion of secondary viral infections [104]). While these two sporulation genes have been found only in *Firmicutes* and *Actinobacteria*, the only vOTU to have *whiB* was linked to an acidobacterial host (vOTU_178 [Fig. 4]). A phylogenetic analysis of the *whiB* AMG grouped it with actinobacterial versions and, more distantly, with another mycobacteriophage (Fig. 4), suggesting either (i) misidentification of host (unlikely, as it was linked to three different acidobacterial hosts, each with zero mismatches of the CRISPR spacer), (ii) that the virus could infect hosts spanning the two phyla (unlikely, as only ~1% of identified virus-host relationships span phyla [45]), or (iii) that the gene was horizontally transferred into the *Acidobacteria*. Identification of these 30 diverse AMGs (carried by 25% of the vOTUs) suggests a viral modulation of host metabolisms across these dynamic environments and supports the findings from bulk metagenome-derived viruses of Emerson et al. (46) at this site. That study's AMGs spanned the same categories as those reported here, except for *whiB*, which was not found, but the study did not discuss them, other than the glycoside hydrolases, one of which was experimentally validated.

Thus far, the limited studies of soil viruses have identified few AMGs relative to studies of marine environments. This may be due to undersampling or to difficulties in identifying AMGs; since AMGs are homologs of host genes, they can be mistaken for microbial contamination (105) and thus are more difficult to discern in bulk-soil metagenomes (whereas marine virology has been dominated by viromes); also, microbial gene function is more poorly understood in soils (106). Alternately, soil viruses could indeed carry fewer AMGs. One could speculate on a link between host lifestyle and the usefulness of carrying AMGs; most known AMGs are for photo- and chemoautotrophs (73, 107, 108), although this may be due to more studies of these metabolisms or phage-host systems. Thus far, soils are described as dominated by heterotrophic bacteria (109–113), and if AMGs were indeed less useful for viruses infecting heterotrophs, that could explain their limited detection in soil viruses. However, a deeper and broader survey of soil viruses will be required to explore this hypothesis.

Evaluating sample storage on vOTU recovery. While our previous research demonstrated that differing storage conditions (frozen versus chilled) of these Arctic soils did not yield different viral abundances (by direct counts [41]), the impact of storage method on viral community structure was unknown. Here, we examined that in the palsa and bog habitats for which viromes were successfully recovered from both storage conditions. Storage impacted recovered community structure only in the bog habitat, with separation among the viromes' reads (Fig. 5A and B) and a broader recovery of vOTUs from the chilled sample (Fig. 5C and D), leading to higher diversity metrics (Fig. S3) and appreciable separation of the recovered chilled-versus-frozen bog vOTU profiles in ordination (Fig. 5E). The greater recovery of vOTUs from the chilled sample was likely partly due to higher DNA input and sequencing depth, which was 107-fold more than bog frozen replicate A (BFA) and 350-fold more than bog frozen replicate B (BFB). This led to 1.6- to 9-fold more reads assembling into contigs (compared to viromes BFA and BFB, respectively [Table 1]) and 3.5- to 9-fold more distinct contigs; while one might expect that as the number of reads increased, a portion would assemble into already-established contigs, that was not observed. This higher proportional diversity in the chilled bog virome than in the two frozen ones could have several potential causes. Freezing might have decreased viral diversity by damaging viral particles, although these viruses regularly undergo freezing (albeit not with the rapidity of liquid nitrogen). Alternatively, there could be a persistent metabolically active microbial community under the chilled conditions with ongoing viral infections, distinct from those in the field community. Finally, there could have been bog-specific induction of temperate viruses under chilled conditions (since this difference was not seen in the palsa samples). The bog habit is very acidic (pH ~4 versus ~6 in palsa and fen [11, 46]), with a dynamic water table, and both of these have been

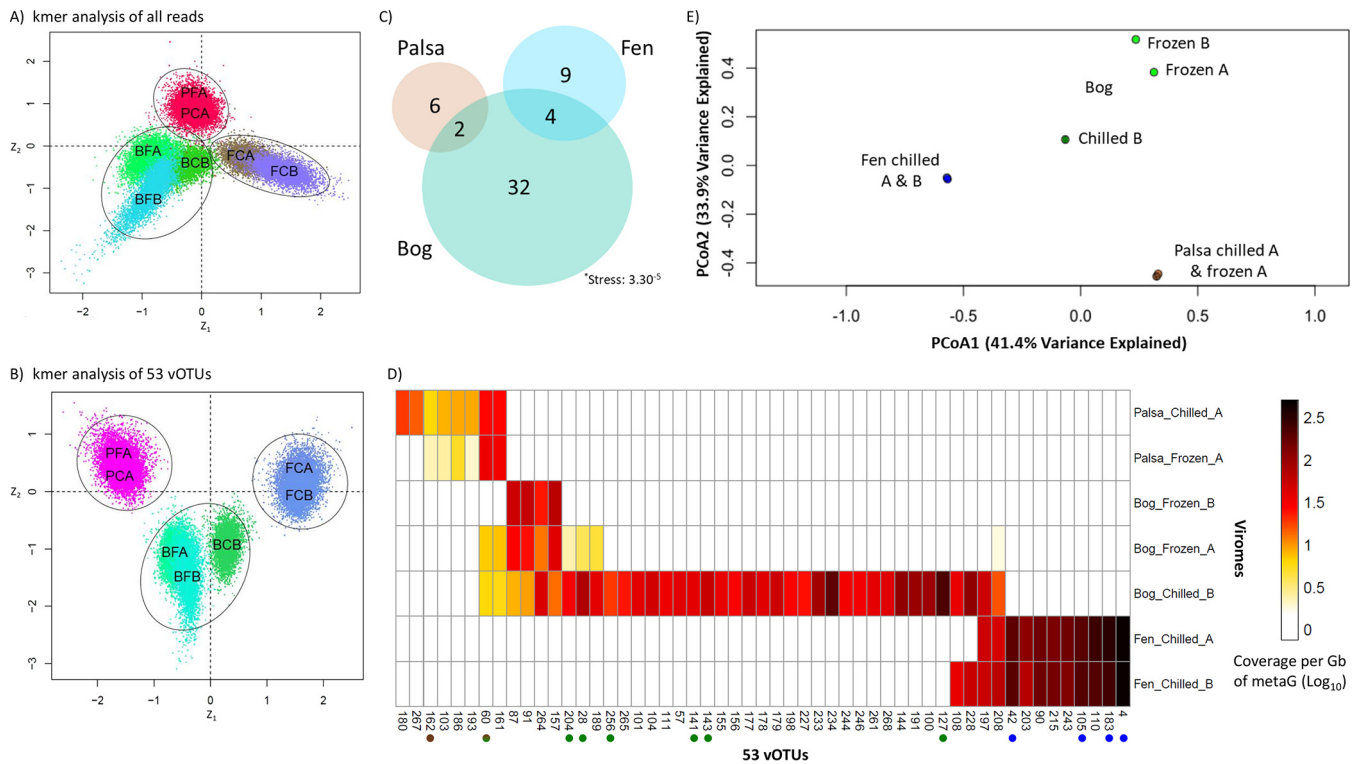


FIG 5 Viral community structure across the thaw gradient. (A and B) Social network analyses of the total reads from the seven viromes (A) and all the reads mapped to the 53 vOTUs (B). Dots in the social networks represent statistical samples taken from the marginal posterior distributions (Bayesian method), and each habitat is denoted by a black circle. (C) Euler diagram relating the seven viromes and their 53 vOTUs. (D) The relative abundance of vOTUs (columns) in the seven viromes (rows). Reads were mapped to this nonredundant set of contigs to estimate their relative abundance. “Chilled” and “Frozen” indicate sample storage at 4°C or flash-freezing in liquid nitrogen and storage at –80°C. “A” and “B” denote technical replicates. Dots after contig names indicate membership in a viral cluster, and fill color indicates habitat specificity. (E) Principal-coordinate analysis of the viromes by normalized relative abundance of the 53 vOTUs.

hypothesized or demonstrated to increase selection for temperate viruses (80, 114–118). In addition, galacturonic acid, a compound produced by the *Sphagnum* moss that dominates the bog habitats and is present at our site, is a known inhibitor of microbial activity and inferred to shape microbial communities (reviewed in references 119 and 120) and is an important bog chemical at this site (11, 121). In addition, of the 19 vOTUs shared between this study and the bulk-soil metagenome study of Emerson et al. (46) (in which the soil was likely to be enriched for temperate viruses based on its majority sampling of microbial DNA), 13 were unique to the bog, and of those, 10 were present in only the chilled rather than frozen viromes and the remaining 3 were enriched in the chilled viromes.

Finally, while the chilled bog sample was an outlier to all other viromes (Fig. 5E), a social network analysis (also known as k-mer analysis) of the total reads (Fig. 5A) and of the reads that mapped to the viromes (Fig. 5B) indicated that habitat remained the primary driver of recovered communities. Because of this, the diversity analyses were redone with the chilled bog sample taken out (Fig. S3B) instead of subsampling the reads (in addition to the virome normalization that we had already done, called “total-sum scaling” [described in references 26, 56, and 74]), because this is a smaller data set (subsampling smaller data sets described further in reference 122) and the storage effect was observed only for the bog.

Habitat specificity of the 53 vOTUs along the thaw gradient. We explored the ecology of the recovered vOTUs across the thaw gradients, by fragment recruitment mapping against (i) the viromes and (ii) bulk-soil metagenomes. Virome mapping revealed that the relative abundance of each habitat’s vOTUs increased along the thaw gradient; relative to the palsa vOTU abundances, bog vOTUs were 3-fold more abundant and fen vOTUs were 12-fold more abundant (Fig. 5D). This is consistent with

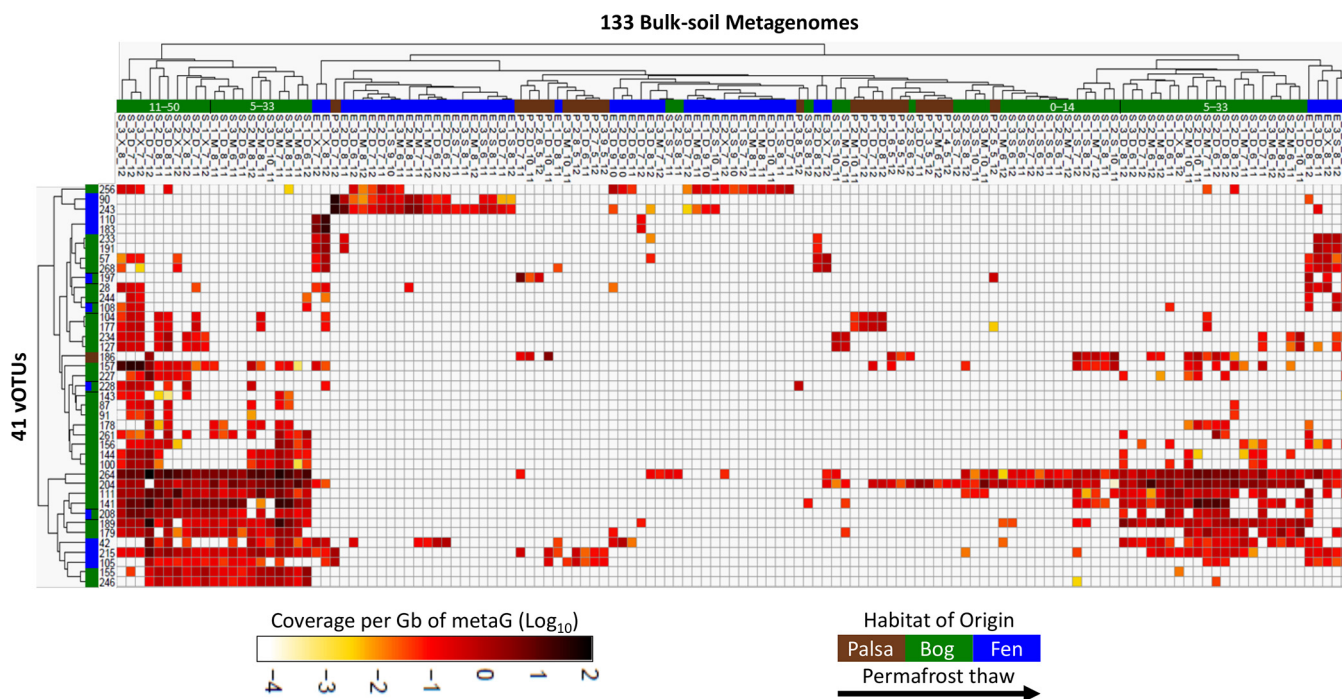


FIG 6 vOTU abundance in 133 bulk-soil metagenomes. The heat map represents abundance of vOTUs (rows) in the bulk-soil metagenomes (columns); metagenome reads were mapped to the nonredundant set of contigs to estimate their relative abundances. Only the 41 vOTUs present in the metagenomes (out of 53) and the 63 bulk-soil metagenomes (out of 214) that contained matches to the vOTUs are shown. Metagenome names denote source: habitat of origin (P, palsa; S, bog; E, fen), soil core replicate (1, 2, or 3), depth (3-cm intervals denoted with respect to geochemical transitions; generally, S = 1 to 4 cm, M = 5 to 14 cm, D = 11 to 33 cm, and X = 30 to 50 cm), month collected (5 to 10 for May to October, respectively), and year collected (2010, 2011, or 2012).

overall increases in virus-like particles with thaw observed previously at the site via direct counts (41). Only a minority (11%) of the vOTUs occurred in more than one habitat, and none were shared between the palsa and fen (Fig. 5C). Consistent with this, principal-coordinate analyses (PCoA; using a Bray-Curtis dissimilarity metric) separated the vOTU-derived community profiles according to habitat type, which also explained ~75% of the variation in the data set (Fig. 5E). Mapping of the 214 bulk-soil metagenomes from the three habitats (17) revealed that a majority (41; 77%) of the vOTUs were present in the bulk-soil metagenomes (Fig. 6), collectively occurring in 62% (123) of them. Of the 41 vOTUs present, most derived from the bog, and their distribution among the 133 metagenomes reflected this, peaking quite dramatically in the bog (Fig. S4). This strong bog signal in the bulk-soil metagenomes—both in the proportion of bog-derived vOTUs present in the bulk metagenomes and in the abundance of all vOTUs in the bog samples—is consistent with the hypothesized higher abundance of temperate viruses in the bog, suggested by the chilled-versus-frozen storage results above. Overall, vOTU abundances in larger and longer-duration bulk-soil metagenomes indicated less vOTU habitat specificity than in the seven viromes: 10% were unique to one habitat, 22% of vOTUs were present in all habitats, 22% were shared between palsa and bog, 27% were shared between palsa and fen, and 68% were shared between bog and fen (Fig. 6). The difference in observations from vOTU read recruitment of viromes and bulk-soil metagenomes could be due to many actual and potential differences, arising from their different source material (but from the same sites) and different methodology, including vOTUs’ actual abundances (they derive from different samples), infection rates, temperate versus lytic states, burst size, and/or virion stability and extractability.

The vOTUs’ habitat preferences observed in both read data sets are consistent with our network analytics (Fig. 2), with the numerous documented physicochemical and biological shifts along the thaw gradient, and with observations of viral habitat specificity at other terrestrial sites. Changes in physicochemistry are known to impact viral

morphology (reviewed in references 37, 124, and 125) and replication strategy (36, 37). In addition, at Stordalen Mire (and at other, similar sites [112]), microbiota are strongly differentiated by thaw-stage habitat, with some limited overlap among “dry” communities (i.e., those above the water table, the palsa and shallow bog) and among “wet” ones (those below the water table, the deeper bog and fen) (15–17). These shifting microbial hosts likely impact viral community structure, creating distinctive viral communities across the palsa, bog, and fen habitats. Expanding from the 53 vOTUs examined here, the recent analysis by Emerson et al. (46) of nearly 2,000 vOTUs recovered from the bulk-soil metagenomes also showed strong habitat specificity among the recovered vOTUs (only 0.1% were shared among all habitats, with <4.5% shared between any two habitats). These findings are also consistent with observations of distinct viral communities from desert, prairie, and rainforests (126) and from grasslands and arctic soils (45). In contrast, an emerging paradigm in the marine field is “seascape ecology” (127), where the majority of taxa are detected across broad geographical areas, as are marine viruses (7, 26). This important difference in habitat specificity between soils and oceans may be due to the greater physical structuring of soil habitats.

Although vOTU richness and diversity appeared to increase along the thaw gradient (roughly equivalent in palsa and bog and ~2-fold higher in fen, omitting the chilled bog sample [Fig. S3B]), this data set captured only a small fraction of the viral diversity (based on a collector’s curve comparison [Fig. S5]), and therefore, the undersampling prevents diversity inferences. Intriguingly, while our virome-derived vOTU richness was lowest in the palsa, the much greater sampling by Emerson et al. (46) recovered the most vOTUs in the palsa, more than double that in the fen (42% versus 18.9% of total vOTUs). This major difference could potentially be due to the known increase in microbial alpha diversity along the thaw gradient (16, 17), causing increased difficulty of viral genome reconstruction in the bulk-soil metagenomes; specifically, this could be due to poorer assembly of temperate phages within an increasingly diverse microbiota or of lytic or free viruses due to concomitantly increasing viral diversity (which is consistent with the increased vOTU richness with thaw in our virome data set). Notably, these diversity inferences are limited to double-stranded DNA (dsDNA) viruses, because our methods did not capture single-stranded DNA (ssDNA) or RNA viruses, a known part of the soil virosphere (128–130).

Challenges in characterizing the soil virosphere. The low yield of viral contigs given the relatively large sequencing depth of the viromes reflects several factors that currently challenge soil viromics. First, resuspending viruses from soils is a challenge due to their adsorption to the soil matrix (43). Second, yields of viral DNA are often very low (due to both low input biomass and potentially low extraction efficiency), requiring amplification; this leads to biases (54, 58–65) or poor assembly and few viral contigs (described further in reference 131). Third, viral contig identification requires a reference database, yet soil viruses are underrepresented in current databases; for example, a majority (85%) of our sequence space was unknown. Fourth, nonviral DNA may coextract (a common feature among marine viromes, as described in reference 34). Last, the optimal approach to identifying ecological units within viral sequence space is unclear.

In this study, DNA yields (and sequencing inputs) decreased along the thaw gradient, as did total reads, but counterintuitively, viral reads increased (Table 1) (the fen had ~5-fold more viral reads than the palsa). This may have been partly due to the shift to a more aquatic-type habitat, for which viruses are better represented in the databases, or to an actual increase in viral DNA (as a portion of total) concomitant with known viral abundance increases (41). A large portion of the assembled reads were nonviral (Table 1), representing either microbial contamination or gene transfer agents (GTAs), i.e., virus-like capsids that package microbial DNA (reviewed in reference 123). Since the viral particle purification protocol involved an 0.45- μm filter followed by CsCl density gradient-based separation of the viral particles (removing free genomic DNA), contam-

ination by microbial DNA seems unlikely. While ultrasmall microbial cells have been found in our soils (46) and other permafrost soils (reviewed in references 132 and 133) and may have passed through the 0.45- μm filters, they would be expected to be removed in the CsCl gradient since their density is similar to that of larger microbial cells and not viruses (reviewed in references 133 to 135). Therefore, to identify GTAs we searched our contigs for 16S rRNA genes and for known GTAs. We found six contigs that had 16S rRNA matches to multiple microbes (136) and 94 contigs with matches to known GTAs (123), together accounting for $\sim 25\%$ of the assembled reads. GTAs may, thus, represent an appreciable and unavoidable contaminant in soil viromes, as has been observed in marine systems (reviewed in reference 123). Against this backdrop of potential contaminant DNA and a preponderance of unknown genes in viral sequence space, identifying ecological units in soil viromes is a challenge. We performed a sensitivity analysis on three ways to characterize the ecological units in our data set: read characterization, contigs, and as vOTUs. While all three methods have validity, there is a higher probability for inclusion of contaminants that can dramatically impact conclusions from the first two approaches. We, therefore, erred on the side of caution and reported our findings in the context of identified vOTUs.

This study's virome-based approach contrasts with that used in the work of Emerson et al. (46), which recovered vOTUs from bulk-soil metagenomes from the same site but with different years, months, depths, and preservation methods (Fig. 7). While the viromes derived from separated viral particles, the bulk-soil metagenomes captured viruses within hosts—i.e., those engaged in active infection and those integrated into hosts—as well as free viruses successfully extracted with the general extraction protocol. This study generated 18 Gb of sequence from 7 viromes, while Emerson et al. (46) analyzed 178 Gb from 190 metagenomes (178 bulk-soil and 12 size-fractionated metagenomes), and based on rarefaction, neither approach captured the total viral diversity in these soils (Fig. S5). The efficiency of vOTU recovery was >2 -fold higher using the virome approach (2.93 vOTUs/Gbp of virome versus 1.30 vOTUs/Gbp of bulk-soil metagenome), suggesting that equivalent virome-focused sequencing effort could yield $>4,300$ vOTUs (although dsDNA viral diversity would likely saturate below that). The efficiency difference is caused not only by enrichment for viral DNA in the viromes but conversely by the higher stringency for vOTU identification required in the complex bulk-soil metagenomes; Emerson et al. (46) required $\geq 95\%$ nucleotide identity across $\geq 90\%$ of each read, and each contig had to have $\geq 70\%$ of the contig covered with $\geq 1\times$ coverage depth, while in these viromes we required $\geq 90\%$ ANI and each contig had to have $\geq 75\%$ of the contig covered with $\geq 1\times$ coverage depth. Of the 19 vOTUs that were shared between the two data sets, the longer, virome-derived sequences defined them. These findings suggest that viromes (which greatly enrich for viral particles) and bulk-soil metagenomes (which are less methodologically intensive and provide simultaneous information on both viruses and microbes) offer complementary views of viral communities in soils, and if only one method can be applied, its selection will depend on the goal of the study.

Over the last 2 decades, viruses have been revealed to be ubiquitous, abundant, and diverse in many habitats, but their role in soils has been underexplored. The observations made here from virome-derived viruses in a model permafrost-thaw ecosystem show that these vOTUs are primarily novel, change with permafrost thaw, and infect hosts highly relevant to C cycling. The next important step is to more comprehensively characterize these viral communities (from more diverse samples and including ssDNA and RNA viruses) and begin quantifying their direct and indirect impacts on C cycling in this changing landscape. This should encompass the complementary information present in virome, bulk metagenomes, and the viral signal from MAGs, analyzed in the context of the abundant metadata available. With increasing characterization of soil viruses and their mechanistic interactions with hosts and quantification of their biogeochemical impacts, soil viral ecology may significantly advance our understanding of terrestrial ecosystem biogeochemical cycling, as has marine viral ecology in the oceans.

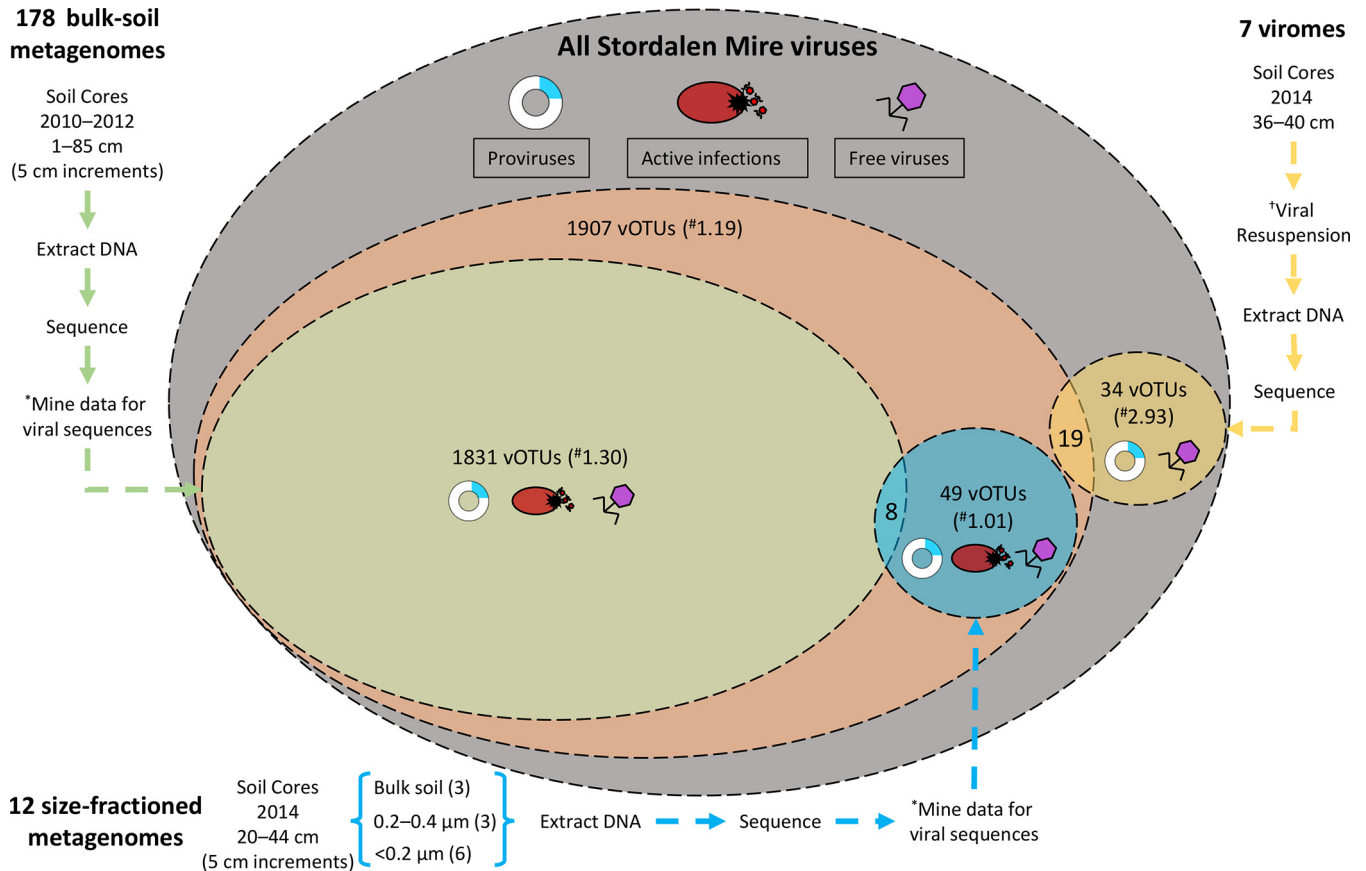


FIG 7 Contrasting Stordalen Mire viruses derived from viromes and bulk-soil metagenomes. Currently, two data sets exist describing Stordalen Mire (SM) archaeal and bacterial viruses. Emerson et al. (46) characterized the viral signal in bulk-soil metagenomes (described in reference 17), while here we characterize viruses from viromes, derived from separated viral particles. There are three possible stages of the viral life cycle at which to capture viruses: proviruses (those integrated into a host genome; blue), active infections (viruses undergoing lytic infection; red), and free viruses (viruses not currently infecting a host; purple). The largest oval represents all the theoretical SM viruses (gray). The next largest oval represents the vOTUs reported in the work of Emerson et al. (46) (orange). Within that oval are the vOTUs derived from bulk-soil metagenomes (green) and from size-fractioned bulk-soil metagenomes also used in that study (blue). The final oval represents the vOTUs identified in this study (yellow circle). An asterisk indicates that the viral signal was mined from bulk-soil metagenomes. A dagger indicates that viruses were resuspended from the soils using a previously optimized protocol (41). A numeral sign indicates the vOTU yield normalized per gigabase-pair of metagenome. The active viruses or proviruses detected in the size-fractioned bulk-soil metagenomes are only those that infect microbial hosts that could pass through the reduced-pore-size filters (more sample information is in reference 46).

MATERIALS AND METHODS

Sample collection. Samples were collected from 16 to 19 July 2014 from peatland cores in the Stordalen Mire field site near Abisko, Sweden (Fig. 1; more site information in references 7 to 13). The soils derived from *palsa* (one stored chilled and the other stored frozen), *bog* (one stored chilled and two stored frozen), and *fen* (both stored chilled) habitats along the Stordalen Mire permafrost thaw gradient. These three subhabitats are common to northern wetlands and together cover ~98% of Stordalen Mire’s nonlake surface (9). The sampled *palsa*, *bog*, and *fen* are directly adjacent, such that all cores were collected within a 120-m total radius. For this work, the cores were subsampled at 36 to 40 cm, and material from each was divided into two sets. Set 1 was chilled and stored at 4°C, and set 2 was flash-frozen in liquid nitrogen and stored at –80°C as described in the work of Trubl et al. (41). Both sets were processed using a viral resuspension method optimized for these soils (41). Briefly, 10 ml of a 1% potassium citrate resuspension buffer amended with 10% phosphate-buffered saline, 5 mM EDTA, and 150 mM magnesium sulfate was added; viruses were physically dispersed using vortexing for 1 min with manual shaking for 30 s (done 3 times) and then shaking of the tubes at 400 rpm for 15 min at 4°C; and finally, the tubes were centrifuged for 20 min at 15,000 × *g* at 4°C to pellet debris, and the supernatant was filtered through an 0.45-μm cellulose acetate vacuum filter (Corning, Corning, NY, USA) into a new 50-ml tube. Three washes were done on each sample. For CsCl density gradient purification of the particles, CsCl density layers of rho 1.2, 1.4, 1.5, and 1.65 were used to establish the gradient; we included a 1.2-g/cm³ CsCl layer to try to remove any small microbial cells that might have come through the 0.45-μm filter (for microbial cell densities, see references 137 and 138; for viral particle densities, see reference 50). We then collected the 1.4- to 1.52-g/cm³ range from the gradient for DNA extraction, to target the dsDNA range (according to reference 50). The viral DNA was extracted using Wizard columns (Promega, Madison, WI; products A7181 and A7211), cleaned up with AMPure beads (Beckman Coulter,

Brea, CA; product A63881), and quantified using a Qubit fluorometer (Invitrogen). DNA libraries were prepared using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA; product FC-131-1024) and sequenced using an Illumina MiSeq (V3; 600 cycles, 6 samples/run, 150-bp paired end) at the University of Arizona Genetics Core facility. Seventeen viral contigs were previously described in the work of Emerson et al. (46) (Fig. 7).

The 214 bulk-soil metagenomes and associated recovered MAGs used here for analyses were described previously (17) and derive from the same sampling sites from 2010 to 2012 and 5-cm increments from 1- to 85-cm depths. They were extracted using a modification of the PowerSoil kit (Qiagen, Hilden, Germany) and sequenced via TruSeq Nano (Illumina) library preparation, or for low-concentration DNA samples, libraries were created using the Nextera XT DNA sample preparation kit (Illumina).

vOTU recovery. Eight viromes were prepared, and seven samples were successfully sequenced (2 palsa, one chilled and one frozen; 3 bog, one chilled and two frozen; and 2 fen, both chilled). The sequences were quality controlled using Trimmomatic (139) (adaptors were removed, reads were trimmed as soon as the per-base quality dropped below 20 on average on 4-nucleotide [nt] sliding windows, and reads shorter than 50 bp were discarded) and then assembled separately with IDBA-UD (140), and contigs were processed with VirSorter to distinguish viral from microbial contigs (virome decontamination mode [69]). The same contigs were also compared by BLAST to a pool of putative laboratory contaminants (i.e., phages cultivated in the lab: *Enterobacteria* phage PhiX17, Alpha3, M13, *Cellulophaga baltica* phages, and *Pseudomonas* phages) that we cultured. All contigs matching these genomes at more than 95% ANI were removed. VirSorted contigs were manually inspected by observing the key features of the viral contigs that VirSorter evaluates (e.g., the presence of a viral hallmark gene places the contigs in VirSorter category 1 or 2, but further inspection is needed to confirm that it is a genuine viral contig and not a GTA or plasmid). To identify GTAs, we searched through all of our contigs assembled by IDBA-UD for (i) taxa related to the 5 types of GTAs (keyword searches were done on *Rhodobacterales*, *Desulfovibrio*, *Brachyspira*, *Methanococcus*, and *Bartonella*) and (ii) microbial DNA from the SILVA rRNA database (release 128 [136]), with all the assembled contigs with $\geq 95\%$ ANI. The percentage of reads that mapped to these contigs is described in Text S1 in the supplemental material.

After verification that the VirSorted contigs were genuine viruses, quality-controlled reads from the seven viromes were pooled and assembled together with IDBA-UD to generate a nonredundant set of contigs. Resulting contigs were rescreened as described above, removing all identifiable contamination. The contigs then underwent further quality checks by (i) removing all contigs of < 10 kb and (ii) using only contigs from VirSorter categories 1 and 2.

To detect putative archaeal viruses, the VirSorter output was used as an input for MARVD (with default settings [141]). The output putative archaeal virus sequences were then filtered to include only those contigs of ≥ 10 kb in size, resulting in the set of putative archaeal vOTUs described here.

Viral genes were annotated using a pipeline described in the work of Daly et al. (100). Briefly, for each contig, open reading frames (ORFs) were freshly predicted using MetaProdigal (142), and sequences were compared to KEGG (143) and UniRef and InterProScan (144) using USEARCH (145), with single and reverse best-hit matches with a bit score greater than 60. AMGs were identified by manual inspection of the protein annotations guided by known resident microbial metabolic functions (identified in reference 17). To determine confidence in functional assignment, representatives for each AMG underwent phylogenetic analyses. First, each sequence was used for a BLAST search and the top 100 hits were investigated to identify main taxon groups. An alignment with the hits and the matching viral sequence (MUSCLE with default parameters [146]) was done with manual curation to refine the alignment (e.g., regions of very low conservation from the beginning or end were removed). FastTree (default parameters with 1,000 bootstraps [147]) was used to make the phylogeny, and iTOL (148) was used to visualize and edit the tree (any distance sequences were removed). To see if this AMG was widespread across the putative soil viruses, a BLASTp search (default settings) of each AMG against all putative viral proteins from our viromes was done. The sequences from identified homologs (based on a bit score of > 70 and an E value of 10^{-4}) were used with the AMG of interest to construct a new phylogenetic tree (same methods as used before). Finally, structures were predicted using I-TASSER (149) for our AMGs of interest and their neighbors. To assess correct structural predictions, AMGs of interest and their neighbors' structures were compared with TM-align (TM-score normalized by the length of the reference protein [150]).

Gene-sharing network construction, analysis, and clustering of viral genomes (fragments). We built a gene-sharing network where the viral genomes and contigs are represented by nodes and significant similarities as edges (74, 75). We downloaded 198,556 protein sequences representing the genomes of 1,999 bacterial and archaeal viruses from NCBI RefSeq (v 75 [151]). Including protein sequences from the 53 Stordalen Mire viral contigs, a total of 199,613 protein sequences were subjected to all-to-all BLASTp searches (default parameters, an E value threshold of 10^{-4} , and a bit score of 50) and defined as protein clusters (PCs) in the same manner as previously described (70). The resulting output was parsed in the form of a matrix comprising the genomes (vOTUs) and PCs. We then determined the similarities between genomes by calculating the probability of finding the number of PCs shared between the genomes and/or vOTUs, based on the hypergeometric formula as previously described (74, 75). A similarity score was obtained by taking the negative logarithm (base 10) of the hypergeometric *P* value multiplied by the total number of pairwise genome (vOTUs) comparisons (i.e., $2,052 \times 2,051$). Genome (vOTU) pairs with a similarity score of ≥ 1 were previously shown to be significantly similar through permutation test of PCs and/or singletons (proteins that do not have close relatives) between

genomes (vOTUs). The resulting network comprising 1,772 viral genomes, including 53 vOTUs and 58,201 edges, was visualized with Cytoscape (version 3.1.1; <http://cytoscape.org/>), using an edge-weighted spring-embedded model, which places the genomes or fragments sharing more PCs closer to each other. There were 398 RefSeq viruses not showing significant similarity to vOTUs that were excluded for clarity. To gain detailed insights into the genetic connections, the network was decomposed into a series of coherent groups of nodes (also known as VCs [70, 72, 73]), with an optimal inflation factor of 1.6. Thus, the discontinuous network structure of individual components, together with the vOTUs, indicates their distinct gene pools (71). To assign vOTUs into VCs, PCs needed to include ≥ 2 genomes and/or genome fragments, then the Markov clustering algorithm was used, and the optimal inflation factor was calculated by exploring values ranging from 1.0 to 5 by steps of 0.2. The taxonomic affiliation was taken from the NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/taxonomy>) and the International Committee on Taxonomy of Viruses (ICTV) taxonomy (ICTV Master Species List v1.3, as of February 2018).

vOTU ecology. Virome reads were mapped back to the nonredundant set of contigs to estimate their coverage, calculated as number of base pairs mapped to each read normalized by the length of the contig, and by the total number of base pairs sequenced in the metagenome in order to be comparable between samples (Bowtie 2, threshold of 90% ANI on the read mapping, and 75% of contig covered to be considered detected [58, 152]). The heat map of the vOTU's relative abundances across the seven viromes, as inferred by read mapping, was constructed in R (CRAN 1.0.8 package pheatmap).

The 214 bulk-soil metagenomes and 1,529 associated recovered MAGs used here for analyses were described in the work of Woodcroft et al. (17). The paired MAG reads were mapped to the viral contigs with Bowtie 2 (as described above for the virome reads). The heat map of the vOTU's relative abundances across the 214 bulk-soil metagenomes, as inferred by read mapping, was constructed in R (CRAN 1.0.8 package pheatmap); only microbial metagenomes with a viral signal were shown.

Viral-host methodologies. We used two different approaches to predict putative hosts for the vOTUs: one relying on CRISPR spacer matches (45, 46, 100, 153) and one relying on direct sequence similarity between virus and host genomes (154). For CRISPR linkages, Crass (v0.3.6, default parameters), a program that searches through raw metagenomic reads for CRISPRs, was used (further information in Table S3) (155). For BLAST, the vOTU nucleotide sequences were compared to the MAGs (17) as described in the work of Emerson et al. (46). Any viral sequences with a bit score of 50, E value threshold of 10^{-3} , and $\geq 70\%$ ANI across $\geq 2,500$ bp were considered for host prediction (described in reference 154).

Phylogenetic analyses to resolve taxonomy. Two phylogenies were constructed. The first had the alignment of the protein sequences that are common to all *Felixounavirinae* and *Vequintavirinae* as well as vOTU_4, and the second had an alignment of select sequences from PC_03881, including vOTU_165. These alignments were generated using the ClustalW implementation in MEGA5 (version 5.2.1; <http://www.megasoftware.net/>). We excluded noninformative positions with the BMGE software package (156). The alignments were then concatenated into a FASTA file, and the maximum likelihood tree was built with MEGA5 using a JTT (Jones-Taylor-Thornton) model for each tree. A bootstrap analysis with 1,000 replications was conducted with uniform rates and a partial depletion of gaps for a 95% site coverage cutoff score.

Data processing and availability. All data (sequences, site information, and supplemental tables and files) are available as a data bundle at the IsoGenie project database under data downloads at <https://isogenie.osu.edu/>. Additionally, viromes were deposited under BioProject identifier (ID) PRJNA445426 and SRA SUB3893166, with the following BioSample accession numbers: SAMN08784142 for palsa chilled replicate A, SAMN08784143 for palsa frozen replicate A, SAMN08784152 for bog frozen replicate A, SAMN08784154 for bog frozen replicate B, SAMN08784153 for bog chilled replicate B, SAMN08784163 for fen chilled replicate A, and SAMN08784165 for fen chilled replicate B. Data were processed using either The Ohio Supercomputer Center services (Columbus, OH) or the CyVerse app (74).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00076-18>.

FIG S1, TIF file, 0.3 MB.

FIG S2, TIF file, 0.5 MB.

FIG S3, TIF file, 0.6 MB.

FIG S4, TIF file, 0.4 MB.

FIG S5, TIF file, 0.6 MB.

FIG S6, TIF file, 0.5 MB.

TABLE S1, XLSX file, 0.04 MB.

TABLE S2, XLSX file, 0.04 MB.

TABLE S3, XLSX file, 0.04 MB.

TABLE S4, XLSX file, 0.01 MB.

TEXT S1, PDF file, 0.3 MB.

ACKNOWLEDGMENTS

We thank Bonnie Poulos and Christine Schirmer for their assistance on different stages of this project. We also thank SWES-MEL, TMPL, The University of Arizona

Genetics Core facility, the MAVERIC lab at the Ohio State University, the Abisko Naturvetenskapliga Station, and the Joint Genome Institute for support. We thank Moira Hough, Robert Jones, and Rachel Wilson for sample collection assistance.

Bioinformatics were supported by The Ohio Supercomputer Center and by the National Science Foundation under award numbers DBI-0735191 and DBI-1265383 (<https://www.cyverse.org>). This study was funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research (grants DE-SC0004632, DE-SC0010580, and DE-SC0016440) and by a Gordon and Betty Moore Foundation Investigator Award (GBMF#3790 to M.B.S.).

REFERENCES

- Allen MR, Barros VR, Broome J, Cramer W, Christ R, Church JA, Clarke L, Dahe Q, Dasgupta P, Dubash NK, Edenhofer O, Elgizouli I, Field CB, Forster P, Friedlingstein P, Fuglestvedt J, Gomez-Echeverri L, Hallegatte S, Hegerl G, Howden M, Jiang K, Jimenez Cisneros B, Kattsov V, Lee H, Mach KJ, Marotzke J, Mastrandrea MD, Meyer L, Minx J, Mulugetta Y, O'Brien K, Oppenheimer M, Pereira JJ, Pichs-Madruga R, Plattner G-K, Pörtner H-O, Power SB, Preston B, Ravindranath NH, Reisinger A, Riahi K, Rusticucci M, Scholes R, Seyboth K, Sokona Y, Stavins R, Stocker TF, Tschakert P, van Vuuren D, van Ypersele J-P. 2014. IPCC fifth assessment synthesis report—climate change 2014 synthesis report. IPCC, Geneva, Switzerland.
- Hugelius G, Strauss J, Zubrzycki S, Harden JW, Schuur EAG, Ping CL, Schirmer L, Grosse G, Michaelson GJ, Koven CD, O'Donnell JA, Eberling B, Mishra U, Camill P, Yu Z, Palmtag J, Kuhry P. 2014. Improved estimates show large circumpolar stocks of permafrost carbon while quantifying uncertainty ranges and identifying remaining data gaps. *Biogeosci Discuss* 11:4771–4822. <https://doi.org/10.5194/bgd-11-4771-2014>.
- Schuur EAG, McGuire AD, Schädel C, Grosse G, Harden JW, Hayes DJ, Hugelius G, Koven CD, Kuhry P, Lawrence DM, Natali SM, Olefeldt D, Romanovsky VE, Schaefer K, Turetsky MR, Treat CC, Vonk JE. 2015. Climate change and the permafrost carbon feedback. *Nature* 520: 171–179. <https://doi.org/10.1038/nature14338>.
- Elberling B, Michelsen A, Schädel C, Schuur EA, Christiansen HH, Berg L, Tamstorf MP, Sigsgaard C. 2013. Long-term CO₂ production following permafrost thaw. *Nat Clim Chang* 3:890–894. <https://doi.org/10.1038/nclimate1955>.
- Shelef E, Rowl JC, Wilson CJ, Hilley GE, Mishra U, Altmann GL, Ping CL. 2017. Large uncertainty in permafrost carbon stocks due to hillslope soil deposits. *Geophys Res Lett* 44:6134–6144. <https://doi.org/10.1002/2017GL073823>.
- Tarnocai C, Canadell JG, Schuur EAG, Kuhry P, Mazhitova G, Zimov S. 2009. Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochem Cycles* 23:GB2023. <https://doi.org/10.1029/2008GB003327>.
- Wilson RM, Fitzhugh L, Whiting GJ, Frolking S, Harrison MD, Dimova N, Burnett WC, Chanton JP. 2017. Greenhouse gas balance over thaw-freeze cycles in discontinuous zone permafrost. *J Geophys Res Biogeosci* 122:387–404. <https://doi.org/10.1002/2016JG003600>.
- Bäckstrand K, Crill PM, Jackowicz-Korczynski M, Mastepanov M, Christensen TR, Bastviken D. 2010. Annual carbon gas budget for a subarctic peatland, northern Sweden. *Biogeosciences* 7:95–108. <https://doi.org/10.5194/bg-7-95-2010>.
- Johansson M, Christensen TR, Akerman HJ, Callaghan TV. 2006. What determines the current presence or absence of permafrost in the Torne-träsk region, a sub-arctic landscape in northern Sweden? *Ambio* 35: 190–197. [https://doi.org/10.1579/0044-7447\(2006\)35\[190:WDTCPJ\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2006)35[190:WDTCPJ]2.0.CO;2).
- Malmer N, Johansson T, Olsrud M, Christensen TR. 2005. Vegetation, climatic changes and net carbon sequestration in a North-Scandinavian subarctic mire over 30 years. *Glob Chang Biol* 11:1895–1909. <https://doi.org/10.1111/j.1365-2486.2005.01042.x>.
- Hodgkins SB, Tfaily MM, McCalley CK, Logan TA, Crill PM, Saleska SR, Rich VI, Chanton JP. 2014. Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proc Natl Acad Sci U S A* 111:5819–5824. <https://doi.org/10.1073/pnas.1314641111>.
- McCalley CK, Woodcroft BJ, Hodgkins SB, Wehr RA, Kim EH, Mondav R, Crill PM, Chanton JP, Rich VI, Tyson GW, Saleska SR. 2014. Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* 514:478–481. <https://doi.org/10.1038/nature13798>.
- Normand AE, Smith AN, Clark MW, Long JR, Reddy KR. 2017. Chemical composition of soil organic matter in a subarctic peatland: influence of shifting vegetation communities. *Soil Sci Soc Am J* 81:41–49. <https://doi.org/10.2136/sssaj2016.05.0148>.
- Torbick N, Persson A, Olefeldt D, Frolking S, Salas W, Hagen S, Crill P, Li C. 2012. High resolution mapping of peatland hydroperiod at a high-latitude Swedish mire. *Remote Sens* 4:1974–1994. <https://doi.org/10.3390/rs4071974>.
- Mondav R, Woodcroft BJ, Kim E-H, McCalley CK, Hodgkins SB, Crill PM, Chanton J, Hurst GB, VerBerkmoes NC, Saleska SR, Hugenholtz P, Rich VI, Tyson GW. 2014. Discovery of a novel methanogen prevalent in thawing permafrost. *Nat Commun* 5:3212. <https://doi.org/10.1038/ncomms4212>.
- Mondav R, McCalley CK, Hodgkins SB, Frolking S, Saleska SR, Rich VI, Chanton JP, Crill PM. 2017. Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. *Environ Microbiol* 19:3201–3218. <https://doi.org/10.1111/1462-2920.13809>.
- Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, Hoelzle RD, Lamberton TO, McCalley CK, Hodgkins SB, Wilson RM, Purvine SO, Nicora CD, Li C, Frolking S, Chanton JP, Crill PM, Saleska SR, Rich VI, Tyson GW. 2018. Genome-centric view of carbon processing in thawing permafrost. *Nature* 560:49–54. <https://doi.org/10.1038/s41586-018-0338-1>.
- Christensen TR, Johansson T, Åkerman HJ, Mastepanov M, Malmer N, Friborg T, Crill P, Svensson BH. 2004. Thawing sub-arctic permafrost: effects on vegetation and methane emissions. *Geophys Res Lett* 31: L04501. <https://doi.org/10.1029/2003GL018680>.
- Christensen TR, Jackowicz-Korczyński M, Aurela M, Crill P, Heliasz M, Mastepanov M, Friborg T. 2012. Monitoring the multi-year carbon balance of a subarctic peatland mire with micrometeorological techniques. *Ambio* 41:207–217. <https://doi.org/10.1007/s13280-012-0302-5>.
- Schädel C, Bader MK-F, Schuur EAG, Biasi C, Bracho R, Čapek P, De Baets S, Diáková K, Ernakovich J, Estop-Aragones C, Graham DE, Hartley IP, Iversen CM, Kane E, Knoblauch C, Lupascu M, Martikainen PJ, Natali SM, Norby RJ, O'Donnell JA, Chowdhury TR, Šantrůčková H, Shaver G, Sloan VL, Treat CC, Turetsky MR, Waldrop MP, Wickland KP. 2016. Potential carbon emissions dominated by carbon dioxide from thawed permafrost soils. *Nat Clim Chang* 6:950. <https://doi.org/10.1038/nclimate3054>.
- Deng J, McCalley C, Frolking S, Chanton J, Crill P, Varner R, Tyson G, Rich V, Saleska S, Hines M, Li C. 2017. Adding stable carbon isotopes improves model representation of the role of microbial communities in peatland methane cycling. *J Adv Model Earth Syst* 9:1412–1430. <https://doi.org/10.1002/2016MS000817>.
- Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399:541–548. <https://doi.org/10.1038/21119>.
- Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–361. <https://doi.org/10.1038/nature04160>.
- Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>.
- Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. 2014. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc Natl Acad Sci U S A* 111: 10714–10719. <https://doi.org/10.1073/pnas.1319778111>.
- Brum JR, Ignacio-Espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, De Vargas C, Gasol JM, Gorsky G, Gregory AC,

- Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandel-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. <https://doi.org/10.1126/science.1261498>.
27. Fridman S, Flores-Urbe J, Larom S, Alalouf O, Liran O, Yacoby I, Salama F, Bailleul B, Rappaport F, Ziv T, Sharon I, Cornejo-Castillo FM, Philofof A, Dupont CL, Sánchez P, Acinas SG, Rohwer FL, Lindell D, Béjà O. 2017. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected prochlorococcus cells. *Nat Microbiol* 2:1350. <https://doi.org/10.1038/s41564-017-0002-9>.
 28. Breitbart M. 2012. Marine viruses: truth or dare. *Annu Rev Mar Sci* 4:425. <https://doi.org/10.1146/annurev-marine-120709-142805>.
 29. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532:465–470. <https://doi.org/10.1038/nature16942>.
 30. Middelboe M, Brussaard CP. 2017. Marine viruses: key players in marine ecosystems. *Viruses* 9:302. <https://doi.org/10.3390/v9100302>.
 31. Yoosof S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H, Mashiyama ST, Joachimiak MP, van Belle C, Chandonia J-M, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC. 2007. The Sorcerer II global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol* 5:e16. <https://doi.org/10.1371/journal.pbio.0050016>.
 32. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulic JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F. 2008. Functional metagenomic profiling of nine biomes. *Nature* 452:629. <https://doi.org/10.1038/nature06810>.
 33. Sharon I, Battchikova N, Aro EM, Giglione C, Meinel T, Glaser F, Pinter RY, Breitbart M, Rohwer F, Béjà O. 2011. Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J* 5:1178. <https://doi.org/10.1038/ismej.2011.2>.
 34. Hurwitz BL, Hallam SJ, Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* 14:R123. <https://doi.org/10.1186/gb-2013-14-11-123>.
 35. Hurwitz BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and taxonomic niche specialization in the ‘core’ and ‘flexible’ Pacific Ocean virome. *ISME J* 9:472–484. <https://doi.org/10.1038/ismej.2014.143>.
 36. Kimura M, Jia ZJ, Nakayama N, Asakawa S. 2008. Ecology of viruses in soils: past, present and future perspectives. *Soil Sci Plant Nutr* 54:1–32. <https://doi.org/10.1111/j.1747-0765.2007.00197.x>.
 37. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. 2017. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. *Annu Rev Virol* 4:201. <https://doi.org/10.1146/annurev-virology-101416-041639>.
 38. Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* 15:579–590. <https://doi.org/10.1038/nrmicro.2017.87>.
 39. Pratama AA, van Elsas JD. 2018. The ‘neglected’ soil virome—potential role and impact. *Trends Microbiol* 26:649–662. <https://doi.org/10.1016/j.tim.2017.12.004>.
 40. Williamson KE, Corzo KA, Drissi CL, Buckingham JM, Thompson CP, Helton RR. 2013. Estimates of viral abundance in soils are strongly influenced by extraction and enumeration methods. *Biol Fertil Soils* 49:857–869. <https://doi.org/10.1007/s00374-013-0780-z>.
 41. Trubl G, Solonenko N, Chittick L, Solonenko SA, Rich VI, Sullivan MB. 2016. Optimization of viral resuspension methods for carbon-rich soils along a permafrost thaw gradient. *PeerJ* 4:e1999. <https://doi.org/10.7717/peerj.1999>.
 42. Narr A, Nawaz A, Wick LY, Harms H, Chatzinotas A. 2017. Soil viral communities vary temporally and along a land use transect as revealed by virus-like particle counting and a modified community fingerprinting approach (FRAPD). *Front Microbiol* 8:1975. <https://doi.org/10.3389/fmicb.2017.01975>.
 43. Goyal SM, Gerba CP. 1979. Comparative adsorption of human enteroviruses, simian rotavirus, and selected bacteriophages to soils. *Appl Environ Microbiol* 38:241–247.
 44. Cresawn SG, Pope WH, Jacobs-Sera D, Bowman CA, Russell DA, Dedrick RM, Adair T, Anders KR, Ball S, Bollivar D, Breitenberger C, Burnett SH, Butela K, Byrnes D, Carzo S, Cornely KA, Cross T, Daniels RL, Dunbar D, Findley AM, Gissendanner CR, Golebiewska UP, Hartzog GA, Hatherill JR, Hughes LE, Jalloh CS, De Los Santos C, Ekanem K, Khambule SL, King RA, King-Smith C, Klyczek K, Krukoni GP, Laing C, Lapin JS, Lopez AJ, Mkhwanazi SM, Molloy SD, Moran D, Munsamy V, Pacey E, Plymale R, Poxleitner M, Reyna N, Schildbach JF, Stuke J, Taylor SE, Ware VC, Wellmann AL, Westholm D, Wodarski D, Zajko M, Zikalala TS, Hendrix RW, Hatfull GF. 2015. Comparative genomics of cluster O mycobacteriophages. *PLoS One* 10:e0118725. <https://doi.org/10.1371/journal.pone.0118725>.
 45. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Hunt-emann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpidis NC. 2016. Uncovering earth’s virome. *Nature* 536:425–430. <https://doi.org/10.1038/nature19094>.
 46. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li CL, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* 3:870–880. <https://doi.org/10.1038/s41564-018-0190-y>.
 47. Goordial J, Davila A, Greer CW, Cannam R, DiRuggiero J, McKay CP, Whyte LG. 2017. Comparative activity and functional ecology of permafrost soils and lithic niches in a hyper-arid polar desert. *Environ Microbiol* 19:443–458. <https://doi.org/10.1111/1462-2920.13353>.
 48. Rosario K, Breitbart M. 2011. Exploring the viral world through metagenomics. *Curr Opin Virol* 1:289–297. <https://doi.org/10.1016/j.coviro.2011.06.004>.
 49. Logares R, Haverkamp TH, Kumar S, Lanzén A, Nederbragt AJ, Quince C, Kauterud H. 2012. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *J Microbiol Methods* 91:106–113. <https://doi.org/10.1016/j.mimet.2012.07.017>.
 50. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4:470–483. <https://doi.org/10.1038/nprot.2009.10>.
 51. Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D. 2014. High-level diversity of tailed phages, eukaryote-associated viruses, and viroplasm-like elements in the metaviromes of Antarctic soils. *Appl Environ Microbiol* 80:6888–6897. <https://doi.org/10.1128/AEM.01525-14>.
 52. John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* 3:195–202. <https://doi.org/10.1111/j.1758-2229.2010.00208.x>.
 53. Duhaime MB, Deng L, Poulos BT, Sullivan MB. 2012. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ Microbiol* 14:2526–2537. <https://doi.org/10.1111/j.1462-2920.2012.02791.x>.
 54. Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, Coleman ML, Breitbart M, Sullivan MB. 2016. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* 4:e2777. <https://doi.org/10.7717/peerj.2777>.
 55. Brum JR, Sullivan MB. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Microbiol* 13:147. <https://doi.org/10.1038/nrmicro3404>.
 56. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817. <https://doi.org/10.7717/peerj.3817>.
 57. Hayes S, Mahony J, Nauta A, van Sinderen D. 2017. Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* 9:127. <https://doi.org/10.3390/v9060127>.
 58. Binga EK, Lasken RS, Neufeld JD. 2008. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* 2:233–241. <https://doi.org/10.1038/ismej.2008.10>.
 59. Yilmaz S, Allgaier M, Hugenholtz P. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods* 7:943–944. <https://doi.org/10.1038/nmeth1210-943>.
 60. Polson SW, Wilhelm SW, Wommack KE. 2011. Unraveling the viral tapestry (from inside the capsid out). *ISME J* 5:165. <https://doi.org/10.1038/ismej.2010.81>.
 61. Kim MS, Whon TW, Bae JW. 2013. Comparative viral metagenomics of

- environmental samples from Korea. *Genomics Inform* 11:121–128. <https://doi.org/10.5808/GI.2013.11.3.121>.
62. Marine R, McCarren C, Vorrasane V, Nasko D, Crowgey E, Polson SW, Wommack KE. 2014. Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* 2:3. <https://doi.org/10.1186/2049-2618-2-3>.
 63. Cremers G, Gambelli L, van Alen T, van Niftrik L, den Camp HJO. 2018. Bioreactor virome metagenomics sequencing using DNA spike-ins. *PeerJ* 6:e4351. <https://doi.org/10.7717/peerj.4351>.
 64. Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D. 2014. Niche-dependent genetic diversity in antarctic metaviromes. *Bacteriophage* 4:e980125. <https://doi.org/10.4161/21597081.2014.980125>.
 65. Adriaenssens EM, Kramer R, Van Goethem MW, Makhalanyane TP, Hogg I, Cowan DA. 2017. Environmental drivers of viral community composition in antarctic soils identified by viromics. *Microbiome* 5:83. <https://doi.org/10.1186/s40168-017-0301-7>.
 66. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, Maitland A, Chittick L, dos Santos F, Weitz JS, Worden AZ, Woyke T, Sullivan MB. 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* 17:930. <https://doi.org/10.1186/s12864-016-3286-x>.
 67. Bobay LM, Ochman H. 2018. Biological species in the viral world. *Proc Natl Acad Sci U S A* 115:6040–6045. <https://doi.org/10.1073/pnas.1717593115>.
 68. Roux S, Adriaenssens EM, Dutlith BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonte JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrahi IK, Ogata H, Paez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodreigues-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton K, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA. Minimum information about an uncultivated virus genome (MIUViG): a community consensus on standards and best practices for describing genome sequences from uncultivated viruses. *Nat Biotechnol*, in press.
 69. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.
 70. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. 2008. Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol Biol Evol* 25:762–777. <https://doi.org/10.1093/molbev/msn023>.
 71. Halary S, Leigh JW, Cheaib B, Lopez P, Baptiste E. 2010. Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci U S A* 107:127–132. <https://doi.org/10.1073/pnas.0908978107>.
 72. Roux S, Hallam SJ, Woyke T, Sullivan MB. 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife* 4:1–20. <https://doi.org/10.7554/eLife.08490>.
 73. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. <https://doi.org/10.1038/nature19366>.
 74. Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB. 2017. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J* 11:7–14. <https://doi.org/10.1038/ismej.2016.89>.
 75. Bolduc B, Jang HB, Doulier G, You ZQ, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that meet archaea and bacteria. *PeerJ* 5:e3243. <https://doi.org/10.7717/peerj.3243>.
 76. Rombouts S, Volckaert A, Venneman S, Declercq B, Vandenheuvel D, Allonsius CN, Van Malderghem C, Jang HB, Briers Y, Noben JP, Klumpff J, Van Vaerenbergh J, Maes M, Lavigne R. 2016. Characterization of novel bacteriophages for biocontrol of bacterial blight in leek caused by *Pseudomonas syringae* pv. *porri*. *Front Microbiol* 7:279. <https://doi.org/10.3389/fmicb.2016.00279>.
 77. Youle M, Haynes M, Rohwer F. 2012. Scratching the surface of biology's dark matter, p 61–81. *In* Witzany G (ed), *Viruses: essential agents of life*. Springer, Dordrecht, Netherlands.
 78. Hatfull GF. 2015. Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J Virol* 89:8107–8110. <https://doi.org/10.1128/JVI.01340-15>.
 79. Waldron PR, Holodniy M. 2015. Peripheral blood mononuclear cell gene expression remains broadly altered years after successful interferon-based hepatitis C virus treatment. *J Immunol Res* 2015:958231. <https://doi.org/10.1155/2015/958231>.
 80. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. 2016. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J* 10:437. <https://doi.org/10.1038/ismej.2015.125>.
 81. Zablocki O, Adriaenssens EM, Cowan D. 2016. Diversity and ecology of viruses in hyperarid desert soils. *Appl Environ Microbiol* 82:770–777. <https://doi.org/10.1128/AEM.02651-15>.
 82. Lamont I, Richardson H, Carter DR, Egan JB. 1993. Genes for the establishment and maintenance of lysogeny by the temperate coliphage 186. *J Bacteriol* 175:5286–5288. <https://doi.org/10.1128/jb.175.16.5286-5288.1993>.
 83. Villafane R, Black J. 1994. Identification of four genes involved in the lysogenic pathway of the *Salmonella newington* bacterial virus ϵ 34. *Arch Virol* 135:179–183. <https://doi.org/10.1007/BF01309776>.
 84. Stewart FM, Levin BR. 1984. The population biology of bacterial viruses: why be temperate. *Theor Popul Biol* 26:93–117. [https://doi.org/10.1016/0040-5809\(84\)90026-1](https://doi.org/10.1016/0040-5809(84)90026-1).
 85. Chibani-Chennoufi S, Bruttin A, Dillmann M-L, Brussow H. 2004. Phage-host interaction: an ecological perspective. *J Bacteriol* 186:3677–3686. <https://doi.org/10.1128/JB.186.12.3677-3686.2004>.
 86. Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE. 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res Microbiol* 159:349–357. <https://doi.org/10.1016/j.resmic.2008.04.010>.
 87. Abedon ST. 2011. Communication among phages, bacteria, and soil environments, p 37–65. *In* Witzany G (ed), *Biocommunication in soil microorganisms*. Springer-Verlag, Berlin, Germany.
 88. Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, Schleper C. 2003. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol Microbiol* 50:563–575. <https://doi.org/10.1046/j.1365-2958.2003.03707.x>.
 89. Foessel BU, Nägele V, Naether A, Wüst PK, Weinert J, Bonkowski M, Lohaus G, Polle A, Alt F, Oelmann Y, Fischer M, Friedrich MW, Overmann J. 2014. Determinants of acidobacteria activity inferred from the relative abundances of 16S rRNA transcripts in German grassland and forest soils. *Environ Microbiol* 16:658–675. <https://doi.org/10.1111/1462-2920.12162>.
 90. Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE. 2016. The ecology of acidobacteria: moving beyond genes and genomes. *Front Microbiol* 7:744. <https://doi.org/10.3389/fmicb.2016.00744>.
 91. Pearce DA, Newsham KK, Thorne MA, Calvo-Bado L, Krsek M, Laskaris P, Hodson A, Wellington EM. 2012. Metagenomic analysis of a southern maritime Antarctic soil. *Front Microbiol* 3:403. <https://doi.org/10.3389/fmicb.2012.00403>.
 92. Janssen PH. 1998. Pathway of glucose catabolism by strain VeGlc2, an anaerobe belonging to the *Verrucomicrobiales* lineage of bacterial descent. *Appl Environ Microbiol* 64:4830–4833.
 93. Kant R, Van Passel MW, Sangwan P, Palva A, Lucas S, Copeland A, Lapidus A, del Rio TG, Dalin E, Tice H, Bruce D. 2011. Genome sequence of *Pedospaera parvula* Ellin514, an aerobic verrucomicrobial isolate from pasture soil. *J Bacteriol* 193:2900–2901. <https://doi.org/10.1128/JB.00299-11>.
 94. Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA, Knight R, Fierer N. 2011. The under-recognized dominance of verrucomicrobia in soil bacterial communities. *Soil Biol Biochem* 43:1450–1455. <https://doi.org/10.1016/j.soilbio.2011.03.012>.
 95. Štursová M, Žifčáková L, Leigh MB, Burgess R, Baldrian P. 2012. Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiol Ecol* 80:735–746. <https://doi.org/10.1111/j.1574-6941.2012.01343.x>.
 96. Soares FL, Jr, Melo IS, Dias ACF, Andreote FD. 2012. Cellulolytic bacteria from soils in harsh environments. *World J Microbiol Biotechnol* 28:2195–2203. <https://doi.org/10.1007/s11274-012-1025-2>.
 97. Schmidt O, Hink L, Horn MA, Drake HL. 2016. Peat: home to novel

- syntrophic species that feed acetate-and hydrogen-scavenging methanogens. *ISME J* 10:1954–1966. <https://doi.org/10.1038/ismej.2015.256>.
98. Wawrik B, Marks CR, Davidova IA, McInerney MJ, Pruitt S, Duncan KE, Sufflita JM, Callaghan AV. 2016. Methanogenic paraffin degradation proceeds via alkane addition to fumarate by ‘Smithella’ spp. mediated by a syntrophic coupling with hydrogenotrophic methanogens. *Environ Microbiol* 18:2604–2619. <https://doi.org/10.1111/1462-2920.13374>.
 99. Juottonen H, Eiler A, Biasi C, Tuittila ES, Yrjälä K, Fritze H. 2017. Distinct anaerobic bacterial consumers of cellobiose-derived carbon in boreal fens with different CO₂/CH₄ production ratios. *Appl Environ Microbiol* 83:e02533-16. <https://doi.org/10.1128/AEM.02533-16>.
 100. Daly RA, Borton MA, Wilkins MJ, Hoyt DW, Kountz DJ, Wolfe RA, Welch SA, Marcus DN, Trexler RV, MacRae JD, Krzycki JA, Cole DR, Mouser PJ, Wrighton KC. 2016. Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nat Microbiol* 1:16146. <https://doi.org/10.1038/nmicrobiol.2016.146>.
 101. Anderson CL, Sullivan MB, Fernando SC. 2017. Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome* 5:155. <https://doi.org/10.1186/s40168-017-0374-3>.
 102. Kormanec J, Homerova D. 1993. Streptomyces aureofaciens whiB gene encoding putative transcription factor essential for differentiation. *Nucleic Acids Res* 21:2512. <https://doi.org/10.1093/nar/21.10.2512>.
 103. Resnekov O, Driks A, Losick R. 1995. Identification and characterization of sporulation gene *spoV*S from *Bacillus subtilis*. *J Bacteriol* 177:5628–5635. <https://doi.org/10.1128/jb.177.19.5628-5635.1995>.
 104. Rybniker J, Nowag A, Van Gumpel E, Nissen N, Robinson N, Plum G, Hartmann P. 2010. Insights into the function of the WhiB-like protein of mycobacteriophage TM4—a transcriptional inhibitor of WhiB2. *Mol Microbiol* 77:642–657. <https://doi.org/10.1111/j.1365-2958.2010.07235.x>.
 105. Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JB. 2016. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* 499:219–229. <https://doi.org/10.1016/j.viro.2016.09.016>.
 106. Jansson JK, Hofmøckel KS. 2018. The soil microbiome—from metagenomics to metaphenomics. *Curr Opin Microbiol* 43:162–168. <https://doi.org/10.1016/j.mib.2018.01.013>.
 107. Thompson MR, Kaminski JJ, Kurt-Jones EA, Fitzgerald KA. 2011. Pattern recognition receptors and the innate immune response to viral infection. *Viruses* 3:920–940. <https://doi.org/10.3390/v3060920>.
 108. Anantharam K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. 2014. Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344:757–760. <https://doi.org/10.1126/science.1252229>.
 109. Martin JK. 1977. Effect of soil moisture on the release of organic carbon from wheat roots. *Soil Biol Biochem* 9:303–304. [https://doi.org/10.1016/0038-0717\(77\)90039-6](https://doi.org/10.1016/0038-0717(77)90039-6).
 110. Floyd MM, Tang J, Kane M, Emerson D. 2005. Captured diversity in a culture collection: case study of the geographic and habitat distributions of environmental isolates held at the American Type Culture Collection. *Appl Environ Microbiol* 71:2813–2823. <https://doi.org/10.1128/AEM.71.6.2813-2823.2005>.
 111. Fierer N, Bradford MA, Jackson RB. 2007. Toward an ecological classification of soil bacteria. *Ecology* 88:1354–1364. <https://doi.org/10.1890/05-1839>.
 112. Jansson JK, Taş N. 2014. The microbial ecology of permafrost. *Nat Rev Microbiol* 12:414–425. <https://doi.org/10.1038/nrmicro3262>.
 113. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. *Science* 359:320–325. <https://doi.org/10.1126/science.aap9516>.
 114. Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S, McDermott T, Young MJ. 2001. Viruses from extreme thermal environments. *Proc Natl Acad Sci U S A* 98:13341–13345. <https://doi.org/10.1073/pnas.231170198>.
 115. Laybourn-Parry J, Marshall WA, Madan NJ. 2007. Viral dynamics and patterns of lysogeny in saline antarctic lakes. *Polar Biol* 30:351–358. <https://doi.org/10.1007/s00300-006-0191-9>.
 116. Le Romancer M, Gaillard M, Geslin C, Prieur D. 2007. Viruses in extreme environments. *Rev Environ Sci Biotechnol* 6:17–31. <https://doi.org/10.1007/s11157-006-0011-2>.
 117. Evans C, Brussaard CP. 2012. Regional variation in lytic and lysogenic viral infection in the Southern Ocean and its contribution to biogeochemical cycling. *Appl Environ Microbiol* 78:6741–6748. <https://doi.org/10.1128/AEM.01388-12>.
 118. Payet JP, Suttle CA. 2013. To kill or not to kill: the balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol Oceanogr* 58:465–474. <https://doi.org/10.4319/lo.2013.58.2.0465>.
 119. Eppinga MB, Rietkerk M, Wassen MJ, De Ruiter PC. 2009. Linking habitat modification to catastrophic shifts and vegetation patterns in bogs. *Plant Ecol* 200:53–68. <https://doi.org/10.1007/s11258-007-9309-6>.
 120. Weston DJ, Turetsky MR, Johnson MG, Granath G, Lindo Z, Belyea LR, Rice SK, Hanson DT, Engelhardt KAM, Schmutz J, Dorrepaal E, Euskirchen ES, Stenøien HK, Szóvényi P, Jackson M, Piatkowski BT, Muchero W, Norby RJ, Kostka JE, Glass JB, Rydin H, Limpens J, Tuittila E-S, Ullrich KK, Carrell A, Benscoter BW, Chen J-G, Oke TA, Nilsson MB, Ranjan P, Jacobson D, Lilleskov EA, Clymo RS, Shaw AJ. 2018. The sphagnum project: enabling ecological and evolutionary insights through a genus-level sequencing project. *New Phytol* 217:16–25. <https://doi.org/10.1111/nph.14860>.
 121. Wilson RM, Tfaily MM, Rich VI, Keller JK, Bridgman SD, Zalman CM, Meredith L, Hanson PJ, Hines M, Pfeifer-Meister L, Saleska SR, Crill P, Cooper WT, Chanton JP, Kostka JE. 2017. Hydrogenation of organic matter as a terminal electron sink sustains high CO₂/CH₄ production ratios during anaerobic decomposition. *Org Geochem* 112:22–32. <https://doi.org/10.1016/j.orggeochem.2017.06.011>.
 122. McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 10:e1003531. <https://doi.org/10.1371/journal.pcbi.1003531>.
 123. Lang AS, Westbye AB, Beatty JT. 2017. The distribution, evolution, and roles of gene transfer agents (GTAs) in prokaryotic genetic exchange. *Annu Rev Virol* 4:87. <https://doi.org/10.1146/annurev-virology-101416-041624>.
 124. Hurst CJ, Gerba CP, Cech I. 1980. Effects of environmental variables and soil characteristics on virus survival in soil. *Appl Environ Microbiol* 40:1067–1079.
 125. Gerba CP. 1984. Applied and theoretical aspects of virus adsorption to surfaces. *Adv Appl Microbiol* 30:133–168. [https://doi.org/10.1016/S0065-2164\(08\)70054-6](https://doi.org/10.1016/S0065-2164(08)70054-6).
 126. Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* 73:7059–7066. <https://doi.org/10.1128/AEM.00358-07>.
 127. Kavanaugh MT, Oliver MJ, Chavez FP, Letelier RM, Muller-Karger FE, Doney SC. 2016. Seascapes as a new venaucular for pelagic ocean monitoring, management and conservation. *ICES J Mar Sci* 73:1839–1850. <https://doi.org/10.1093/icesjms/fsw086>.
 128. Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G. 2013. Are we missing half of the viruses in the ocean? *ISME J* 7:672. <https://doi.org/10.1038/ismej.2012.121>.
 129. Greninger AL. 2017. A decade of RNA virus metagenomics is (not) enough. *Virus Res* 244:218–229. <https://doi.org/10.1038/nature16942>.
 130. Zhang YZ, Shi M, Holmes EC. 2018. Using metagenomics to characterize an expanding virosphere. *Cell* 172:1168–1172. <https://doi.org/10.1016/j.cell.2018.02.043>.
 131. Rinke C, Low S, Woodcroft BJ, Raina JB, Skarshewski A, Le XH, Butler MK, Stocker R, Seymour J, Tyson GW, Hugenholtz P. 2016. Validation of picogram-and femtogram-input DNA libraries for microscale metagenomics. *PeerJ* 4:e2486. <https://doi.org/10.7717/peerj.2486>.
 132. Kuhn E, Ichimura AS, Peng V, Fritsen CH, Trubl G, Doran PT, Murray AE. 2014. Brine assemblages of ultrasmall microbial cells within the ice cover of Lake Vida, Antarctica. *Appl Environ Microbiol* 80:3687–3698. <https://doi.org/10.1128/AEM.00276-14>.
 133. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, Singh A, Williams KH, Siegerist CE, Tringe SG, Downing KH, Comolli LR, Banfield JF. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun* 6:6372. <https://doi.org/10.1038/ncomms7372>.
 134. Solden L, Lloyd K, Wrighton K. 2016. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* 31:217–226. <https://doi.org/10.1016/j.mib.2016.04.020>.
 135. Sariaslani S, Gadd GM (ed). 2017. *Advances in applied microbiology*, vol 101. Elsevier Academic Press, New York, NY.
 136. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
 137. Bakken LR, Olsen RA. 1983. Buoyant densities and dry-matter contents

- of microorganisms: conversion of a measured biovolume into biomass. *Appl Environ Microbiol* 45:1188–1195.
138. Pollard EC, Grady LJ. 1967. CsCl density gradient centrifugation studies of intact bacterial cells. *Biophys J* 7:205. [https://doi.org/10.1016/S0006-3495\(67\)86584-6](https://doi.org/10.1016/S0006-3495(67)86584-6).
 139. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 140. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>.
 141. Vik DR, Roux S, Brum JR, Bolduc B, Emerson JB, Padilla CC, Stewart FJ, Sullivan MB. 2017. Putative archaeal viruses from the mesopelagic ocean. *PeerJ* 5:e3428. <https://doi.org/10.7717/peerj.3428>.
 142. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. 2012. Gene and translation initiation site prediction in metagenomics sequences. *Bioinformatics* 28:2223–2230. <https://doi.org/10.1093/bioinformatics/bts429>.
 143. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. <https://doi.org/10.1093/nar/28.1.27>.
 144. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res* 33:W116–W120. <https://doi.org/10.1093/nar/gki442>.
 145. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
 146. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
 147. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650. <https://doi.org/10.1093/molbev/msp077>.
 148. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128. <https://doi.org/10.1093/bioinformatics/btl529>.
 149. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER suite: protein structure and function prediction. *Nat Methods* 12:7. <https://doi.org/10.1038/nmeth.3213>.
 150. Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309. <https://doi.org/10.1093/nar/gki524>.
 151. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. *Nucleic Acids Res* 43:D571–D577. <https://doi.org/10.1093/nar/gku1207>.
 152. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
 153. Sanguino L, Franqueville L, Vogel TM, Larose C. 2015. Linking environmental prokaryotic viruses and their host through CRISPRs. *FEMS Microbiol Ecol* 91:fiv046. <https://doi.org/10.1093/femsec/fiv046>.
 154. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev* 40:258–272. <https://doi.org/10.1093/femsre/fuv048>.
 155. Skennerton CT, Imelfort M, Tyson GW. 2013. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res* 41:e105. <https://doi.org/10.1093/nar/gkt183>.
 156. Criscuolo A, Gribaldo S. 2010. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210. <https://doi.org/10.1186/1471-2148-10-210>.
 157. Shindell DT, Faluvegi G, Koch DM, Schmidt GA, Unger N, Bauer SE. 2009. Improved attribution of climate forcing to emissions. *Science* 326:716–718. <https://doi.org/10.1126/science.1174760>.
 158. Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicroa CD, Purvine SO, Hoyt DW, Schückel J, Jorgensen B, Willats W, Spalinger DE, Firkins JL, Lipton MS, Sullivan MB, Pope PB, Wrighton WC. Interspecies cross-feedings orchestrate carbon degradation in the rumen ecosystem. *Nat Microbiol*, in press.