Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing


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 fermenting abundant microbial lineages from Acidobacteria, Verrucomicrobia, and Deltaproteobacteria, 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). 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 (infected by lysing dominant microbial hosts) and bottom-up (infected from carriage of auxiliary metabolic genes) controls. This work serves as a foundation which future studies can build upon to understand the role of viruses in permafrost thawing.


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

September/October 2018 Volume 3 Issue 5 e00076-18
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 perennally frozen ground, i.e., permafrost, stores 30 to 50% of global soil carbon (C; ~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 palsa, 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 ~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-sequencing (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 (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 those 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
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 an 0.45-μm 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 FIG 1]

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.

TABLE 1 Soil virome read information

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

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 ≧90% average nucleotide identity between the read and the contig, and then for a contig to be considered detected, reads had to cover ≧75% of the contig.
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× 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 ~4% of VCs being globally ubiquitous (73). Most (~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., ~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 ~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...
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.

>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.
**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 chemooorganoheterotrophs 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 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).
predicted host was *Pedosphaera parvula*, within the phylum *Verrucomicrobia*, which is ubiquitous in soil, is abundant across our soils ([H11011](3%) in palsa and ([H11011](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 super-infection 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 phylogenic 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 chemoheterotrophs (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
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. 5B) 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
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 ~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-fractioned 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 ≥95% nucleotide identity across ≥90% of each read, and each contig had to have ≥70% of the contig covered with ≥1× coverage depth, while in these viromes we required ≥90% ANI and each contig had to have ≥75% of the contig covered with ≥1× 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.
**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 ρ 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 psal, one chilled and one frozen; 3 bog, one chilled and two frozen; and 2 fen, both chilled). The sequences were quality controlled using Trimomatic (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 to BLAST to a pool of putative laboratory contaminants (i.e., phages cultivated in the lab: Enterobacteria phage PhiX17, Alpha3, M13, Cellulophaga baltica phages, and Pseudalteromonas 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 phage 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, Desulfovibrion, Brachyspira, Methanococcus, and Bartonella) and (ii) microbial DNA from the SILVA rRNA database (release 128 [136]), with all the assembled contigs with ≥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 re-screened 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 MAViV (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 taxonomic 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^-15, 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 (vOTU) comparisons (i.e., 2,052 × 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
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); 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, 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\(^{-5}\), and ≥70% ANI across ≥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 Felixouvirinae and Vequentivirinae 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
REFERENCES


Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. 2017. Viruses in 
doi.org/10.1038/ismej.2011.2. 
https://doi.org/10.1002/1529-8817. 
Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome Biol 14:R123. 
Hurst BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and 
taxonomic niche specialisation in the ‘core’ and ‘flexible’ Pacific Ocean 
Whyte LG. 2017. Comparative activity and functional ecology of per-
mafrost soils and lithic niches in a hyper-arid polar desert. Environ Microbiol 
Rosario K, Breitbart M. 2011. Exploring the viral world through meta-
2011.06.004. 
Logares R, Haverkamp TH, Kumar S, Lanzen A, Nederbragt AJ, Quince C, 
doi.org/10.1016/j.mimet.2012.07.017. 
Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009. Labora-
https://doi.org/10.1038/nprot.2009.10. 
Zabikski O, van Zyl L, Aravindan SM, Rubagotti E, Tuffin M, Cary S, 
Courd WA. 2014. High-level diversity of tailed phages, eukaryote-
associated viruses, and virophage-like elements in the metaviromes of 
John SG, Mendez CB, Deng L, Polos B, Kaufman AKM, Kern S, Brum J, 
Duhaime MB, Deng L, Polos BT, Sullivan MB. 2012. Towards quanti-
Roux S, Solonenko NE, Dang VT, Polos BT, Schwemck MD, Goldsmith 
doi.org/10.1038/nrmicro3404. 
doi.org/10.7717/peerj.3817. 
Binga EK, Lasken RS, Neufeld JD. 2008. Something from (almost) nothing: the impact of multiple displacement amplification on micro-
Yilmaz S, Allgaier M, Hugenholtz P. 2010. Multiple displacement am-
Poison SW, Wilhelm SW, Wommack KE. 2011. Unraveling the viral 
1038/ismej.2010.81. 
Kim MS, Whon TW, Bae JW. 2013. Comparative viral metagenomics of


Downloaded from https://m Systems.asm.org on September 13, 2019 at UNIVERSITY OF ARIZONA LIBRARY


