Title: Soil viruses are underexplored players in ecosystem carbon processing

Running title: Quantitatively-derived soil viral metagenomes

Gareth Trubl[1], Ho Bin Jang[1], Simon Roux[1,†], Joanne B. Emerson[1,‡], Natalie Solonenko[1], Dean R. Vik[1], Lindsey Solden[1], Jared Ellenbogen[1], Alexander T. Runyon [1], Benjamin Bolduc[1], Ben J. Woodcroft[2], Scott R. Saleska[3], Gene W. Tyson[2], Kelly C. Wrighton[1], Matthew B. Sullivan[1,4], & Virginia I. Rich[1,#]

[1]Department of Microbiology, The Ohio State University, Columbus, OH, USA

[2]Australian Centre for Ecogenomics, The University of Queensland, St. Lucia, Queensland, Australia

[3]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

[4]Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA

†Current address: United States Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA, USA.

‡Current address: Department of Plant Pathology, University of California, Davis, Davis, CA, USA

Summary

Rapidly thawing permafrost harbors ~30–50% 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 (vOTUs) recovered from seven quantitatively-derived (i.e. not multiple-displacement-amplified) viral-particle metagenomes (viromes) along a permafrost thaw gradient. 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 dramatically different distributions along the thaw gradient habitats, and a shift from soil-virus-like assemblages in the dry palsas to aquatic-virus-like in the inundated fen. Seventeen vOTUs were linked to microbial hosts (*in silico*), implicating viruses in infecting abundant microbial lineages from *Acidobacteria*, *Verrucomicrobia*, and *Deltaproteoacteria*, including those encoding key biogeochemical functions such as organic matter degradation. Thirty-one auxiliary metabolic genes (AMGs) were identified, and suggested viral-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 (Emerson et al. *in press*, Nature Microbiology). We compared the ecology of the recovered vOTUs along a permafrost thaw gradient, and found: (1) habitat specificity, (2) a shift in viral community identity from soil-like to aquatic-like viruses, (3) infection of dominant microbial hosts, and (4) encoding 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 encoding auxiliary metabolic genes) controls. This work serves as a foundation upon which future studies can build upon to increase our understanding of the soil virosphere and how viruses affect soil ecosystem services.

Introduction

Anthropogenic climate change is elevating global temperatures, most rapidly at the poles (1). High-latitude perennially-frozen ground, i.e. permafrost, stores 30–50% of global soil carbon (C; ~1300 Pg; 2, 3) and is thawing at a rate of $\geq$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. 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 (7). The Mire contains a mosaic of thaw stages (8), from intact permafrost palsas, to partially-thawed moss-dominated bogs, to fully-thawed sedge-dominated fens (9–12). Thaw shifts hydrology (13), altering plant communities (12), and shifting belowground organic matter (OM) towards more labile forms (10, 12), with concomitant shifts in microbiota (14–16), and C gas release (7, 9, 17–

19). Of particular note is the thaw-associated increase in CH_4 emissions, due to its 33-times greater climate forcing potential than CO_2 (per kg, at a 100-year time-scale; 20), and the associated shifts in key methanogens. These include novel methanogenic lineages (14) with high predictive value for the character of the emitted CH_4 (11). 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 ~one-third of ocean microorganisms day^{-1}, 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 viral-encoded "auxiliary metabolic genes" (AMGs; 28, 30) including those involved in 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 (10^7–10^9 virus particles per gram of soil; 37, 40–42), knowledge of soil viral ecology has come mainly from the fraction that desorb easily from soils (<10% in 43) and the much smaller subset that have been isolated (44).

One approach to studying soil viruses has been to bypasses 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

92 knowledge of soil viruses and have highlighted the magnitude of genetic novelty these entities
93 may represent. An analysis of 3,042 publicly-available assembled metagenomes spanning 10
94 ecotypes (19% from soils) increased by 16-fold the total number of known viral genes, doubled
95 the number of microbial phyla with evidence of viral infection, and revealed that the vast
96 majority of viruses appeared to be habitat-specific (45). This approach was also applied to 178
97 metagenomes from the thawing permafrost gradient of Stordalen Mire (46), where viral linkages
98 to potential hosts were appreciably advanced by the parallel recovery of 1,529 microbial
99 metagenome-assembled genomes (MAGs; 16). This effort recovered ~2000 thaw-gradient
100 viruses, more than doubling known viral genera in Refseq, identified linkages to abundant
101 microbial hosts encoding important C-processing metabolisms such as methanogenesis, and
102 demonstrated that CH_4 dynamics was best predicted by viruses of methanogens and
103 methanotrophs (46). Viral analyses of bulk-soil metagenomes have thus powerfully expanded
104 knowledge of soil viruses and highlighted the large amount of genetic novelty they represent.
105 However this approach is by nature inefficient at capturing viral signal, with typically <2% of
106 reads identified as viral (46, 47). The small amount of viral DNA present in bulk-soil extracts
107 can lead to poor or no assembly of viral sequences in the resulting metagenomes, and omission
108 from downstream analyses (discussed further in 37, 39, 48, 49). In addition, viruses that are
109 captured in bulk-soil metagenomes likely represent a subset of the viral community, since >90%
110 of free viruses adsorb to soil (43), and so depending on the specific soil, communities, and
111 extraction conditions, bulk-soil metagenomes are likely be depleted for some free viruses and
112 enriched for actively reproducing and temperate viruses.
113 Examination of free viruses, while potentially a more efficient and comprehensive
114 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*. 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 (51), with now relatively well-established downstream pipelines for quantitative sample-to-sequence (52) and sequence-to-ecological-inference (53, 54) processing, collectively resulting in great advances in marine viromics (55). 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 datasets both stochastically and systematically biased and non-quantitative (53, 56–61). The few polar soil viromes have been from Antarctic soils, and further demonstrated the genetic novelty of this gene pool, while suggesting resident viral communities were dominated by tailed-viruses, had high habitat specificity, and were structured by pH (62–64).

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 vOTUs (approximately representing species-level taxonomy; 65). 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 Emerson et al. (46).

Results and Discussion

Viruses in complex soils

Using recently develop 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 2–26 million reads, with an average of 18% of the reads assembling into 28,025 total contigs across the dataset. Among these, VirSorter predicted that 393 contigs were viruses (VirSorter categories 1, 2, and 3; per 66; see Methods; Table 2). After manual inspection, three putative plasmids were identified and removed (i.e. contigs 5, 394, and 3167; Table 2), along with two putatively archaeal viruses (vOTUs 165 and 225; analyzed separately, see supplementary information). Finally, ten additional contigs that did not meet our threshold for read recruitment (i.e. 90% average nucleotide identity across 75% of contig covered) were removed, resulting in 378 putative virus sequences (Table 2). Of these, 53 bacteriophage (phage) were considered well-sampled 'viral populations' (54) also known as viral operational taxonomic units (vOTUs) as they had contig lengths $\geq$10 kb (average 19.6 kb, range: 10.3 kb–129.6 kb), were most robustly viral (VirSorter category 1 or 2; 66), and were relatively well-covered contigs (averaged 74x coverage, Table 1). These 53 viral populations 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. We therefore applied a gene-sharing network where nodes were genomes and edges between nodes indicated the gene content similarities, and accommodating fragmented genomes of varying sizes (67–72). In such networks, viruses sharing a high number of genes localize into viral clusters (VCs) which represent approximately genus-level taxonomy (69, 72). We represented relationships across the 53 vOTUs with 2,010 known bacterial and archaeal viruses (RefSeq, version 75) as a weighted network (Fig. 2). Only 15% of the Mire vOTUs had similarity to RefSeq viruses (Fig. 2). Three vOTUs fell into 3 VCs comprised of viruses belonging to the *Felixounavirinae* and *Vequintavirinae* (VC10), *Tevenvirinae* and *Eucampyvirinae* (VC3), and the *Bcep22virus, F116virus* and *Kpp25virus* (VC4) (Fig. 2). 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 (73), 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.2). 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. 2). 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 (70). 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

183 due to a limited number of common genes between them, i.e. ~70% of the total proteins in these
184 viromes are unique (Fig. 2), reflecting the relative novelty of these viruses and the
185 undersampling of soil viruses (39).

186 Annotation of the 53 vOTUs resulted in only ~30% of the genes being annotated, which
187 is not atypical; >60% of genes encoded in uncultivated viruses have typically been classified as
188 unknown in other studies (46, 66, 74–78). Of genes with annotations, we first considered those
189 involved in lysogeny, to provide insight into the viruses' replication cycle. Only three viruses
190 encoded an integrase gene (other characteristic lysogeny genes were not detected; 79, 80; Table
191 S1), suggesting they could be temperate viruses, two of which were from the bog habitat. It had
192 been proposed that since soils are structured and considered harsh environments, a majority of
193 soil viruses would be temperate viruses (81). Although our dataset is small, a dominance of
194 temperate viruses is not observed here. We hypothesize that the low encounter rate produced by
195 the highly structured soil environment could, rather than selecting for temperate phage, select for
196 efficient virulent viruses (concept derived from 82–84). Recent analyses of the viral signal mined
197 from bulk-soil metagenomes from this site provides more evidence for our hypothesis of
198 efficient virulent viruses, because >50% of the identified viruses were likely not temperate
199 (based on the fact they were not detected as prophage; 46). As a more comprehensive portrait of
200 soil viruses grows, spanning various habitats, this hypothesis can be further tested. Beyond
201 integrase genes, the remaining annotated genes spanned known viral genes and host-like genes.
202 Viral genes included those involved in structure and replication, and their taxonomic affiliations
203 were unknown or highly variable, supporting the quite limited affiliation of these vOTUs with
204 known viruses. Host-like genes included AMGs, which are described in greater detail in the next
205 section.

206 *Host-linked viruses are predicted to infect key C cycling microbes*

207 In order to examine these viruses' impacts on the Mire's resident microbial communities
208 and processes, we sought to link them to their hosts via emerging standard *in silico* host
209 prediction methods, significantly empowered by the recent recovery of 1,529 MAGs from the
210 site (508 from palsa, 588 from bog, and 433 from fen; 16). Tentative bacterial hosts were
211 identified for 17 of the 53 vOTUs (Fig. 3; Table S2): these hosts spanned four genera among
212 three phyla (*Verrucomicrobia: Pedosphaera*, *Acidobacteria: Acidobacterium* and *Candidatus*
213 Solibacter, and *Deltaproteobacteria: Smithella*). Eight viruses were linked to more than one host,
214 but always within the same species. The four predicted microbial hosts are among the most
215 abundant in the microbial communities, and have notable roles in C cycling (15; 16). Three are
216 acidophilic, obligately aerobic chemoorganoheterotrophs and include the Mire's dominant
217 polysaccharide-degrading lineage (*Acidobacteria)*, and the fourth is an obligate anaerobe shown
218 to be syntrophic with methanogens (*Smithella)*. *Acidobacterium* is a highly abundant, diverse,
219 and ubiquitous soil microbe (85–87), and a member of the most abundant phylum in Stordalen
220 Mire. The relative abundance of this phylum peaked in the bog at 29%, but still had a
221 considerably high relative abundance in the other two habitats (5% in palsa and 3% fen) (16). It
222 is a versatile carbohydrate utilizer, and has recently been identified as the primary degrader of
223 large polysaccharides in the palsa and bog habitats in the Mire, and is also an acetogen (16).
224 Seven vOTUs were inferred to infect *Acidobacterium*, implicating these viruses in directly
225 modulating a key stage of soil organic matter decomposition. The second identified
226 *Acidobacterial* host was in the newly proposed species *Candidatus* Solibacter usitatus, another
227 carbohydrate degrader (88). The third predicted host was *Pedosphaera parvula*, within the
228 phylum *Verrucomicrobia* which is ubiquitous in soil, abundant across our soils (~3% in palsa

and ~7% in bog and fen habitats, based on metagenomic relative abundance; 16), utilizes cellulose and sugars (89–93) and in this habitat, this organism could be acetogenic (16). Lastly, vOTU_28 was linked to the *Deltaproteobacteria Smithella* sp. SDB, another acidophilic chemoorganoheterotroph, but an obligate anaerobe, with a known syntrophic relationship with methanogens (94, 95). 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 (96).

We next sought to examine viral AMGs for connections to C cycling. To more robustly identify AMGs than the standard protein family-based search approach, we used a custom-built in-house pipeline previously described in Daly et al. (97), and further tailored to identify putative AMGs based on the metabolisms described in the 1,529 MAGs recently reported from these same soils (16). From this, we identified 34 AMGs from 13 vOTUs (Fig. 4; Table S1; Table S3), encompassing C acquisition and processing (three in polysaccharide-binding, one involved in polysaccharide degradation, and 23 in central C metabolism) and sporulation. Glycoside hydrolases that help breakdown complex OM are abundant in resident microbiota (16) and may be especially useful in this high OM environment; notably to our knowledge they have not been found in marine viromes, but have been found in soil (at our site; 46) and rumen (98; Solden et al. *submitted—99*). In addition, central C metabolism genes in viruses may increase nucleotide and energy production during infection, and have been increasingly observed as AMGs (31, 32, 33, 34, 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 (100, 101). A WhiB-like protein has been previously identified in mycobacteriophage TM4 (WhiBTM4), and experimentally shown to not only transcriptionally

252 regulate host septation, but also cause superinfection exclusion (i.e. exclusion of secondary viral
253 infections; 102). While these two sporulation genes have only been found in *Firmicutes* and
254 *Actinobacteria*, the only vOTU to have *whiB* was linked to an acidobacterial host (vOTU_178;
255 Fig. 4). A phylogenic analysis of the *whiB* AMG grouped it with actinobacterial versions and
256 more distantly with another mycobacteriophage (Fig. 4), suggesting either (1) misidentification
257 of host (unlikely, as it was linked to three different acidobacterial hosts, each with zero
258 mismatches of the CRISPR spacer), (2) the virus could infect hosts spanning both phyla
259 (unlikely, as only ~1% of identified virus-host relationships span phyla; 45), or (3) the gene was
260 horizontally transferred into the *Acidobacteria*. Identification of these 34 diverse AMGs
261 (encoded by 25% of the vOTUs) suggests a viral modulation of host metabolisms across these
262 dynamic environments, and supports the findings from bulk metagenome-derived viruses of
263 Emerson et al. (46) at this site. That study's AMGs spanned the same categories as those
264 reported here, except for *whiB* which was not found, but did not discuss them other than the
265 glycoside hydrolases, one of which was experimentally validated.

266 Thus far, the limited studies of soil viruses have identified few AMGs relative to studies
267 of marine environments. This may be due to under-sampling, or difficulties in identifying
268 AMGs; since AMGs are homologs of host genes, they can be mistaken for microbial
269 contamination (103) and thus are more difficult to discern in bulk-soil metagenomes (whereas
270 marine virology has been dominated by viromes); also, microbial gene function is more poorly
271 understood in soils (104). Alternately, soil viruses could indeed encode fewer AMGs. One could
272 speculate a link between host lifestyle and the usefulness of encoding AMGs; most known
273 AMGs are for photo- and chemo-autotrophs (70, 105, 106), although this may be due to more
274 studies of these metabolisms or phage-host systems. Thus far, soils are described as dominated

by heterotrophic bacteria (107–111), and if AMGs were indeed less useful for viruses encoding 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.

Sample storage impacts 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 successful from both storage conditions. Storage impacted recovered community structure only in the bog habitat, with dramatically broader recovery of vOTUs from the chilled sample (Fig. 5A/B), leading to higher diversity metrics (Fig. S4), and appreciable separation of the recovered chilled-vs-frozen bog vOTU profiles in ordination (Fig. 5C). The greater vOTUs recovery 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 3.5–9-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 relative to 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; 10, 46), with a dynamic water table, and each of these has been hypothesized or demonstrated to increase selection for temperate viruses (77, 112–116). In addition, of the 19 vOTUs shared between this study and the bulk-soil metagenome study of Emerson et al. (46; which 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 only present in 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 (dendrogram, Fig. S5A), a social network analysis of the reads that mapped to the viromes (Fig. S5B & C) 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. S2B) instead of subsampling the reads, because this is a smaller dataset (subsampling smaller datasets described further in 117) 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 the (i) 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's abundances, bog vOTUs were 3-fold more abundant and fen vOTUs were 12-fold more abundant (Fig. 5A)**.** This is consistent with overall increases in viral-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. 5B). Consistent with this, principal coordinates 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 dataset (Fig. 5C). Mapping of the 214 bulk-soil metagenomes from the three habitats (16) revealed that a majority (41; 77%) of the vOTUs were present in the bulk-soil metagenomes (Fig. 6), collectively occurring in 62% (133) 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 proportion of bog-derived vOTU's present in the bulk metagenomes, and in 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% between palsa and fen, and 68% between bog and fen (Fig. 6). The difference in observations from vOTU read recruitment of viromes versus 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 datasets is consistent 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 37, 118, 119) and replication strategy (36, 37). In addition, at Stordalen Mire (and at other similar sites; 110), 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) (14, 15, 16). These shifting microbial hosts likely impact viral community structure. Expanding from the 53 vOTUs examined here, Emerson et al.'s (46) recent analysis 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 (120), and from grasslands and arctic soils (45). In contrast, an emerging paradigm in the marine field is 'seascape ecology' (121), where the majority of taxa are detected across broad geographical areas, as are marine virus (26, 70). 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. S4), this dataset only captured a small fractioned of the viral diversity (based on the collector's curve from 46), and therefore the undersampling prevents diversity inferences. Intriguingly, while our virome-derived vOTU richness was lowest in the palsa, Emerson et al.'s (46) much greater sampling recovered the most vOTUs in the palsa, more than double that in the fen (42% vs. 18.9% of total vOTUs). This major difference could potentially be due to the known increase in microbial alpha diversity along the thaw gradient (15, 16), 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

367 increased vOTU richness with thaw in our virome dataset). Notably, neither this dataset nor that
368 of Emerson et al. (46) captured ssDNA or RNA viruses, which potentially represent up to half of
369 viral particles (122–124).

370 *Challenges in characterizing the soil virosphere*

371 The low yield of viral contigs given the relatively large sequencing depth of the viromes
372 reflects several factors that currently challenge soil viromics. First, resuspending viruses from
373 soils is a challenge due to their adsorption to the soil matrix (43). Second, yields of viral DNA
374 are often very low (due to both low input biomass and potentially low extraction efficiency),
375 requiring amplification; this leads to biases (53, 56–61) or poor assembly and few viral contigs
376 (described further in 125). Third, viral contig identification requires a reference database, yet soil
377 viruses are underrepresented in current databases; for example, a majority (85%) of our sequence
378 space was unknown. Fourth, non-viral DNA may co-extract. Lastly, the optimal approach to
379 identifying ecological units within viral sequence space is unclear.

380 In this study, DNA yields (and sequencing inputs) decreased along the thaw gradient, as
381 did total reads, but counterintuitively viral reads increased (Table 1; the fen had ~5-fold more
382 viral reads than the palsa). This may have been partly due to the shift to a more aquatic-type
383 habitat, for which viruses are better represented in the databases, or to an actual increase in viral
384 DNA (as a portion of total) concomitant with known viral abundance increases (41). A large
385 portion of the assembled reads were non-viral (Table 1), representing either microbial
386 contamination, or gene transfer agents (GTAs), i.e., viral-like capsids that package microbial
387 DNA (reviewed in 126). Since the viral particle purification protocol involved 0.2 µM filter
388 followed by CsCl density gradient-based separation of the viral particles (removing free genomic
389 DNA), contamination by microbial DNA seems unlikely. While ultra-small microbial cells have

been found in our soils (46) and other permafrost soils (reviewed in 127, 128), and may have passed through the 0.2 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 128–130). 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 (131), and 94 contigs with matches to known GTAs (126), 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 126). In 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 dataset: reads characterization, contigs, and as vOTUs (data not shown). 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 (Fig. 7) with that used in Emerson et al. (46), which recovered vOTUs from bulk-soil metagenomes from the same site, but different years, months, depths, and preservation methods. 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 bulk metagenomes, and neither approach captured the total viral diversity in these soils based on rarefaction (46). The efficiency of vOTU recovery

was, unsurprisingly, >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 diversity would likely saturate below that). Of the 19 vOTUs that were shared between the two datasets, 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) can offer complementary views of viral communities in soils and the optimal method 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 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 viral signal from MAGs, analyzed in the context of the abundant metadata available. With increasing characterization of soil viruses, 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.

Methods and Materials

Sample collection

435 Samples were collected from July 16–19, 2014 from peatland cores in the Stordalen Mire
436 field site near Abisko, Sweden (Fig. 1; more site information in 7, 10, 12). The soils derived
437 from palsa (one stored chilled and the other stored frozen), bog (one stored chilled and two
438 stored frozen), and fen (both stored chilled) habitats along the Stordalen Mire permafrost thaw
439 gradient. These three sub-habitats are common to northern wetlands, and together cover ~98% of
440 Stordalen Mire's non-lake surface (8). The sampled palsa, bog, and fen are directly adjacent,
441 such that all cores were collected within a 120 m total radius. For this work, the cores were
442 subsampled at 36–40 cm, and material from each was divided into two sets. Set 1 was chilled
443 and stored at 4°C, and set 2 was flash-frozen in liquid nitrogen and stored at −80°C as described
444 in Trubl et al. (41). Both sets were processed using a viral resuspension method optimized for
445 these soils (41). For CsCl density gradient purification of the particles, CsCl density layers of rho
446 1.2, 1.4, 1.5, and 1.65 were used to establish the gradient; we included a 1.2 g/cm^3 CsCl layer to
447 try to remove any small microbial cells that might have come through the 0.2um filter (for
448 microbial cell densities see 132, 133; for viral particle densities see 50). We then collected the
449 1.4-1.52 g/cm^3 range from the gradient for DNA extraction, to target the dsDNA range (per 50).
450 The viral DNA was extracted using Wizard columns (Promega, Madison, WI, products A7181
451 and A7211), and cleaned up with AMPure beads (Beckman Coulter, Brea, CA, product A63881).
452 DNA libraries were prepared using Nextera XT DNA Library Preparation Kit (Illumina, San
453 Diego, CA, product FC-131-1024) and sequenced using an Illumina MiSeq (V3 600 cycle, 6
454 samples/run, 150 bp paired end) at the University of Arizona Genetics Core facility (UAGC).
455 Seventeen viral contigs were previously described in Emerson et al. (46) (Fig. 7).
456 The 214 bulk-soil metagenomes and associated recovered MAGs used here for analyses
457 were described in Woodcroft et al. (16), and derive from the same sampling sites from 2010-

2012, and 1–85 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), as described in Woodcroft et al (16).

vOTU recovery

Eight viromes were prepped 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 (134; adaptors were removed, reads were trimmed as soon as the per-base quality dropped below 20 on average on 4 nt sliding windows, and reads shorter than 50 bp were discarded), then assembled separately with IDBA-UD (135), and contigs were processed with VirSorter to distinguish viral from microbial contigs (virome decontamination mode; 66). 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 *Pseudoalteromonas* phages). All contigs matching these genomes at more than 95% average nucleotide identity (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 categories 1 or 2, but further inspection is needed to confirm 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 (1) taxa related to the 5 types of GTAs (keyword searches were: *Rhodobacterales*, *Desulfovibrio*, *Brachyspira*, *Methanococcus*, and *Bartonella*) and (2) microbial DNA the SILVA ribosomal RNA database (release 128; 131), with all the assembled contigs with ≥95% ANI. The percent of reads that mapped to these contigs was calculated as previously described.

After having verified 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 non-redundant 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 <10 kb and (ii) only using 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; 136). The output putative archaeal virus sequences were then filtered to include only those contigs ≥10 kb in size resulting in the set of putative archaeal vOTUs described here.

Viral genes were annotated using a pipeline described in Daly et al. (97). Briefly, for each contig, ORFs were freshly predicted using MetaProdigal (137) and sequences were compared to KEGG (138), UniRef and InterproScan (139) using USEARCH (140), with single and reverse best-hit matches greater than a 60 bitscore. AMGs were identified by manual inspection of the protein annotations guided by known resident microbial metabolic functions (identified in 16). To determine confidence in functional assignment, representatives for each AMGs underwent phylogenetic analyses. First each sequence was BLASTed and the top 100 hits were investigated to identify main taxa groups. An alignment with the hits and the matching viral sequence (MUSCLE with default parameters; 141) 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 1000 bootstraps; 142) was used to make the phylogeny and iTol (143) was used to visualize and edit the tree (any distance sequences were removed). To see if this AMG was wide-spread across the putative soil viruses, a BLASTp (default settings) of each AMG against all putative viral proteins from our viromes was done. The sequences from

identified homologs (based on a bitscore >70 and an e value of 10^{-4}) were used with the AMG of interest to construct a new phylogenic tree (same methods as before). Finally, structures were predicted using i-TASSER (144) 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 length of the reference protein; 145).

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 (71, 72). We downloaded 198,556 protein sequences representing the genomes of 1,999 bacterial and archaeal viruses from NCBI RefSeq (v 75; 146). 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, with an e-value threshold of 10^{-4}, and defined as protein clusters (PCs) in the same manner as previously described (67). Based on the number of PCs shared between the genomes and/or genome fragments, a similarity score was calculated using vConTACT (71, 72). The resulting network 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. 398 RefSeq viruses not showing significant similarity to viral contigs were excluded for clarity. The resulting network was composed of 1,722 viral genomes including 53 contigs and 58,201 edges. To gain detailed insights into the genetic connections, the network was decomposed into a series of coherent groups of nodes (aka VCs; 69, 71, 72), with an optimal inflation factor of 1.6. Thus, the discontinuous network structure of individual components, together with the isolated contigs, indicates their distinct gene pools (68). To assign contigs into VCs, PCs needed to include ≥2 genomes and/or genome fragments, then Markov clustering (MCL) 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 (http://www.ncbi.nlm.nih.gov/taxonomy).

vOTU ecology

Virome reads were mapped back to the non-redundant set of contigs to estimate their coverage, calculated as number of bp mapped to each read normalized by the length of the contig, and by the total number of bp sequenced in the metagenome in order to be comparable between samples (Bowtie 2, threshold of 90% average nucleotide identity on the read mapping, and 75% of contig covered to be considered as detected; 54, 147). 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 Woodcroft et al. (16). The paired MAG reads were mapped to the viral contigs with Bowtie2 (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, 97, 148) and one on direct sequence similarity between virus and host genomes (149). 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

S2; 150). For BLAST, the vOTU nucleotide sequences were compared to the MAGs (16) as described in Emerson et al. (46). Any viral sequences with a bit score of 50, E-value threshold of 10^{-3}, and ≥70% average nucleotide identity across ≥2500 bp were considered for host prediction (described in 151).

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 non-informative positions with the BMGE software package (152). The alignments were then concatenated into a FASTA file and the maximum likelihood tree was built with MEGA5 using 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.

Accession numbers

All data (sequences, site information, 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 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

570 chilled replicate B, SAMN08784163 for Fen chilled replicate A, and SAMN08784165 for Fen
571 chilled replicate B.

572 **Acknowledgments**

573 We thank Bonnie Poulos and Christine Schirmer for their assistance on different stages of this
574 project. We also thank SWES-MEL, TMPL, and The University of Arizona Genetics Core
575 facility, MAVERIC lab at the Ohio State University, the Abisko Naturvetenskapliga Station, and
576 the Joint Genome Institute for support. We thank Moira Hough, Robert Jones, and Rachel
577 Wilson for sample collection assistance. Bioinformatics were supported by The Ohio
578 Supercomputer Center and by the National Science Foundation under Award Numbers DBI-
579 0735191 and DBI-1265383; URL: www.cyverse.org. This study was funded by the Genomic
580 Science Program of the United States Department of Energy Office of Biological and
581 Environmental Research, (grants DE-SC0004632, DE-SC0010580, and DE-SC0016440), and by
582 a Gordon and Betty Moore Foundation Investigator Award (GBMF#3790 to MBS). We thank
583 Dr. Michael Palace (palace@guero.sr.unh.edu) for generating and allowing us to use the
584 unmanned aerial vehicle (UAV) image in Fig. S1.

585

586 **References**

587 1. Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L.,
588 Dahe, Q., Dasgupta, P., Dubash, N.K. and Edenhofer, O., 2014. IPCC fifth assessment
589 synthesis report-climate change 2014 synthesis report.
590 2. Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J.W., Schuur, E., Ping, C.L., Schirrmeister,
591 L., Grosse, G., Michaelson, G.J., Koven, C.D. and O'Donnell, J.A., 2014. Estimated stocks

of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences*, *11*(23), pp.6573-6593.

3. Schuur, E.A.G., McGuire, A.D., Schädel, C., Grosse, G., Harden, J.W., Hayes, D.J., Hugelius, G., Koven, C.D., Kuhry, P., Lawrence, D.M. and Natali, S.M., 2015. Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), pp.171-179.

4. Elberling, B., Michelsen, A., Schädel, C., Schuur, E.A., Christiansen, H.H., Berg, L., Tamstorf, M.P. and Sigsgaard, C., 2013. Long-term CO2 production following permafrost thaw. *Nature Climate Change*, *3*(10), pp.890-894.

5. Shelef, E., Rowl, J.C., Wilson, C.J., Hilley, G.E., Mishra, U., Altmann, G.L. and Ping, C.L., Large Uncertainty in Permafrost Carbon Stocks due to Hillslope Soil Deposits. *Geophysical Research Letters*.

6. Tarnocai, C., Canadell, J.G., Schuur, E.A.G., Kuhry, P., Mazhitova, G. and Zimov, S., 2009. Soil organic carbon pools in the northern circumpolar permafrost region. *Global biogeochemical cycles*, *23*(2).

7. Bäckstrand, K., Crill, P.M., Jackowicz-Korczynski, M., Mastepanov, M., Christensen, T.R. and Bastviken, D., 2010. Annual carbon gas budget for a subarctic peatland, Northern Sweden. *Biogeosciences*, *7*(1), pp.95-108.

8. Johansson, M., Christensen, T.R., Akerman, H.J. and Callaghan, T.V., 2006. What determines the current presence or absence of permafrost in the Torneträsk Region, a sub-Arctic landscape in Northern Sweden?. *AMBIO: A Journal of the Human Environment*, *35*(4), pp.190-197.

9. Malmer, N., Johansson, T., Olsrud, M. and Christensen, T.R., 2005. Vegetation, climatic changes and net carbon sequestration in a North□Scandinavian subarctic mire over 30 years. *Global Change Biology*, *11*(11), pp.1895-1909.

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

11. McCalley, C.K., Woodcroft, B.J., Hodgkins, S.B., Wehr, R.A., Kim, E.H., Mondav, R., Crill, P.M., Chanton, J.P., Rich, V.I., Tyson, G.W. and Saleska, S.R., 2014. Methane dynamics regulated by microbial community response to permafrost thaw. *Nature*, *514*(7523), pp.478-481.

12. Normand, A.E., Smith, A.N., Clark, M.W., Long, J.R. and Reddy, K.R., 2017. Chemical Composition of Soil Organic Matter in a Subarctic Peatland: Influence of Shifting Vegetation Communities. *Soil Science Society of America Journal*, *81*(1), pp.41-49.

13. Torbick, N., Persson, A., Olefeldt, D., Frolking, S., Salas, W., Hagen, S., Crill, P. and Li, C., 2012. High resolution mapping of peatland hydroperiod at a high-latitude Swedish mire. *Remote Sensing*, *4*(7), pp.1974-1994.

14. Mondav, R., Woodcroft, B.J., Kim, E.H., McCalley, C.K., Hodgkins, S.B., Crill, P.M., Chanton, J., Hurst, G.B., VerBerkmoes, N.C., Saleska, S.R. and Hugenholtz, P., 2014. Discovery of a novel methanogen prevalent in thawing permafrost. *Nature communications*, *5*, p.3212.

15. Mondav, R., McCalley, C.K., Hodgkins, S.B., Frolking, S., Saleska, S.R., Rich, V.I., Chanton, J.P. and Crill, P.M., 2017. Microbial network, phylogenetic diversity and

community membership in the active layer across a permafrost thaw gradient. *Environmental Microbiology*.

16. Woodcroft, B. J. , Singleton, C. M., Boyd, J. A. , Evans, P. N. , Hoelzle, R. D., Lamberton, T. O., McCalley, C. K., Hodgkins, S. B. , Wilson, R. M., Chanton, J. P. , Crill, P. M., Saleska, S. R., Rich, V. I., Tyson, G. W. (*in press*). Genome-centric metagenomic insights into microbial carbon processing across a permafrost thaw gradient.

17. Christensen, T.R., Johansson, T., Åkerman, H.J., Mastepanov, M., Malmer, N., Friborg, T., Crill, P. and Svensson, B.H., 2004. Thawing sub□arctic permafrost: Effects on vegetation and methane emissions. *Geophysical research letters*, *31*(4).

18. Christensen, T.R., Jackowicz-Korczyński, M., Aurela, M., Crill, P., Heliasz, M., Mastepanov, M. and Friborg, T., 2012. Monitoring the multi-year carbon balance of a subarctic palsa mire with micrometeorological techniques. *Ambio*, *41*(3), pp.207-217.

19. Schädel, C., Bader, M.K.F., Schuur, E.A., Biasi, C., Bracho, R., Čapek, P., De Baets, S., Diáková, K., Ernakovich, J., Estop-Aragones, C. and Graham, D.E., 2016. Potential carbon emissions dominated by carbon dioxide from thawed permafrost soils. *Nature Climate Change*.

20. Shindell, D.T., Faluvegi, G., Koch, D.M., Schmidt, G.A., Unger, N. and Bauer, S.E., 2009. Improved attribution of climate forcing to emissions. *Science*, *326*(5953), pp.716-718.

21. 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, *Journal of Advances in Modeling Earth Systems*. 9: 1412–1430. DOI: 10.1002/2016MS000817

658 22. Fuhrman, J.A., 1999. Marine viruses and their biogeochemical and ecological
659 effects. *Nature*, *399*(6736), pp.541-548.
660 23. Suttle, C.A., 2005. Viruses in the sea. *Nature*, *437*(7057), pp.356-361.
661 24. Suttle, C.A., 2007. Marine viruses—major players in the global ecosystem. *Nature Reviews*
662 *Microbiology*, *5*(10), pp.801-812.
663 25. Hurwitz, B.L., Westveld, A.H., Brum, J.R. and Sullivan, M.B., 2014. Modeling ecological
664 drivers in marine viral communities using comparative metagenomics and network
665 analyses. *Proceedings of the National Academy of Sciences*, *111*(29), pp.10714-10719.
666 26. Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A.,
667 Chaffron, S., Cruaud, C., De Vargas, C., Gasol, J.M. and Gorsky, G., 2015. Patterns and
668 ecological drivers of ocean viral communities. *Science*, *348*(6237), p.1261498.
669 27. Fridman, S., Flores-Uribe, J., Larom, S., Alalouf, O., Liran, O., Yacoby, I., Salama, F.,
670 Bailleul, B., Rappaport, F., Ziv, T. and Sharon, I., 2017. A myovirus encoding both
671 photosystem I and II proteins enhances cyclic electron flow in infected Prochlorococcus
672 cells. *Nature microbiology*, *2*(10), p.1350.
673 28. Breitbart, M., 2012. Marine viruses: truth or dare. *Marine Science*, *4*.
674 29. Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic,
675 S., Berline, L., Brum, J.R. and Coelho, L.P., 2016. Plankton networks driving carbon
676 export in the oligotrophic ocean. *Nature*.
677 30. Middelboe, M. and Brussaard, C.P., 2017. Marine Viruses: Key Players in Marine
678 Ecosystems. *Viruses* **2017**, *9*, 302.
679 31. Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K.,
680 Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W. and Jaroszewski, L., 2007. The Sorcerer

II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS biology*, *5*(3), p.e16.

32. Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L. and McDaniel, L., 2008. Functional metagenomic profiling of nine biomes. *Nature*, *452*(7187), p.629.

33. Sharon, I., Battchikova, N., Aro, E.M., Giglione, C., Meinnel, T., Glaser, F., Pinter, R.Y., Breitbart, M., Rohwer, F. and Béjà, O., 2011. Comparative metagenomics of microbial traits within oceanic viral communities. *The ISME journal*, *5*(7), p.1178.

34. Hurwitz, B.L., Hallam, S.J. and Sullivan, M.B., 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome biology*, *14*(11), p.R123.

35. Hurwitz, B. L., Brum, J. R. & Sullivan, M. B. 2015. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J.* **9**, 472–484.

36. Kimura, M., Jia, Z.J., Nakayama, N. and Asakawa, S., 2008. Ecology of viruses in soils: past, present and future perspectives. *Soil Science and Plant Nutrition*, *54*(1), pp.1-32.

37. Williamson, K.E., Fuhrmann, J.J., Wommack, K.E. and Radosevich, M., 2017. Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annual Review of Virology*, *4*(1).

38. Fierer, N., 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), pp.579-590.

39. Pratama, A.A. and van Elsas, J.D., 2018. The 'Neglected'Soil Virome–Potential Role and Impact. *Trends in Microbiology*.

703 40. Williamson, K.E., Corzo, K.A., Drissi, C.L., Buckingham, J.M., Thompson, C.P. and
704 Helton, R.R., 2013. Estimates of viral abundance in soils are strongly influenced by
705 extraction and enumeration methods. *Biology and Fertility of Soils*, *49*(7), pp.857-869.
706 Rohwer, F. and Thurber, R.V., 2009. Viruses manipulate the marine
707 environment. *Nature*, *459*(7244), p.207.
708 41. Trubl, G., Solonenko, N., Chittick, L., Solonenko, S.A., Rich, V.I. and Sullivan, M.B.,
709 2016. Optimization of viral resuspension methods for carbon-rich soils along a permafrost
710 thaw gradient. *PeerJ*, *4*, p.e1999. Sime-Ngando, T. and Colombet, J., 2009. Virus and
711 prophages in aquatic ecosystems. *Canadian journal of microbiology*, *55*(2), pp.95-109
712 42. Narr, A., Nawaz, A., Wick, L.Y., Harms, H. and Chatzinotas, A., 2017. Soil Viral
713 Communities Vary Temporally and along a Land Use Transect as Revealed by Virus-Like
714 Particle Counting and a Modified Community Fingerprinting Approach
715 (fRAPD). *Frontiers in Microbiology*, *8*, p.1975.
716 43. Goyal, S.M. and Gerba, C.P., 1979. Comparative adsorption of human enteroviruses,
717 simian rotavirus, and selected bacteriophages to soils. *Applied and Environmental*
718 *Microbiology*, *38*(2), pp.241-247.
719 44. Cresawn, S.G., Pope, W.H., Jacobs-Sera, D., Bowman, C.A., Russell, D.A., Dedrick, R.M.,
720 Adair, T., Anders, K.R., Ball, S., Bollivar, D. and Breitenberger, C., 2015. Comparative
721 genomics of cluster O mycobacteriophages. *PLoS One*, *10*(3), p.e0118725.Weinbauer,
722 M.G. and Rassoulzadegan, F., 2004. Are viruses driving microbial diversification and
723 diversity?. *Environmental microbiology*, *6*(1), pp.1-11

45. Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N. and Kyrpides, N.C., 2016. Uncovering Earth's virome. *Nature*, *536*(7617), pp.425-430.

46. Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H-B., Singleton, C.M., Solden, L. M., Naas, A. E., Boyd, J. A., Hodgkins, S. B., Wilson, R. M., Trubl, G., Li, L., Frolking, S., Pope, P. B., Wrighton, K. C., Crill, P. M., Chanton, J. P., Saleska, S. R., Tyson, G. W., Rich V. I., Sullivan, M. B. *In press, Nature Microbiology*. Host-linked soil viral ecology along a permafrost thaw gradient.

47. Goordial, J., Davila, A., Greer, C.W., Cannam, R., DiRuggiero, J., McKay, C.P. and Whyte, L.G., 2017. Comparative activity and functional ecology of permafrost soils and lithic niches in a hyper□arid polar desert. *Environmental microbiology*, *19*(2), pp.443-458.

48. Rosario, K. and Breitbart, M., 2011. Exploring the viral world through metagenomics. *Current opinion in virology*, *1*(4), pp.289-297.

49. Logares, R., Haverkamp, T.H., Kumar, S., Lanzén, A., Nederbragt, A.J., Quince, C. and Kauserud, H., 2012. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *Journal of microbiological methods*, *91*(1), pp.106-113.

50. Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L. and Rohwer, F., 2009. Laboratory procedures to generate viral metagenomes. *Nature protocols*, *4*(4), pp.470-483.

51. John, S.G., Mendez, C.B., Deng, L., Poulos, B., Kauffman, A.K.M., Kern, S., Brum, J., Polz, M.F., Boyle, E.A. and Sullivan, M.B., 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environmental microbiology reports*, *3*(2), pp.195-202.

747 52. Duhaime, M.B., Deng, L., Poulos, B.T. and Sullivan, M.B., 2012. Towards quantitative
748 metagenomics of wild viruses and other ultra□low concentration DNA samples: a rigorous
749 assessment and optimization of the linker amplification method. *Environmental*
750 *Microbiology*, *14*(9), pp.2526-2537Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M. and
751 Chisholm, S.W., 2005.
752 53. Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenck, S.M., Goldsmith, D.B.,
753 Coleman, M.L., Breitbart, M. and Sullivan, M.B., 2016. Towards quantitative viromics for
754 both double-stranded and single-stranded DNA viruses. *PeerJ*, *4*, p.e2777.
755 54. Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A. and Sullivan, M.B., 2017. Benchmarking
756 viromics: an in silico evaluation of metagenome-enabled estimates of viral community
757 composition and diversity. *PeerJ*, *5*, p.e3817.
758 55. Hayes, S., Mahony, J., Nauta, A. and van Sinderen, D., 2017. f. *Viruses*, *9*(6), p.127.
759 56. Binga, E.K., Lasken, R.S. and Neufeld, J.D., 2008. Something from (almost) nothing: the
760 impact of multiple displacement amplification on microbial ecology. *The ISME*
761 *journal*, *2*(3), pp.233-241.
762 57. Yilmaz, S., Allgaier, M. and Hugenholtz, P., 2010. Multiple displacement amplification
763 compromises quantitative analysis of metagenomes. *Nature methods*, *7*(12), pp.943-944.
764 58. Polson, S.W., Wilhelm, S.W. and Wommack, K.E., 2011. Unraveling the viral tapestry
765 (from inside the capsid out). *The ISME journal*, *5*(2), p.165.
766 59. Kim, M.S., Whon, T.W. and Bae, J.W., 2013. Comparative viral metagenomics of
767 environmental samples from Korea. *Genomics & informatics*, *11*(3), pp.121-128.
768 60. Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W. and
769 Wommack, K.E., 2014. Caught in the middle with multiple displacement amplification: the

myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome*, *2*(1), p.3.

61. Cremers, G., Gambelli, L., van Alen, T., van Niftrik, L. and den Camp, H.J.O., 2018. Bioreactor virome metagenomics sequencing using DNA spike-ins. *PeerJ*, *6*, p.e4351.

62. Zablocki, O., van Zyl, L., Adriaenssens, E.M., Rubagotti, E., Tuffin, M., Cary, S.C. and Cowan, D., 2014. High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of Antarctic soils. *Applied and environmental microbiology*,*80*(22), pp.6888-6897.

63. Zablocki, O., van Zyl, L., Adriaenssens, E.M., Rubagotti, E., Tuffin, M., Cary, S.C. and Cowan, D., 2014. Niche-dependent genetic diversity in Antarctic metaviromes. *Bacteriophage*, *4*(4), p.e980125.

64. Adriaenssens, E.M., Kramer, R., Van Goethem, M.W., Makhalanyane, T.P., Hogg, I. and Cowan, D.A., 2017. Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome*, *5*(1), p.83.

65. Gregory, A.C., Solonenko, S.A., Ignacio-Espinoza, J.C., LaButti, K., Copeland, A., Sudek, S., Maitland, A., Chittick, L., dos Santos, F., Weitz, J.S. and Worden, A.Z., 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC genomics*, *17*(1), p.930.

66. Roux, S. Enault, F. Hurwitz, B.L. and Sullivan, M.B. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ*, *3*, p.e985.

67. 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.

68. Halary, S., Leigh, J.W., Cheaib, B., Lopez, P., Bapteste, E. 2010. Network analyses structure genetic diversity in independent genetic worlds. Proc Natl Acad Sci U S A 107: 127-132.

69. Roux, S., Hallam, S.J., Woyke, T., Sullivan, M.B. 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife 4: 1-20.

70. Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., Solonenko, N., Lara, E., Poulain, J. and Pesant, S., 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*.

71. Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B.L. and Sullivan, M.B., 2016. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *The ISME Journal*

72. Bolduc, B., Jang, H.B., Doulcier, G., You, Z.Q., Roux, S. and Sullivan, M.B., 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, p.e3243.

73. Rombouts, S., Volckaert, A., Venneman, S., Declercq, B., Vandenheuvel, D., Allonsius, C.N., Van Malderghem, C., Jang, H.B., Briers, Y., Noben, JP., Klumpp, 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.

74. Youle, M., Haynes, M. and Rohwer, F., 2012. Scratching the surface of biology's dark matter. In *Viruses: Essential agents of life* (pp. 61-81). Springer Netherlands.

75. Hatfull, G.F. 2015. Dark matter of the biosphere: the amazing world of bacteriophage diversity. *Journal of virology*, *89*(16), pp.8107-8110.

816 76. Waldron, P.R. and Holodniy, M., 2015. Peripheral blood mononuclear cell gene expression
817 remains broadly altered years after successful interferon-based Hepatitis C Virus
818 treatment. *Journal of immunology research*.
819 77. Brum, J.R., Hurwitz, B.L., Schofield, O., Ducklow, H.W. and Sullivan, M.B., 2016.
820 Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial
821 dynamics. *The ISME journal*, *10*(2), p.437.
822 78. Zablocki, O., Adriaenssens, E.M. and Cowan, D., 2016. Diversity and ecology of viruses in
823 hyperarid desert soils. *Applied and environmental microbiology*, *82*(3), pp.770-777.
824 79. Lamont, I., Richardson, H., Carter, D.R. and Egan, J.B., 1993. Genes for the establishment
825 and maintenance of lysogeny by the temperate coliphage 186. *Journal of*
826 *bacteriology*, *175*(16), pp.5286-5288.
827 80. Villafane, R. and Black, J., 1994. Identification of four genes involved in the lysogenic
828 pathway of theSalmonella newington bacterial virus ε34. *Archives of virology*, *135*(1-2),
829 pp.179-183.
830 81. Stewart, F.M. and Levin, B.R. 1984. The population biology of bacterial viruses: why be
831 temperate. *Theoretical population biology*, *26*(1), pp.93-117.
832 82. Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.L. and Brüssow, H., 2004. Phage-host
833 interaction: an ecological perspective. *Journal of bacteriology*, *186*(12), pp.3677-3686
834 83. Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T. and Wommack, K.E.,
835 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic
836 environments. *Research in Microbiology*, *159*(5), pp.349-357.
837 84. Abedon, S.T., 2011. Communication among phages, bacteria, and soil environments.
838 In *Biocommunication in soil microorganisms* (pp. 37-65). Springer Berlin Heidelberg

85. Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S.C., Treusch, A.H., Eck, J. and Schleper, C., 2003. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Molecular microbiology*, *50*(2), pp.563-575.

86. Foesel, B.U., Nägele, V., Naether, A., Wüst, P.K., Weinert, J., Bonkowski, M., Lohaus, G., Polle, A., Alt, F., Oelmann, Y. and Fischer, M., 2014. Determinants of Acidobacteria activity inferred from the relative abundances of 16S rRNA transcripts in German grassland and forest soils. *Environmental microbiology*, *16*(3), pp.658-675.

87. Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A. and Kuramae, E.E., 2016. The ecology of Acidobacteria: moving beyond genes and genomes. *Frontiers in Microbiology*, *7*.

88. Pearce, D.A., Newsham, K.K., Thorne, M.A., Calvo-Bado, L., Krsek, M., Laskaris, P., Hodson, A. and Wellington, E.M., 2012. Metagenomic analysis of a southern maritime Antarctic soil.

89. Janssen, P.H., 1998. Pathway of glucose catabolism by strain VeGlc2, an anaerobe belonging to the *Verrucomicrobiales* lineage of bacterial descent. *Applied and environmental microbiology*, *64*(12), pp.4830-4833.

90. Kant, R., Van Passel, M.W., Sangwan, P., Palva, A., Lucas, S., Copeland, A., Lapidus, A., del Rio, T.G., Dalin, E., Tice, H. and Bruce, D., 2011. Genome sequence of Pedosphaera parvula Ellin514, an aerobic verrucomicrobial isolate from pasture soil. *Journal of bacteriology*.

860 91. Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A.,
861 Knight, R. and Fierer, N., 2011. The under-recognized dominance of Verrucomicrobia in
862 soil bacterial communities. *Soil Biology and Biochemistry*, *43*(7), pp.1450-1455.
863 92. Štursová, M., Žifčáková, L., Leigh, M.B., Burgess, R. and Baldrian, P., 2012. Cellulose
864 utilization in forest litter and soil: identification of bacterial and fungal
865 decomposers. *FEMS Microbiology Ecology*, *80*(3), pp.735-746.
866 93. Soares Jr, F.L., Melo, I.S., Dias, A.C.F. and Andreote, F.D., 2012. Cellulolytic bacteria
867 from soils in harsh environments. *World Journal of Microbiology and*
868 *Biotechnology*, *28*(5), pp.2195-2203.
869 94. Schmidt, O., Hink, L., Horn, M.A. and Drake, H.L., 2016. Peat: home to novel syntrophic
870 species that feed acetate-and hydrogen-scavenging methanogens. *The ISME journal*, *10*(8),
871 pp.1954-1966.
872 95. Wawrik, B., Marks, C.R., Davidova, I.A., McInerney, M.J., Pruitt, S., Duncan, K.E.,
873 Suflita, J.M. and Callaghan, A.V., 2016. Methanogenic paraffin degradation proceeds via
874 alkane addition to fumarate by 'Smithella'spp. mediated by a syntrophic coupling with
875 hydrogenotrophic methanogens. *Environmental microbiology*, *18*(8), pp.2604-2619.
876 96. Juottonen, H., Eiler, A., Biasi, C., Tuittila, E.S., Yrjälä, K. and Fritze, H., 2017. Distinct
877 anaerobic bacterial consumers of cellobiose-derived carbon in boreal fens with different
878 CO_2/CH_4 production ratios. *Applied and environmental microbiology*, *83*(4), pp.e02533-16.
879 97. Daly, R.A., Borton, M.A., Wilkins, M.J., Hoyt, D.W., Kountz, D.J., Wolfe, R.A., Welch,
880 S.A., Marcus, D.N., Trexler, R.V., MacRae, J.D. and Krzycki, J.A. 2016. Microbial
881 metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nature*
882 *Microbiology*, *1*, p.16146.

98. Anderson, C.L., Sullivan, M.B. and Fernando, S.C., 2017. Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome*, *5*(1), p.155

99. Solden LM, Roux S, Daly RA, Collis WB, Naas AE, Nicora CD, Purvine SO, Hoyt DW, Schuckel J, Jorgensen B, Willats W, Spalinger DE, Firkins JL, Lipton MS, Sullivan MB, Pope PB, Wrighton KC. Decrypting carbon degradation and phage infection networks in the rumen ecosystem. Submitted to Nature Microbiology.

100. Kormanec, J. and Homerova, D., 1993. Streptomyces aureofaciens whiB gene encoding putative transcription factor essential for differentiation. *Nucleic acids research*, *21*(10), p.2512.

101. Resnekov, O., Driks, A. and Losick, R., 1995. Identification and characterization of sporulation gene spoVS from Bacillus subtilis. *Journal of bacteriology*, *177*(19), pp.5628-5635.

102. Rybniker, J., Nowag, A., Van Gumpel, E., Nissen, N., Robinson, N., Plum, G. and Hartmann, P., 2010. Insights into the function of the WhiB□like protein of mycobacteriophage TM4–a transcriptional inhibitor of WhiB2. *Molecular microbiology*, *77*(3), pp.642-657.

103. Crummett, L.T., Puxty, R.J., Weihe, C., Marston, M.F. and Martiny, J.B., 2016. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology*, *499*, pp.219-229.

104. Jansson, J.K. and Hofmockel, K.S., 2018. The soil microbiome—from metagenomics to metaphenomics. *Current opinion in microbiology*, *43*, pp.162-168.

105. Thompson, M.R., Kaminski, J.J., Kurt-Jones, E.A. and Fitzgerald, K.A., 2011. Pattern recognition receptors and the innate immune response to viral infection. *Viruses*, *3*(6), pp.920-940.

106. Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K.A., Toner, B.M. and Dick, G.J., 2014. Sulfur oxidation genes in diverse deep-sea viruses. *Science*, *344*(6185), pp.757-760.

107. Martin, J.K., 1977. Effect of soil moisture on the release of organic carbon from wheat roots. *Soil Biology and Biochemistry*, *9*(4), pp.303-304.

108. Floyd, M. M., J. Tang, M. Kane, and D. Emerson. 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.

109. Fierer, N., Bradford, M.A. and Jackson, R.B., 2007. Toward an ecological classification of soil bacteria. *Ecology*, *88*(6), pp.1354-1364.

110. Jansson, J.K. and Taş, N., 2014. The microbial ecology of permafrost. *Nature reviews Microbiology*, *12*(6), pp.414-425.

111. Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K. and Fierer, N., 2018. A global atlas of the dominant bacteria found in soil. *Science*, *359*(6373), pp.320-325.

112. Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T. and Young, M.J., 2001. Viruses from extreme thermal environments. *Proceedings of the National Academy of Sciences*, *98*(23), pp.13341-13345.

113. Laybourn-Parry, J., Marshall, W.A. and Madan, N.J., 2007. Viral dynamics and patterns of lysogeny in saline Antarctic lakes. *Polar Biology*, *30*(3), pp.351-358.

926 114. Le Romancer, M., Gaillard, M., Geslin, C. and Prieur, D., 2007. Viruses in extreme
927 environments. *Reviews in Environmental Science and Bio/Technology*, *6*(1-3), pp.17-31.
928 115. Evans, C. and Brussaard, C.P., 2012. Regional variation in lytic and lysogenic viral
929 infection in the Southern Ocean and its contribution to biogeochemical cycling. *Applied*
930 *and environmental microbiology*, *78*(18), pp.6741-6748.
931 116. Payet, J.P. and Suttle, C.A., 2013. To kill or not to kill: the balance between lytic and
932 lysogenic viral infection is driven by trophic status. *Limnol. Oceanogr*, *58*(2), pp.465-474.
933 117. McMurdie, P.J. and Holmes, S., 2014. Waste not, want not: why rarefying microbiome data
934 is inadmissible. *PLoS computational biology*, *10*(4), p.e1003531.
935 118. Hurst, C.J., Gerba, C.P. and Cech, I., 1980. Effects of environmental variables and soil
936 characteristics on virus survival in soil. *Applied and environmental microbiology*, *40*(6),
937 pp.1067-1079.
938 119. Gerba, C.P., 1984. Applied and theoretical aspects of virus adsorption to
939 surfaces. *Advances in applied microbiology*, *30*, pp.133-168.
940 120. Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M.,
941 Edwards, R.A., Felts, B., Rayhawk, S. and Knight, R., 2007. Metagenomic and small-
942 subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses
943 in soil. *Applied and environmental microbiology*, *73*(21), pp.7059-7066.
944 121. Kavanaugh, M.T., Oliver, M.J., Chavez, F.P., Letelier, R.M., Muller-Karger, F.E. and
945 Doney, S.C., 2016. Seascapes as a new vernacular for pelagic ocean monitoring,
946 management and conservation. *ICES Journal of Marine Science*, *73*(7), pp.1839-1850.
947 122. Steward, G.F., Culley, A.I., Mueller, J.A., Wood-Charlson, E.M., Belcaid, M. and Poisson,
948 G., 2013. Are we missing half of the viruses in the ocean?. The ISME journal, 7(3), p.672.

123. Greninger, A.L., 2017. A decade of RNA virus metagenomics is (not) enough. *Virus Research*.

124. Zhang, Y.Z., Shi, M. and Holmes, E.C., 2018. Using Metagenomics to Characterize an Expanding Virosphere. *Cell*, *172*(6), pp.1168-1172.

125. Rinke, C., Low, S., Woodcroft, B.J., Raina, J.B., Skarshewski, A., Le, X.H., Butler, M.K., Stocker, R., Seymour, J., Tyson, G.W. and Hugenholtz, P., 2016. Validation of picogram- and femtogram-input DNA libraries for microscale metagenomics. *PeerJ*, *4*, p.e2486.

126. Lang, A.S., Westbye, A.B. and Beatty, J.T., 2017. The Distribution, Evolution, and Roles of Gene Transfer Agents (GTAs) in Prokaryotic Genetic Exchange. *Annual review of virology*, *4*(1).

127. Kuhn, E., Ichimura, A.S., Peng, V., Fritsen, C.H., Trubl, G., Doran, P.T. and Murray, A.E., 2014. Brine assemblages of ultrasmall microbial cells within the ice cover of Lake Vida, Antarctica. *Applied and environmental microbiology*, *80*(12), pp.3687-3698.

128. Luef, B., Frischkorn, K.R., Wrighton, K.C., Holman, H.Y.N., Birarda, G., Thomas, B.C., Singh, A., Williams, K.H., Siegerist, C.E., Tringe, S.G. and Downing, K.H., 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature communications*, *6*, p.6372.

129. Solden, L., Lloyd, K. and Wrighton, K., 2016. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Current opinion in microbiology*, *31*, pp.217-226.

130. Sariaslani, Sima and Gadd, Geoffrey Michael. Advances in applied microbiology. Vol. 101. Elsevier academic press, 2017

131. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, *41*(D1), pp.D590-D596.

132. Bakken, L.R. and Olsen, R.A., 1983. Buoyant densities and dry-matter contents of microorganisms: conversion of a measured biovolume into biomass. *Applied and Environmental Microbiology*, *45*(4), pp.1188-1195.

133. Pollard, E.C. and Grady, L.J., 1967. CsCl density gradient centrifugation studies of intact bacterial cells. *Biophysical journal*, *7*(2), p.205.

134. Bolger, A.M. Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, p.btu170.

135. Peng, Y. Leung, H.C. Yiu, S.M. and Chin, F.Y. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), pp.1420-1428.

136. Vik, D.R., Roux, S., Brum, J.R., Bolduc, B., Emerson, J.B., Padilla, C.C., Stewart, F.J. and Sullivan, M.B., 2017. Putative archaeal viruses from the mesopelagic ocean. *PeerJ*, *5*, p.e3428.

137. Hyatt, D., LoCascio, P.F., Hauser, L.J. and Uberbacher, E.C., 2012. Gene and translation initiation site prediction in metagenomics sequences. *Bioinformatics*, *28*(17), pp.2223-2230.

138. Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), pp.27-30.

991 139. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez,
992 R., 2005. InterProScan: protein domains identifier. *Nucleic acids research*, *33*(suppl_2),
993 pp.W116-W120.
994 140. Edgar, R.C., 2010. Search and clustering orders of magnitude faster than
995 BLAST. *Bioinformatics*, *26*(19), pp.2460-2461.
996 141. Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high
997 throughput. *Nucleic acids research*, *32*(5), pp.1792-1797.
998 142. Price, M.N., Dehal, P.S. and Arkin, A.P., 2009. FastTree: computing large minimum
999 evolution trees with profiles instead of a distance matrix. *Molecular biology and*
1000 *evolution*, *26*(7), pp.1641-1650.
1001 143. Letunic, I. and Bork, P., 2006. Interactive Tree Of Life (iTOL): an online tool for
1002 phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), pp.127-128.
1003 144. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y., 2015. The I-TASSER Suite:
1004 protein structure and function prediction. *Nature methods*, *12*(1), p.7.
1005 145. Zhang, Y. and Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based
1006 on the TM-score. *Nucleic acids research*, *33*(7), pp.2302-2309.
1007 146. Brister, J.R., Ako-Adjei, D., Bao, Y. and Blinkova, O., 2014. NCBI viral genomes
1008 resource. *Nucleic acids research*, *43*(D1), pp.D571-D577.
1009 147. Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature*
1010 *methods*, *9*(4), pp.357-359.
1011 148. Sanguino, L., Franqueville, L., Vogel, T.M. and Larose, C., 2015. Linking environmental
1012 prokaryotic viruses and their host through CRISPRs. *FEMS microbiology ecology*, *91*(5),
1013 p.fiv046.

149. Emerson, J.B., Andrade, K., Thomas, B.C., Norman, A., Allen, E.E., Heidelberg, K.B. and Banfield, J.F., 2013. Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea*.

150. Skennerton, C.T., Imelfort, M. and Tyson, G.W., 2013. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic acids research*, p.gkt183.

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

152. 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.

Table legends

Table 1. Soil viromes read information. The 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 non-redundant set of contigs to estimate their relative abundance, calculated as number of bp mapped to each read normalized by the length of the contig and the total number of bp 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 as detected reads had to cover ≥75% of the contig.

Table 2. Soil viruses' bioinformatics information. All 393 putative soil viruses are listed (378 after VirSorter/MArVD and manual inspection). For the vOTUs, the virome(s) in which it originated from, its genomic information, and its coverage is provided. For the other putative soil viral contigs, the origin virome(s) is provided, and contig length are provided. Additionally, the three mobile genetic elements and ten viral contigs with no coverage are reported with their virome(s) of origin (if applicable) and contig length. No contigs were chimeric (i.e. constructed with reads coming from multiple viromes). A † denotes the contig did not meet our threshold for read mapping (i.e. reads recruited to contigs only if they had 90% ANI and then if $\geq$ 70% of the contig was covered) and therefore could not be counted as detected.

Figure legends

Figure 1. Overview of sample-to-ecology methods pipeline. Sampling of the thaw chronosequence at Stordalen Mire (68°21 N, 19°03 E, 359 m a.s.l.). The underlying image was collected via unmanned aerial vehicle (UAV) and extensively manually curated for GPS accuracy (generated by Dr. Michael Palace). Sampling locations were mapped onto this image based on their GPS coordinates. Soil cores were taken in July of 2014. Viruses were resuspended as previously described in Trubl et al. (41). Viromes were generated using samples from 36–40 cm. Identified vOTUs were further characterized using geochemical data and metagenome-assembled genomes (MAGs; 16) from Stordalen Mire. Additionally, these vOTUs were compared to the vOTUs from bulk-soil-derived viromes (46).

Figure 2. Relating Stordalen Mire viruses to known viral sequence space. 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 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 (slightly off because it's a 2-dimensional representation of a 3-D space). Dashed lines indicate two network regions of consistent known taxonomy, allowing assignment of contigs 4, 143, and 28. The pie chart represents the number of the Stordalen Mire viral proteins (i) that are recovered by protein clusters (PCs) (yellow and red) and singletons (gray) and (ii) that are shared with RefSeq viruses (yellow) or not (red and gray). Proteins of viral genomes/vOTUs in the dataset were grouped into PCs through all-to-all BlastP comparisons (E-value cut-off $<10^{-4}$) followed by Markov clustering algorithm-based clustering (see Materials and Methods). Proteins that were not grouped into PCs are designated as singletons.

Figure 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; see Table S2) and 2 by BLAST (dotted line). Node shape denotes organism (oval for microbe and triangle for virus). Viral nodes are color coded by habitat of origin (green for bog and blue for fen).

Figure 4. Characterization of select AMGs. FastTree phylogenies were constructed for select AMGs (one from each group and one example for central carbon metabolism), and their structures and those of their nearest neighbors were predicted using iTasser (detailed in Table S3). Tree lineages are shaded blue for bacteria and red for viruses. "vOTU" sequences are from the 53-vOTU virome-derived dataset, while "SoilVir contig" represent homologues from Emerson et al.'s (46) bulk metagenome-derived 1907 vOTUs. AMGs are color coded by

function: green for central carbon metabolism, blue for polysaccharide degradation, yellow for polysaccharide binding, and purple for sporulation. The first predicted model for each soil virus is shown and was used for the TM-align comparison. Structures are ordered from left to right based on the appearance of their sequence in the tree from top to bottom.

Figure 5. Three views into viral community structure across the thaw gradient. (A) The relative abundance of vOTUs (columns) in the seven viromes (rows). Reads were mapped to this non-redundant set of contigs to estimate their relative abundance (calculated as bp mapped to each read, normalized by contig length and the total bp in each metagenome). Red color gradient indicates $\log_{10}$ coverage per Gbp of metagenome. 'Chilled' and 'Frozen' indicate sample storage at 4°C or flash frozen in liquid nitrogen and stored at −80°C. "A" and "B" denote technical replicates. Dots after contig names indicate membership in a viral cluster, filled dots denote cluster is novel, and fill color indicates habitat specificity (palsa = brown, bog = green, fen = blue). (B) Euler diagram relating the seven viromes and their 53 vOTUs. (C) Principal coordinate analysis of the viromes by normalized relative abundance of the 53 vOTUs.

Figure 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 non-redundant set of contigs to estimate their relative abundances (calculated as bp mapped to each read, normalized by contig length and the total bp in each metagenome). Red color gradient indicates $\log_{10}$ coverage per Gbp of metagenome. 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 (palsa = P, bog = S, fen = E); soil core replicate (1–3); depth (3 cm intervals denoted with respect to

geochemical transitions, see 46; generally, S = 1–4 cm, M = 5–14 cm, D = 11–33 cm, and X = 30–50 cm); month collected (5–10 as May–October); year collected (2010, 11, 12). See Emerson et al. (46) and Woodcroft et al. (16) for more metagenome and sample details.

Figure 7. Contrasting Stordalen Mire viruses derived from viromes and bulk-soil metagenomes. Currently, two datasets exist describing Stordalen Mire (SM) Archaeal and Bacterial viruses. Emerson et al. (46) characterized the viral signal in bulk-soil metagenomes (described in 16), 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 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). Nineteen vOTUs are shared between the two datasets (57, 90, 110, 111, 141, 144, 155, 157, 179, 183, 186, 189, 197, 204, 208, 243, 246, 261, and 264). Also shown are the methods that produced each dataset. * denotes the viral signal was mined from bulk-soil metagenomes. † denotes that viruses were resuspended from the soils using a previously optimized protocol (41). # denotes the vOTU yield normalized per Gbp 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 in 46).

Supplementary Table legends

Table S1. Virally-encoded auxiliary metabolic genes and other genes of interest. Genes were annotated and AMGs identified by running assembled contigs through a pipeline developed by

the Wrighton lab at The Ohio State University previously described in Daly et al. (23). The habitat that the vOTU was derived from is listed. Predicted genes that are AMGs or integrase-related are bolded and unannotated genes are no present. Additionally, the PhoH-like protein is bolded due to its highly debated function as a phosphate starvation gene (reviewed in 24).

Table S2. Viral-host linkages supporting information. NCBI BLAST linkages were determined based on queries and CRISPR information was provided using Crass software. Host genomes IDs were assigned from the Joint Genome Institute's Integrated Microbial Genomes Database. Microbial bins were pulled from Woodcroft et al. (*1*).

Table S3. Structural comparison between select AMGs and phylogenetic neighbors. Predicted structures for AMGs and neighbors were determined and a comparison of the first model of their predicted structure was performed using TM-align. Structure similarity between two proteins is rated on a scale of 0.0–1.0, with TM-scores < 0.30 suggest random structural similarity and scores = 0.5 suggest similar folds and scores near 1 suggest a perfect match between two structures.

Table S4. Codon usage frequency. The codon usage frequency was determined for the 53 vOTUs and the linked microbial bins.

Supplementary Figure legends

Figure S1. Phylogenetic analysis of vOTU 4. Phylogenetic relationships between vOTU_4 and its related viruses. A maximum-likelihood tree was constructed upon a concatenation of two structural proteins (major capsid protein and baseplate protein) that are common to the *Felixounvirinae* and *Vequintavirinae* viruses. The numbers at the branch represent the bootstrapping probabilities from 1000 replicates. Edges with bootstrap values above 75% are represented. The scale bar indicates the number of substitution per site.

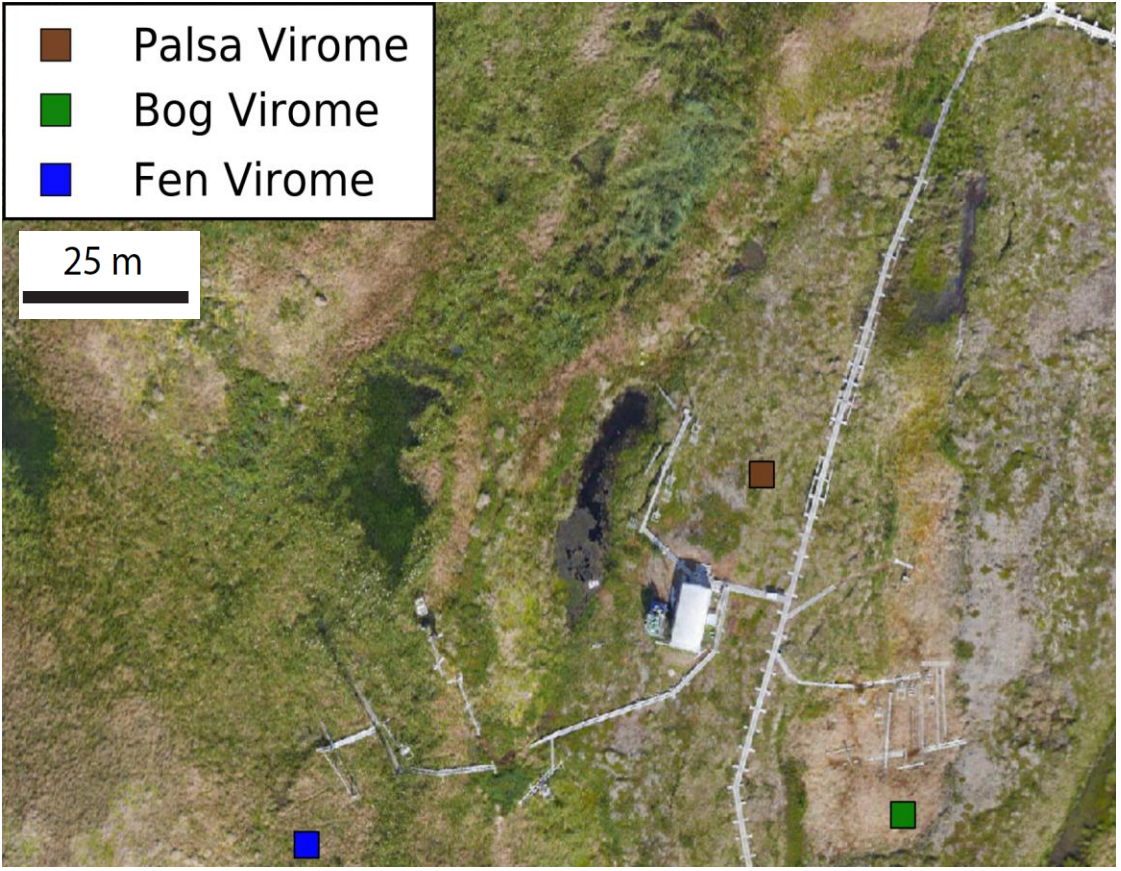
Figure S2. Viral biodiversity increases with permafrost thaw. Richness, Shannon's Diversity index and Pielou's evenness index were calculated for each virome and the viromes were plotted by habitat. Chilled samples are denoted with a lighter color and frozen samples denoted with a darker color. (A) The diversity indices for all seven viromes. (B) The diversity indices of six viromes (bog chilled B was removed).

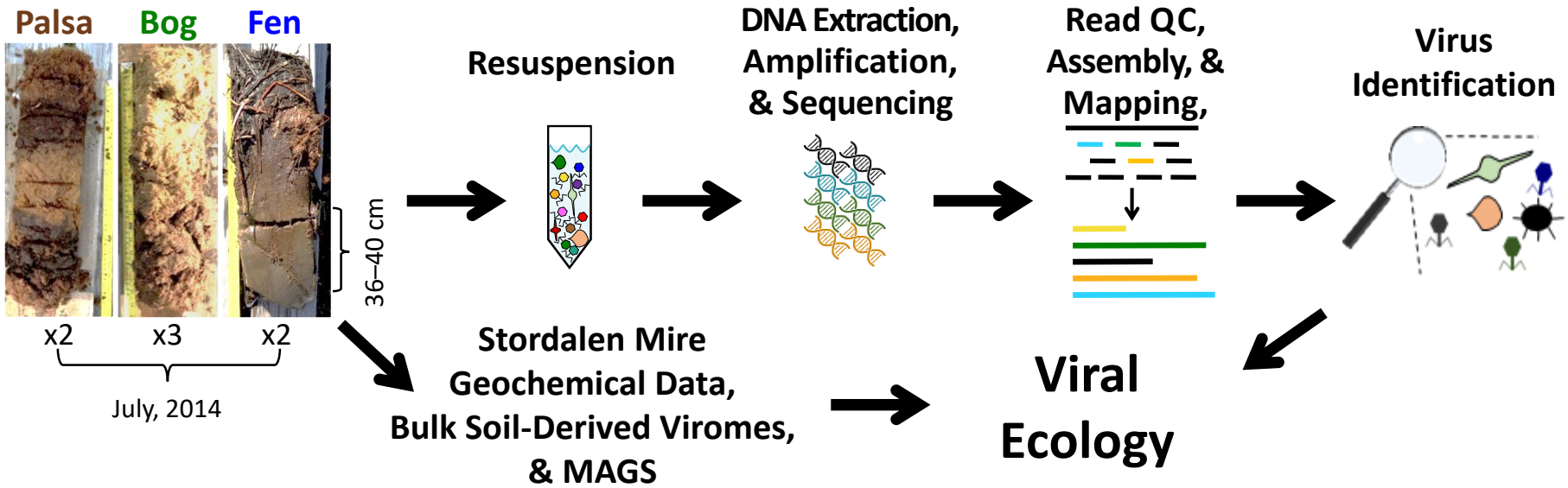
Figure S3. Visualizing relationships among soil viral communities. The y-axis is a measure of Bray-Curtis dissimilarity, with an average dissimilarity used for viromes (i.e. dissimilarities are averaged at each step between viromes for the agglomerative method). Bootstraps n=1000; two types of *p*-values: Approximately Unbiased (AU) *p*-value in blue and Bootstrap Probability (BP) value in purple. AU *p*-value, which is computed by multiscale bootstrap resampling, is a better approximation to unbiased *p*-value than BP value computed by normal bootstrap resampling. Social networks of: (A) the 53 vOTU sequences and (B) all the reads mapped to the 53 vOTUs from the seven viromes with clusters circled in black. Dots in the social networks represent statistical samples taken from the marginal posterior distributions (Bayesian Method).

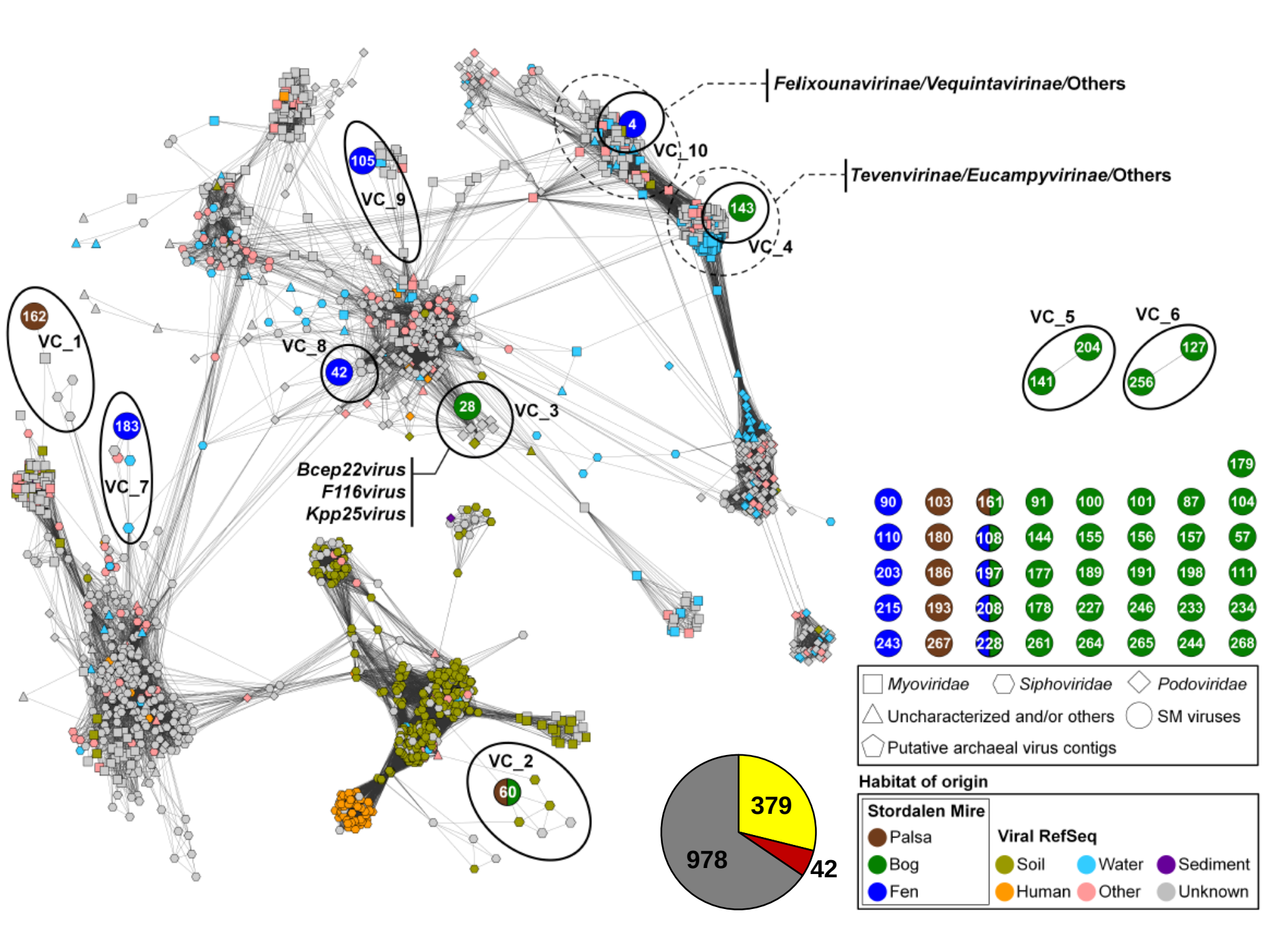
Figure S4. Identified Viral Signal in the MAGs. (A) The stacked bar chart shows the percent of viral signal occurrences from the 53 vOTUs collected in 2014 in the 133 bulk-soil

metagenomes that had a signal collected from 2010–2012. In 2010, only fen samples were collected for microbial metagenomes. Viral signal occurrences were normalized by the number of viromes constructed for each habitat and the number of metagenomes for each habitat. The total number of occurrences for each year is italicized. (B) The number of occurrences (presented as a percentage) of a 'viral signal' in a bulk-soil metagenome partitioned by the origin of the bulk-soil metagenome and the vOTU.

Figure S5. Codon usage frequency for the linked viruses and their microbial hosts. Principal Coordinates Analysis of the codon usage frequency of microbial hosts and their linked viruses, using the Bray-Curtis dissimilarity metric. Microbial hosts are denoted by circles and colored by phylum (A) or genus/species (B). The associated viruses have a matching color to its host and are denoted with a square. (C) The average dissimilarity metric between the viral contigs linked to potential microbial hosts is plotted against each viruses' contig length ($x10^3$). Average dissimilarity distance was used with viral contigs with multiple hosts.

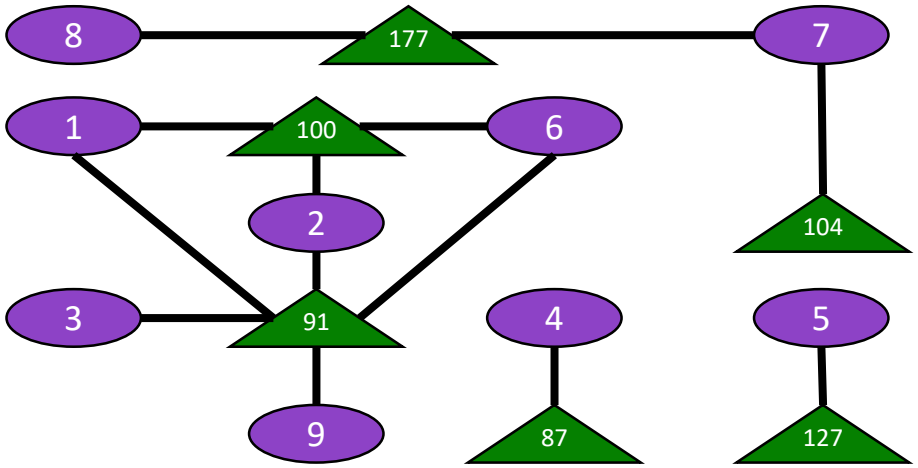

Palsa Virome
Bog Virome
Fen Virome
25 m

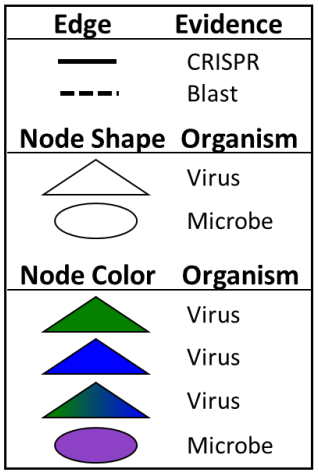

Palsa
Bog
Fen
36–40 cm
x2
x3
x2
July, 2014
Resuspension
DNA Extraction, Amplification, & Sequencing
Read QC, Assembly, & Mapping,
Virus Identification
Stordalen Mire Geochemical Data, Bulk Soil-Derived Viromes, & MAGS
Viral Ecology


Felixounavirinae/Vequintavirinae/Others
Tevenvirinae/Eucampyvirinae/Others
VC_10
VC_9
VC_4
VC_8
VC_1
VC_7
VC_3
VC_2
VC_5
VC_6
Bcep22virus
F116virus
Kpp25virus
4
105
143
162
183
42
28
60
204
141
127
256
90 103 161 91 100 101 87 179
110 180 108 144 155 156 157 57
203 186 197 177 189 191 198 111
215 193 208 178 227 246 233 234
243 267 228 261 264 265 244 268
Myoviridae
Siphoviridae
Podoviridae
Uncharacterized and/or others
SM viruses
Putative archaeal virus contigs
Habitat of origin
Stordalen Mire
Palsa
Bog
Fen
Viral RefSeq
Soil
Water
Sediment
Human
Other
Unknown
379
978
42

Verrucomicrobia

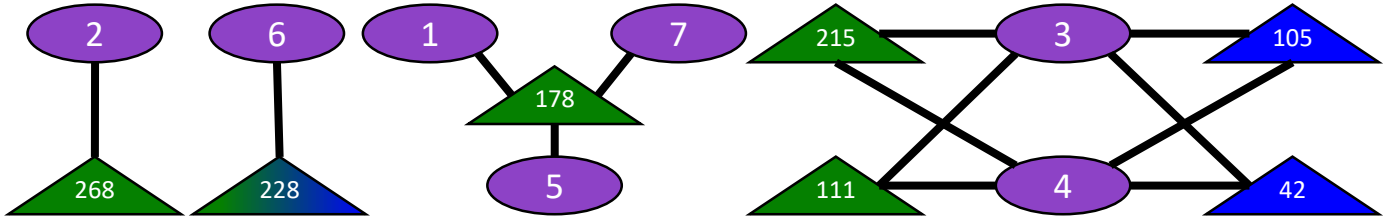
(*Verrucomicrobiae;*
Verrucomicrobiales;
Verrucomicrobia subdivision 3;
Pedosphaera;
Pedosphaera parvula)



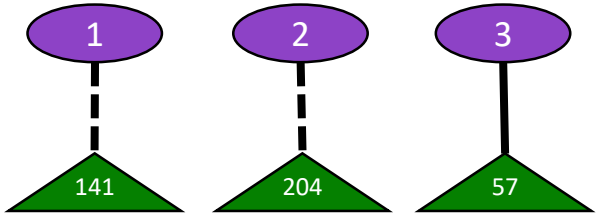


Acidobacteria

(*Acidobacteriales;*
Acidobacteriaceae;
Acidobacterium)

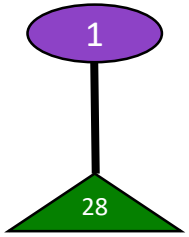


(*Solibacteres;*
Solibacterales;
Solibacteraceae;
Candidatus Solibacter;
Solibacter usitatus)



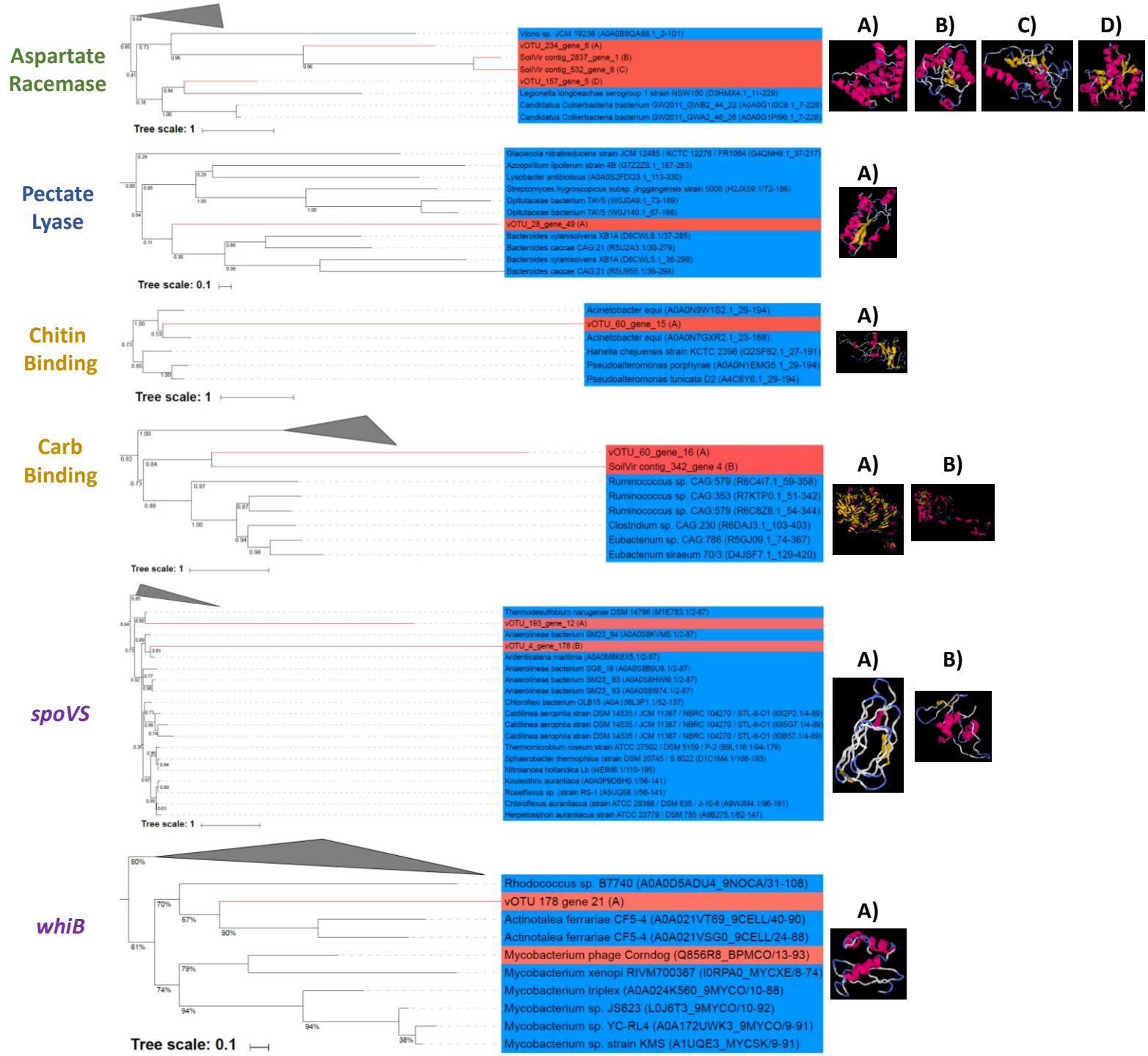
Proteobacteria

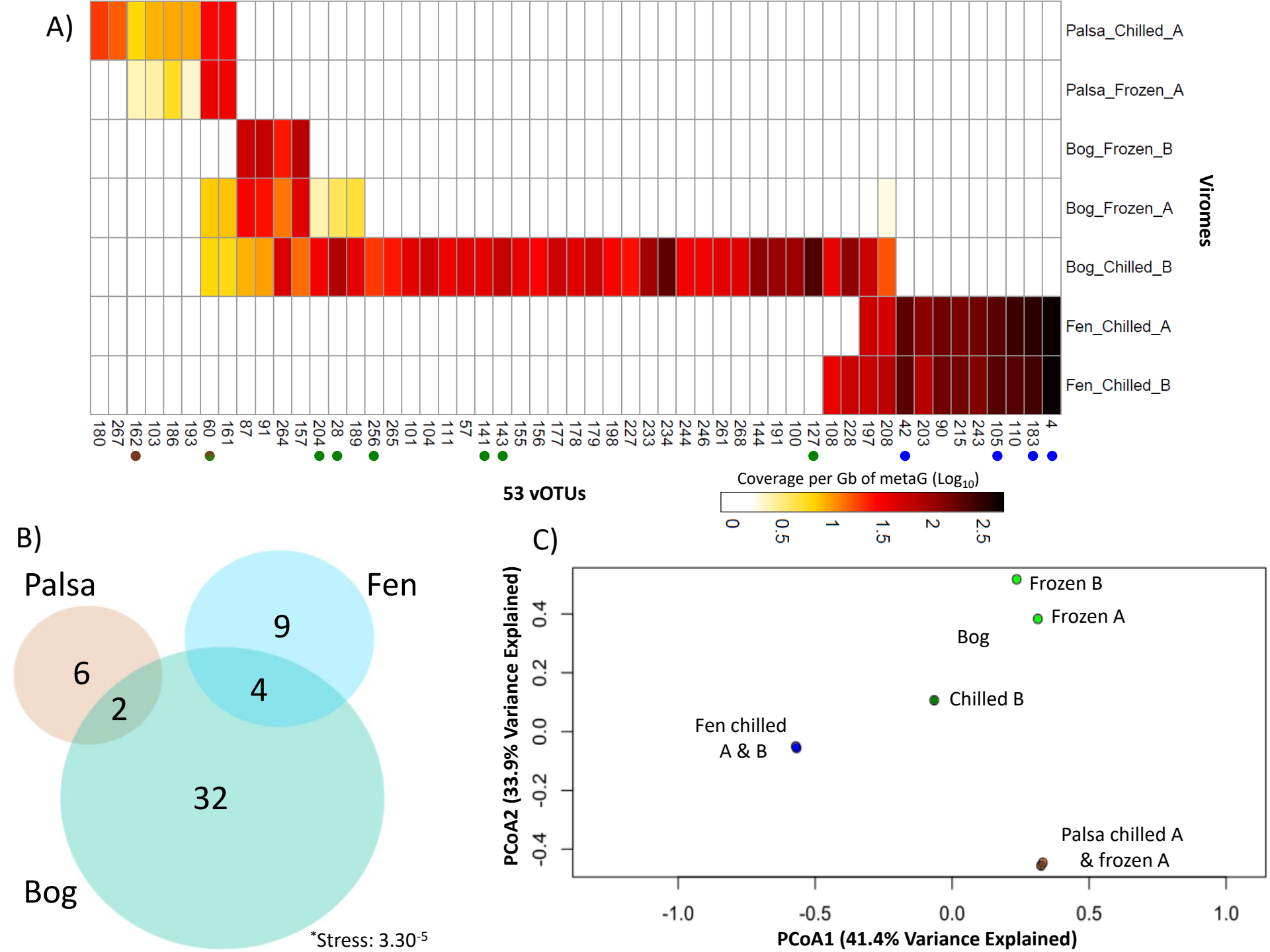
(*Deltaproteobacteria;*
Syntrophobacterales;
Syntrophaceae;
Smithella sp. SDB)

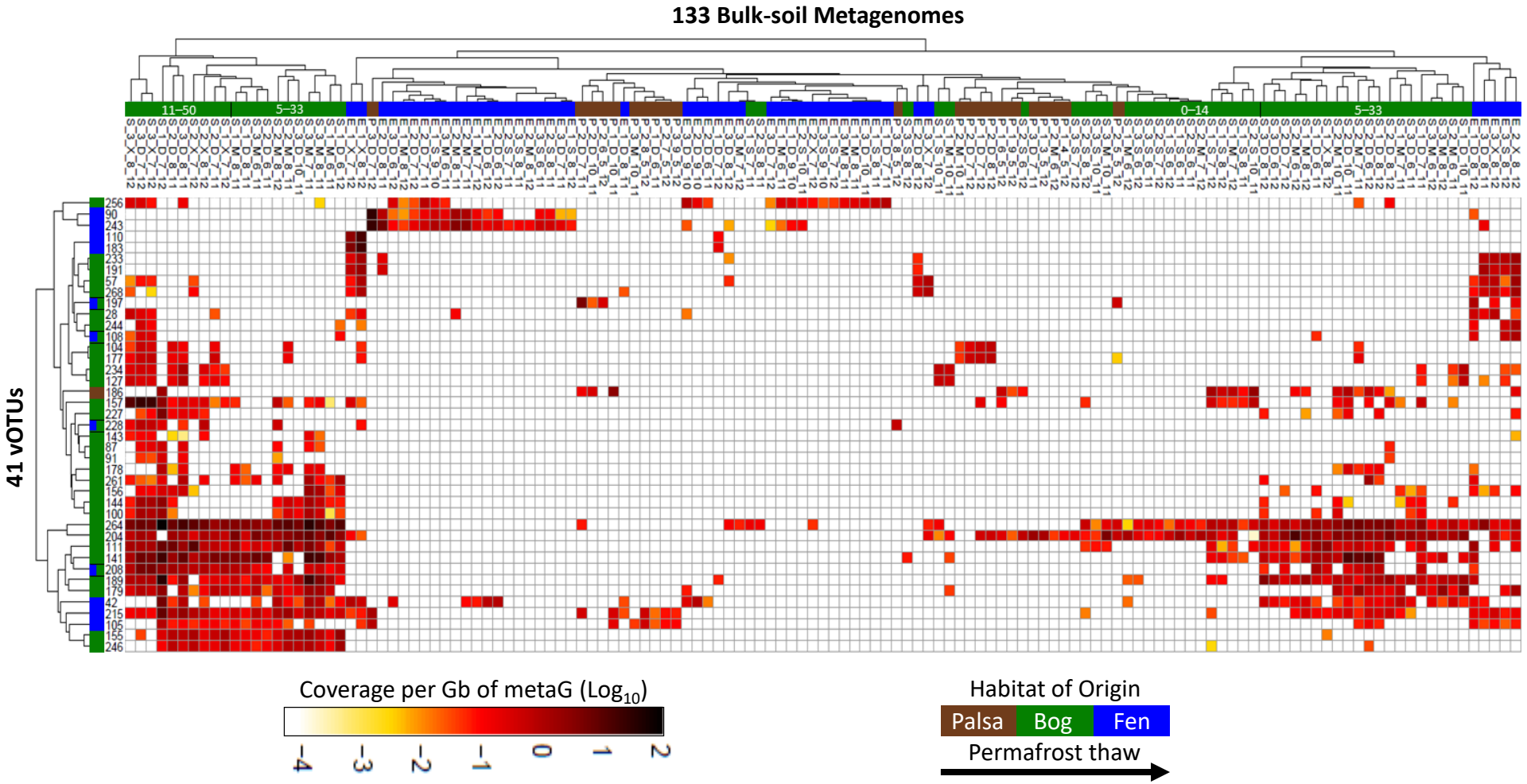


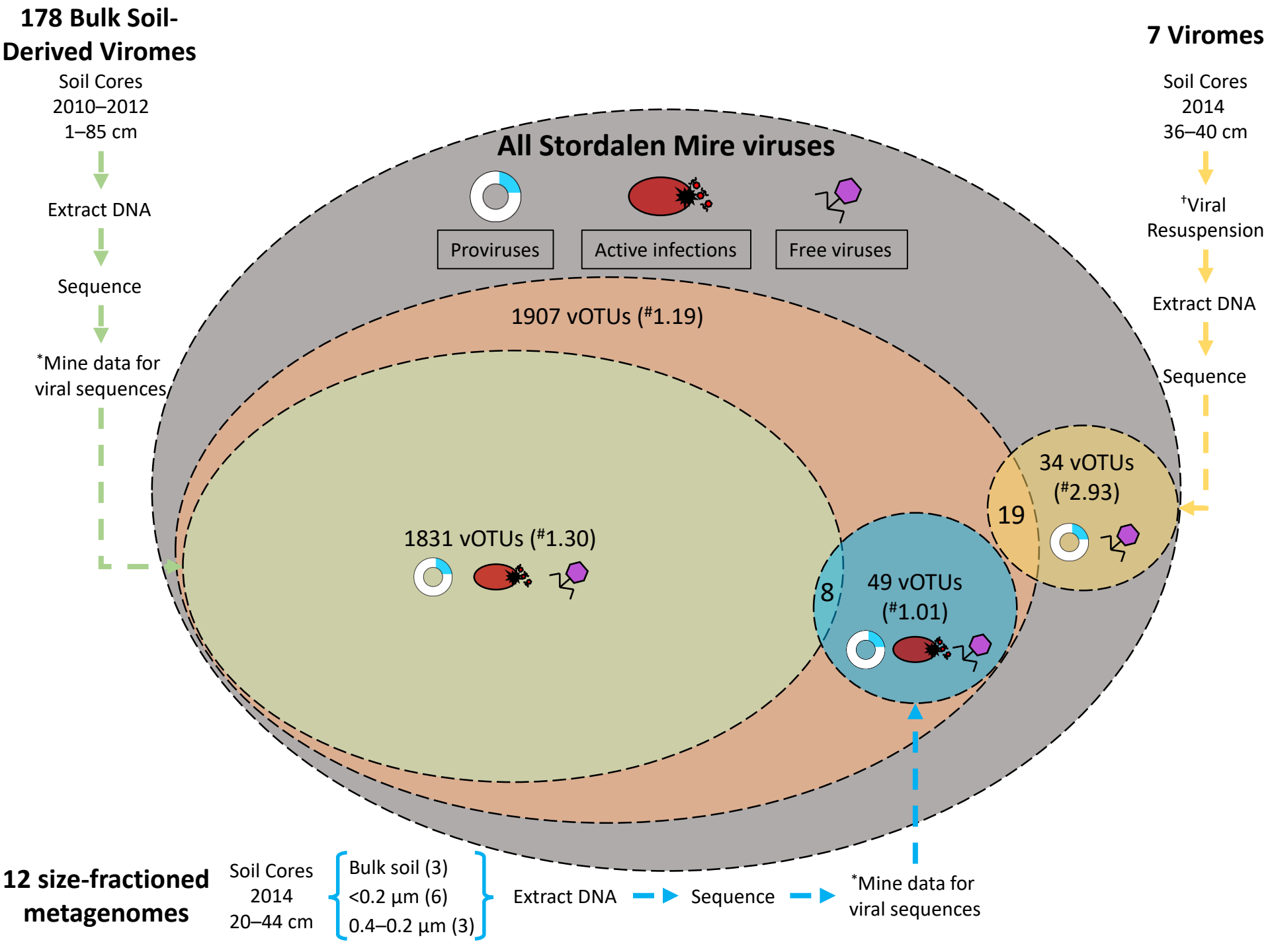
Organism
- Bacteria
- Virus

AMG Function
- Central Carbon Metabolism
- Polysaccharide Degradation
- Polysaccharide Binding
- Sporulation




A)
Palsa_Chilled_A
Palsa_Frozen_A
Bog_Frozen_B
Bog_Frozen_A
Bog_Chilled_B
Fen_Chilled_A
Fen_Chilled_B
Viromes
180 267 162 103 186 193 60 161 87 91 264 157 204 28 189 256 265 101 104 111 57 141 143 155 156 177 178 179 198 227 233 234 244 246 261 268 144 191 100 127 108 228 197 208 42 203 90 215 243 105 110 183 4
53 vOTUs
Coverage per Gb of metaG (Log10)
0 0.5 1 1.5 2 2.5
B)
Palsa
Fen
Bog
6
2
9
4
32
*Stress: 3.30-5
C)
Frozen B
Frozen A
Bog
Chilled B
Fen chilled
A & B
Palsa chilled A
& frozen A
PCoA2 (33.9% Variance Explained)
0.4 0.2 0.0 -0.2 -0.4
PCoA1 (41.4% Variance Explained)
-1.0 -0.5 0.0 0.5 1.0


133 Bulk-soil Metagenomes
11–50
5–33
0–14
5–33
41 vOTUs
Coverage per Gb of metaG (Log10)
−4
−3
−2
−1
0
1
2
Habitat of Origin
Palsa
Bog
Fen
Permafrost thaw


178 Bulk Soil-Derived Viromes
Soil Cores
2010–2012
1–85 cm
Extract DNA
Sequence
*Mine data for viral sequences
7 Viromes
Soil Cores
2014
36–40 cm
†Viral Resuspension
Extract DNA
Sequence
All Stordalen Mire viruses
Proviruses
Active infections
Free viruses
1907 vOTUs (#1.19)
1831 vOTUs (#1.30)
8
49 vOTUs (#1.01)
19
34 vOTUs (#2.93)
12 size-fractioned metagenomes
Soil Cores
2014
20–44 cm
Bulk soil (3)
<0.2 µm (6)
0.4–0.2 µm (3)
Extract DNA
Sequence
*Mine data for viral sequences