Full title: RNA virus ecogenomics along a subarctic permafrost thaw gradient

Akbar Adjie Pratama[1,2,3†], Guillermo Domínguez-Huerta[1,2,3†], James M. Wainaina[1,2,3], Benjamin Bolduc[1,2,3], Funing Tian[1,2,3], Jared Ellenbogen[2,4], Jiarong Guo[1,2], EMERGE Field Team, EMERGE Coordinators, Kelly C. Wrighton[2,4], Ahmed A. Zayed[1,2,3] & Matthew B. Sullivan[1,2,3,5*]

[1] Department of Microbiology, The Ohio State University, Columbus, Ohio, USA
[2] EMERGE Biology Integration Institute,
[3] Center of Microbiome Science, The Ohio State University, Columbus, Ohio, USA
[4] Department of Soil and Crop Science, Colorado State University, Fort Collins, Colorado, USA
[5] Department of Civil, Environmental, and Geodetic Engineering and Evolution, Ecology, and Organismal Biology at The Ohio State University, Ohio State University, Columbus, Ohio, USA

Current addresses:
Akbar Adjie Pratama
Friedrich Schiller University Jena, Institute of Biodiversity, Aquatic Geomicrobiology, Jena, Germany
Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena, Jena, Germany
Guillermo Domínguez-Huerta
Instituto Español de Oceanografía (IEO-CSIC), Centro Oceanográfico de Málaga, Málaga, Spain
James M. Wainaina
Department of Biology Woods Hole Oceanographic Institution, Woods Hole, MA, USA

[†] These authors contributed equally to this work.
[*]Corresponding author: sullivan.948@osu.edu

ABSTRACT

Climate change thaws permafrost, which releases greenhouse gases partly from dormant microorganisms awakening and metabolizing organic matter. Though DNA viruses that infect these soil microbes have been studied, little is known on soil RNA viruses, which typically infect microeukaryotes. Here we identify and characterize 2,651 RNA viruses from a 4-year time series of bulk soil metatranscriptomes derived from the climatically fragile Stordalen Mire ecosystem —a long studied permafrost peatland. RNA virus diversity was structured by habitat (palsa, bog, and fen), and these patterns correlated with pH and carbon dioxide and methane emissions. Further, host prediction, virus-encoded metabolite-transforming and information-processing functions suggested roles in ecosystem-scale carbon fluxes and contribute to greenhouse gases emissions. Together, these RNA virus ecogenomic data in permafrost provide essential baseline information for integration into predictive models to support hypothesis testing.

Climate change accelerates permafrost thaw through reduced ice cover lessening albedo (*1*). To study this, a model ecosystem – Stordalen Mire in northern Sweden – was established decades ago and revealed that permafrost frozen for thousands of years (*2*) now thaws near-annually (*3*). Concomitant with such thaw, dormant Stordalen Mire microorganisms are also annually awakened and in turn metabolize organic matter that accelerate greenhouse gas (GHG) emissions (*4*, *5*). Globally , such processes could unlock permafrost soil carbon reserves three times greater than the carbon currently in the atmosphere (*6*, *7*), though how specific soils and their microbiomes respond are among the greatest consequential unknowns in climate models (*8*). Thus, mechanistic understanding of the interplay between thawing permafrost, microbial communities (bacteria and archaea) and their viruses, and carbon dynamics represent a key grand challenge for climate science.

Previous studies in Stordalen Mire have characterized three distinct habitats representing a thaw gradient: palsas (partially frozen), bogs (intermediate thaw), and fully-thawed fens – using metagenomic sequencing, functional genomics, and ecosystem measurements (*5*, *8–10*). These studies identified dominant microbial phyla (e.g., Acidobacteria and Actinobacteria) and demonstrated that microbial communities varied across the permafrost thaw gradient. Additionally, key microbial metabolic pathways, including polysaccharide degradation and methane metabolism, revealed their roles in altering ecosystem greenhouse gas (GHG) outputs across the thawing gradients (*5*). Among the notable findings, the syntrophy between *Candidatus* Methanoflorens and *Candidatus* Acidiflorens has been shown to be ubiquitous for regulating carbon dynamics in such soils (*5*, *9*, *11*). Further, double-stranded (ds)DNA viruses – important in marine microbial biogeochemistry (via lysis, gene transfer, and metabolic reprogramming) (*12–14*) and frequently overlooked in microbiome studies and climate change modeling (*15*, *16*) – also appear important to GHG and ecosystem processes at Stordalen Mire. Specifically, ecogenomic analyses predict that dsDNA viruses are diverse and modulate microbial ecosystem outputs via infecting key microbial hosts and encoding auxiliary metabolic genes (AMGs) (*17–19*).

Despite these advances, the roles of other biological entities, such as mobile genetic elements and RNA viruses, remain largely unexplored in permafrost ecosystems. Recent reviews suggest soil RNA viruses influence food web structure, energy flow, and nutrient cycling, potentially shaping soil health and fertility (*20*). Given the unique environmental conditions of permafrost ecosystems, RNA viruses may play an even more pronounced role in shaping microbial communities and regulating carbon cycling and GHG. Here, we address this gap by investigating RNA viruses (the focus here) and mobile genetic elements (see Companion Guo et al., mobilome) in Stordalen Mire. By leveraging four years of bulk soil metatranscriptomes, we provide the first comprehensive analysis of RNA viruses across the permafrost thaw gradient, revealing their diversity, distribution, and functional potential in regulating ecosystem processes. Specifically, we screened the data to identify 25,337 viral contigs that were dereplicated into 2,651 virus operational taxonomic units (vOTUs), which we classified and ecologically contextualized. Similar to their dsDNA counterparts, Mire RNA viruses were largely novel and likely play diverse and significant roles in ecosystem processes. Our analyses revealed that RNA virus distribution is influenced by environmental factors such as pH and CH_4 concentrations. Moreover, detailed examination of host-virus interactions and functional potential suggested these viruses play a crucial role in regulating carbon cycling and greenhouse gas emissions at the ecosystem scale. Studying RNA viruses also offers a unique lens into the underexplored eukaryotic communities in these environments, further underscoring their importance in permafrost regions.

Stordalen Mire hosts thousands of novel RNA viruses

To identify RNA viruses in Stordalen Mire, here we focused on RNA viruses of the kingdom *Orthornavirae*, which replicate their genomes via RNA-directed RNA polymerase (RdRp). Hence, we screened the RdRp gene across ≈630 Gbp of bulk soil RNA metatranscriptomes sequencing data (*n*=55) collected from Stordalen Mire. These samples were obtained from three canonical permafrost thaw habitats (palsa, bog, and fen) and spanned across the soil active layers over the four years (2010, 2011, 2012 (*5*, *17*) and 2016 (*10*, *19*)). To enhance the detection of divergent RNA viruses, we iteratively searched and refined RdRp hidden Markov models (HMMs), a strategy previously successful (*21*, *22*) (**fig. S1**). Together these efforts identified 25,337 RdRp-containing contigs (RdRps length mean ± s.d.: 185.47 ± 112.30) and removed 3,225 false positives (see Methods). Of the resultant 22,112 RdRps that could be used for ecological analyses, 1,702 were complete or near-complete such that they could be used in deeper taxonomic analyses (**fig. S2**).

To determine the megataxa (phyla and classes) of RNA viruses present in this ecosystem, we utilized a previously benchmarked approach based on Markov clustering algorithm (MCL) analysis of RdRps (*22*). First, because global phylogenies of the >3-billion-year-old RdRp-encoding gene are challenging (*23*), we applied machine learning to pre-sort RdRp sequences into five orthornaviraen phyla that are recognized by the International Committee on Taxonomy of Viruses, or ICTV (*Duplornaviricota, Kitrinoviricota, Lenarviricota*, *Negarnaviricota,* and *Pisuviricota*) (*23*, *24*). Note that our study did not consider the recently established sixth RdRp-encoding phylum, *Ambiviricota* (*23*, *25*), as analyses were closed before ambiviricots were described. For the machine learning pre-sorting step, we specifically applied a network-based, iterative clustering approach to the 1,702 complete/near-complete RdRp sequences and another 704,003 complete RdRp references sequences derived as follows: 6,732 from the global *Tara* Oceans project (*22*), 3,325 from coastal seawater near the Yangtze River estuary (*26*), and 695,648 from GenBank (version 245). This merged network analysis of near-complete RdRp sequences revealed a 96% adjusted rand index (98.7% agreement, the number of correct assignments per total number of assignments to ICTV taxonomy) and 18 distinct "megaclusters" (**fig. S3; Table S5**). Of the ten megaclusters that contained Stordalen Mire sequences, seven matched ICTV- taxa within the five established phyla of *Duplornaviricota, Kitrinoviricota, Lenarviricota*, *Negarnaviricota,* and *Pisuviricota* (**Fig. 1A**). In addition to the established phyla, lineages of *incertae sedis* viruses, i.e., Permutotertaviridae ('permutotetra-like') and 'kitrino-like' viruses, were included as well. These 5 phyla include a total of 20 classes and Stordalen Mire sequences are assigned to 16 of them (**Table S5**), and at vOTUs/"species"-level, RNA viruses are all different (**fig. S2F**).

Beyond ICTV-recognized taxa, from our analysis, we also identified the recently reported megataxa '*Taraviricota*' (*25*) of Stordalen Mire RNA viruses. This new phyla is not currently ICTV-recognized due to a lack of complete genomes, but other studies support the taxonomic assertions (*27*, *28*) (**fig. S3**; see **Table S5**). Moreover, we identified four novel virus operational units (vOTUs) encoding unclassified RdRps that clustered within *Lenarviricota* as a novel class tentatively named novel class "*Stormiviricetes*" (**fig. 1A, Fig. S3**). Their genomes ranged from 2.5 to 5.8 kb in length, and the longest one encoded two partially-overlapping open reading frames (**fig S4**). Given they are lenarviricots, we presume these are positive-sense, single-stranded RNA viruses.

To assess the similarity of Stordalen Mire RNA viruses with known reference sequences, we performed phylogenetic distance analysis using UniFrac. Phylogenetic analyses revealed permafrost virus-enriched clusters within the phyla *Lenarviricota, Kitrinoviricota, Negarnaviricota, Duplornaviricota (chrymotiviricetes),* and *Pisuviricota*, often forming phylogenetically distinct blocks (UniFrac distance >0.9) with limited reference sequences outside Stordalen (**Fig. 1B; fig. S5-S12**). Aligned to the previous literature (*29*, *30*), which found

139 that permafrost hosts harbor phylogenetically distinct RNA viruses—including reoviruses
140 (Duplornaviricota), hypoviruses (Pisuviricota), rhabdoviruses (Negarnaviricota), and
141 leviviricetes—our findings reinforce the uniqueness of RNA viruses in thawing permafrost (*29*).
142 Comparing these patterns to other soil ecosystems, we find distinct differences in RNA virus
143 composition. For example, in agricultural soils, leviviricetes (*Lenaviricota*) are most
144 represented (*31*), whereas soils with added root litter show increased levels of barnavirids
145 (*Pisuviricota*), megabirnavirids, and quadrivirids (*Duplornaviricota*), along with mitovirids
146 (*Lenarviricota*) and leviviricetes (*32*).
147
148 **Ecological patterns of Stordalen Mire RNA viruses**
149 We next sought to assess the ecological distributions of RNA viruses. We used a recent
150 community consensus definition (*33*) for OTUs of RNA viruses, preclustered at 90% average
151 nucleotide identity across 80% of the shorter sequence length and 1 kb minimum contig length
152 (materials and methods) (*21*, *22*). Clustering using these parameters collapsed the 22,112
153 Stordalen Mire RdRp sequences to 2,651 vOTUs (**fig. S13-S15**). Compared to other studies, this
154 number is relatively large and suggests high soil RNA viral diversity considering it derives from
155 a single sampling location and yet is nearly half that (2,651 vs 5,504 vOTUs) found throughout
156 the global oceans (*22*). Consistent with such high diversity, individual samples contributed from
157 ~83 to 166 vOTUs on average, and accumulation curves revealed sampling was not yet saturated
158 (**fig. S13C)**; though most large-scale RNA virus surveys of marine, soil, and engineered
159 environments fail to saturate RNA viral diversity (*27*, *34*, *35*).
160 To assess this further within soils - a biome where microbiomes are considered
161 particularly diverse (*36*, *37*) - we collated and reanalyzed all available soil metatranscriptomes
162 using our iterative HMM RdRp search strategy. This revealed the following number of RNA
163 virus vOTUs: 83,027 vOTUs across 288 diverse Chinese soil samples (*38*), 3,161 vOTUs across
164 83 Warwick-England crop agricultural soil samples (*31*), 2,112 vOTUs across 48 California
165 rhizosphere soil samples (*39*), and 1,412 vOTUs across 16 Welsh upland peatland samples (*32*).
166 The comparison, based on the vOTU clustering at 90% nucleotide identity and 80% coverage,
167 across these different soil RNA studies revealed that only 62 vOTUs overlapped to Chinese soils
168 and 28 vOTUs to those Welsh upland peatland (**fig. S16C**). Further, 98% of the Stordalen Mire
169 vOTUs were unique across these soil studies, similar to the high proportion of study-specific
170 unique viruses detected in other soil datasets (**fig. S6C**).
171 We next explored vOTU-based community structure and patterns as follows. First, we
172 explored beta-diversity of RNA vOTU relative abundances via ordination and hierarchical
173 clustering of Bray-Curtis dissimilarity, which revealed three distinct meta-communities that
174 matched the major soil types (palsa, bog, and fen) (PERMANOVA, $p = 0.001$; **Fig. 2B, fig.**
175 **S13-S15**). This approach added RNA viruses to habitat-specific clustering of plants,
176 microbiomes, and dsDNA viruses at this site (*17*, *19*, *40*), similar to RNA virus community
177 differentiation across soil types observed in other studies (*31*, *32*, *38*, *39*, *41*). This structuring
178 in Stordalen Mire may be shaped by the distinct plant communities associated with different
179 thawing stages, as Sphagnum mosses, hare's-tail cottongrass (*Eriophorum vaginatum*), and
180 common cottongrass (*Eriophorum angustifolium*) dominate the palsas, bogs, and fens,
181 respectively (*42–47*).
182 Next, we assessed alpha-diversity (Shannon's *H*) of RNA viruses and compared it to
183 broader Stordalen Mire taxa including dsDNA viruses and microbial communities (eukaryotes
184 and prokaryotes). Mire RNA viral diversity increased along the thaw gradient, with the highest
185 diversity observed in fens followed by bogs and the lowest at palsas, just as seen for gene-
186 marker-inferred alpha-diversity for eukaryotes assayed here (*18S rRNA*; **Fig. 2F**) and
187 prokaryotes assayed elsewhere (16S rRNA; complementary paper). Interestingly, dsDNA virus
188 diversity peaks in bogs but does not differ significantly from that in fens (**Fig. 2D**). However,

alpha diversity measured for plant diversity exhibited distinct trends, no significant difference across palsa, bog and fen and peaks in palsas for plants, respectively (**Fig. 2F, 2H**). These differences likely reflect the varying resolution of the markers used, *COX1* and *RPS3* provide higher taxonomic resolution (*39*, *48*), whereas 18S rRNA is more conserved and better suited for broad taxonomic overviews but lacks the resolution to capture finer-scale variation (*49*). Though prior soil RNA virus studies have documented RNA virus diversity (*31*, *32*, *38*, *39*, *41*), few assessed broader communities taxa within their sites (*31*, *39*). Interestingly, while dsDNA virus diversity peaked in bogs, its Shannon's *H* in fens was not statistically significant. Given that fens represent fully thawed, inundated permafrost systems, these results suggest that permafrost thaw is associated with increased RNA viruses, eukaryotes, and microbial diversity. However, the broader ecological consequences of these changes remain uncertain.

Ecological drivers of orthornaviraens in thawing permafrost

We next applied ecological driver analyses to understand how biotic and abiotic factors impact Stordalen Mire RNA viral community structure across habitats and environmental factors. Myriad analyses revealed that Stordalen Mire RNA virus community structure was most impacted by pH, followed distantly by $\delta^{13}C$ (CH_4), $\delta^{13}C$ (CO_2), and total relative abundance of organisms (eukaryote) (goodness of fit and Mantel tests in **Fig 3B-C**, relative importance analysis in **fig. S17,** and redundancy analysis in **fig. S18**). Previously, prokaryotic dsDNA viruses have shown pH-influenced community structure at local and global scales (*50*). In thawing permafrost systems, pH can impact ecological parameters such as microbial activity, nutrient availability, and plant species composition by modulating the decomposition of newly thawed organic matter. Thawing also leads to organic acid accumulation in low-pH bogs, while the higher pH in fens promotes bicarbonate formation, making them more alkaline (*51*, *52*), which in turn could have an indirect effect on RNA virus communities via host composition selection (as seen prior for prokaryotes in this ecosystem, Emerson et al (*17*)). Alternatively, pH might directly alter virus community structure through conformational changes critical to virus integrity and infectivity, as known for viral pathogens (*53*). As pH emerges as a large-scale prominent driver of RNA virus community structure, supporting prior findings for prokaryotic dsDNA viruses (*50*) linking RNA virus diversity to pH, nutrients, and organic content in soil (*38*).

Having better understood drivers of community structure, we next sought to assess which vOTUs best predicted ecosystem features of interest across Stordalen Mire. To this end, we used weighted gene/genome correlation network analysis (WGCNA) to group vOTUs into modules and assess modular correlations to environmental factors as developed for ocean biota (*54*) including application for identifying key ocean DNA and RNA viruses that best predict carbon flux (*21*, *55*). Of 13 environmental traits assessed, four factors, including carbon (%), nitrogen (%), isotopic signature of carbon dioxide, and carbon in peat, are not the best predictors of RNA virus diversity. Instead, air and soil temperature, pH, CH_4, CO_2, and the relative abundance of microbes across the thawing permafrost were prominent predictors (**Fig. 3C**). Moreover, our analysis also showed 12 of 20 vOTU modules correlated to these environmental factors ($p \leq 0.05$). Among these, the WGCNA gray module strongly correlated with CH_4 concentrations (corr. = 0.87, p = 7.3e-57; **Fig. 3D**). While dsDNA viruses have been speculated to indirectly impact methane cycling via methane-related AMGs (*17*, *19*, *56*) or microbial lysis dynamics (*57*, *58*), our result provides a connection between RNA viruses and CH_4 concentrations in thawing permafrost systems. However, the mechanism of RNA viruses influencing CH_4 in thawing permafrost remains unclear, possibly through host interactions. Toward the latter consideration, our analysis also found the WGCNA yellow module showed the strongest correlation with organismal abundance (corr. = 0.94, p = 4e-69; **Fig. 3E**). This extends long-standing theories such as "kill the winner," (*59*) which suggests that host-driven

239 shifts in microbial ecosystems could also indirectly influence RNA viruses and their ecological
240 roles, including potential links to methane-related processes. A subsequent correlation analysis
241 of RNA virus richness and diversity (Shannon's *H*) with host abundance supported this
242 hypothesis, showing a significant positive relationship (**Fig. 3F**). Together, these findings raise
243 new questions about the ecological roles of RNA viruses in rapidly changing environments and
244 underscore the need for future mechanistic studies to explore their contributions to methane-
245 related biogeochemical processes.
246
247 **RNA viruses infect diverse eukaryotic hosts that are key components of food webs and**
248 **biogeochemistry in Stordalen Mire**
249 Past studies have begun to uncover RNA viruses that affect multiple trophic levels
250 (bacteria, protists, fungi, and invertebrates), as well as inferred carbon cycling impact through
251 host lysis (*32*, *39*). However, soil RNA virus ecological roles are largely underexplored.
252 To provide context for the organisms inhabiting the ecosystem, we screened for hallmark
253 genes for eukaryotes (*COX1*, *RPS3*, 18S rRNA) and prokaryotes (*RPS3*, 16S rRNA) (**Fig. 4**, **fig.**
254 **S19**) and estimated their abundances across the metagenomes (**fig. S19**). Bacteria and archaea
255 are well known at the site due to past genome-resolved studies (*5*), but known RNA viruses
256 more commonly infect eukaryotes (*21*, *60*), so we expected the same here. Since the eukaryotic
257 communities at Stordalen Mire are not yet characterized, we describe their patterns here. Despite
258 discrepancies found in the organisms inferred from the different gene screenings —likely due
259 to bioinformatic or biological biases) (**fig. S19** and **fig. S20**—, they suggested a diverse
260 taxonomic repertoire of protists, fungi, plants, and invertebrates with habitat-dependent
261 abundance profiles as follows. Palsas were enriched in plants (monocotyledons of Poales, likely
262 hare's-tail cottongrass, *E. vaginatum* (*42*)), providing primary production that supports
263 heterotrophs with diverse ecological roles: saprophytic and mycorrhizal (Agaricales and
264 Helotiales) and pathogenic fungi (Mycosphaerellales), litter-decay insects (Entomobryomorph
265 and Poduromorph springtails), rodents, and parasitic apicomplex protists (Eugregarinorida).
266 Wetter bogs were dominated by fungi (Heliotales), tardigrades (Parachela), bacteriovorous
267 rotifers (Adinetida, Ploima) and ciliates (Haptorida). Inundated fens were dominated by
268 oligochaetean worms (Haplotaxida and Lumbriculidae), likely feeding on the rich bacteria and
269 archaea biomass fueled by peatland carbon, and monocotyledons of Poales (likely, *Carex* sp.
270 and *Eriophorum* sp.). In general, plants, fungi and herbivorous invertebrates were well
271 distributed along the thawing gradient. In bogs and fens, prokaryotic life is grazed by
272 bacterivorous protists and invertebrates (*61*), with protists possibly being outcompeted by
273 invertebrates (**fig. S20**) as previously reported in inundated environments (*62*). With this
274 mapped backdrop of potential host taxa across the Mire, we next assessed whether our newly
275 identified RNA viruses might infect any of these organism taxa as predicted by the following *in*
276 *silico* approaches. Unlike DNA viruses, RNA virus host prediction is not as well developed
277 because orthornaviraens lack programmed endogenization into the host genome. Thus, as done
278 previously (Zayed et al (*22*)), we combined host predictions from multiple approaches based on
279 different types of features derived from RNA virus contigs as follows: (*i*) prior host information
280 of known virus taxa classified using the RdRp sequences (**fig.S21,23**), (*ii*) similarity to closely-
281 related endogenous virus elements that reside in known cellular genomes (**fig. S21**), (*iii*)
282 prediction of alternative genetic codes that suggest possible host organisms and subcellular
283 compartments (**fig. S21**), and (*iv*) prediction of Shine-Dalgarno motifs and CRISPR spacers that
284 suggest bacterial hosts (**fig. S21, 22**). Where more than one line of evidence existed, the final
285 host prediction was assigned using a hierarchical scoring approach that differentially weighted
286 the predictions according to confidence (see materials and methods).
287 Overall, the approach worked well to provide at least some taxon rank host prediction
288 for the vast majority (1,871 or 70.9%) of the 2,561 vOTUs (**Fig. 4C**, individual host inference

approach results are in **fig. S21, S22** and **Table S6**). These efforts revealed that Stordalen Mire RNA viruses most commonly infected fungi (e.g., Microascales, Pezizales), invertebrates (e.g., Hymenoptera, Rhabditida), and protists (e.g., Ciliophora, Trypanosomatida). Similarly to what we report here, RNA viruses that putatively infect fungi are consistently found as abundant in diverse soils (*32*, *38*, *39*, *63*) and those that infect other eukaryotic hosts (protists and invertebrates) are more rare (*32*, *38*), although this fact may derived from methodological differences. Notably, by integrating RdRp assignment to *Leviviricetes* and prediction of Shine–Dalgarno motifs and CRISPR spacers, only 69 vOTUs (2.7%) putatively infect bacteria, and these were not particularly abundant as they proportionally comprised less than 2% of the vOTU abundances (**fig. S19, S20**). By contrast, prior soil surveys have observed more than an order of magnitude higher relative abundances of bacterial RNA viruses, with more than a third of the leviviricete vOTUs (*32*). However, the predominance of RNA viruses likely infecting eukaryotes is consistent with findings in the global oceans (*21*, *22*, *60*).

Beyond these qualitative inferences, we next sought to more quantitatively assess virus-host interaction biology across the Mire by assessing vOTU abundances in light of their predicted hosts (**fig. S3**). In palsas, RNA viruses commonly infected hymenopteran insects (19.4%), unknown fungi (7.5%), mortierellalian fungi (6.6%), and microascalian fungi (4.0%) (**fig. S19**, **S20**), all of which ultimately derived carbon from plants that support herbivorous, parasitic, symbiotic, and saprotrophic life around them in this habitat (*5*, *31*). In bogs, RNA viruses infected unknown fungi (24.7%), plants, lepidopterans (9.1%), hymenopterans (5.8%), and protists (ciliates and trypanosomatids, 5.6%) (**fig. S19, fig. S20**). By contrast, in the fully inundated fens, RNA viruses mostly infected microascalian fungi (33.1%), unknown plants and lepidopterans (19.5%), hymenopterans (5.7%), priapulids (5.3%), and rhabditids (4.9%) (**fig. S19, S20**). In wetter habitats like bogs and fens, peatlands provide a rich source of biogenic carbon, supporting abundant bacteria and archaea (which increase CO_2 and CH_4 emissions), which in turns fuels active grazing. This shift in the carbon sources could explain the higher frequency of protists in bogs and non-arthropod invertebrates in fens compared to bogs.

Extrapolating from what is known about these predicted hosts, we propose the following ways RNA virus infection might modulate key ecological roles that impact soil carbon cycling. First, impacts on primary production of the Mire ecosystem through infection of plants and photosynthetic protists. Second, infection of fungi could impact their multiple roles in rhizosphere building, saprotrophic degradation of litter and peatland and parasitism. Third, among other functions, bacterial and archaeal predation could be influenced by the infection of protists and non-arthropod invertebrates (e.g., priapulids, nematodes). Critically, because RNA viruses that infect eukaryotes are known to utilize a range of infection strategies from cryptic or persistent to acute infections, associated with different levels of impact on the hosts, this viral infection continuum (*64*, *65*) leaves open many questions about the ecological footprint of RNA viruses. Even after assuming acute infection for all RNA viruses, hence suppressing the ecological roles of their hosts, the effects on the carbon cycling often remain complex to assess. Plant RNA viruses could likely reduce photosynthetic primary production and increase dead plant material, thus decreasing CO_2 absorption and increasing respiration by saprotrophs. However, given the diverse ecological roles of fungi, the impact of their RNA viruses across the three habitats could range from decreased CO_2 levels by suppressing phytopathogenic and saprotrophic fungi in litter and peatlands to reduced CO_2 assimilation due to hampered mycorrhizal fungi. Furthermore, the higher prevalence of RNA viruses infecting potential grazers in bogs and fens could reduce pressure on bacteria and archaea, thereby enhancing the mobilization of carbon from peatlands into CO_2 and CH_4 in these wetter habitats (*66*).

Auxiliary metabolic genes of Stordalen Mire RNA viruses likely suppress ecological activities performed by their hosts via molecular shutoff of gene expression

339 Another axis for understanding the ecological impact of RNA viruses would be to assess how
340 they might metabolically reprogram their hosts during infection. While impossible to assess in
341 complex natural communities at any relevant scale, one window into the metabolisms viruses
342 could impact during infection include auxiliary metabolic genes, or AMGs (*67*). While well-
343 established for dsDNA viruses (*17–19*), previously identified AMGs in RNA viruses included
344 diverse protein domains and motifs linked to both metabolite-transforming and information-
345 processing pathways, requiring functional annotation with multiple databases for accurate
346 detection. To this end, we annotated amino acid sequences encoded by RNA virus contigs using
347 multiple databases (NCBI CDD, NCBI NR, KEGG, UniRef90, PFAM, dbCAN, MEROPS; see
348 Methods) and revealed 52 AMGs across 51 RNA virus contigs longer than 500 nts, and 33
349 AMGs across ~1.2% of the vOTUs (*n*=32) (**S8 Table**). As previously reported, AMGs encoded
350 cellular-like protein domains and motifs (instead of complete ORFs), likely shaped by purifying
351 selection on RNA virus genomes (facing strong streamlining constraints) after acquiring AMGs
352 through illegitimate recombination with host mRNA (*21*, *34*). As a control, we searched the
353 vOTU sequences against co-sampled cellular-fraction metagenomics, short-reads and long-
354 reads, found no hits, which we interpret as consistent with the genes being RNA virus derived
355 rather than from host genomes or endogenous virus elements (EVEs).
356 What kinds of AMGs were observed? Functional annotation revealed AMG-encoded
357 proteins were associated with transcription, translation, posttranslational modifications,
358 membrane proteins, and miscellaneous metabolic pathways (**Fig. 5**), with many functions
359 similar to those for RNA viruses from oceanic plankton and land invertebrates (*22*, *24*, *68*).
360 Closer examination of the AMG sequences revealed that AMGs encoded protein domains (e.g.,
361 NADAR) and motifs (e.g., motifs for SUMOylation), which could indicate the selection of
362 specific catalytic activities with cellular proteins (**S8 Table**). To explore deeper into the effects
363 of these AMGs on the host cells, we identified a conservative subset of 16 AMG-encoded
364 protein sequences (indicated with double asterisks in **Fig. 5**; **S8 Table**) that were not broken by
365 contig ends and had alignments with cdd profiles of cellular domains.
366 Focusing just on these conservative 16 AMGs, activities predicted from the specific
367 motifs found suggested translational and posttranslational interferencemechanisms as follows.
368 *First*, five RNA viruses that encode proteins with two RNA-binding zinc knuckles (motifs of
369 the domains AIR1 and UMSBP), that would possibly block host mRNA translation to redirect
370 it toward virus RNA translation. Given that the only putative host of these viruses is a
371 trypanosomid, suppression of gene expression by blocking translation in infected cells would
372 hamper their parasitic activity on invertebrate hosts, which could increase grazing on bacteria
373 and archaea (and hence, decrease microbial respiration-derived CO_2 and methanogenesis) in
374 fens. *Second*, five viruses encoded specific motifs within the domains DnaJ (HSP70-interacting
375 motif (*69*, *70*)) and Siz/PIAS RING (CHC2- and C3H2C3-type motifs (*71*)), which putatively
376 increase protein turnover via ubiquitination and SUMOylation, likely providing amino acids for
377 building virus proteins. Although hosts (hymenopterans) could be assigned only to one virus,
378 unbalanced increase of protein degradation over synthesis would lead the host cell to a catabolic
379 state, hence diminishing the associated ecological roles. *Finally*, seven viruses encoded NUDIX
380 hydrolases, and two viruses encoded NADAR. Both domains have been proposed to counter the
381 antiviral defense based on NAD^+-dependent ADP ribosylation of virus macromolecules (*72*):
382 NUDIX cleaves the required NAD^+ for this system and NADAR deconjugates ADP-ribosyl
383 moieties from virus proteins and RNAs. Only one virus-encoding NUDIX was assigned to a
384 specific host taxon (dinoflagellates of the order Suessiales). Depletion of NAD^+, a cofactor
385 required in many essential cellular processes, would disrupt the functions of dinoflagellates
386 (photosynthesis and grazing), with concomitant impact on biogeochemical cycling by
387 decreasing CO_2 absorption or increasing CH_4 methanogenesis. Together, these interference

mechanisms boost virus fitness while decreasing host fitness with subsequent suppression of their ecological roles in the Stordalen mire with diverse impacts on carbon cycling.

In summary, Stordalen More RNA viruses likely impact ecosystem processes through infection of biogeochemically important microbial eukaryotic hosts and AMGs that specifically target processes underpinning the cellular physiology that dictates biogeochemical cycling. However, quantifying these impacts awaits perturbation experiments and/or modeling designed leveraging these initial empirical observations.

CONCLUSIONS

Our study provides a baseline for understanding ecological patterns and the footprint of soil RNA viruses in thawing permafrost, emphasizing their potential implications amid Arctic environmental and climate change. By exploring diversity patterns, we posit a near-future scenario in which permafrost thawing and proliferation of palsas will conceivably activate and expand the existing RNA virus communities, leading to increased diversity and complexity compared to current levels. We evaluated the potential roles of RNA viruses to modulate the mineralization rates of peatland carbon into CO_2 and methane through infection. We suggest that pathogenesis and killing along with AMG-driven suppression of gene expression by RNA viruses have the potential to impact greenhouse gas emissions. Fungal hosts assume extremely diverse ecological roles, from being saprotrophs feeding on plant litter or peatland to biotic interactions (e.g., being parasites of plants or invertebrates and mutualists in lichens). Infection of saprotrophic, parasitic, and mutualistic fungi and heterotrophic protists has the potential to increase CO_2 emissions from peatland. It could also increase methanogenesis by infecting (*i*) grazers (e.g., priapulids, nematodes, arthropods, ciliates), (*ii*) invertebrates stimulating microbiota, and (*iii*) anaerobic protists and ciliates (in particular, those associated with methanogenic archaeal endosymbionts). The manner, intensity, and direction of such effects in diverse hosts need to be addressed to allow the enable footprint of RNA viruses to be integrated into predictive ecological models. This knowledge will empower us to improve the accuracy of ecosystem models and improve the prediction of the ecological implications of climate change.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under Award Number DE-SC0023307 to M.B.S. Funding was provided by DOE grants to MBS. This research is a contribution of the EMERGE Biology Integration Institute, funded by the National Science Foundation, Biology Integration Institutes Program, Award #2022070. The authors thank the Swedish Polar Research Secretariat and SITES (Swedish Research Council grant 4.3-2021-00164) for the support of the work done at the Abisko Scientific Research Station. This work was supported in part through a Laulima Government Solutions, LLC, prime contract with the U.S. National Institute of Allergy and Infectious Diseases (NIAID) under Contract No. HHSN272201800013C. K.C.W. funding was provided by EMERGE (#2022070), NSF Cyverse [#2149505 ,Collaborative Research: Updating iVirus -- the CyVerse-powered analytical toolkit for viruses of microbes], and was partially funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, grants #DE-SC0023456.

A.A.P is currently funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2051 – Project-ID 390713860. G.D.H is currently funded by the Spanish Ministry for Ecological Transition and the Demographic Challenge.

Data availability

438 The sequencing data described in this article, including metagenomic and metatranscriptomic
439 data also described in Woodcroft et al., Emerson, et al, Sun et al. and McGivern et al. (*5*, *10*, *17*,
440 *19*), are submitted under NCBI BioProject accession number PRJNA386568. The
441 metagenomics assembled genomes (MAGs) used in this study were acquired from a previous
442 study (Woodcroft et al.,(*5*)) and deposited at DDBJ/ENA/GenBank under the accession numbers
443 provided in **Table S6**.

444 **SUPPLEMENTARY**
445
446 **Methods**
447 **Sampling, purification of nucleic acids, library preparation, and short-read sequencing**
448 A total of 55 metatranscriptomics were used in this study (**S1 Table**). Twenty-eight
449 metatranscriptomics were collected during 2010, 2011, and 2012 (*5*, *11*); and 27
450 metatranscriptomics samples were collected in 2016 (*10*) (≈0.63 terabases [Tb] of data). A
451 detailed description of active permafrost sampling strategies and protocols was previously
452 described (*5*, *11*). Briefly, triplicate samples were taken from the active layer across the thaw
453 permafrost gradient comprising palsa (intact permafrost), bog (partially thawed), and fen (fully
454 thawed). The soil cores were taken at different depths of 2.5–48.5 cm (**S2 Table**). Soil cores
455 were kept at saturation with three volumes of LifeGuard solution (MoBio Laboratories) and
456 stored at -80°C until processing.
457 The protocols for extraction and purification of RNA for 2010–2012 samples were described
458 previously (*5*). The total RNA of samples was subjected to removal of residual DNA using
459 DNAse I (Roche) following the manufacturer's instructions. Next, ScriptSeq Complete
460 (Bacterial)-low input kits (Epicentre) were used for library preparation. To check the quality of
461 RNA and libraries during processing, Agilent 2100 Bioanalyzer and Agilent 2200 Tapestation
462 (Agilent Technologies) with Qubit (ThermoFisher Scientific) were used. Before the deeper
463 NextSeq (Illumina) sequencing, the resulting metatranscriptomic libraries were assessed on
464 either a HiSeq (Illumina) or MiSeq (Illumina) to evaluate library quality. For 2016 samples, the
465 TruSeq Stranded mRNA Library Prep (Illumina cat. No. 20020594, 20020595) was used for
466 library preparation. Using this, polyA-containing mRNA molecules were purified in two rounds
467 with oligo-dT magnetic beads. During the second elution process of the polyA RNA, the RNA
468 was fragmented and primed for cDNA synthesis.
469
470 **Metatranscriptome assembly, virus identification, and evaluation of completeness of**
471 **putative RdRps**
472 The description of the metatranscriptome read assembly, virus identification, and taxonomic
473 classification methods have been detailed previously (*21*, *22*). Briefly, the raw-read RNA
474 sequences downloaded from the National Institutes of Health (NIH) National Center for
475 Biotechnology Information (NCBI) Sequence Read Archive (SRA) database
476 (https://www.ncbi.nlm.nih.gov/sra) were first processed to eliminate adapter and low-quality
477 sequences by using Trimmomatic v0.36 (*73*) with default parameters. PhiX contamination was
478 also removed by BBDuk (bbduk.sh; ktrim=*r*, minlen=40, minlenfraction=0.6, mink=11, tbo tpe
479 k=23, hdist=1, hdist2=1, ftm=5). Next, *de novo* assembly was performed using MEGAHIT
480 v1.1.3 (*74*) with default settings. To avoid problems associated with alternative genetic code
481 and non-canonical translation events, EMBOSS Transeq was used (instead of protein prediction
482 with prodigal) to yield RdRp footprints.
483 Approaches to identify the RNA virus genomes or transcripts from assembled
484 metatranscriptomes and the authentication of virus RdRp were described previously (*22*). To
485 improve the identification of the divergent RdRp, the Hidden Markov Model (HMM) was
486 applied over 10 iterations to search for virus RdRp domains (**fig. S1**). To evaluate the
487 authenticity, thus true positive hits of virus RdRp sequences, competitive HMM searches were
488 conducted of 70 RdRps derived from the lineages defined previously (*24*), eight recently
489 discovered lineages (*22*), and 16,268 PfamA (v33.0) non-RdRp HMMs. Further, the putative
490 RdRp hits with the best match to an RdRp HMM and a bit-score ≥30 was kept as true positives
491 for proteins containing the virus RdRp domain. In addition, to confirm that the RNA virus
492 contigs were not from endogenous viruses, the identified sequences were searched against the
493 available metagenomes.

494 To estimate RdRp domain completeness, based on previous efforts (*22*), the extraction of
495 footprint amino acid sequences of the RdRp domain was performed considering the challenging
496 cases of the usage of alternative genetic codes and non-canonical translation events (e.g.,
497 ribosomal frameshifting). Extraction and reconstruction of RdRp domain sequences were
498 performed by running another iteration of the search for the translated amino acid against the
499 70 RdRp profile HMMs, and the hits with a bit score ≥30 and with aligning regions of ≥90% of
500 the average length of the best-matching HMMs (to be considered as "complete" or "near-
501 complete" virus RdRp domain sequences). The 3,560 extracted RdRps that were shorter than
502 100 amino acids were further inspected for RdRp annotations, having detected only 18 of them
503 for which no virus signal was apparent, hence further supporting that 99.5% of hits are RdRp-
504 encoding sequences derived from RNA viruses. Lastly, to estimate genome completeness for
505 the recovered viral contigs, CheckV (*75*) was used with its default parameters (**fig. S2**), which
506 compared amino acid identity and HMMs of complete viral genomes derived from NIH NCBI
507 GenBank and environmental samples.
508
509 **Virus operational taxonomic units (vOTU) and calculations of vOTU relative abundances**
510 **and diversity metrics**
511 To analyze the diversity and ecological pattern of soil RNA viruses, they were first designated
512 as vOTUs or an approximate "species-rank" ecological unit; the detailed analysis of the RNA
513 virus sequence space to establish a "universal" cutoff has been reported previously (*21*, *22*).
514 Using the previously recommended cutoff, 90% average nucleotide identity (ANI) across 80%
515 of the shorter sequence length (*21*, *22*), resulted in 9,567 total vOTUs, of which 2,561 were ≥1
516 kb (**S3 Table**). Further, to assess the robustness of this cutoff in the soil system, an investigation
517 of the ecological patterns was performed across an ANI range of 60–100% (and 80% allele
518 frequency [AF]), including consensus cutoff recommended by (*33*), 95% ANI and 85% AF (**Fig.**
519 **S3**).
520 To generate the relative abundance table of the designated vOTUs, the previously described
521 approach was followed (*21*, *22*). Before the mapping, virus contigs were trimmed off their
522 polyA/polyT sequences; this was done to avoid inflated abundance and better estimate
523 horizontal coverage for polyA-tailed viruses, respectively. Further, a quality-controlled
524 sequence read of metatranscriptomic samples to the vOTUs was recruited, and trimmed mean
525 abundances were calculated with CoverM (https://github.com/wwood/CoverM). The trimmed
526 mean abundance table was then normalized by dividing abundances by the number of quality-
527 controlled reads per sample per 1 Tb to get the final relative abundance table (**Table S4**).
528 The ecological analysis was done using samples collected in 2012 and 2016 to minimize
529 bias introduced by samples variably from 2010 and 2011. Different diversity metrics were
530 calculated, including α-diversity (Shannon's H, Simpson, Inverse–Simpson, Richness, and
531 Perilous J Evenness) and β-diversity (Bray–Curtis dissimilarity method), using R vegan package
532 (*76*). vOTUs that were not present in any samples and samples with no vOTUs present were
533 excluded. The abundance table of vOTUs was log-transformed (function 'decostand', method
534 = 'log') before the calculation of the Bray–Curtis dissimilarity analysis (function 'vegdist',
535 method 'bray'). This was then used to generate principal coordinate analysis (pCoA, function
536 'capscale' of the vegan package). The emerging metacommunities in pCoA were statistically
537 verified using permanova tests (function 'adonis') and the confidence intervals were plotted
538 using the standard deviation method (function 'ordiellipse') at 95 and 97.5% confidence limits.
539 The same Bray–Curtis dissimilarity matrices were used for hierarchical clustering (function
540 'pvclust', method.dist ='co'' and method.hclust = 'average') using 1,000 bootstrap iterations.
541 Only the approximately unbiased (AU) bootstrap values were reported on the dendrograms. The
542 heatmaps of the hierarchical clustering analysis were generated using the heatmap3 package
543 (*77*) (**fig. S3–S5**).

544
545 **Comparing Stordalen Mire RNA virus to other soil RNA virus studies**
546 Metatranscriptome sequences were downloaded from recently published soil RNA virus studies,
547 which included examination of California grassland (*39*), grassland and coastal soils (*32*), crop
548 field soil (*31*), Kansas grassland (*41*), and various soil types across China (*38*). Similar pipelines,
549 including metatranscriptome assembly, virus identification, and evaluation of completeness of
550 putative RdRps, were run as described above. To compare the virus sequences across different
551 soil studies, the identified RNA virus contigs were clustered at 90% ANI and 80% AF to get
552 vOTUs. For RdRp-level comparison, only those "complete"/"near-complete" (≥90% of the
553 average length of the best-matching HMMs) were used. The RdRps were clustered using
554 USEARCH (-cluster_fast) (*78*) at 75% Identity of Amino Acid (IAA), representing taxonomic
555 rank between family and genera (*24*). Subsequently, those clustered at 75%, were considered
556 shared among studies. These results were plotted in UpSet using the UpSetR package in R (*79*)
557 (**fig. S6**). To compare the identified orthornavirans from this study to other soil studies,
558 taxonomy assignment was done using RdRp-scan (*80*).
559
560 **RdRp-based taxonomic annotation of RNA viruses**
561 The taxonomic classification analysis was conducted using both the network-based, iterative
562 clustering approach and the phylogenetic described previously (full details in Zayed et al., 2022)
563 (*22*) (**fig. S7**). Briefly, a total of 705,705 complete and near-complete RdRp domains—including
564 1,702 from the current Stordalen Mire soil study; 6,732 from a *Tara* Ocean study (*21*, *22*); 3,325
565 from coastal ocean RNA viromes (*24*); and 695,648 from GenBank [version 245]—were used
566 in the analysis. First, these sequences were pre-clustered at 50% amino acid identity using
567 Uclust v10.0.240 (usearch -cluster_fast -id 0.50 -sort length) (*78*). Then, these centroid
568 sequences (n = 6,553) were used in pairwise comparisons using blastp v2.10.0+ (-gapopen 9 -
569 gapextend 1 -word_size 3 -threshold 10) (*81*). Further, the e-values for each pair were extracted
570 and negative-long10-transformed in MCL v14-137 (-stream-mirror -stream-neg-log10 -stream-
571 tf 'ceil(200)') (*82*). The transformed e-values were used in an MCL network for iterative
572 clustering, changing the granularity parameter at each iteration (range 0.1–8). From this point,
573 'adj.rand.index' value was used to evaluate all the cluster sets from MCL and concurrently
574 compared to the previously established ICTV taxonomy (**Table S5**). The clusters with the
575 highest rand index value at the phylum and class ranks were picked and the taxonomic
576 delineations were then assigned for sequences within each cluster.
577 Moreover, phylogenetic trees were generated for each assigned taxonomic network-derived
578 cluster (at the phylum and class ranks) as established previously (*22*). Briefly, from each
579 resultant network-derived major cluster, sequences from each cluster were aligned separately
580 using the E-INS-I strategy over 1,000 iterations in MAFFT v7.017 (*83*). Aligned sequences were
581 subsequently trimmed using Trimal (*84*). Sites with ≥20% gaps were removed, and the
582 alignments were manually inspected. Before phylogenetic analysis, sequences were screened
583 for possible recombination events using 3Seq (*85*), with a recombinant event determined by a
584 Bonferroni-corrected p-value cutoff of 0.05. Phylogenetic relationships of sequences within a
585 cluster were first assigned the appropriate evolutionary model using ModelFinder (*86*). Then, a
586 subsequent maximum-likelihood phylogenetic tree was generated using bootstrap support
587 generated for 1,000 iterations in IQ-TREE (*87*). Given the high number of reference sequences
588 within the phylum-level phylogenetic trees, a subset of these trees was created to aid in
589 visualization. Redundant reference sequences were reduced while maintaining the diversity
590 captured in the original tree using Treemer, applying stop option -X (50–200) depending on the
591 phylum (*88*). For **Fig. 1**, sequences used to build the trees were preclustered at 50% identity,
592 and clades supported by 100% bootstrap values were collapsed.

To taxonomically assign sequence to a family rank, a two-pronged approach was used. First, the class-rank phylogenetic trees were used to establish the placement of sequences on the tree, requiring a sequence to fall within a clade or sister clade to be assigned to the same taxon. Then, sequences were assigned tentative names identical to the reference sequences of the established clades with a '-like' suffix added. Further, the prior analysis was complemented by the taxonomic assignment from the conservative MCL clusters. Subsequently, these two approaches were combined and the more specific taxonomic assignment of the two methods (e.g., without the '-like' suffix or higher-resolution taxonomic assignment) was picked as the final putative classification for the novel sequences.

Host inference of hosts and cellular organisms' analysis

To infer the virus-host associations, a few independent approaches—including RdRp-based taxonomy, endogenous virus elements (EVEs), RdRp-based against the cellular non-redundant (nr) database, alternative genetic code, Shine–Dalgarno motifs, and CRISPR-Cas spacers analysis—were used. The following is the detailed approach: (*i*) The RdRp-based taxonomy approach was used to infer the potential virus–host associations. For this approach, the known RNA virus taxa (phyla, classes, and families) that were assigned to the vOTUs after clustering the RdRp domain protein sequence similarity network were used to retrieve the information on putative hosts. (*ii*) EVEs looked for closely-related RdRp-encoded cellular genomes, and the RdRp hits (from virus contigs) suggested that a virus contig can have a putative host taxonomically related to one of the cellular genomes (*68*). All available assembly genomes from cellular organisms in GenBank version 205 (last accessed May 2021) were downloaded; this included archaea (4.89 Gbp), bacteria (234 Gbp), fungi (157 Gbp), invertebrates (618 Gbp), plants (683 Gbp), protozoa (41.6 Gbp), and vertebrates (2.44 Tbp). A more sensitive diamond (version 0.9.36) blastx algorithm against the aforementioned databases was used (--evalue 1e-20–-sensitive–-out–-outfmt 6). The threshold for the search was set to 100 amino acids for length and 1 × 10−20 for e-value. The results were also checked manually to accurately assign the putative host taxonomy. (*iii*) Host information was inferred from cellular nr databases. The query of virus RdRp footprints was then used against this database. The threshold for the search was set to 100 amino acids for length and 1×10^{-5} for e-value. Only the top match hits were used, and the results were manually checked. (*iv*) Using genetic code, the translated sequences were searched once more against the 84 RdRp profile HMMs, and those with a bit score ≥30 and aligning regions of ≥90% of the average length of the best-matching HMMs were considered proteins containing "complete" virus RdRp domain sequences. Sequences containing more than one hit in the same or different frames (presumably non-canonical translation cases) were resolved manually by joining the aligning regions into the same protein sequence. (*v*) Shine–Dalgarno motifs, following the same approach described by Menardo et al., 2018 (*88*). (*vi*) To predict potential prokaryotic hosts using the CRISPR-Cas spacer approach, blastn was used against ≈12 million spacer sequences (*89*), and only blastn hits with 0 or 1 mismatch over the whole spacer length were considered. Furthermore, spacers were extracted from 1,529 previously described metagenomics-assembled genomes (MAGs) (*5*) using MinCed (https://github.com/ctSkennerton/minced). A minimum of two repeats was required for a CRISPR to be included in the analysis. Further, as in the preceding analysis, only blastn hits with 0 or 1 mismatch over the whole spacer length were considered. The best hit sequence alignment was then visualized using mView (https://www.ebi.ac.uk/Tools/msa/mview/). Moreover, the taxonomic assignment of the identified orthornavirans, which was performed using RdRp-scan (*80*), was also utilized to infer the hosts. Additionally, to identify the RNA virus infecting bacteria, bacteriolytics analysis was conducted using HMM searches, utilizing the HMM profiles previously provided (*34*) and only considering hits with a bit score ≥30 for further analysis (**Fig. S9–S12**). Leviviricetes require

643 bacterial pili to attach to and infect cells (*90*). Therefore, we performed competitive searches
644 using HMMsearch with nine type-IV pilus (T4P) profiles (PF13681.9, PF11104.11, PF16074.8,
645 PF17456.5, PF11773.11, PF18222.4, PF04350.16, PF04351.16, and PF04964.17) and 13 non-
646 T4P profiles containing the keyword "pilus" (PF07434.14, PF09589.13, PF14855.9,
647 PF09476.13, PF16976.8, PF09460.13, PF13629.9, PF13400.9, PF16968.8, PF09977.12,
648 PF05946.15, PF06340.14, and PF10671.12). The best match to a T4P HMM profile and a bit
649 score ≥50 were considered true positives for proteins containing the bacterial T4P domain. To
650 ensure a true positive match, random hits were manually checked across the datasets with
651 BLASTP using RefSeq and nr databases. The integrated host interference analysis can be seen
652 in **S7 Table**.
653 Next, to get a more comprehensive picture of the eukaryotic community and due to the lack
654 of this information in Stordalen Mire, a further investigation focused on cellular organism
655 community structure using HMM marker genes—cytochrome c oxidase subunit 1 mitochondrial
656 gene (*COX1*); a component of the eukaryotic 40S ribosomal subunit, ribosomal protein S3
657 (*RPS3*) (*39*); and small subunit ribosomal RNA (SSU rRNA). For *COX1* and *RPS3* HMM
658 profiles were used to screen the metatranscriptomic data. Hits with an e-value = 10–5 and a bit
659 score ≥30 were extracted, and the top match for each marker gene was retained. The *COX1* and
660 *RPS3* sequences were dereplicated and clustered at 98 and 99% amino acid identity,
661 respectively, using USEARCH (-cluster_fast) (*78*). The reconstruction analysis was then
662 performed for the second time by running the translated contigs against the marker genes
663 HMMs. The hits with a bit score ≥30 and aligning regions of ≥90% of the average length of the
664 best-matching HMMs (to be considered as "complete" or "near-complete" virus RdRp domain
665 sequences) were retained. Further, a manual check was also done to verify the taxonomy of the
666 identified community. Additionally, SSU rRNA—including 28S and 18S (eukaryotes); 23S,
667 16S, and 23S (bacteria); and 16S (archaea)—were extracted from assembled contigs and
668 subsequently blasted against the Silva database (*91*) to determine their cellular organism
669 taxonomy. The vOTUs of the cellular organisms were then extracted from the aforementioned
670 dereplicated marker genes.
671
672 **Ecological drivers' analysis**
673 To examine the relationships between the relative abundance of orthornaviraens in different
674 habitats and environmental factors, a redundancy analysis (RDA) was performed using the
675 vegan package in R (*76*) and Hellinger. To ensure a suitable ratio of explanatory variables to
676 samples, variance inflation factors (VIF) were used and included soil depth (avg.), air
677 temperature (^{o}C), soil temperature (^{o}C), porewater pH, carbon dioxide (CO_2; mM), methane
678 (CH_4; mM), carbon (C; %), nitrogen (N; %), δ^{13}C of CO_2, δ^{13}C of CH_4, δ^{15}N of peat, relative
679 abundance of cellular organisms (*COX1* and *RPS3*), and relative abundance of cellular organism
680 (SSU rRNA) in the analysis.
681 The environmental variables were fitted to the first two dimensions of the PCoA using a
682 generalized additive model (function envfit; permutations = 9,999 and na.rm = TRUE) and R
683 vegan package (*76*). Then, they were correlated with all the pCoA dimensions using a mantel
684 test (function mantel; permutations = 9,999 and method = "spearman") after scaling (function
685 scale) and calculating their distance matrices (function vegdist; method = "Euclidean" and na.rm
686 = TRUE). Furthermore, Spearman's correlation analysis was performed (function cor; use =
687 "pairwise.complete.obs") and plotted using ggplot2. Finally, Spearman's pairwise correlation
688 was performed with vegan package in R. The "rcorr" function and the Hellinger transformed
689 data were used as input. Correlograms were generated using the corrplot package to visualize
690 the correlation results; only variables with significant correlations highlighted at a significance
691 level of 0.05 were shown.

692 To quantify the relative contributions of environmental variables to the changes in virus
693 community (the first coordinate of a pCoA; pCo1), a relative-importance analysis test was
694 performed. This analysis utilized the "relaimpo" v2.2–348 package (*92*), using commonly used
695 multiple-linear regression models (i.e., LMG, First, Last, Betasq, and Pratt) with 1,000
696 bootstraps. The relaimpo package provided methods for estimating the contributions of
697 explanatory variables within a multi-linear regression model. The Lindeman–Melinda–Gold
698 model (LMG) calculated the marginal contribution of each predictor variable while accounting
699 for the presence of all other predictor variables in the model. The model "last" (also called
700 usefulness) calculated the importance of each variable as the reduction in the residual sum of
701 squares (RSS) when that variable was added last to a model that already contained all the other
702 predictor variables. The model "first" calculated the importance of each variable as the reduction
703 in the RSS when all the other predictor variables were added to a model that already contained
704 that variable. Model "Betasq" calculated the importance of each variable as the square of its
705 standardized regression coefficient. Model "Prat" calculated the importance of each variable as
706 the sum of squared correlations between that variable and all other variables in the model (**fig.
707 S16**).
708
709 **Functional annotation of RNA virus genomes**
710 Functional aspects of the protein sequences encoded by RNA virus nucleotide sequences (partial
711 or complete genomes and transcripts) at 95% ANI and 80% AF were studied by integrating
712 annotations from several tools. Nucleotide sequences of contigs coding RdRp were searched
713 against (*i*) the conserved domain database (CDD) by using CDD search (*93*), HMMER (*94*),
714 and a minimum e-value of 0.01 and (*ii*) against the NCBI nr database using blastx and a
715 minimum bit score of 50. Protein sequences predicted from the RdRp-coding contigs by using
716 Prodigal (apply corresponding genetic codes previously inferred) were (*i*) searched against
717 KEGG, UniRef, CAZY, MEROPS, Pfam, and VOG databases by using DRAM (*95*). Virus
718 coding sequences whose functional annotation suggested a "cellular" origin were considered
719 "auxiliary metabolic genes" (AMGs) after manual inspection if none of the following cases was
720 accomplished: (*i*) the putative cellular domains overlapped typical RNA viral domains (e.g.,
721 RdRp, capsid, helicase, protease), (*ii*) the putatively cellular domain was equivalent to known
722 RNA viral domains, (*iii*) blastx hits also included viral proteins (to avoid poor annotations in
723 the NCBI nr database), or (*iv*) blastx hits were not hypothetical proteins of unknown functions.
724 To determine redundancies and avoid different annotation names for the same gene function,
725 AMG sequences not at the contig ends were considered "complete" and compared against each
726 other through blastp (cutoffs: e-value of 0.001, 20% IAA, and query coverage of 20%). To avoid
727 the possibility of these AMGs being derived from sequenced genomic DNA carrying EVEs,
728 RNA virus contigs were searched against all the available assembled and co-sampled DNA
729 sequencing data (as described in the sections on metatranscriptome assembly, virus
730 identification, and evaluation of completeness of putative RdRps). The final list of RNA virus
731 AMGs can be seen in **Table S8**.

REFERENCES

1. J. Obu, How Much of the Earth's Surface is Underlain by Permafrost? *JGR Earth Surface* **126**, e2021JF006123 (2021).

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

3. T. Johansson, N. Malmer, P. M. Crill, T. Friborg, J. H. Åkerman, M. Mastepanov, T. R. Christensen, Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Global Change Biology* **12**, 2352–2369 (2006).

4. R. Cavicchioli, W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd, A. T. Classen, T. W. Crowther, R. Danovaro, C. M. Foreman, J. Huisman, D. A. Hutchins, J. K. Jansson, D. M. Karl, B. Koskella, D. B. Mark Welch, J. B. H. Martiny, M. A. Moran, V. J. Orphan, D. S. Reay, J. V. Remais, V. I. Rich, B. K. Singh, L. Y. Stein, F. J. Stewart, M. B. Sullivan, M. J. H. Van Oppen, S. C. Weaver, E. A. Webb, N. S. Webster, Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol* **17**, 569–586 (2019).

5. B. J. Woodcroft, C. M. Singleton, J. A. Boyd, P. N. Evans, J. B. Emerson, A. A. F. Zayed, R. D. Hoelzle, T. O. Lamberton, C. K. McCalley, S. B. Hodgkins, R. M. Wilson, S. O. Purvine, C. D. Nicora, C. Li, S. Frolking, J. P. Chanton, P. M. Crill, S. R. Saleska, V. I. Rich, G. W. Tyson, Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).

6. G. Hugelius, J. Strauss, S. Zubrzycki, J. W. Harden, E. A. G. Schuur, C.-L. Ping, L. Schirrmeister, G. Grosse, G. J. Michaelson, C. D. Koven, J. A. O'Donnell, B. Elberling, U. Mishra, P. Camill, Z. Yu, J. Palmtag, P. Kuhry, Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* **11**, 6573–6593 (2014).

7. B. K. Biskaborn, S. L. Smith, J. Noetzli, H. Matthes, G. Vieira, D. A. Streletskiy, P. Schoeneich, V. E. Romanovsky, A. G. Lewkowicz, A. Abramov, M. Allard, J. Boike, W. L. Cable, H. H. Christiansen, R. Delaloye, B. Diekmann, D. Drozdov, B. Etzelmüller, G. Grosse, M. Guglielmin, T. Ingeman-Nielsen, K. Isaksen, M. Ishikawa, M. Johansson, H. Johannsson, A. Joo, D. Kaverin, A. Kholodov, P. Konstantinov, T. Kröger, C. Lambiel, J.-P. Lanckman, D. Luo, G. Malkova, I. Meiklejohn, N. Moskalenko, M. Oliva, M. Phillips, M. Ramos, A. B. K. Sannel, D. Sergeev, C. Seybold, P. Skryabin, A. Vasiliev, Q. Wu, K. Yoshikawa, M. Zheleznyak, H. Lantuit, Permafrost is warming at a global scale. *Nat Commun* **10**, 264 (2019).

8. R. K. Varner, P. M. Crill, S. Frolking, C. K. McCalley, S. A. Burke, J. P. Chanton, M. E. Holmes, Isogenie Project Coordinators, S. Saleska, M. W. Palace, Permafrost thaw driven changes in hydrology and vegetation cover increase trace gas emissions and climate forcing in Stordalen Mire from 1970 to 2014. *Phil. Trans. R. Soc. A.* **380**, 20210022 (2022).

9. J. B. Ellenbogen, M. A. Borton, B. B. McGivern, D. R. Cronin, D. W. Hoyt, V. Freire-Zapata, C. K. McCalley, R. K. Varner, P. M. Crill, R. A. Wehr, J. P. Chanton, B. J. Woodcroft, M. M. Tfaily, G. W. Tyson, V. I. Rich, K. C. Wrighton, Methylotrophy in the

Mire: direct and indirect routes for methane production in thawing permafrost. *mSystems* **9**, e00698-23 (2024).

10. B. B. McGivern, D. R. Cronin, J. B. Ellenbogen, M. A. Borton, E. L. Knutson, V. Freire-Zapata, J. A. Bouranis, L. Bernhardt, A. I. Hernandez, R. M. Flynn, R. Woyda, A. B. Cory, R. M. Wilson, J. P. Chanton, B. J. Woodcroft, J. G. Ernakovich, M. M. Tfaily, M. B. Sullivan, G. W. Tyson, V. I. Rich, A. E. Hagerman, K. C. Wrighton, Microbial polyphenol metabolism is part of the thawing permafrost carbon cycle. *Nat Microbiol* **9**, 1454–1466 (2024).

11. C. M. Singleton, C. K. McCalley, B. J. Woodcroft, J. A. Boyd, P. N. Evans, S. B. Hodgkins, J. P. Chanton, S. Frolking, P. M. Crill, S. R. Saleska, V. I. Rich, G. W. Tyson, Methanotrophy across a natural permafrost thaw environment. *ISME J* **12**, 2544–2558 (2018).

12. S. Roux, B. E. Dutilh, J. R. Brum, S. Sunagawa, M. B. Duhaime, A. Loy, B. T. Poulos, N. Solonenko, E. Lara, J. Poulain, S. Pesant, S. Kandels-Lewis, C. Dimier, M. Picheral, S. Searson, C. Cruaud, A. Alberti, C. M. Duarte, J. M. Gasol, D. Vaqué, Tara Oceans Coordinators, P. Bork, S. G. Acinas, P. Wincker, M. B. Sullivan, Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).

13. C. Howard-Varona, M. M. Lindback, G. E. Bastien, N. Solonenko, A. A. Zayed, H. Jang, B. Andreopoulos, H. M. Brewer, T. Glavina Del Rio, J. N. Adkins, S. Paul, M. B. Sullivan, M. B. Duhaime, Phage-specific metabolic reprogramming of virocells. *The ISME Journal* **14**, 881–895 (2020).

14. F. Tian, J. M. Wainaina, C. Howard-Varona, G. Domínguez-Huerta, B. Bolduc, M. C. Gazitúa, G. Smith, M. R. Gittrich, O. Zablocki, D. R. Cronin, D. Eveillard, S. J. Hallam, M. B. Sullivan, Prokaryotic-virus-encoded auxiliary metabolic genes throughout the global oceans. *Microbiome* **12**, 159 (2024).

15. S. Roux, J. B. Emerson, Diversity in the soil virosphere: to infinity and beyond? *Trends in Microbiology* **30**, 1025–1035 (2022).

16. J. K. Jansson, K. S. Hofmockel, Soil microbiomes and climate change. *Nat Rev Microbiol* **18**, 35–46 (2020).

17. J. B. Emerson, S. Roux, J. R. Brum, B. Bolduc, B. J. Woodcroft, H. B. Jang, C. M. Singleton, L. M. Solden, A. E. Naas, J. A. Boyd, S. B. Hodgkins, R. M. Wilson, G. Trubl, C. Li, S. Frolking, P. B. Pope, K. C. Wrighton, P. M. Crill, J. P. Chanton, S. R. Saleska, G. W. Tyson, V. I. Rich, M. B. Sullivan, Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* **3**, 870–880 (2018).

18. G. Trubl, H. B. Jang, S. Roux, J. B. Emerson, N. Solonenko, D. R. Vik, L. Solden, J. Ellenbogen, A. T. Runyon, B. Bolduc, B. J. Woodcroft, S. R. Saleska, G. W. Tyson, K. C. Wrighton, M. B. Sullivan, V. I. Rich, Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems* **3**, e00076-18 (2018).

19. C. L. Sun, A. A. Pratama, M. C. Gazitúa, D. Cronin, B. B. McGivern, J. M. Wainaina, D. R. Vik, A. A. Zayed, B. Bolduc, IsoGenie Project Field Teams 2010-2017, the IsoGenie

and EMERGE Projects Coordinators, K. C. Wrighton, V. I. Rich, M. B. Sullivan, Virus ecology and 7-year temporal dynamics across a permafrost thaw gradient. *Environmental Microbiology* **26**, e16665 (2024).

20. C. Carreira, C. Lønborg, B. Acharya, L. Aryal, Z. Buivydaite, F. Borim Corrêa, T. Chen, C. Lorenzen Elberg, J. B. Emerson, L. Hillary, R. B. Khadka, V. Langlois, K. Mason-Jones, T. Netherway, S. Sutela, G. Trubl, A. Wa Kang'eri, R. Wang, R. A. White, A. Winding, T. Zhao, R. Sapkota, Integrating viruses into soil food web biogeochemistry. *Nat Microbiol*, doi: 10.1038/s41564-024-01767-x (2024).

21. G. Dominguez-Huerta, A. A. Zayed, J. M. Wainaina, J. Guo, F. Tian, A. A. Pratama, B. Bolduc, M. Mohssen, O. Zablocki, E. Pelletier, E. Delage, A. Alberti, J.-M. Aury, Q. Carradec, J. H. Kuhn, H. Ogata, P. Wincker, A. Culley, S. Chaffron, M. B. Sullivan, Diversity and ecological footprint of Global Ocean RNA viruses. **376**, 1202–1208 (2022).

22. A. A. Zayed, J. M. Wainaina, G. Dominguez-Huerta, E. Pelletier, J. Guo, M. Mohssen, F. Tian, A. A. Pratama, B. Bolduc, O. Zablocki, D. Cronin, L. Solden, E. Delage, A. Alberti, J.-M. Aury, Q. Carradec, C. Da Silva, K. Labadie, J. Poulain, H.-J. Ruscheweyh, G. Salazar, E. Shatoff, Tara Oceans Coordinators‡, R. Bundschuh, K. Fredrick, L. S. Kubatko, S. Chaffron, A. I. Culley, S. Sunagawa, J. H. Kuhn, P. Wincker, M. B. Sullivan, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. De Vargas, G. Gorsky, L. Guidi, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, N. Poulton, S. Pesant, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sungawa, P. Wincker, Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science* **376**, 156–162 (2022).

23. E. V. Koonin, J. H. Kuhn, V. V. Dolja, M. Krupovic, Megataxonomy and global ecology of the virosphere. *The ISME Journal* **18**, wrad042 (2024).

24. Y. I. Wolf, D. Kazlauskas, J. Iranzo, A. Lucía-Sanz, J. H. Kuhn, M. Krupovic, V. V. Dolja, E. V. Koonin, Origins and Evolution of the Global RNA Virome. *mBio* **9**, e02329-18 (2018).

25. J. H. Kuhn, L. Botella, M. De La Peña, E. J. Vainio, M. Krupovic, B. D. Lee, B. Navarro, S. Sabanadzovic, P. Simmonds, M. Turina, *Ambiviricota* , a novel ribovirian phylum for viruses with viroid-like properties. *J Virol* **98**, e00831-24 (2024).

26. Y. I. Wolf, S. Silas, Y. Wang, S. Wu, M. Bocek, D. Kazlauskas, M. Krupovic, A. Fire, V. V. Dolja, E. V. Koonin, Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* **5**, 1262–1270 (2020).

27. X. Hou, Y. He, P. Fang, S.-Q. Mei, Z. Xu, W.-C. Wu, J.-H. Tian, S. Zhang, Z.-Y. Zeng, Q.-Y. Gou, G.-Y. Xin, S.-J. Le, Y.-Y. Xia, Y.-L. Zhou, F.-M. Hui, Y.-F. Pan, J.-S. Eden, Z.-H. Yang, C. Han, Y.-L. Shu, D. Guo, J. Li, E. C. Holmes, Z.-R. Li, M. Shi, Using artificial intelligence to document the hidden RNA virosphere. *Cell* **187**, 6929-6942.e16 (2024).

28. S. Urayama, A. Fukudome, M. Hirai, T. Okumura, Y. Nishimura, Y. Takaki, N. Kurosawa, E. V. Koonin, M. Krupovic, T. Nunoura, Double-stranded RNA sequencing

858 reveals distinct riboviruses associated with thermoacidophilic bacteria from hot springs
859 in Japan. *Nat Microbiol* **9**, 514–523 (2024).

860 29. R. Wu, E. M. Bottos, V. G. Danna, J. C. Stegen, J. K. Jansson, M. R. Davison, RNA
861 Viruses Linked to Eukaryotic Hosts in Thawed Permafrost. *mSystems* **7**, e00582-22
862 (2022).

863 30. J.-M. Alempic, A. Lartigue, A. E. Goncharov, G. Grosse, J. Strauss, A. N. Tikhonov, A.
864 N. Fedorov, O. Poirot, M. Legendre, S. Santini, C. Abergel, J.-M. Claverie, An Update
865 on Eukaryotic Viruses Revived from Ancient Permafrost. *Viruses* **15**, 564 (2023).

866 31. G. Muscatt, S. Hilton, S. Raguideau, G. Teakle, I. D. E. A. Lidbury, E. M. H. Wellington,
867 C. Quince, A. Millard, G. D. Bending, E. Jameson, Crop management shapes the
868 diversity and activity of DNA and RNA viruses in the rhizosphere. *Microbiome* **10**, 181
869 (2022).

870 32. L. S. Hillary, E. M. Adriaenssens, D. L. Jones, J. E. McDonald, RNA-viromics reveals
871 diverse communities of soil RNA viruses with the potential to affect grassland
872 ecosystems across multiple trophic levels. *ISME COMMUN.* **2**, 34 (2022).

873 33. S. Roux, E. M. Adriaenssens, B. E. Dutilh, E. V. Koonin, A. M. Kropinski, M. Krupovic,
874 J. H. Kuhn, R. Lavigne, J. R. Brister, A. Varsani, C. Amid, R. K. Aziz, S. R.
875 Bordenstein, P. Bork, M. Breitbart, G. R. Cochrane, R. A. Daly, C. Desnues, M. B.
876 Duhaime, J. B. Emerson, F. Enault, J. A. Fuhrman, P. Hingamp, P. Hugenholtz, B. L.
877 Hurwitz, N. N. Ivanova, J. M. Labonté, K.-B. Lee, R. R. Malmstrom, M. Martinez-
878 Garcia, I. K. Mizrachi, H. Ogata, D. Páez-Espino, M.-A. Petit, C. Putonti, T. Rattei, A.
879 Reyes, F. Rodriguez-Valera, K. Rosario, L. Schriml, F. Schulz, G. F. Steward, M. B.
880 Sullivan, S. Sunagawa, C. A. Suttle, B. Temperton, S. G. Tringe, R. V. Thurber, N. S.
881 Webster, K. L. Whiteson, S. W. Wilhelm, K. E. Wommack, T. Woyke, K. C. Wrighton,
882 P. Yilmaz, T. Yoshida, M. J. Young, N. Yutin, L. Z. Allen, N. C. Kyrpides, E. A. Eloe-
883 Fadrosh, Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat*
884 *Biotechnol* **37**, 29–37 (2019).

885 34. U. Neri, Y. I. Wolf, S. Roux, A. P. Camargo, B. Lee, D. Kazlauskas, I. M. Chen, N.
886 Ivanova, L. Zeigler Allen, D. Paez-Espino, D. A. Bryant, D. Bhaya, M. Krupovic, V. V.
887 Dolja, N. C. Kyrpides, E. V. Koonin, U. Gophna, A. B. Narrowe, A. J. Probst, A.
888 Sczyrba, A. Kohler, A. Séguin, A. Shade, B. J. Campbell, B. D. Lindahl, B. K. Reese, B.
889 M. Roque, C. DeRito, C. Averill, D. Cullen, D. A. C. Beck, D. A. Walsh, D. M. Ward,
890 D. Wu, E. Eloe-Fadrosh, E. L. Brodie, E. B. Young, E. A. Lilleskov, F. J. Castillo, F. M.
891 Martin, G. R. LeCleir, G. T. Attwood, H. Cadillo-Quiroz, H. M. Simon, I. Hewson, I. V.
892 Grigoriev, J. M. Tiedje, J. K. Jansson, J. Lee, J. S. VanderGheynst, J. Dangl, J. S.
893 Bowman, J. L. Blanchard, J. L. Bowen, J. Xu, J. F. Banfield, J. W. Deming, J. E. Kostka,
894 J. M. Gladden, J. Z. Rapp, J. Sharpe, K. D. McMahon, K. K. Treseder, K. D. Bidle, K. C.
895 Wrighton, K. Thamatrakoln, K. Nusslein, L. K. Meredith, L. Ramirez, M. Buee, M.
896 Huntemann, M. G. Kalyuzhnaya, M. P. Waldrop, M. B. Sullivan, M. O. Schrenk, M.
897 Hess, M. A. Vega, M. A. O'Malley, M. Medina, N. E. Gilbert, N. Delherbe, O. U.
898 Mason, P. Dijkstra, P. F. Chuckran, P. Baldrian, P. Constant, R. Stepanauskas, R. A.
899 Daly, R. Lamendella, R. J. Gruninger, R. M. McKay, S. Hylander, S. L. Lebeis, S. P.
900 Esser, S. G. Acinas, S. S. Wilhelm, S. W. Singer, S. S. Tringe, T. Woyke, T. B. K.
901 Reddy, T. H. Bell, T. Mock, T. McAllister, V. Thiel, V. J. Denef, W.-T. Liu, W.

Martens-Habbena, X.-J. Allen Liu, Z. S. Cooper, Z. Wang, Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell* **185**, 4023-4037.e18 (2022).

35. G. Dominguez-Huerta, J. M. Wainaina, A. A. Zayed, A. I. Culley, J. H. Kuhn, M. B. Sullivan, The RNA virosphere: How big and diverse is it? *Environmental Microbiology* **25**, 209–215 (2023).

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

37. N. Fierer, R. B. Jackson, The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 626–631 (2006).

38. Y.-M. Chen, S. Sadiq, J.-H. Tian, X. Chen, X.-D. Lin, J.-J. Shen, H. Chen, Z.-Y. Hao, M. Wille, Z.-C. Zhou, J. Wu, F. Li, H.-W. Wang, W.-D. Yang, Q.-Y. Xu, W. Wang, W.-H. Gao, E. C. Holmes, Y.-Z. Zhang, RNA viromes from terrestrial sites across China expand environmental viral diversity. *Nat Microbiol* **7**, 1312–1323 (2022).

39. E. P. Starr, E. E. Nuccio, J. Pett-Ridge, J. F. Banfield, M. K. Firestone, Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25900–25908 (2019).

40. M. E. Holmes, P. M. Crill, W. C. Burnett, C. K. McCalley, R. M. Wilson, S. Frolking, K. -Y. Chang, W. J. Riley, R. K. Varner, S. B. Hodgkins, IsoGenie Project Coordinators, IsoGenie Field Team, A. P. McNichol, S. R. Saleska, V. I. Rich, J. P. Chanton, Carbon Accumulation, Flux, and Fate in Stordalen Mire, a Permafrost Peatland in Transition. *Global Biogeochemical Cycles* **36**, e2021GB007113 (2022).

41. R. Wu, M. R. Davison, Y. Gao, C. D. Nicora, J. E. Mcdermott, K. E. Burnum-Johnson, K. S. Hofmockel, J. K. Jansson, Moisture modulates soil reservoirs of active DNA and RNA viruses. *Commun Biol* **4**, 992 (2021).

42. M. Hough, A. McClure, B. Bolduc, E. Dorrepaal, S. Saleska, V. Klepac-Ceraj, V. Rich, Biotic and Environmental Drivers of Plant Microbiomes Across a Permafrost Thaw Gradient. *Front. Microbiol.* **11**, 796 (2020).

43. M. P. Waldrop, C. L. Chabot, S. Liebner, S. Holm, M. W. Snyder, M. Dillon, S. R. Dudgeon, T. A. Douglas, M.-C. Leewis, K. M. Walter Anthony, J. W. McFarland, C. D. Arp, A. C. Bondurant, N. Taş, R. Mackelprang, Permafrost microbial communities and functional genes are structured by latitudinal and soil geochemical gradients. *The ISME Journal* **17**, 1224–1235 (2023).

44. R. Mackelprang, A. Burkert, M. Haw, T. Mahendrarajah, C. H. Conaway, T. A. Douglas, M. P. Waldrop, Microbial survival strategies in ancient permafrost: insights from metagenomics. *The ISME Journal* **11**, 2305–2318 (2017).

45. J. G. Ernakovich, R. A. Barbato, V. I. Rich, C. Schädel, R. E. Hewitt, S. J. Doherty, E. D. Whalen, B. W. Abbott, J. Barta, C. Biasi, C. L. Chabot, J. Hultman, C. Knoblauch, M. C. Y. L. Vetter, M. Leewis, S. Liebner, R. Mackelprang, T. C. Onstott, A. Richter, U. M. E. Schütte, H. M. P. Siljanen, N. Taş, I. Timling, T. A. Vishnivetskaya, M. P. Waldrop, M.

Winkel, Microbiome assembly in thawing permafrost and its feedbacks to climate. *Global Change Biology* **28**, 5007–5026 (2022).

46. M. Schostag, A. Priemé, S. Jacquiod, J. Russel, F. Ekelund, C. S. Jacobsen, Bacterial and protozoan dynamics upon thawing and freezing of an active layer permafrost soil. *The ISME Journal* **13**, 1345–1359 (2019).

47. E. A. G. Schuur, B. W. Abbott, R. Commane, J. Ernakovich, E. Euskirchen, G. Hugelius, G. Grosse, M. Jones, C. Koven, V. Leshyk, D. Lawrence, M. M. Loranty, M. Mauritz, D. Olefeldt, S. Natali, H. Rodenhizer, V. Salmon, C. Schädel, J. Strauss, C. Treat, M. Turetsky, Permafrost and Climate Change: Carbon Cycle Feedbacks From the Warming Arctic. *Annu. Rev. Environ. Resour.* **47**, 343–371 (2022).

48. A. Bucklin, D. Steinke, L. Blanco-Bercial, DNA Barcoding of Marine Metazoa. *Annu. Rev. Mar. Sci.* **3**, 471–508 (2011).

49. J. Pawlowski, S. Audic, S. Adl, D. Bass, L. Belbahri, C. Berney, S. S. Bowser, I. Cepicka, J. Decelle, M. Dunthorn, A. M. Fiore-Donno, G. H. Gile, M. Holzmann, R. Jahn, M. Jirků, P. J. Keeling, M. Kostka, A. Kudryavtsev, E. Lara, J. Lukeš, D. G. Mann, E. A. D. Mitchell, F. Nitsche, M. Romeralo, G. W. Saunders, A. G. B. Simpson, A. V. Smirnov, J. L. Spouge, R. F. Stern, T. Stoeck, J. Zimmermann, D. Schindel, C. De Vargas, CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biol* **10**, e1001419 (2012).

50. S. Lee, J. W. Sorensen, R. L. Walker, J. B. Emerson, G. W. Nicol, C. Hazard, Soil pH influences the structure of virus communities at local and global scales. *Soil Biology and Biochemistry* **166**, 108569 (2022).

51. R. M. Wilson, M. A. Hough, B. A. Verbeke, S. B. Hodgkins, J. P. Chanton, S. D. Saleska, V. I. Rich, M. M. Tfaily, G. Tyson, M. B. Sullivan, E. Brodie, W. J. Riley, B. Woodcroft, C. McCalley, S. C. Dominguez, P. M. Crill, R. K. Varner, S. Frolking, W. T. Cooper, Plant organic matter inputs exert a strong control on soil organic matter decomposition in a thawing permafrost peatland. *Science of The Total Environment* **820**, 152757 (2022).

52. A. Fofana, D. Anderson, C. K. McCalley, S. Hodgkins, R. M. Wilson, D. Cronin, N. Raab, M. Torabi, R. K. Varner, P. Crill, S. R. Saleska, J. P. Chanton, M. M. Tfaily, V. I. Rich, Mapping substrate use across a permafrost thaw gradient. *Soil Biology and Biochemistry* **175**, 108809 (2022).

53. A. J. B. Kreutzberger, A. Sanyal, A. Saminathan, L.-M. Bloyet, S. Stumpf, Z. Liu, R. Ojha, M. T. Patjas, A. Geneid, G. Scanavachi, C. A. Doyle, E. Somerville, R. B. D. C. Correia, G. Di Caprio, S. Toppila-Salmi, A. Mäkitie, V. Kiessling, O. Vapalahti, S. P. J. Whelan, G. Balistreri, T. Kirchhausen, SARS-CoV-2 requires acidic pH to infect cells. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2209514119 (2022).

54. L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, Tara Oceans Consortium Coordinators, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. De

Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470 (2016).

55. A. C. Gregory, A. A. Zayed, N. Conceição-Neto, B. Temperton, B. Bolduc, A. Alberti, M. Ardyna, K. Arkhipova, M. Carmichael, C. Cruaud, C. Dimier, G. Domínguez-Huerta, J. Ferland, S. Kandels, Y. Liu, C. Marec, S. Pesant, M. Picheral, S. Pisarev, J. Poulain, J.-É. Tremblay, D. Vik, M. Babin, C. Bowler, A. I. Culley, C. De Vargas, B. E. Dutilh, D. Iudicone, L. Karp-Boss, S. Roux, S. Sunagawa, P. Wincker, M. B. Sullivan, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. De Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109-1123.e14 (2019).

56. Z.-P. Zhong, J. Du, S. Köstlbacher, P. Pjevac, S. Orlić, M. B. Sullivan, Viral potential to modulate microbial methane metabolism varies by habitat. *Nat Commun* **15**, 1857 (2024).

57. C. A. Suttle, Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **5**, 801–812 (2007).

58. R. Wu, G. Trubl, N. Taş, J. K. Jansson, Permafrost as a potential pathogen reservoir. *One Earth* **5**, 351–360 (2022).

59. T. Thingstad, R. Lignell, Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).

60. E. V. Koonin, V. V. Dolja, M. Krupovic, Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **479–480**, 2–25 (2015).

61. B. H. Svensson, T. Rosswall, Energy Flow through the Subarctic Mire at Stordalen. *Ecological Bulletins* **30**.

62. I. Bielańska-Grajner, T. Mieczan, A. Cudak, Co-Occurrence of Ciliates and Rotifers in Peat Mosses. *Pol. J. Environ. Stud.* **20**, 533–540 (2011).

63. X. Lu, Z. Dai, J. Xue, W. Li, P. Ni, J. Xu, C. Zhou, W. Zhang, Discovery of novel RNA viruses through analysis of fungi-associated next-generation sequencing data. *BMC Genomics* **25**, 517 (2024).

64. A. M. S. Correa, C. Howard-Varona, S. R. Coy, A. Buchan, M. B. Sullivan, J. S. Weitz, Revisiting the rules of life for viruses of microorganisms. *Nat Rev Microbiol* **19**, 501–513 (2021).

65. M. J. Roossinck, Lifestyles of plant viruses. *Phil. Trans. R. Soc. B* **365**, 1899–1905 (2010).

66. C. R. Perryman, C. K. McCalley, A. Malhotra, M. F. Fahnestock, N. N. Kashi, J. G. Bryce, R. Giesler, R. K. Varner, Thaw Transitions and Redox Conditions Drive Methane Oxidation in a Permafrost Peatland. *J. Geophys. Res. Biogeosci.* **125** (2020).

1023 67. Y. Zhao, Z. Zhang, P. Liu, M. Feng, R. Wen, Y. Liu, Functional and evolutionary
1024 characterization of potential auxiliary metabolic genes (AMGs) of the global RNA
1025 virome. [Preprint] (2024). https://doi.org/10.1101/2024.10.16.618626.

1026 68. M. Shi, X.-D. Lin, J.-H. Tian, L.-J. Chen, X. Chen, C.-X. Li, X.-C. Qin, J. Li, J.-P. Cao,
1027 J.-S. Eden, J. Buchmann, W. Wang, J. Xu, E. C. Holmes, Y.-Z. Zhang, Redefining the
1028 invertebrate RNA virosphere. *Nature* **540**, 539–543 (2016).

1029 69. P. Genevaux, F. Schwager, C. Georgopoulos, W. L. Kelley, Scanning Mutagenesis
1030 Identifies Amino Acid Residues Essential for the *in Vivo* Activity of the *Escherichia coli*
1031 DnaJ (Hsp40) J-Domain. *Genetics* **162**, 1045–1053 (2002).

1032 70. A. Lubkowska, W. Pluta, A. Strońska, A. Lalko, Role of Heat Shock Proteins (HSP70
1033 and HSP90) in Viral Infection. *IJMS* **22**, 9366 (2021).

1034 71. T. Matsui, Y. Kodera, H. Endoh, E. Miyauchi, H. Komatsu, K. Sato, T. Tanaka, T.
1035 Kohno, T. Maeda, RNA Recognition Mechanism of the Minimal Active Domain of the
1036 Human Immunodeficiency Virus Type-2 Nucleocapsid Protein. *The Journal of*
1037 *Biochemistry* **141**, 269–277 (2007).

1038 72. L. M. Iyer, A. M. Burroughs, V. Anantharaman, L. Aravind, Apprehending the NAD+–
1039 ADPr-Dependent Systems in the Virus World. *Viruses* **14**, 1977 (2022).

1040 73. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
1041 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

1042 74. D. Li, R. Luo, C.-M. Liu, C.-M. Leung, H.-F. Ting, K. Sadakane, H. Yamashita, T.-W.
1043 Lam, MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced
1044 methodologies and community practices. *Methods* **102**, 3–11 (2016).

1045 75. S. Nayfach, A. P. Camargo, F. Schulz, E. Eloe-Fadrosh, S. Roux, N. C. Kyrpides,
1046 CheckV assesses the quality and completeness of metagenome-assembled viral genomes.
1047 *Nat Biotechnol* **39**, 578–585 (2021).

1048 76. P. Dixon, VEGAN, a package of R functions for community ecology. *Journal of*
1049 *Vegetation Science* **14**, 927–930 (2003).

1050 77. S. Zhao, Y. Guo, Q. Sheng, Y. Shyr, Heatmap3: an improved heatmap package with
1051 more powerful and convenient features. *BMC Bioinformatics* **15**, P16, 1471-2105-15-
1052 S10-P16 (2014).

1053 78. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST.
1054 *Bioinformatics* **26**, 2460–2461 (2010).

1055 79. J. R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of
1056 intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

1057 80. J. Charon, J. P. Buchmann, S. Sadiq, E. C. Holmes, RdRp-scan: A bioinformatic resource
1058 to identify and annotate divergent RNA viruses in metagenomic sequence data. *Virus*
1059 *Evolution* **8**, veac082 (2022).

1060 81. S. Altschul, Gapped BLAST and PSI-BLAST: a new generation of protein database
1061 search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

1062 82. A. J. Enright, An efficient algorithm for large-scale detection of protein families. *Nucleic*
1063 *Acids Research* **30**, 1575–1584 (2002).

1064 83. K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast
1065 Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).

1066 84. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated
1067 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973
1068 (2009).

1069 85. M. F. Boni, D. Posada, M. W. Feldman, An Exact Nonparametric Method for Inferring
1070 Mosaic Structure in Sequence Triplets. *Genetics* **176**, 1035–1047 (2007).

1071 86. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. Von Haeseler, L. S. Jermiin,
1072 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**,
1073 587–589 (2017).

1074 87. L.-T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A Fast and
1075 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.
1076 *Molecular Biology and Evolution* **32**, 268–274 (2015).

1077 88. F. Menardo, C. Loiseau, D. Brites, M. Coscolla, S. M. Gygli, L. K. Rutaihwa, A.
1078 Trauner, C. Beisel, S. Borrell, S. Gagneux, Treemmer: a tool to reduce large
1079 phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics* **19**, 164
1080 (2018).

1081 89. M. B. Dion, P.-L. Plante, E. Zufferey, S. A. Shah, J. Corbeil, S. Moineau, Streamlining
1082 CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic*
1083 *Acids Research* **49**, 3127–3138 (2021).

1084 90. J. Rumnieks, K. Tars, Diversity of pili-specific bacteriophages: genome sequence of
1085 IncM plasmid-dependent RNA phage M. *BMC Microbiol* **12**, 277 (2012).

1086 91. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O.
1087 Glöckner, The SILVA ribosomal RNA gene database project: improved data processing
1088 and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012).

1089 92. U. Grömping, Relative Importance for Linear Regression in *R* : The Package **relaimpo**.
1090 *J. Stat. Soft.* **17** (2006).

1091 93. A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R.
1092 C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S.
1093 Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, S. H. Bryant, CDD:
1094 NCBI's conserved domain database. *Nucleic Acids Research* **43**, D222–D226 (2015).

1095 94. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity
1096 searching. *Nucleic Acids Research* **39**, W29–W37 (2011).

95. M. Shaffer, M. A. Borton, B. B. McGivern, A. A. Zayed, S. L. La Rosa, L. M. Solden, P. Liu, A. B. Narrowe, J. Rodríguez-Ramos, B. Bolduc, M. C. Gazitúa, R. A. Daly, G. J. Smith, D. R. Vik, P. B. Pope, M. B. Sullivan, S. Roux, K. C. Wrighton, DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research* **48**, 8883–8900 (2020).

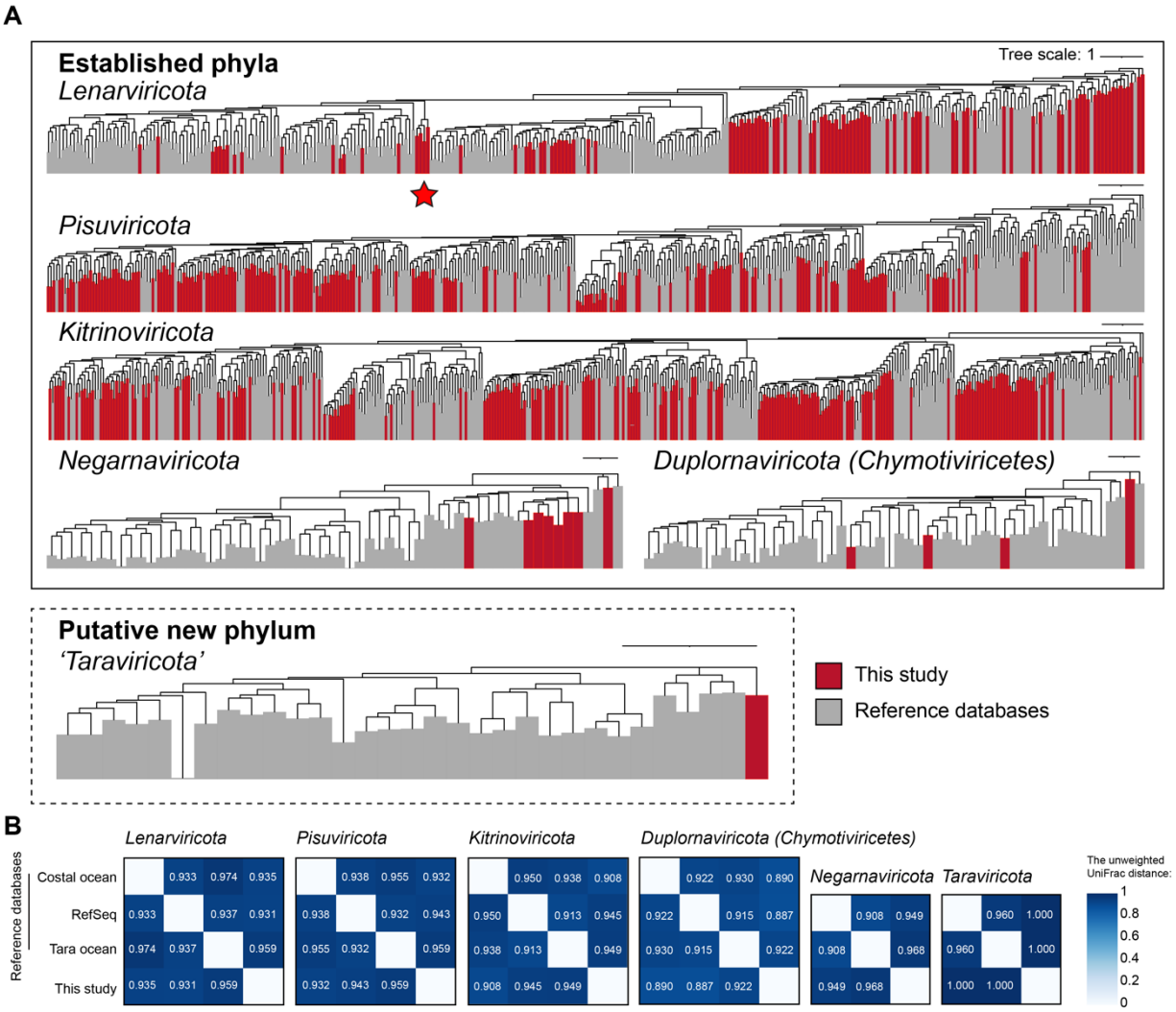

Figure 1. Phylum-ranked, RdRp-based orthornaviran phylogenies.
(A) Six maximum-likelihood phylogenetic trees were built from an RdRp-guided taxonomy analysis of near-complete RdRp sequences. Five trees depict ICTV-established phyla, and one corresponds to a previously suggested phylum, '*Taraviricota*' (Zayed et al.,(*22*)). The red color indicates Stordalen Mire orthornaviran sequences discovered in this study, whereas grey indicates previously known reference sequences (including open-ocean and coastal-ocean RNA viromes). The scale bar indicatesone amino acid residue substitution per site. The ☆ represents a likely new class, Candidatus 'Stormiviricetes'. RdRp, RNA-directed RNA polymerase; ICTV, International Committee on Taxonomy of Viruses. **(B)** Heatmap showing the unweighted UniFrac distances between RNA virus phylum-rank trees across datasets from the coastal ocean (*24*), RefSeq, *Tara* Oceans (*21*, *22*), and Stordalen Mire (this study). Phyla analyzed include *Lenarviricota, Pisuviricota, Kitrinoviricota, Duplornaviricota* (Chymotiviricetes), *Negarnaviricota*, and '*Taraviricota'*. Higher values (darker cells) indicate greater phylogenetic dissimilarity between datasets, whereas lower values (lighter cells) indicate greater similarity.

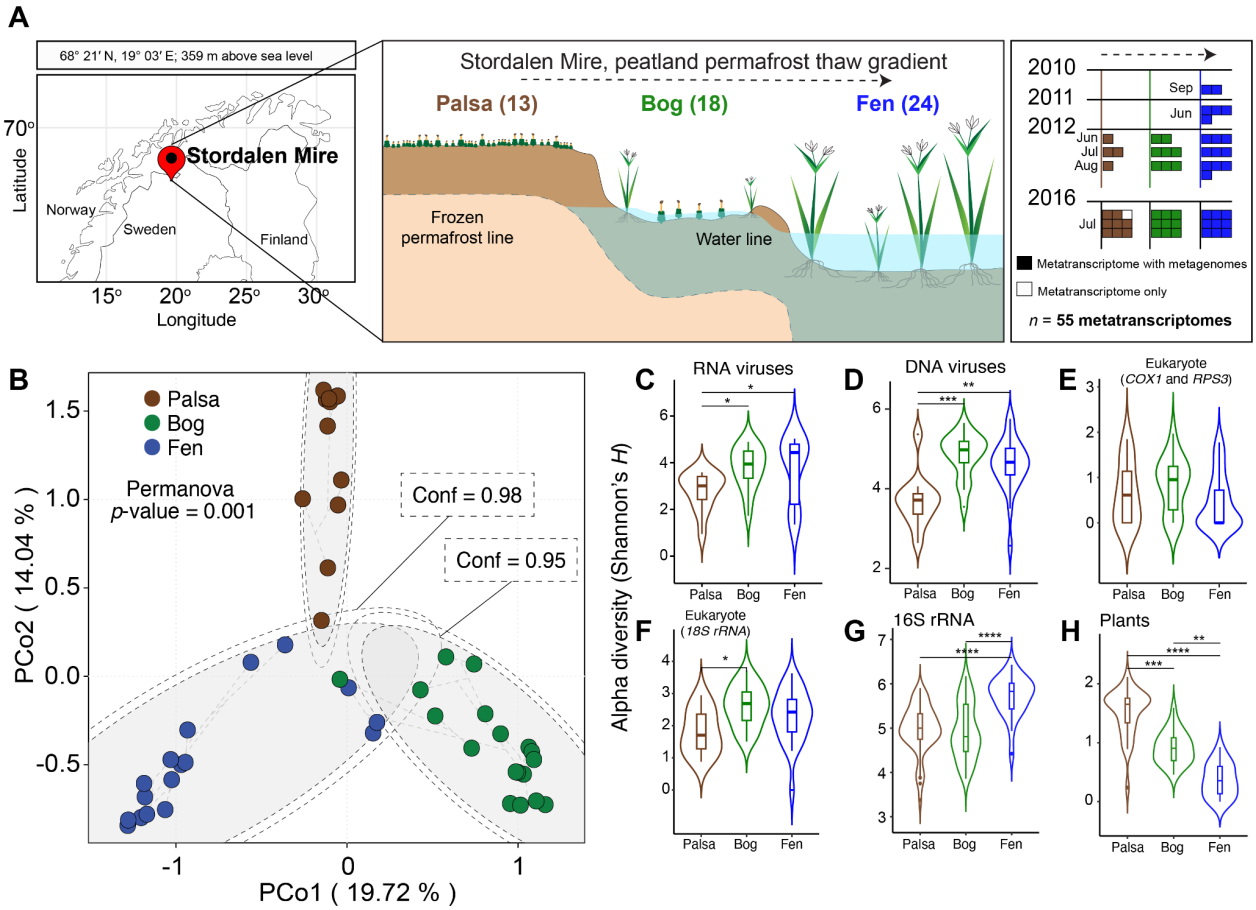

Figure 2. Cross-habitat soil sampling of Stordalen Mire in northern Sweden. **(A)** Soil metatranscriptomes were collected from the active layer across the peatland permafrost thawing gradient with intact palsa (approximately 30 cm), bog (intermediate thawing, approximately 60 cm), and fen (fully thawed) habitats. **(B)** Principal component analysis (PCoA) of a Bray–Curtis dissimilarity matrix calculated from all virus operational units (vOTUs; i.e., species-rank units) uncovered in this study. Dot colors correspond to the colours of the thawing gradient (brown for palsa; green for bog; and blue for fen. Alpha-diversity (Shannon's *H*) of **(C)** RNA viruses, **(D)** dsDNA viruses, **(E)** Eukaryote and prokaryote (based on *COX1* and *RPS3*), **(F)** Prokaryote (based on 18S rRNA), **(G)** 16S rRNA, and **(H)** Plants. Violin plots with boxplots show medians and quartiles. Statistical analysis was performed using the Kruskal–Wallis test with *post hoc* Dunn test and *p*-adjusted Bonferroni. Only the significant values are shown and donated: * *p*-value ≤0.05; ** *p*-value ≤0.01, *** *p*-value ≤0.001, and **** *p*-value ≤0.0001. **Fig. 2** G and H on the complementary papers.

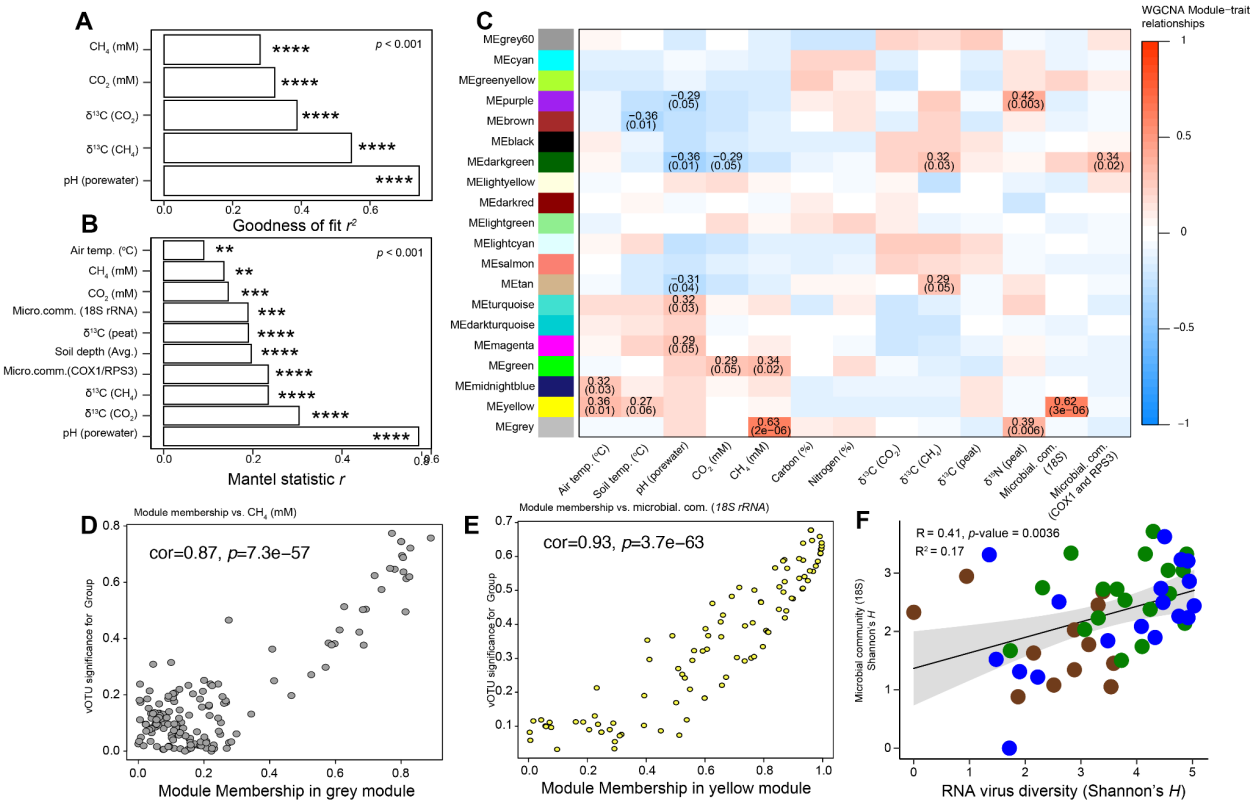

Figure 3. Ecological drivers of Stordalen Mire orthornaviraens.
(**A, B**) The first two dimensions (with a goodness of fit r^2 derived from a generalized additive model) and all dimensions (determined through a Mantel test based on Spearman's correlation) revealed potential ecological drivers and predictors of orthornaviran beta-diversity. pH was consistently identified as the strongest predictor of beta-diversity. Only the significant values are shown and donated as follows: * *p*-value ≤0.05; ** *p* -value ≤0.01, and *** *p* -value ≤0.001. (**C**) Heat map of module-trait relationships derived from Weighted Gene Co-Expression Network Analysis (WGCNA). Each row corresponds to viral operational taxonomic units (vOTUs), grouped into modules (labeled by module colors), and each column represents an environmental trait. The color scale indicates the correlation coefficients. The correlation values and corresponding *p*-values (in parentheses) are shown within each cell; only *p*-values ≤0.05 are shown. (**D, E**) Correlation analysis between module membership in the grey and yellow modules vs. methane concentration (CH_4 mM) and microbial community (*18S rRNA*) relative abundance, respectively. Each point represents a vOTU within the grey and yellow modules. (**F**) Spearman's correlation analysis between orthornaviran and host cellular organism diversity (Shannon's *H*).

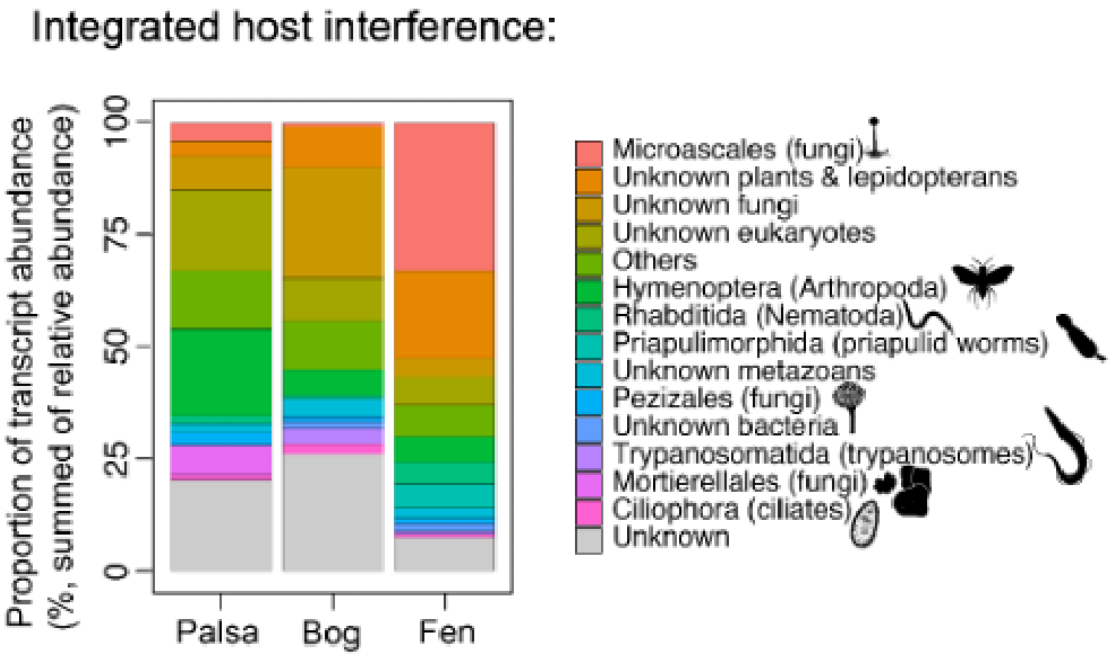

Figure 4. Integrated host inferences, and the co-phylogeny relationship between Stordalen Mire orthornaviraens and their predicted cellular hosts.
(A) Taxonomic assignment of the prokaryotes and eukaryotes in Stordalen Mire identified via *COX1* and *RPS3* gene and 18S rRNA analysis. Cytochrome c oxidase subunit 1 mitochondrial gene (*COX1*); a component of the eukaryotic 40S ribosomal subunit, ribosomal protein S3 (*RPS3*); and small subunit ribosomal RNA (*18S rRNA*).

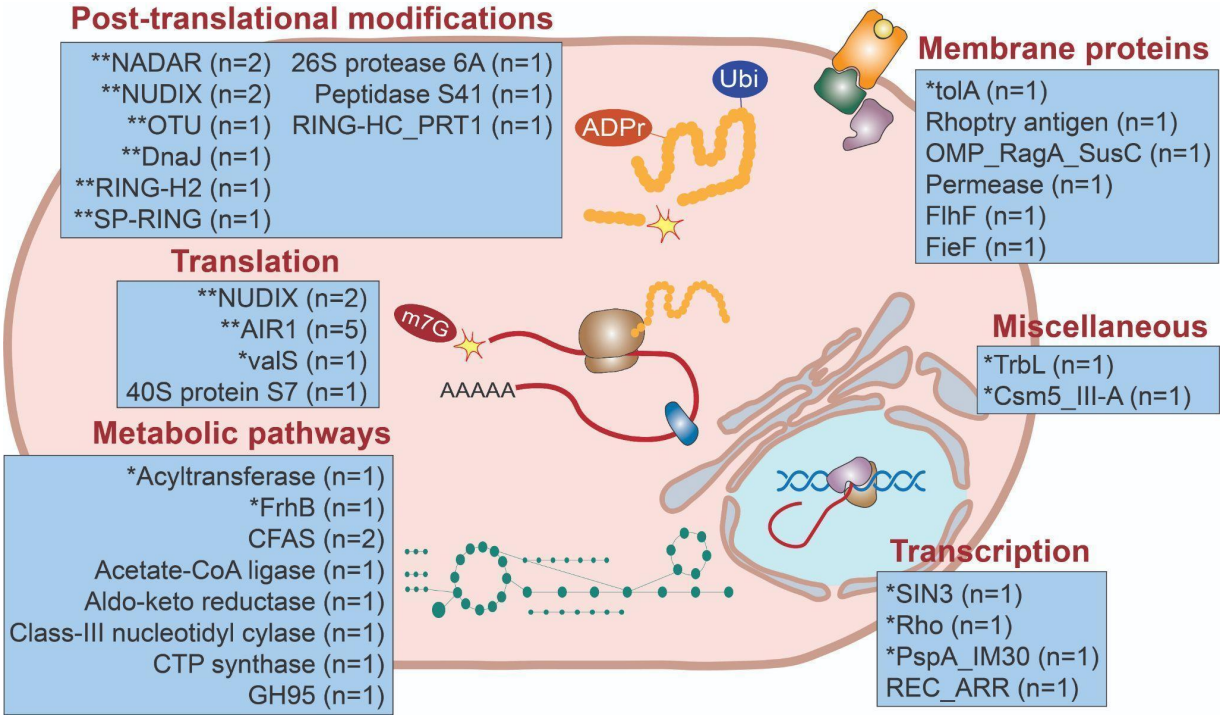

Figure 5. Functional diversity of AMGs carried by Stordalen Mire orthornaviraens. Schematic representation of the hypothesized effects of roles played in the manipulation of host metabolism by orthornaviraen AMGs, which are separated according to functional categories (red text) predicted from using solely functional domains (single asterisk, "permissive" set) and motifs within domains as well (double asterisk, "conservative" set). Blue boxes contain the corresponding protein annotation based on the conserved domain database (CDD), with the corresponding number of vOTUs listed in parentheses. Double asterisks indicate that AMGs that were not located at the contig ends and for which function could be identified studying motifs on the CDD, whereas those with single asterisks indicate that AMGs were not located at contig ends and for which functions could not be identified based on motifs. The absence of an asterisk indicates that AMGs were considered "broken" because given that they are located at contig ends, limiting our ability to predict their functions. AMGs that were found in viral contigs of 500–1,000 nt (**Table S8**) are not reported in this figure. RING-HC_PRT1, RING finger, HC subclass, found in thale cress (*Arabidopsis thaliana* (L.) Heynh.)) proteolysis 1 protein; valS, valine-tRNA ligase; FrhB, coenzyme F420 hydrogenase subunit beta; CFAS, cyclopropane fatty acid synthetase; GH95, glycoside hydrolase family 95; TolA, Tol proteins necessary for the uptake of all the group A colicins; TonB-linked outer membrane protein SusD/RagB family; FlhF, flagellar biosynthesis protein; FieF, divalent metal cation (Fe/Co/Zn/Cd) transporter; TrbL, conjugal transfer protein TrbL; Csm5_III-A, CRISPR system type III-A Csm effector Csm5; SIN3, transcriptional regulatory protein SIN3; Rho, transcription termination factor Rho; PspA_IM3, PspA a protein that suppresses σ^{54}-dependent transcription; REC_ARR, phosphoacceptor receiver domain of type B Arabidopsis response regulators; NADAR, *Escherichia coli* swarming motility protein YbiA and related proteins; NUDIX, hydrolases of Nucleoside Diphosphate linked to another moiety (X); AIR1, arginine methyltransferase-Interacting RING finger protein.