Viral spillover risk in High Arctic increases with melting glaciers

Audrée Lemieux[1], Graham A. Colby[1], Alexandre J. Poulain[1], Stéphane Aris-Brosou[1,2,✉]

[1]Department of Biology, University of Ottawa, Ottawa, ON K1N 6N5, Canada

[2]Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON K1N 6N5, Canada

[✉]e-mail: `sarisbro@uottawa.ca`

Abstract

While many viruses have a single natural host, host restriction can be incomplete, hereby leading to spillovers to other host species, potentially causing significant diseases as it is the case with the Influenza A, Ebola, or the SARS-CoV-2 viruses. However, such spillover risks are difficult to quantify. As climate change is rapidly transforming environments, it is becoming critical to quantify the potential for spillovers, in an unbiased manner. For this, we resorted to a metagenomics approach, and focused on two environments in the High Arctic, soil and lake sediments from Lake Hazen. We used DNA and RNA sequencing to reconstruct the lake's virosphere and its range of eukaryotic hosts, and show that spillover risk is higher in lake sediments than in soil and increased with runoff from glacier melt - a proxy for climate change. Should climate change also shift species range of potential vectors northwards, the High Arctic could become fertile ground for emerging pandemics.

Viruses are ubiquitous and are often described as the most abundant replicators on Earth (Dávila-Ramos et al., 2019; Koonin et al., 2015; Zhang et al., 2018). In spite of having highly diverse genomes, viruses are not independent "organisms" or replicators (Dawkins, 1989), as they need to infect a host's cell in order to replicate. These virus/host relationships seem relatively stable within superkingdoms, so that viruses can be classified as *archaeal*, *bacterial* (also known as *bacteriophages*), and *eukaryotic* (Malik et al., 2017; Nasir et al., 2014, 2017), but host changes are not uncommon below this rank. Indeed, many species causing significant diseases can infect more than one host, as it is the case with the Influenza A (Long et al., 2019), Ebola (Rewar & Mirdha, 2014), and SARS-CoV-2 (Zhou & Shi, 2021) viruses. While these three viruses have non-human wild animal reservoirs as natural hosts, others have a broader host range, or their reservoir is more challenging to identify. Such host restrictions (or alternatively spillover capacities or risks) are to date poorly defined pragmatically and hence, difficult to assess without resorting to expert opinion (Grange et al., 2021).

Numerous factors are thought to influence such a viral spillover risk. For instance, viral particles need to attach themselves to specific receptors on their host's cell to invade it (Longdon et al., 2014; Maginnis, 2018; Parrish et al., 2008). The conservation of those receptors across multiple species allows these hosts to be more predisposed to becoming infected by the same virus (Parrish et al., 2008; Woolhouse et al., 2005). Indeed, from an evolutionary standpoint, viruses are more prone to infecting hosts that are phylogenetically close to their natural host (Longdon et al., 2014, 2018), potentially because it is easier for them to infect and colonize species that are genetically similar (Geoghegan et al., 2017). Alternatively, but not exclusively, high mutation rates might explain why RNA viruses change host more often than other viruses (Longdon et al., 2014), as most lack proofreading mechanisms (Parrish et al., 2008).

To further investigate the role of host relatedness in viral host switch events, multiple studies have focused on the cophylogenetic relationships between viruses and their hosts (Bellec et al., 2014; Bennett et al., 2020; Jackson & Charleston, 2004; Madinda et al., 2016; Olival et al., 2017). Conceptually, if both viruses and their hosts cospeciate, the topologies of their respective phylogenetic trees should be identical or *congruent*. On the other hand, the occurrence of spillovers would result in incongruent virus/host phylogenies.

To date only few studies have characterized the communities and genomes of viruses in extreme environments (Dávila-Ramos et al., 2019), and even fewer, if any, describe their spillover risk. The High Arctic is of special interest as it is particularly affected by climate change, warming faster than the rest of the world (Colby et al., 2020; Lehnherr et al., 2018; Ruuskanen et al., 2020). Warming climate and rapid transitions of the environment increase the risks of spillover events by varying the global distributions and dynamics of viruses, and their reservoirs and vectors (Parkinson et al., 2014; Waits et al., 2018), as shown for arboviruses (Ciota & Keyel, 2019) and the Hendra virus (Martin et al., 2018). Furthermore, as the climate changes, the metabolic activity of the Arctic's microbiosphere also shifts, which in turns affects numerous ecosystem processes such as the emergence of new pathogens (Messan et al., 2020). It has now become critical to quantify the risk of these spillovers.

To address these outstanding questions, we resorted to a combination of metagenomics and of cophylogenetic modeling to quantify viral spillover risk by sampling in an unbiased manner both the virosphere and its range of hosts (Zhang et al., 2018). Focusing on eukaryotic hosts, which are critically affected by viral spillovers (Carrasco-Hernandez et al., 2017), we contrasted two local environments, lake sediments and soil samples of a High Arctic lake, Lake Hazen, to test how viral spillover risk is affected by glacier runoff, and hence potentially by global warming, which is expected to increase runoff with increasing

glacier melt at Lake Hazen (Colby et al., 2020; Lehnherr et al., 2018). We show here that the risk increases with warming climate, but is likely to remain low in the absence of "bridge vectors."

Results and Discussion

Plant and fungal viruses are overrepresented. Based on our most sensitive annotation pipeline (Supplementary Methods), viruses represented less than 1% of all contigs (See Supplementary Methods). RNA viruses of all kinds (*i.e.*, dsRNA, +ssRNA, and -ssRNA viruses) were found to be significantly more abundant than DNA viruses in all samples ($> 70.5\%$; binomial tests, $P < 2.48 \times 10^{-7}$; Figure 1, Table 1), with 7.3% to 36.5% of contigs *per* sampling site that had no annotations (binomial tests, $P < 3.82 \times 10^{-5}$; Table S2). This dominance of RNA viruses is not unexpected, as fungi biomass for instance surpasses that of bacteria in Arctic environments by 1-2 orders of magnitude (Schmidt & Bölter, 2002), and eukaryotes are known to be the main targets of RNA viruses (Koonin et al., 2015; Malik et al., 2017; Nasir et al., 2014, 2017). Our results are however difficult to compare with previous studies in the High Arctic, as those were based solely on DNA metagenomics sequencing (Aguirre de Cárcer et al., 2015; Gregory et al., 2019; Labbé et al., 2020), probably because RNA viruses are thought to be unstable (Yau & Seth-Pasricha, 2019), or due to inadequate sampling strategies to extract RNA viruses (Wang, 2020). Our results show nonetheless that it is possible to recover RNA viruses from RNA-seq metagenomics.

All viral types confounded, those infecting plants and/or fungi were overrepresented compared to those infecting animals and protists ($> 68.8\%$, binomial tests, $P < 2.48 \times 10^{-7}$; Table 1). This overrepresentation might reflect a preservation bias, due to the

constitutive defenses found in plants and fungi offered by their waxy epidermal cuticles and cell walls (Kozieł et al., 2021), even if most plant viruses lack a protective lipoprotein envelope as found in animal viruses (Stavolone & Lionetti, 2017). But irrespective of such a preservation bias, this imbalance could imply a high spillover potential among plants and fungi in the High Arctic for two reasons. First, RNA viruses are the most likely pathogens to switch hosts, due to their high rates of evolution (Aris-Brosou et al., 2019; Longdon et al., 2014). Second, plant biomass has been increasing over the past two decades in the High Arctic due to regional warming (Hudson & Henry, 2009), and is likely to keep doing as warming continues.

Spillover risk increases with glacier runoff. Given these viral and eukaryotic host representations, can spillover risk be assessed in these environments? To address this question, we resorted to the novel global-fit model Random TaPas. The stability of its results was assessed by running this algorithm three times, and by combining the results for the normalized Gini coefficients ($G^\star \in [0, 1]$), a direct measure of spillover risk.

In all samples, when the runoff volume was negligible (the C sites; Figure 2a), spillover risk was high, with median $G^\star$ exceeding 0.675 and thus the 2/3 threshold, and was significantly higher in soil than in lake sediments for both GD and PACo (Dunn test, Benjamini-Hochberg [BH] correction, $P < 0.001$). However, in the presence of a low runoff volume (the L sites), spillover risk was higher in lake sediments than in soil for GD, but lower for PACo (Dunn test, BH correction, $P < 0.001$; Figure 2b). Finally, in the high runoff regime (the H sites), for both GD and PACo, spillover risk was higher in lake sediments than in soil (Dunn test, BH correction, $P < 0.001$; Figure 2c). Altogether, these results show that as runoff volume increases from almost non-existent to high, spillover risk shifted from higher in soil, to higher in lake sediments.

This pattern is consistent with the predictions of the Coevolution Effect hypothesis (Zohdy et al., 2019), and provides us with a mechanism explaining the observed increase in spillover risk with runoff. Lake Hazen was recently found to have undergone a dramatic change in sedimentation rates since 2007 compared to the previous 300 years: an increase in glacial runoff drives sediment delivery to the lake, leading to increased turbidity that perturbs anoxic bottom water known from the historical record (Lehnherr et al., 2018). Not only this, but turbidity also varies within the water column throughout the season (St. Pierre et al., 2019), hence fragmenting the lake habitat every year, and more so since 2007. This fragmentation of the aquatic habitat creates conditions that are, under the Coevolution Effect, favorable to spillover. Fragmentation creates barriers to gene flow, that increases genetic drift within finite populations, accelerating the coevolution of viruses and of their hosts. This acceleration leads to viral diversification which, should it be combined with "bridge vectors" (such as mosquitoes in terrestrial systems), increases spillover risk (Zohdy et al., 2019). Lake sediments are environmental archives; over time, they can preserve genetic material from aquatic organisms but also, and probably to a lesser extent, genetic material from its drainage basin. The co-evolutionary signal detected in lake sediments reflects interactions that may have happened in the fragmented aquatic habitat but also elsewhere in the drainage basin. Regardless of where the interaction occurred, our results show that spillover risk increases with runoff, a proxy of climate warming (Figure 2).

Spillovers might already be happening. To go one step further and identify the viruses most at risk of spillover, we focused on the model predictions made by Random TaPas. Under the null model, the occurrence of each virus/host association is evenly distributed on their cophylogeny (when sub-cophylogenies are drawn randomly, from a

uniform law). Departures from an even distribution are measured by the residuals of the linear fit. Positive residuals indicate a more frequent association than expected, that is pairs of host/virus species that contribute the most to the cophylogenetic signal. On the other hand, negative residuals indicate a less frequent association than expected, and hence pairs of host/virus species that contribute little to the cophylogenetic signal, because they tend to create incongruent phylogenies, a signature of spillover risk.

For both soil and lake sediments, the magnitude of the largest residuals tended either to decrease (Soil; Figure 3a) or to stay the same (Sediment; Figure 3b). This means that with increasing runoff, the strength of the cophylogenetic signal may remain steady, or may even weaken. On the other hand, the magnitude of the most negative residuals either remained globally unchanged (Soil; Figure 3a), or tended to become more negative (Sediment; Figure 3b). This latter pattern indicates that as runoff increases, the strength of the cophylogenetic signal deteriorates, potentially implying a higher spillover risk in lake sediments.

With this, Random TaPas can identify the viruses driving the spillover signal. Our results show that certain nucleocytoplasmic large DNA viruses (NCLDVs), such as the cowpox, horsepox, monkeypox, and vaccinia viruses, infect more than one species when the runoff is high (H sites), including humans (Figure 4). This indicates that spillovers are already occurring with these viral species. Among those, giant viruses (*e.g.*, *Megavirus chiliensis* and *Acanthamoeba polyphaga mimivirus*) are known to infect a wide range of eukaryotes from protists to animals (Schulz et al., 2020), which could impact the eukaryotic communities in the High Arctic (Figure 4, S8). As plant viruses represented $\sim 44\%$ of viruses infecting more than one host, this indicates that plants could also be affected.

To our knowledge, this is the first attempt to assess the complete virosphere of both DNA and RNA viruses, and their spillover capacity in any given environments, leading us

to show that increased glacier runoff, a direct consequence of climate change, is expected to increase viral spillover risk. However, as this is the first study applying the Random TaPas algorithm, we do not have yet any comparators in order to gauge the efficacy of $G^\star$ in assessing spillover capacity, both qualitatively and quantitatively.

Altogether, we provided here a proof of concept for a novel and unbiased approach to assessing spillover risk. This is not the same as predicting spillovers or even pandemics, because as long as "bridge vectors" (Zohdy et al., 2019) are not present in the environment, the likelihood of dramatic events probably remains low. But as climate change leads to shifts in species ranges and distributions, new interactions can emerge (Wallingford et al., 2020), bringing in vectors that can mediate viral spillovers (Rocklöv & Dubrow, 2020). This twofold effect of climate change, both increasing spillover risk and leading to a northward shift in species ranges (Osland et al., 2021), could have dramatic effect in the High Arctic. Disentangling this risk from actual spillovers and pandemics will be a critical endeavor to pursue in parallel with surveillance activities, in order to mitigate the impact of future major outbreaks.

Materials and Methods

Data acquisition. Between May 10 and June 10, 2017, sediments and soil cores were collected from Lake Hazen (82°N, 71°W; Quttinirpaaq National Park, northern Ellesmere Island, Nunavut, Canada), the largest High Arctic lake by volume in the world (Lehnherr et al., 2018). Sampling took place as the lake was still completely covered in ice (Table S1), as detailed in Colby et al. Colby et al. (2020). The sediment accumulation at the bottom of the Lake is caused by both allochthonous and autochthonous processes. The former are characterized by meltwaters that flow between late June and the end of August, and

run from the outlet glaciers along the northwestern shoreline through poorly consolidated river valleys, while the latter refer to the sedimentation process within the lake.

To contrast soil and sediment sites, core samples were paired, whenever possible, between these two environments. Soil samples were taken at three locations (Figure S2; C-Soil, L-Soil, and H-Soil) in the dried streambeds of the tributaries, on the northern shore, upstream of the lake and its sediments. The corresponding paired lake sediment samples were also cored at three locations, separated into hydrological regimes by seasonal runoff volume: negligible, low, and high runoff (Figure S2; C-Sed, L-Sed, and H-Sed). Specifically, the C (for *Control*) sites were both far from the direct influence of glacial inflows, while L sites were at a variable distance from Blister Creek, a small glacial inflow, and the H sites were located adjacent to several larger glacial inflows (Abbé River and Snow Goose). The water depth at L-Sed and H-Sed was respectively 50 m and 21 m, and the overlying water depth for site C-Sed was 50 m.

Before sample collection, all equipment was sterilized with 10% bleach and 90% ethanol, and non-powdered latex gloves were worn to minimize contamination. Three cores of $\sim$ 30 cm length were sampled at each location, and the top 5 and 10 cm of each sediment and soil core, respectively, were then collected and homogenized for genetic analysis as detailed in Colby et al. Colby et al. (2020). DNA was extracted on each core using the DNeasy PowerSoil Pro Kit, and RNA with the RNeasy PowerSoil Total RNA Kit (MO BIO Laboratories Inc, Carlsbad, CA, USA), following the kit guidelines, except that the elution volume was 30 µL. DNA and RNA were thereby extracted three times per sampling site, and elution volumes were combined for a total volume of 90 µL instead of 100 µL.

To sequence both DNA and RNA, a total of 12 metagenomic libraries were prepared ($n = 6$ for DNA, $n = 6$ for RNA), two for each sampling site, and run on an Illumina

HiSeq 2500 platform (Illumina, San Diego, CA, USA) at Génome Québec, using Illumina's TruSeq LT adapters (forward: `AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC`, and backward: `AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT`) in a paired-end 125 bp configuration. Each library was replicated ($n = 2$ for DNA, $n = 3$ for RNA) for each sample.

Data preprocessing and taxonomic assignments. A first quality assessment of the raw sequencing data was made using FastQC v0.11.8 (Andrews et al., 2010). Trimmomatic v0.36 (Bolger et al., 2014) was then employed to trim adapters and low-quality reads and bases using the following parameters: `phred33, ILLUMINACLIP:adapters/TruSeq3-PE-2.fa:3:26:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, CROP:105, HEADCROP:15, AVGQUAL:20, MINLEN:36`. A second round of quality check was performed with FastQC to ensure that Illumina's adapter sequences and unpaired reads were properly removed. Reads assembly into contigs was done *de novo* with both SPAdes v3.13.1 (Prjibelski et al., 2020) and metaSPAdes v3.13.1 (Nurk et al., 2017) for DNA, and with Trinity v2.9.0 (Grabherr et al., 2011), rnaSPAdes v3.13.1 (Bushmanova et al., 2019), and metaSPAdes for RNA (See Supplementary text). In all cases, the pipelines were used with their default settings.

Once assembled, a high-level (superkingdom) taxonomic assignment was determined based on BLASTn v2.10.0 (Zhang et al., 2000) searches. Those were performed at a stringent 10^{-19} E-value threshold against the partially non-redundant nucleotide (nr/nt) database from NCBI v5 (NCBI, 1988) (`ftp.ncbi.nlm.nih.gov/blast/db/nt*tar.gz`; downloaded on June 17, 2020). The proportions of taxonomic annotations ("Archaea," "Bacteria," "Eukaryota," or "Viruses") were calculated, and a 95% consensus was taken to assign a superkingdom rank for each contig. When no such 95% consensus could be determined, the contigs were classified as "Other."

To refine the taxonomic assignment of "viruses," GenBank's viral nucleotide sequences v238.0 (Sayers et al., 2019) were retrieved (`ftp.ncbi.nlm.nih.gov/genbank/gbvrl*seq.gz`; downloaded on July 23, 2020), concatenated, converted into FASTA with `seqret` v6.6.0 (Rice et al., 2000), and used to create a local database for BLASTn alignments. For each sampling location, after combining the DNA and RNA contigs classified as viral in the previous step, BLASTn searches were again conducted at the same stringent 10^{-19} E-value threshold, and the accession numbers of all the High-scoring Segment Pairs (HSPs) were used to retrieve their corresponding taxonomy identifiers (IDs) and their full taxonomic lineages with the R package `taxonomizr` v0.5.3 (Sherrill-Mix, 2019). The viral contigs were also mapped with Bowtie2 v2.3.5.1 (Langmead & Salzberg, 2012) using default settings to compare BLASTn and Bowtie2 efficiencies in refining these taxonomic annotations. As searches were found to be more sensitive with BLASTn than with Bowtie2 (see Supplementary text), only BLASTn results are shown hereafter, as our goal was to find as many similar sequences as possible in more than one species to eventually infer the virosphere from the virome.

Eukaryotic contigs were processed as above, based off the nr/nt database. Results were filtered by keeping a maximum of 12 HSPs whose E-value $< 10^{-100}$ *per* contig, for which lineages were obtained.

All samples were filtered to remove non-eukaryotic and uncultured hosts as well as viral and eukaryotic sequences with no taxonomy information. The ViralZone (Hulo et al., 2011) and International Committee on Taxonomy of Viruses (ICTV) (Lefkowitz et al., 2018) databases were consulted to obtain host range information on each found viruses. These taxonomic assignments were then used to retrieve their phylogenetic placements according to the Tree of Life (ToL) (`tolweb.org`), hence generating two trees: one for viruses and one for eukaryotes. For this, we used the `classification` and `class2tree`

functions from the R package `taxize` v0.9.99 (S. Chamberlain et al., 2020; S. A. Chamberlain & Szocs, 2013). In each environment, vertices of the viral and eukaryotic trees were then put in relation with each other according to the Virus-Host DB (downloaded on Sep. 29, 2020) (Mihara et al., 2016). These relations were saved in a binary association matrix (0: no infection; 1: infection), one for each environment. To simplify downstream computations without losing any information, only eukaryotic hosts associated to at least one virus were kept in the non-viral tree.

Spillover quantification. To quantify viral spillovers based on the viral and eukaryotic hosts identified, we employed the Random Tanglegram Partitions algorithm (Random TaPas) (Balbuena et al., 2020). This algorithm computes the cophylogenetic signal or congruence between two phylogenetic trees, the viral and the host trees. When congruence is large, or "perfect," the two trees are identical and hence, there is strong cophylogenetic signal – and absence of spillover. On the other hand, weak congruence is evidence for the existence of spillovers. Random TaPas quantifies congruence in two ways: a geodesic distance (GD) (Schardl et al., 2008), or a Procrustes distance (Procrustes Approach to Cophylogeny: PACo) (Balbuena et al., 2013), the latter measuring the distance between two trees geometrically transformed to make them as identical as possible. To partially account for phylogenetic non-independence when measuring congruence, Random TaPas further implements a resampling scheme where $N = 10^4$ subtrees of about 20% of the total number of virus/hosts links are randomly selected. This selection is used to generate a distribution of the empirical frequency of each association, measured by either GD or PACo.

Each empirical frequency is then regressed against a uniform distribution, and the residuals are used in two ways: (i) to quantify co-speciation, which is inversely propor-

tional to spillover risk; and (ii) to identify those virus/host pairs that contributed the least to the cophylogenetic signal, *i.e.*, the most to spillover risk. This risk is finally quantified by the shape of the distribution of residuals (for GD or PACo), with a normalized Gini coefficient ($G^\star$) that takes its values between 0 (perfect congruence, no spillover) to 1 (maximal spillover risk), with a defined threshold of 2/3 indicating a "large" value of $G^\star$. To account for phylogenetic uncertainty, the process is repeated $n = 1,000$ times, each replicate being a random resolution of the multifurcating virus/host trees of life into a fully bifurcating tree.

Data availability. The raw data used in this study can be found at `www.ncbi.nlm.nih.gov/bioproject/556841` (DNA-Seq) and at `www.ncbi.nlm.nih.gov/bioproject/PRJNA746497/` (RNA-Seq). The code developed for this work is available from `github.com/sarisbro/data`.

Acknowledgements. We thank Frances Pick for helpful comments on an early version of this work. This study was financially supported by the Natural Sciences Research Council of Canada (A.J.P., S.A.B.) and by the University of Ottawa (A.L., G.A.C.).

Author contributions. S.A.B. and A.J.P. designed research; G.A.C. collected and processed the samples; A.L. performed all analyses and wrote the original draft; A.L. and S.A.B. wrote the manuscript with contributions and suggestions from G.A.C. and A.J.P.; and S.A.B. and A.J.P. supervised this study and acquired funding.

Competing interests

The authors declare no competing interests.

References

Aguirre de Cárcer, D., López-Bueno, A., Pearce, D. A., & Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Sci. Ad.*, *1*(5). doi: 10.1126/sciadv.1400127

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., & Wingett, S. (2010). *FastQC.* Retrieved from `https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`

Aris-Brosou, S., Parent, L., & Ibeh, N. (2019). Viral long-term evolutionary strategies favor stability over proliferation. *Viruses*, *11*(8). doi: 10.3390/v11080677

Balbuena, J. A., Míguez-Lozano, R., & Blasco-Costa, I. (2013). PACo: a novel procrustes application to cophylogenetic analysis. *PLoS ONE*, *8*(4), e61048. doi: 10.1371/journal.pone.0061048

Balbuena, J. A., Pérez-Escobar, Ó. A., Llopis-Belenguer, C., & Blasco-Costa, I. (2020). Random tanglegram partitions (Random TaPas): an alexandrian approach to the cophylogenetic Gordian knot. *Syst. Bio.*, *69*(6), 1212–1230. doi: 10.1093/sysbio/syaa033

Bellec, L., Clerissi, C., Edern, R., Foulon, E., Simon, N., Grimsley, N., & Desdevises, Y. (2014). Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. *BMC Evol. Biol.*, *14*(1), 59. doi: 10.1186/1471-2148-14-59

Bennett, A. J., Paskey, A. C., Kuhn, J. H., Bishop-Lilly, K. A., & Goldberg, T. L. (2020). Diversity, transmission, and cophylogeny of Ledanteviruses (*Rhabdoviridae*: *Ledantevirus*) and Nycteribiid bat flies parasitizing Angolan soft-furred fruit bats in Bundibugyo District, Uganda. *Microorganisms*, *8*(5), 750. doi: 10.3390/microorganisms8050750

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, *8*(9), 1–13. doi: 10.1093/gigascience/giz100

Carrasco-Hernandez, R., Jácome, R., López Vidal, Y., & Ponce de León, S. (2017). Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR J.*, *58*(3), 343–358. doi: 10.1093/ilar/ilx026

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., … Grenié, M. (2020). *taxize: Taxonomic information from around the web.* Retrieved from `https://github.com/ropensci/taxize`

Chamberlain, S. A., & Szocs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*, *2*(191). doi: 10.12688/f1000research.2-191.v2

Ciota, A. T., & Keyel, A. C. (2019). The role of temperature in transmission of zoonotic arboviruses. *Viruses*, *11*(11). doi: 10.3390/v11111013

Colby, G. A., Ruuskanen, M. O., St.Pierre, K. A., St.Louis, V. L., Poulain, A. J., & Aris-Brosou, S. (2020). Warming climate is reducing the diversity of dominant microbes in the largest High Arctic lake. *Front. Microbiol.*, *11*, 2316. doi: 10.3389/fmicb.2020.561194

Dávila-Ramos, S., Castelán-Sánchez, H. G., Martínez-ávila, L., Sánchez-Carbente, M. D. R., Peralta, R., Hernández-Mendoza, A., … Batista-García, R. A. (2019). A review on viral metagenomics in extreme environments. *Front. Microbiol.*, *10*, 2403. doi: 10.3389/fmicb.2019.02403

Dawkins, R. (1989). *The selfish gene.* Oxford University Press.

Geoghegan, J. L., Duchêne, S., & Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.*, *13*(2), 1–17. doi: 10.1371/journal.ppat.1006215

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, *29*(7), 644–652. doi: 10.1038/nbt.1883

Grange, Z. L., Goldstein, T., Johnson, C. K., Anthony, S., Gilardi, K., Daszak, P., … Mazet, J. A. K. (2021). Ranking the risk of animal-to-human spillover for newly discovered viruses. *P. Natl. Acad. Sci.*, *118*(15). doi: 10.1073/pnas.2002324118

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., … Sullivan, M. B. (2019). Marine DNA viral macro- and microdiversity from pole to pole. *Cell*, *177*(5), 1109–1123.e14. doi: 10.1016/j.cell.2019.03.040

Hudson, J. M. G., & Henry, G. H. R. (2009). Increased plant biomass in a High Arctic heath community from 1981 to 2008. *Ecology*, *90*(10), 2657–2663. doi: 10.1890/09-0102.1

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., & Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.*, *39*(Database issue), D576–D582. doi: 10.1093/nar/gkq901

Jackson, A. P., & Charleston, M. A. (2004). A cophylogenetic perspective of RNA–virus evolution. *Mol. Biol. Evol.*, *21*(1), 45–57. doi: 10.1093/molbev/msg232

Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology*, *479-480*, 2–25. doi: 10.1016/j.virol.2015.02.039

Kozieł, E., Otulak-Kozieł, K., & Bujarski, J. J. (2021). Plant cell wall as a key player during resistant and susceptible plant-virus interactions. *Front. Microbiol.*, *12*, 495. doi: 10.3389/fmicb.2021.656809

Labbé, M., Girard, C., Vincent, W. F., & Culley, A. I. (2020). Extreme viral partitioning in a marine-derived High Arctic lake. *mSphere*, *5*(3), e00334-20. doi: 10.1128/mSphere.00334-20

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, *9*, 357–359. doi: 10.1038/nmeth.1923

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.*, *46*(D1), D708–D717. doi: 10.1093/nar/gkx932

Lehnherr, I., St. Louis, V. L., Sharp, M., Gardner, A. S., Smol, J. P., Schiff, S. L., ... Talbot, C. H. (2018). The world's largest High Arctic lake responds rapidly to climate warming. *Nat. Commun.*, *9*(1), 1290. doi: 10.1038/s41467-018-03685-z

Long, J. S., Mistry, B., Haslam, S. M., & Barclay, W. S. (2019). Host and viral determinants of Influenza A virus species specificity. *Nat. Rev. Microbiol.*, *17*(2), 67–81. doi: 10.1038/s41579-018-0115-z

Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J., & Jiggins, F. M. (2014). The evolution and genetics of virus host shifts. *PLoS Pathog.*, *10*(11), 1–8. doi: 10.1371/journal.ppat.1004395

Longdon, B., Day, J. P., Alves, J. M., Smith, S. C. L., Houslay, T. M., McGonigle, J. E., ... Jiggins, F. M. (2018). Host shifts result in parallel genetic changes when viruses evolve in closely related species. *PLoS Pathog.*, *14*(4), 1–14. doi: 10.1371/journal.ppat.1006951

Madinda, N. F., Ehlers, B., Wertheim, J. O., Akoua-Koffi, C., Bergl, R. A., Boesch, C., ... Calvignac-Spencer, S. (2016). Assessing host-virus codivergence for close relatives of Merkel cell Polyomavirus infecting African great apes. *J. Virol.*, *90*(19), 8531–8541. doi: 10.1128/JVI.00247-16

Maginnis, M. S. (2018). Virus–receptor interactions: the key to cellular invasion. *J. Mol. Biol.*, *430*(17), 2590–2611. doi: 10.1016/j.jmb.2018.06.024

Malik, S. S., Azem-e Zahra, S., Kim, K. M., Caetano-Anollés, G., & Nasir, A. (2017). Do viruses exchange genes across superkingdoms of life? *Front. Microbiol.*, *8*, 2110. doi: 10.3389/fmicb.2017.02110

Martin, G., Yanez-Arenas, C., Chen, C., Plowright, R. K., Webb, R. J., & Skerratt, L. F. (2018). Climate change could increase the geographic extent of Hendra Virus spillover risk. *Ecohealth*, *15*(3), 509–525. doi: 10.1007/s10393-018-1322-9

Messan, K. S., Jones, R. M., Doherty, S. J., Foley, K., Douglas, T. A., & Barbato, R. A. (2020). The role of changing temperature in microbial metabolic processes during permafrost thaw. *PLoS ONE*, *15*(4), e0232169. doi: 10.1371/journal.pone.0232169

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., ... Ogata, H. (2016). Linking virus genomes with host taxonomy. *Viruses*, *8*(3), 66. doi: 10.3390/v8030066

Nasir, A., Forterre, P., Kim, K. M., & Caetano-Anollés, G. (2014). The distribution and impact of viral lineages in domains of life. *Front. Microbiol.*, *5*, 194. doi: 10.3389/fmicb.2014.00194

Nasir, A., Kim, K. M., & Caetano-Anollés, G. (2017). Long-term evolution of viruses: a Janus-faced balance. *BioEssays*, *39*(8), 1700026. doi: 10.1002/bies.201700026

Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] - [cited 2021 Jan 20]. Available from `https://www.ncbi.nlm.nih.gov/nucleotide/`. (n.d.).

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.*, *27*(5), 824-834. doi: 10.1101/gr.213959.116

Olival, K. J., Hosseini, P. R., Zambrana-Torrelio, C., Ross, N., Bogich, T. L., & Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature*, *546*(7660), 646–650. doi: 10.1038/nature22975

Osland, M. J., Stevens, P. W., Lamont, M. M., Brusca, R. C., Hart, K. M., Waddle, J. H., ... Seminoff, J. A. (2021). Tropicalization of temperate ecosystems in north america: The northward range expansion of tropical organisms in response to warming winter temperatures. *Glob. Change Biol.*, *27*(13), 3009–3034. doi: 10.1111/gcb.15563

Parkinson, A. J., Evengard, B., Semenza, J. C., Ogden, N., Børresen, M. L., Berner, J., ... Albihn, A. (2014). Climate change and infectious diseases in the Arctic: establishment of a circumpolar working group. *Int. J. Circumpol. Heal.*, *73*, 25163. doi: 10.3402/ijch.v73.25163

Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E.-C., Burke, D. S., Calisher, C. H., ... Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. R.*, *72*(3), 457–470. doi: 10.1128/MMBR.00004-08

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using spades de novo assembler. *Curr. Protoc. Bioinformatics*. doi: 10.1002/cpbi.102

Rewar, S., & Mirdha, D. (2014). Transmission of Ebola virus disease: an overview. *Ann. Glob. Health*, *80*(6), 444–451. doi: 10.1016/j.aogh.2015.02.005

Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: The european molecular biology open software suite. *Trends Genet.*, *16*(6), 276–277. doi: 10.1016/S0168-9525(00)02024-2

Rocklöv, J., & Dubrow, R. (2020). Climate change: an enduring challenge for vector-borne disease prevention and control. *Nat. Immunol.*, *21*(5), 479–483. doi: 10.1038/s41590-020-0648-y

Ruuskanen, M. O., Colby, G., St.Pierre, K. A., St.Louis, V. L., Aris-Brosou, S., & Poulain, A. J. (2020). Microbial genomes retrieved from High Arctic lake sediments encode for adaptation to cold and oligotrophic environments. *Limnol. Oceanogr.*, *65*(S1), S233–S247. doi: 10.1002/lno.11334

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Res.*, *47*(D1), D94–D99. doi: 10.1093/nar/gky989

Schardl, C. L., Craven, K. D., Speakman, S., Stromberg, A., Lindstrom, A., & Yoshida, R. (2008). A novel test for host-symbiont codivergence indicates ancient origin of fungal endophytes in grasses. *Syst. Bio.*, *57*(3), 483–498. doi: 10.1080/10635150802172184

Schmidt, N., & Bölter, M. (2002). Fungal and bacterial biomass in tundra soils along an arctic transect from Taimyr Peninsula, central Siberia. *Polar Biol.*, *25*(12), 871–877. doi: 10.1007/s00300-002-0422-7

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., ... Woyke, T. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature*, *578*(7795), 432–436. doi: 10.1038/s41586-020-1957-x

Sherrill-Mix, S. (2019). *taxonomizr: functions to work with NCBI accessions and taxonomy.* Retrieved from `https://cran.r-project.org/package=taxonomizr`

Stavolone, L., & Lionetti, V. (2017). Extracellular matrix in plants and animals: Hooks and locks for viruses. *Front. Microbiol.*, *8*, 1760. doi: 10.3389/fmicb.2017.01760

St. Pierre, K. A., St. Louis, V. L., Lehnherr, I., Schiff, S. L., Muir, D. C. G., Poulain, A. J., ... Gardner, A. S. (2019). Contemporary limnology of the rapidly changing glacierized watershed of the world's largest High Arctic lake. *Sci. Rep.*, *9*(1), 4447. doi: 10.1038/s41598-019-39918-4

Waits, A., Emelyanova, A., Oksanen, A., Abass, K., & Rautio, A. (2018). Human infectious diseases and the changing climate in the Arctic. *Environ. Int.*, *121*, 703–713. doi: 10.1016/j.envint.2018.09.042

Wallingford, P. D., Morelli, T. L., Allen, J. M., Beaury, E. M., Blumenthal, D. M., Bradley, B. A., ... Sorte, C. J. B. (2020). Adjusting the lens of invasion biology to focus on the impacts of climate-driven range shifts. *Nat. Clim. Change*, *10*(5), 398–405. doi: 10.1038/s41558-020-0768-2

Wang, D. (2020). 5 challenges in understanding the role of the virome in health and disease. *PLoS Pathog.*, *16*(3), 1–6. doi: 10.1371/journal.ppat.1008318

Woolhouse, M. E. J., Haydon, D. T., & Antia, R. (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.*, *20*(5), 238–244. doi: 10.1016/j.tree.2005.02.009

Yau, S., & Seth-Pasricha, M. (2019). Viruses of polar aquatic environments. *Viruses*, *11*(2). doi: 10.3390/v11020189

Zhang, Y. Z., Shi, M., & Holmes, E. C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell*, *172*(6), 1168–1172. doi: 10.1016/j.cell.2018.02.043

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.*, *7*(1-2), 203–214. doi: 10.1089/10665270050081478

Zhou, P., & Shi, Z.-L. (2021). SARS-CoV-2 spillover events. *Science*, *371*(6525), 120–122. doi: 10.1126/science.abf6097

Zohdy, S., Schwartz, T. S., & Oaks, J. R. (2019). The Coevolution Effect as a driver of spillover. *Trends Parasitol.*, *35*(6), 399–408. doi: 10.1016/j.pt.2019.03.010

Tables

Table 1. Abundance of the viral families of all the viral HSPs. The host range information was obtained from the ViralZone and International Committee on Taxonomy of Viruses (ICTV) databases. Viruses with a missing family were excluded from this table.

Viral family	Known eukaryotic host range	Count by site					
		Control		Low runoff		High runoff	
		C-Soil	C	L-Soil	L1	H-Soil	H1
dsDNA viruses							
Alloherpesviridae	Fish						1
Ascoviridae	Insects: mainly Noctuids SfAV: *Spodoptera* species only						4
Asfarviridae	Pigs warthogs bushpigs Vector: Argasid ticks						2
Baculoviridae	Arthropods: Lepidoptora, Hymenoptera, Diptera Crustacean: Decapoda (Shrimps)		3				
Herpesviridae	Vertebrates			2	4	2	
Iridoviridae	Insects		11		13		159
Lavidaviridae	Protists infected by *Mimivirus*				1		8
Marseilleviridae	Amoeba						3
Mimiviridae	Amoeba	86	105	60	167	82	886
Papillomaviridae	Vertebrates						4
Phycodnaviridae	Alga	1	69	1	108	37	1,858
Pithoviridae	Amoeba	1	2		4		2
Poxviridae	Human, vertebrates, and arthropods		2		18	20	173
dsRNA viruses							
Chrysoviridae	Fungi	12		9		116	69
Endornaviridae	Plants, fungi, and oomycetes	33	20	115	21	4	212
Hypoviridae	Fungi			2			4
Megabirnaviridae	Fungi	1		3			
Partitiviridae	Fungi and plants	206	384	112	304	352	1,140
Picobirnaviridae	Vertebrates and invertebrates		3		3		36
Reoviridae	Vertebrates, invertebrates, plants, and fungi		1		4		106
Totiviridae	*Totivirus*: Fungi	4	44	45	124	19	957

	Victorivirus: Fungi						
-ssRNA viruses							
Fimoviridae	Plant: European mountain ash	6					
Mymonaviridae	*Sclerotinia sclerotiorum* fungi	1		3		6	
Phasmaviridae	Insects (mosquitos, cockroaches, water striders, psyllids, odonates, and drosophilids)	3					2
Phenuiviridae	RVFV: ruminants, camels, and humans Vector: Mosquitoes	3		8		7	
+ssRNA viruses							
Alphaflexiviridae	Plants and fungi	20			8		50
Barnaviridae	Cultivated mushroom (*Agaricus bisporus*)	1					
Betaflexiviridae	Plants and fungi	11		103	1,716		4,631
Botourmiaviridae	Plants and fungi	52		7		142	
Bromoviridae	Plants			2	4		8
Caliciviridae	Vertebrates		6		7	2	7
Closteroviridae	Plants						2
Deltaflexiviridae	Fungi and plants	2	30			27	
Dicistroviridae	Invertebrates				31		6
Iflaviridae	Insects	2	3		3		3
Luteoviridae	Plants		2				
Marnaviridae	Phytoplankton *Heterosigma akashiwo*		7		23		5
Mitoviridae	Fungi	66	4		8	26	37
Narnaviridae	Fungi	59	3	26		207	3
Nodaviridae	Vertebrates and invertebrates		3		4		
Picornaviridae	Vertebrates				8		9
Potyviridae	Plants				12		33
Solemoviridae	Plants (few species of Gramineae)	114		54			26
Tombusviridae	Plants		21		9	8	40
Virgaviridae	Plants					4	16
Total		684	723	552	2,604	1,061	10,502

Figures

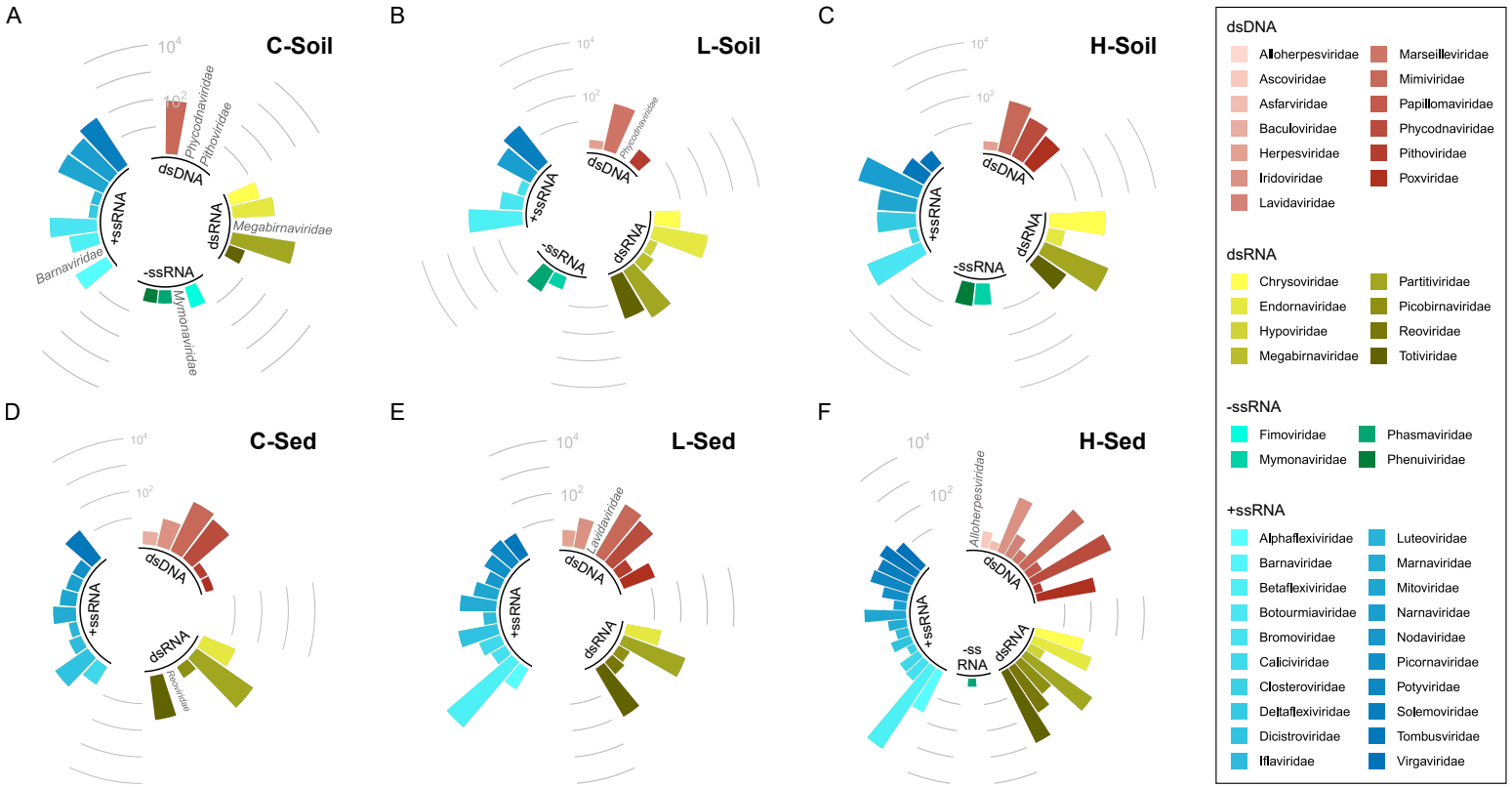

Figure 1. Abundance count of the viral families. (*A*) C-Soil; (*B*) L-Soil; (*C*) H-Soil; (*D*) C-Sed; (*E*) L-Sed; and (*F*) H-Sed sites. Abundances were $\log_{10}$-transformed. Viruses with a missing family were excluded from this analysis. The data used for this figure can be found in Table 1.

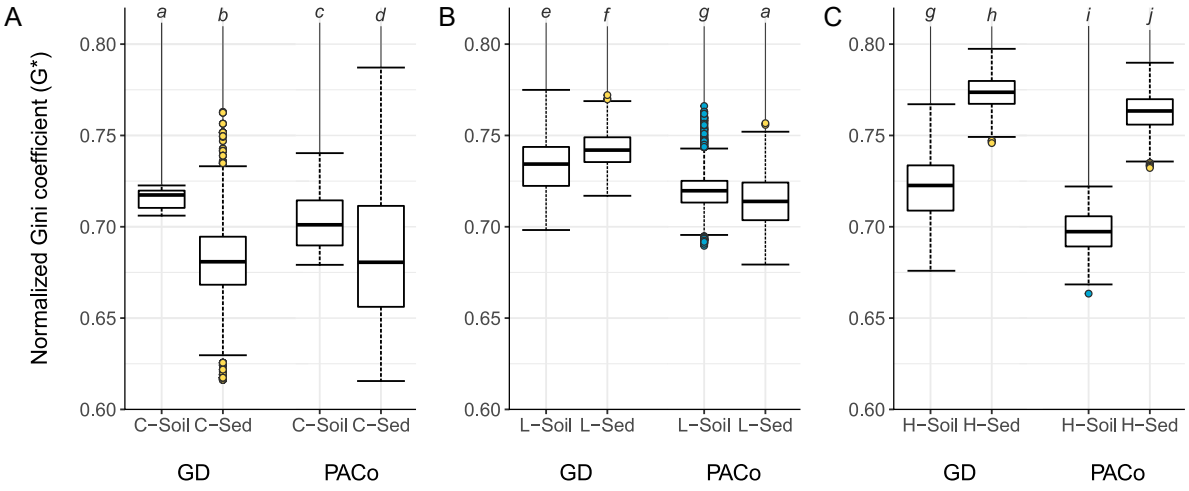

Figure 2. Normalized Gini coefficients ($G^{\star}$) obtained with Random TaPas ($n = 3$ runs). The values are separated by runoff volume: (A) control; (B) low runoff; and (C) high runoff. The two global-fit models used were GD (geodesic distances in tree space) and PACo (Procruses Approach to Cophylogeny). Significant results (Dunn test, BH correction) are marked with letters from a to j ($\alpha = 0.05$). Blue represents the soil and yellow, the lake sediments.

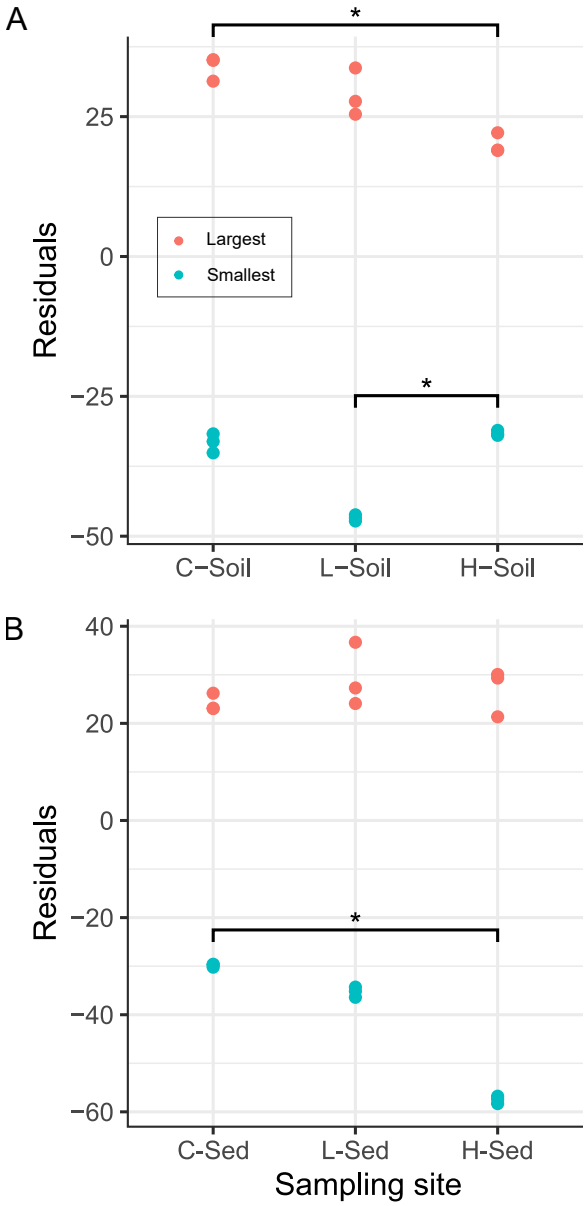

Figure 3. Largest and smallest residuals per sampling site for (*A*) soil and (*B*) lake sediments samples. Residuals were computed by Random TaPas ($n = 3$ runs) using GD (geodesic distances in tree space). Significant results (Dunn test, BH correction) are marked with an asterisk (*) ($\alpha = 0.05$). Red represents the largest and blue, the smallest residuals. Figure S6 further shows these results to be robust to the distance used to compare trees.

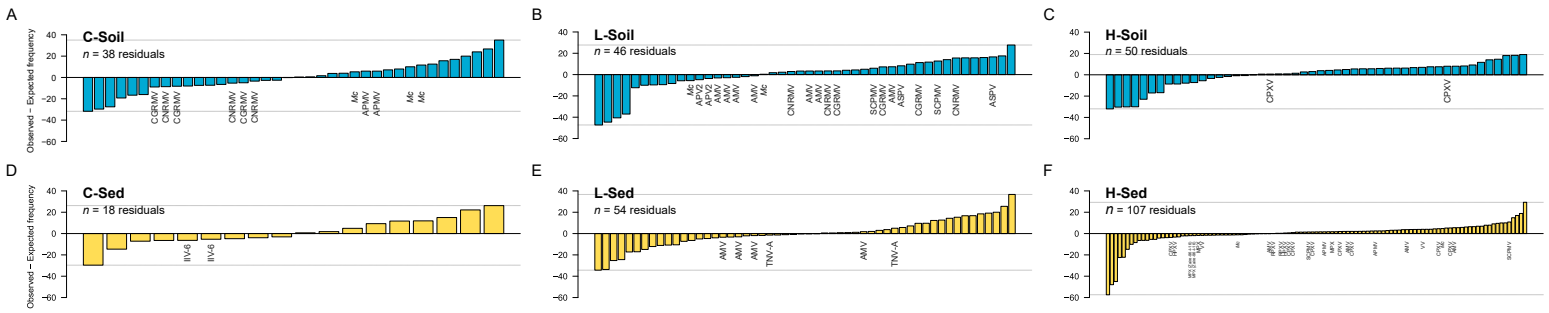

Figure 4. Distribution of the residuals computed by Random TaPas ($n = 1$ run) using GD (geodesic distances in tree space). (*A*) C-Soil; (*B*) L-Soil; (*C*) H-Soil; (*D*) C-Sed; (*E*) L-Sed; and (*F*) H-Sed sites. Viruses infecting > 1 host were annotated according to the species initials or the acronym: CGRMV (*Cherry green ring mottle virus*), CNRMV (*Cherry necrotic rusty mottle virus*), *Mc* (*Megavirus chiliensis*), APMV (*Acanthamoeba polyphaga mimivirus*), AMV (*Alfalfa mosaic virus*), ASPV (*Apple stem pitting virus*), APV2 (*Asian prunus virus 2*), SCPMV (*Southern cowpea mosaic virus*), CPXV (*Cowpox virus*), IIV6 (*Invertebrate iridescent virus 6*), TNV-A (*Tobacco necrosis virus A*), HPXV (*Horsepox virus*), VV (*Vaccinia virus*), RPXV (*Rabbitpox virus*), MPXV (*Monkeypox virus*), and MPXV (Zaire-96-I-16; *Monkeypox virus strain Zaire-96-I-16*). Blue residuals represent the soil, and yellow the lake sediments. The full names of the virus/host associations can be found in Figure S7.