Viral spillover risk in High Arctic increases in a glacierised watershed

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, Canada

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

Author for correspondence:

Stéphane Aris-Brosou

e-mail: `sarisbro@uottawa.ca`

Keywords:

spillovers, host restrictions, cophylogeny, viral metagenomics, High Arctic

Abstract

While many viruses have a single natural host, host restriction can be incomplete, hereby leading to spillovers to other host species. 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. To address this issue, we resorted to an unbiased metagenomics approach, and focused on two environments, soil and lake sediments from Lake Hazen, the largest High Arctic freshwater lake in the world. We used DNA and RNA sequencing to reconstruct the lake's virosphere and its range of eukaryotic hosts, and estimated the spillover risk by measuring the congruence between the viral and the eukaryotic host phylogenetic trees. We 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 viral vectors and reservoirs northwards, the High Arctic could become fertile ground for emerging pandemics.

1 Introduction

Viruses are ubiquitous and are often described as the most abundant replicators on Earth [1–3]. In spite of having highly diverse genomes, viruses are not independent "organisms" or replicators [4], as they need to infect a host's cell in order to replicate. These virus/host relationships seem relatively stable within superkingdoms, and can hence be classified as *archaeal*, *bacterial* (also known as *bacteriophages*), and *eukaryotic* viruses [5–7]. However, below this rank, viruses may infect a novel host from a reservoir host by being able to transmit sustainably in this new host, a process known as viral spillover [8, 9]. Indeed, in the past years, many viruses such as the Influenza A [10], Ebola [11], and SARS-CoV-2 [12] viruses spilled over to humans and caused significant diseases. 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. For instance, iridoviruses are known to infect both invertebrates and vertebrates [13], and *Picornavirales* are found in vertebrates, insects, plants, and protists [2]. Such host restrictions (or alternatively, spillover risks) are to date poorly defined and hence, difficult to assess without resorting to expert opinion [14].

Numerous factors can 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 [15–17]. The conservation of those receptors across multiple species allows these hosts to be more predisposed to becoming infected by the same virus [17, 18]. Indeed, from an evolutionary standpoint, viruses are more prone to infecting hosts that are phylogenetically close to their natural host [15, 19], potentially because it is easier for them to infect and colonize species that are genetically similar [20]. Alternatively, but not exclusively, high mutation rates might explain why RNA viruses spill over more often than other viruses [15], as

most lack proofreading mechanisms, making them more variable and likely to adapt to a new host [17].

While more studies are starting to characterize the communities and genomes of viruses in extreme environments [21–23], only few, 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 [24–27]. 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 [28, 29], as shown for arboviruses [30] and the Hendra virus [31]. 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 [32]. It has now become critical to quantify the risk of these spillovers. An intuitive approach to do this is to focus on the cophylogenetic relationships between viruses and their hosts [33–37]. 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, so it can be postulated that measuring phylogenetic congruency can be used to assess spillover risk.

To test this hypothesis in the context of a changing High Arctic environment, we resorted to a combination of metagenomics and of cophylogenetic modelling by sampling, in an unbiased manner, both the virosphere and its range of hosts [3], focusing on eukaryotes, which are critically affected by viral spillovers [38]. We contrasted two local environments, lake sediments and soil samples of 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 this specific lake [24, 25]. While microbial eukaryotes have been identified in Lake Hazen and other Arctic freshwater ecosystems [39–42], the

Arctic multicellular macro-eukaryotes have yet to be sufficiently characterized. We show here that the risk of spillovers increases with warming climate, but is likely to remain low in the absence of "bridge vectors" and reservoirs.

2 Methods

(a) Data acquisition

An overview of data acquisition and analytical pipeline is shown in figure S1. Between the 10th of May and the 10th of June, 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, and the largest freshwater ecosystem in the High Arctic [25]. Sampling took place as the lake was still completely covered in ice (table S1), as previously described [24]. The sediment accumulation at the bottom of the Lake is caused by both allochthonous and autochthonous processes. The former are characterised 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 sterilised with 10% bleach and 90% ethanol, and non-powdered latex gloves were worn to minimise 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. 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. Further details, such as DNA and RNA yields following extractions, can be found in Colby et al. [24].

(b) Data preprocessing and taxonomic assignments

A first quality assessment of the raw sequencing data was made using FastQC v0.11.8 [43]. Trimmomatic v0.36 [44] 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 [45] and metaSPAdes v3.13.1 [46] for DNA, and with Trinity v2.9.0 [47], rnaSPAdes v3.13.1 [48], and metaSPAdes for RNA. The choice of an assembly tool was based on (i) the number of contigs generated, (ii) the taxonomic annotations, (iii) the time of assembly, and (iv) the contig lengths (see electronic supplementary material). 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 [49] searches. Those were performed at a stringent 10^{-19} E-value threshold against the partially non-redundant nucleotide (nr/nt) database from NCBI v5 [50] (`ftp.ncbi.nlm.nih.gov/blast/db/nt*tar.gz`; downloaded on June 17, 2020). We chose this threshold to increase the significance of our hits, as our preliminary results showed less ambiguity with smaller E-values, starting at a 10^{-19} cut-off. 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 [51] were retrieved (`ftp.ncbi.nlm.nih.gov/genbank/gbvrl*seq.gz`; downloaded on 23$^{\text{rd}}$ of July, 2020), concatenated, converted into FASTA with `seqret` v6.6.0 [52], 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 [53]. The viral contigs were also mapped with Bowtie2 v2.3.5.1 [54], 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 electronic supplementary material), 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. To increase specificity considering that > 100 hits were found *per* contig, 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 [55] and International Committee on Taxonomy of Viruses (ICTV) [56] databases were consulted to obtain host range information on each viral family. 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 known viruses and one for known eukaryotes. For this, we used the `classification` and `class2tree` functions from the R package `taxize` v0.9.99 [57, 58]. 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 the 29$^{\text{th}}$ of September, 2020) [59]. 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.

(c) Spillover quantification

To quantify viral spillovers based on the viral and eukaryotic hosts identified, we employed the Random Tanglegram Partitions algorithm (Random TaPas) [60]. This algorithm computes the cophylogenetic signal or congruence between two phylogenetic trees, the viral and the host trees, with the normalised Gini coefficient ($G^\star$). 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) [61], or a Procrustes distance (Procrustes Approach to Cophylogeny: PACo) [62], 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 proportional 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 $G^\star$ 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 $G^\star$ or large incongruence. 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.

3 Results and Discussion

(a) Plant and fungal viruses are overrepresented

Based on our most sensitive annotation pipeline (see electronic supplementary material), viruses represented less than 1% of all contigs, and our samples were dominated by bacteria, with low proportions of eukaryotes (proportions of bacterial and eukaryotic contigs being respectively $> 89.2\%$ and $< 6.4\%$, in 11 out of 12 samples) (see electronic supplementary material). These results could be due to our extraction process, which might have been biased towards microbial nucleic acids. For instance, an overrepresentation of bacteria was also found in a shotgun-metagenomics based study that also used soil extraction kits [63]. To assess the impact of this potential bias, the extraction process should be taken into consideration by future studies.

RNA viral contigs of all kinds (*i.e.*, dsRNA, +ssRNA, and -ssRNA viruses) were found to be significantly more abundant than DNA viral contigs in all samples, as 70.5% to 87.9% of viral families had a RNA genome (binomial tests, $P < 2.48 \times 10^{-7}$; figure 1, table 1). 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 [64], and eukaryotes are known to be the main targets of RNA viruses [2, 5–7].

Our results are however difficult to compare with previous studies in the High Arctic, as most were solely based on DNA metagenomics sequencing [22, 65, 66], probably because RNA viruses are thought to be unstable [23], or due to inadequate sampling strategies to extract RNA viruses [67]. Two studies have been able to recover RNA viruses but one had not intended to characterise the RNA viral community, rather randomly finding sequences related to ssRNA viruses [68], and while the other also identified RNA and DNA viruses from RNA-seq, the environments were slightly different: although they included a

freshwater lake, more abundant in ssDNA phages, the Baltic Sea contains varying levels of salinity [69] unlike Laze Hazen. Nonetheless, our results and those of this previous study [69] both show that it is possible to recover RNA viruses from RNA-seq metagenomics.

All viral genomes confounded, in all samples, known plants and/or fungi viral families were overrepresented compared to those infecting animals and protists, as proportions of the former ranged between 69.8% to 87.1% (binomial tests, $P < 2.48 \times 10^{-7}$; table 1). This overrepresentation might reflect a preservation bias, due to the constitutive defences found in plants and fungi offered by their waxy epidermal cuticles and cell walls [70], even if most plant viruses lack a protective lipoprotein envelope as found in animal viruses [71]. 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 [15, 72]. Second, plant biomass has been increasing over the past two decades in the High Arctic due to regional warming [73], and is likely to keep doing as warming continues.

(b) 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, which computes the congruence between the virus and the eukaryotic host trees, with large and weak congruent topologies indicating low and high spillover risk, respectively. The stability of its results was assessed by running this algorithm three times, and by combining the results for the normalised Gini coefficients ($G^\star \in [0, 1]$), a direct measure of spillover risk (see Methods).

When the runoff volume was negligible (the C sites; figure 2*a*), spillover risk's median $G^\star$ ranged between 0.675 and 0.725, thus exceeding 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, with $G^\star \in [0.70, 0.75]$ (Dunn test, BH correction, $P < 0.001$; figure 2*b*). Finally, in the high runoff regime (the H sites), for both GD and PACo, spillover risk was higher in lake sediments than in soil, with values of $G^\star > 0.75$ (Dunn test, BH correction, $P < 0.001$; figure 2*c*). Altogether, these results show that as runoff volume increases from almost non-existent to high, spillover risk increased with runoff, and shifted from higher in soil, to higher in lake sediments.

This pattern is consistent with the predictions of the Coevolution Effect hypothesis [74], 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 [25]. Not only this, but turbidity also varies within the water column throughout the season [75], 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, favourable 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) and/or invasive reservoir species, increases spillover risk [74]. 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 coevolutionary 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).

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 of Lake Hazen. 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. Additional studies including more runs of the algorithm and multiple environmental settings of the High Arctic would be necessary to further reinforce our results, and to calibrate the "true" risk of viral spillovers.

(c) 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 3*a*) or to stay the same (Sediment; figure 3*b*). 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 3*a*, 6*a*), or tended to become more negative (Sediment; figure 3*b*, 6*b*). 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. For both GD (figure 4) and PACo (figure S7), the 5 most negative residuals of each sample ($n = 60$) suggest that viruses are most likely to spill over in fungi ($n = 19$), plants ($n = 16$), and protists ($n = 15$; including different species of microalgaes), the other 10 species being mostly insects (animals: $n = 8$; oomycetes: $n = 2$).

Altogether, we provided here 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" and/or invasive reservoir species [74] 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 [76], bringing in vectors that can mediate viral spillovers [77]. This twofold effect of climate change, both increasing spillover risk and leading to a northward shift in species ranges [78], could have dramatic effect in the High Arctic. Disentangling this risk from actual spillovers and pandemics will be a critical endeavour to pursue in parallel with surveillance activities, in order to mitigate the impact of spillovers on economy and health-related aspects of human life, or on other species [9].

Data accessibility. 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`.

Authors' 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. We declare we have no competing interests.

Funding. This work was supported by the Natural Sciences and Engineering Research Council of Canada and by the University of Ottawa.

Acknowledgements. We thank Frances Pick for her helpful comments on an early version of this work.

References

1. Dávila-Ramos, S., Castelán-Sánchez, H. G., Martínez-ávila, L., Sánchez-Carbente, M. D. R., Peralta, R., Hernández-Mendoza, A., Dobson, A. D., Gonzalez, R. A., Pastor, N. & Batista-García, R. A., 2019 A review on viral metagenomics in extreme environments. *Frontiers in Microbiology* **10**, 2403. (doi:10.3389/fmicb.2019.02403).

2. 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).

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

4. Dawkins, R., 1989 *The Selfish Gene*. Oxford University Press.

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

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

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

8. Alexander, K. A., Carlson, C. J., Lewis, B. L., Getz, W. M., Marathe, M. V., Eubank, S. G., Sanderson, C. E. & Blackburn, J. K., 2018 *The Ecology of Pathogen Spillover and Disease Emergence at the Human-Wildlife-Environment Interface*, pp. 267–298. Springer International Publishing. (doi:10.1007/978-3-319-92373-4_8).

9. Becker, D. J., Washburne, A. D., Faust, C. L., Pulliam, J. R. C., Mordecai, E. A., Lloyd-Smith, J. O. & Plowright, R. K., 2019 Dynamic and integrative approaches to understanding pathogen spillover. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190014. (doi:10.1098/rstb.2019.0014).

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

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

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

13. İnce, İ. A., Özcan, O., Ilter-Akulke, A. Z., Scully, E. D. & Özgen, A., 2018 Invertebrate iridoviruses: A glance over the last decade. *Viruses* **10**. (doi:10.3390/v10040161).

14. Grange, Z. L., Goldstein, T., Johnson, C. K., Anthony, S., Gilardi, K., Daszak, P., Olival, K. J., O'Rourke, T., Murray, S., Olson, S. H. *et al.*, 2021 Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proceedings of the National Academy of Sciences* **118**. (doi:10.1073/pnas.2002324118).

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

16. Maginnis, M. S., 2018 Virus–receptor interactions: the key to cellular invasion. *Journal of Molecular Biology* **430**, 2590–2611. (doi:10.1016/j.jmb.2018.06.024).

17. Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E.-C., Burke, D. S., Calisher, C. H., Laughlin, C. A., Saif, L. J. & Daszak, P., 2008 Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews : MMBR* **72**, 457–470. (doi:10.1128/MMBR.00004-08).

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

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

20. 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 Pathogens* **13**, 1–17. (doi:10.1371/journal.ppat.1006215).

21. Gil, J. F., Mesa, V., Estrada-Ortiz, N., Lopez-Obando, M., Gómez, A. & Plácido, J., 2021 Viruses in extreme environments, current overview, and biotechnological potential. *Viruses* **13**. (doi:10.3390/v13010081).

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

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

24. 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. *Frontiers in Microbiology* **11**, 2316. (doi: 10.3389/fmicb.2020.561194).

25. Lehnherr, I., St. Louis, V. L., Sharp, M., Gardner, A. S., Smol, J. P., Schiff, S. L., Muir, D. C. G., Mortimer, C. A., Michelutti, N., Tarnocai, C. *et al.*, 2018 The world's largest High Arctic lake responds rapidly to climate warming. *Nature Communications* **9**, 1290. (doi:10.1038/s41467-018-03685-z).

403 26. Ruuskanen, M. O., Colby, G., St.Pierre, K. A., St.Louis, V. L., Aris-Brosou, S.
404 & Poulain, A. J., 2020 Microbial genomes retrieved from High Arctic lake sedi-
405 ments encode for adaptation to cold and oligotrophic environments. *Limnology and*
406 *Oceanography* **65**, S233–S247. (doi:10.1002/lno.11334).

407 27. Meredith, M., Sommerkorn, M., S., C., Derksen, C., Ekaykin, A., A., H., Kofi-
408 nas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M. *et al.*, 2019 *Polar*
409 *Regions*, pp. 203–320.

410 28. Parkinson, A. J., Evengard, B., Semenza, J. C., Ogden, N., Børresen, M. L., Berner,
411 J., Brubaker, M., Sjöstedt, A., Evander, M., Hondula, D. M. *et al.*, 2014 Climate
412 change and infectious diseases in the Arctic: establishment of a circumpolar working
413 group. *International Journal of Circumpolar Health* **73**, 25163. (doi:10.3402/ijch.
414 v73.25163).

415 29. Waits, A., Emelyanova, A., Oksanen, A., Abass, K. & Rautio, A., 2018 Human in-
416 fectious diseases and the changing climate in the Arctic. *Environment International*
417 **121**, 703–713. (doi:10.1016/j.envint.2018.09.042).

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

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

423 32. Messan, K. S., Jones, R. M., Doherty, S. J., Foley, K., Douglas, T. A. & Barbato,
424 R. A., 2020 The role of changing temperature in microbial metabolic processes
425 during permafrost thaw. *PLoS ONE* **15**, e0232169. (doi:10.1371/journal.pone.
426 0232169).

427 33. Bellec, L., Clerissi, C., Edern, R., Foulon, E., Simon, N., Grimsley, N. & Desdevises,
428 Y., 2014 Cophylogenetic interactions between marine viruses and eukaryotic pico-
429 phytoplankton. *BMC Evolutionary Biology* **14**, 59. (doi:10.1186/1471-2148-14-59).

430 34. Bennett, A. J., Paskey, A. C., Kuhn, J. H., Bishop-Lilly, K. A. & Goldberg,
431 T. L., 2020 Diversity, transmission, and cophylogeny of Ledanteviruses (*Rhab-*
432 *doviridae*: *Ledantevirus*) and Nycteribiid bat flies parasitizing Angolan soft-
433 furred fruit bats in Bundibugyo District, Uganda. *Microorganisms* **8**, 750. (doi:
434 10.3390/microorganisms8050750).

435 35. Jackson, A. P. & Charleston, M. A., 2004 A cophylogenetic perspective of RNA–
436 virus evolution. *Molecular Biology and Evolution* **21**, 45–57. (doi:10.1093/molbev/
437 msg232).

36. Madinda, N. F., Ehlers, B., Wertheim, J. O., Akoua-Koffi, C., Bergl, R. A., Boesch, C., Akonkwa, D. B. M., Eckardt, W., Fruth, B., Gillespie, T. R. *et al.*, 2016 Assessing host-virus codivergence for close relatives of Merkel cell Polyomavirus infecting African great apes. *Journal of Virology* **90**, 8531–8541. (doi:10.1128/JVI.00247-16).

37. 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**, 646–650. (doi:10.1038/nature22975).

38. 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 Journal* **58**, 343–358. (doi:10.1093/ilar/ilx026).

39. Grossart, H.-P., Massana, R., McMahon, K. D. & Walsh, D. A., 2020 Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. *Limnology and Oceanography* **65**, S2–S20. (doi:10.1002/lno.11382).

40. Jacquemot, L., Kalenitchenko, D., Matthes, L. C., Vigneron, A., Mundy, C. J., Tremblay, J.-É. & Lovejoy, C., 2021 Protist communities along freshwater–marine transition zones in Hudson Bay (Canada). *Elementa: Science of the Anthropocene* **9**. (doi:10.1525/elementa.2021.00111).

41. Marquardt, M., Vader, A., Stübner, E. I., Reigstad, M., Gabrielsen, T. M. & Stams, A. J. M., 2016 Strong seasonality of marine microbial eukaryotes in a High-Arctic fjord (Isfjorden, in West Spitsbergen, Norway). *Applied and Environmental Microbiology* **82**, 1868–1880. (doi:10.1128/AEM.03208-15).

42. Azovsky, A. I., Tikhonenkov, D. V. & Mazei, Y. A., 2016 An estimation of the global diversity and distribution of the smallest eukaryotes: Biogeography of marine benthic heterotrophic flagellates. *Protist* **167**, 411–424. (doi:10.1016/j.protis.2016.07.001).

43. Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C. & Wingett, S., 2010. FastQC.

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

45. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A., 2020 Using SPAdes *de novo* assembler. *Current Protocols in Bioinformatics* (doi:10.1002/cpbi.102).

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

47. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652. (doi:10.1038/nbt.1883).

48. 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**, 1–13. (doi:10.1093/gigascience/giz100).

49. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W., 2000 A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**, 203–214. (doi: 10.1089/10665270050081478).

50. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] - [cited 2021 Jul 29]. Available from `https://www.ncbi.nlm.nih.gov/nucleotide/`.

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

52. Rice, P., Longden, L. & Bleasby, A., 2000 EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277. (doi:10.1016/S0168-9525(00)02024-2).

53. Sherrill-Mix, S., 2019. taxonomizr: functions to work with NCBI accessions and taxonomy.

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

55. 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 Research* **39**, D576–D582. (doi:10.1093/nar/gkq901).

56. 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 Research* **46**, D708–D717. (doi: 10.1093/nar/gkx932).

57. Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J. *et al.*, 2020. taxize: Taxonomic information from around the web.

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

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

60. 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. *Systematic Biology* **69**, 1212–1230. (doi: 10.1093/sysbio/syaa033).

61. 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. *Systematic Biology* **57**, 483–498. (doi: 10.1080/10635150802172184).

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

63. Cissell, E. & McCoy, S., 2021 Shotgun metagenomic sequencing reveals the full taxonomic, trophic, and functional diversity of a coral reef benthic cyanobacterial mat from Bonaire, Caribbean Netherlands. *Science of the Total Environment* **755**, 142719. (doi:10.1016/j.scitotenv.2020.142719).

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

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

66. Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C. *et al.*, 2019 Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123.e14. (doi:10.1016/j.cell.2019.03.040).

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

68. de Cárcer, D. A., López-Bueno, A., Pearce, D. A. & Alcamí, A., 2015 Biodiversity and distribution of polar freshwater DNA viruses. *Science Advances* **1**, e1400127. (doi:10.1126/sciadv.1400127).

69. Allen, L. Z., McCrow, J. P., Ininbergs, K., Dupont, C. L., Badger, J. H., Hoffman, J. M., Ekman, M., Allen, A. E., Bergman, B., Venter, J. C. *et al.*, 2017 The Baltic Sea virome: Diversity and transcriptional activity of DNA and RNA viruses. *mSystems* **2**, e00125–16. (doi:10.1128/mSystems.00125-16).

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

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

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

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

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

75. St. Pierre, K. A., St. Louis, V. L., Lehnherr, I., Schiff, S. L., Muir, D. C. G., Poulain, A. J., Smol, J. P., Talbot, C., Ma, M., Findlay, D. L. *et al.*, 2019 Contemporary limnology of the rapidly changing glacierized watershed of the world's largest High Arctic lake. *Scientific Reports* **9**, 4447. (doi:10.1038/s41598-019-39918-4).

76. Wallingford, P. D., Morelli, T. L., Allen, J. M., Beaury, E. M., Blumenthal, D. M., Bradley, B. A., Dukes, J. S., Early, R., Fusco, E. J., Goldberg, D. E. *et al.*, 2020 Adjusting the lens of invasion biology to focus on the impacts of climate-driven range shifts. *Nature Climate Change* **10**, 398–405. (doi:10.1038/s41558-020-0768-2).

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

78. Osland, M. J., Stevens, P. W., Lamont, M. M., Brusca, R. C., Hart, K. M., Waddle, J. H., Langtimm, C. A., Williams, C. M., Keim, B. D., Terando, A. J. *et al.*, 2021 Tropicalization of temperate ecosystems in north america: The northward range

573 expansion of tropical organisms in response to warming winter temperatures. *Global*
574 *Change Biology* **27**, 3009–3034. (doi:10.1111/gcb.15563).

Tables

Table 1. Abundance of the viral families of the viral HSPs. The host range information was obtained from the ViralZone and International Committee on Taxonomy of Viruses (ICTV) databases. Viruses with no or unknown family were excluded from this table.

Viral family	Known eukaryotic host range	Count of HSP 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

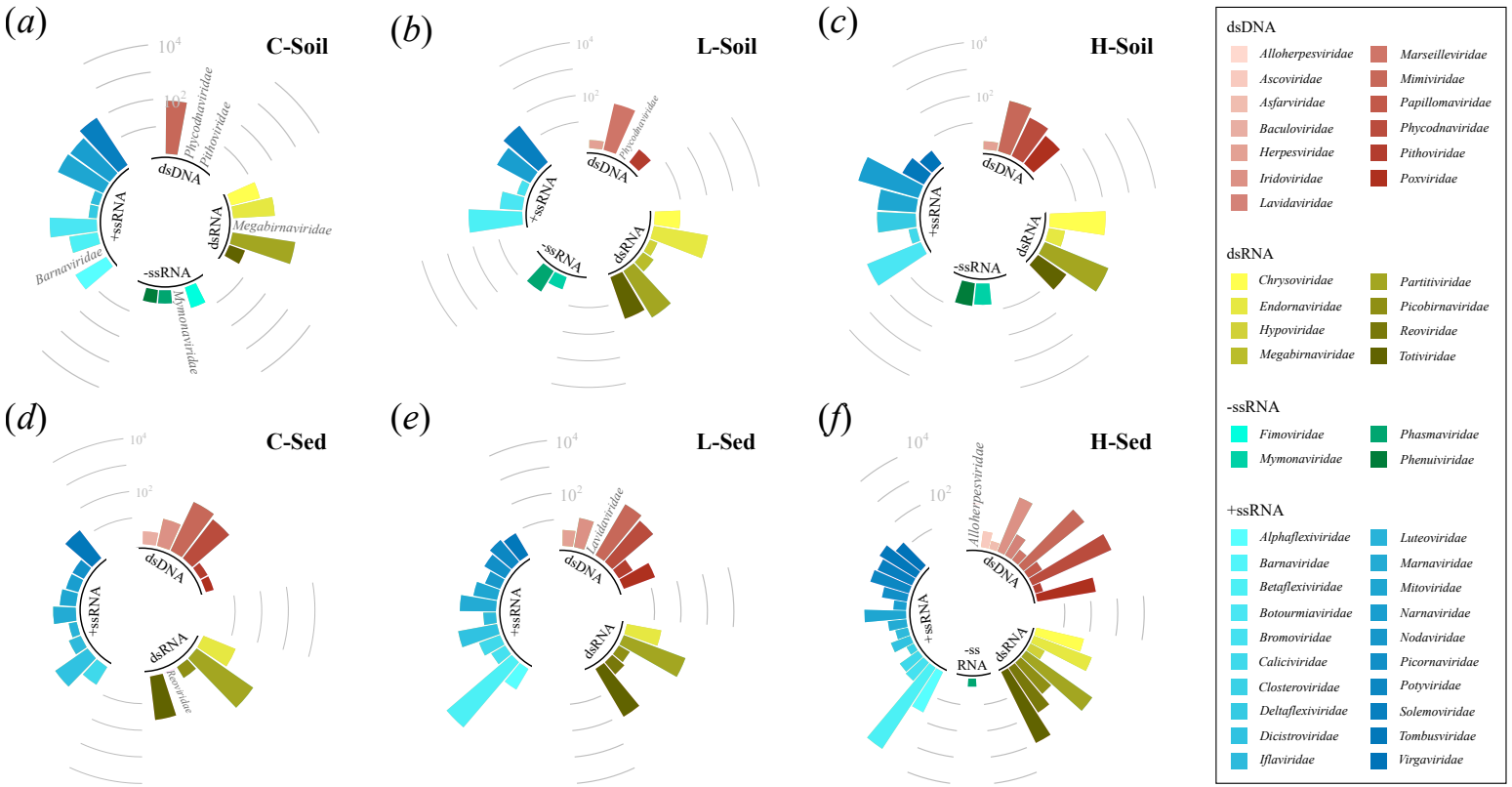

Figure 1. Abundance count of the viral families. (*a*) C-Soil; (*b*) L-Soil; (*c*) H-Soil; (*d*) C-Sed; (*e*) L-Sed; and (*f*) H-Sed sites. Abundances were $\log_{10}$-transformed. Viruses with a missing family were excluded from this analysis. The data used for this figure can be found in table 1.

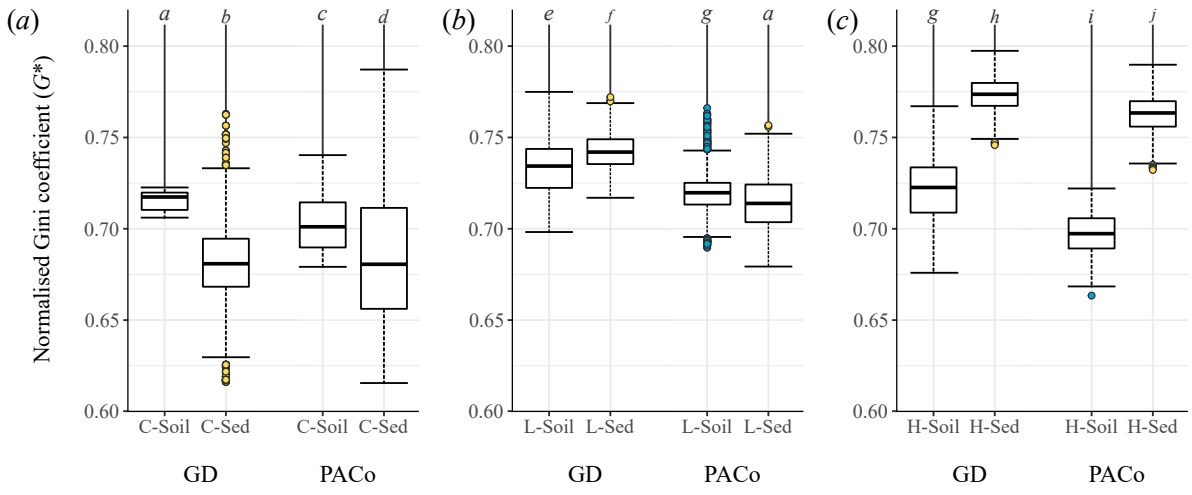

Figure 2. Normalised Gini coefficients ($G^\star$) obtained with Random TaPas ($n = 3$ runs). The values are separated by runoff volume: (a) control; (b) low runoff; and (c) high runoff. The two global-fit models used were GD (geodesic distances in tree space) and PACo (Procruses Approach to Cophylogeny). 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.

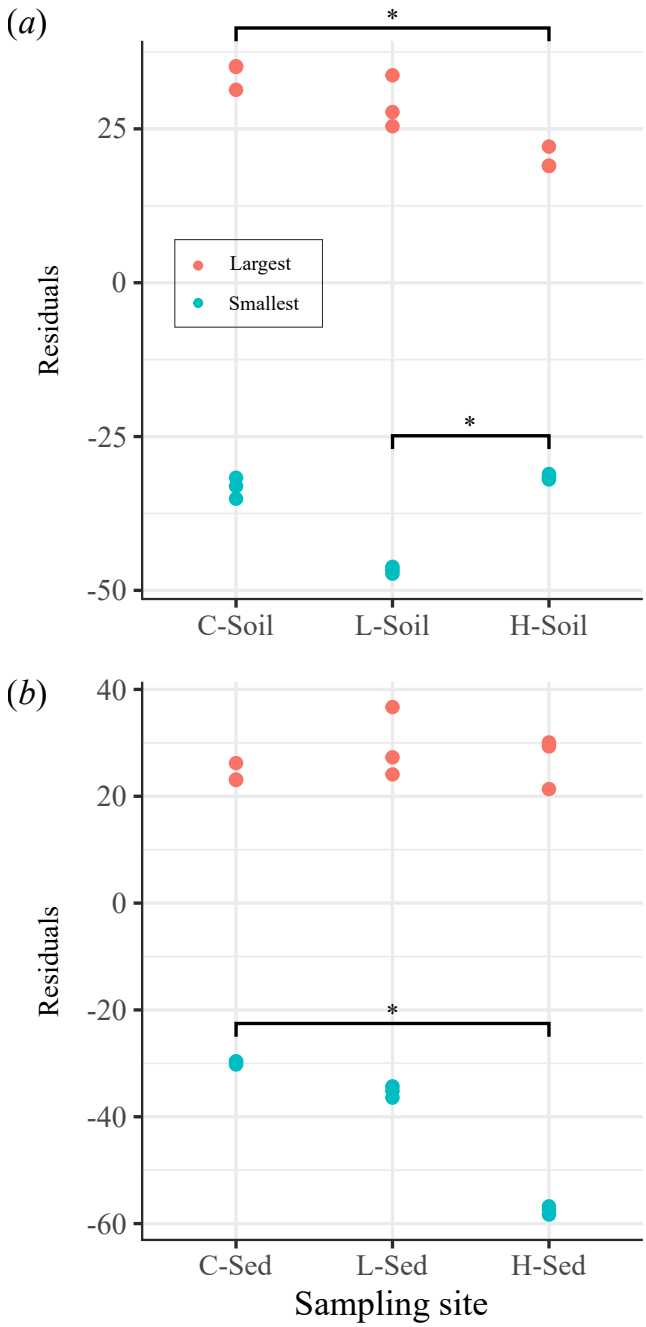

Figure 3. Largest and smallest residuals per sampling site for (a) soil and (b lake sediments samples. Residuals were computed by Random TaPas ($n = 3$ runs) using GD (geodesic distances in tree space). Significant results (Dunn test, BH correction) are marked with an asterisk (*) ($\alpha = 0.05$). 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.

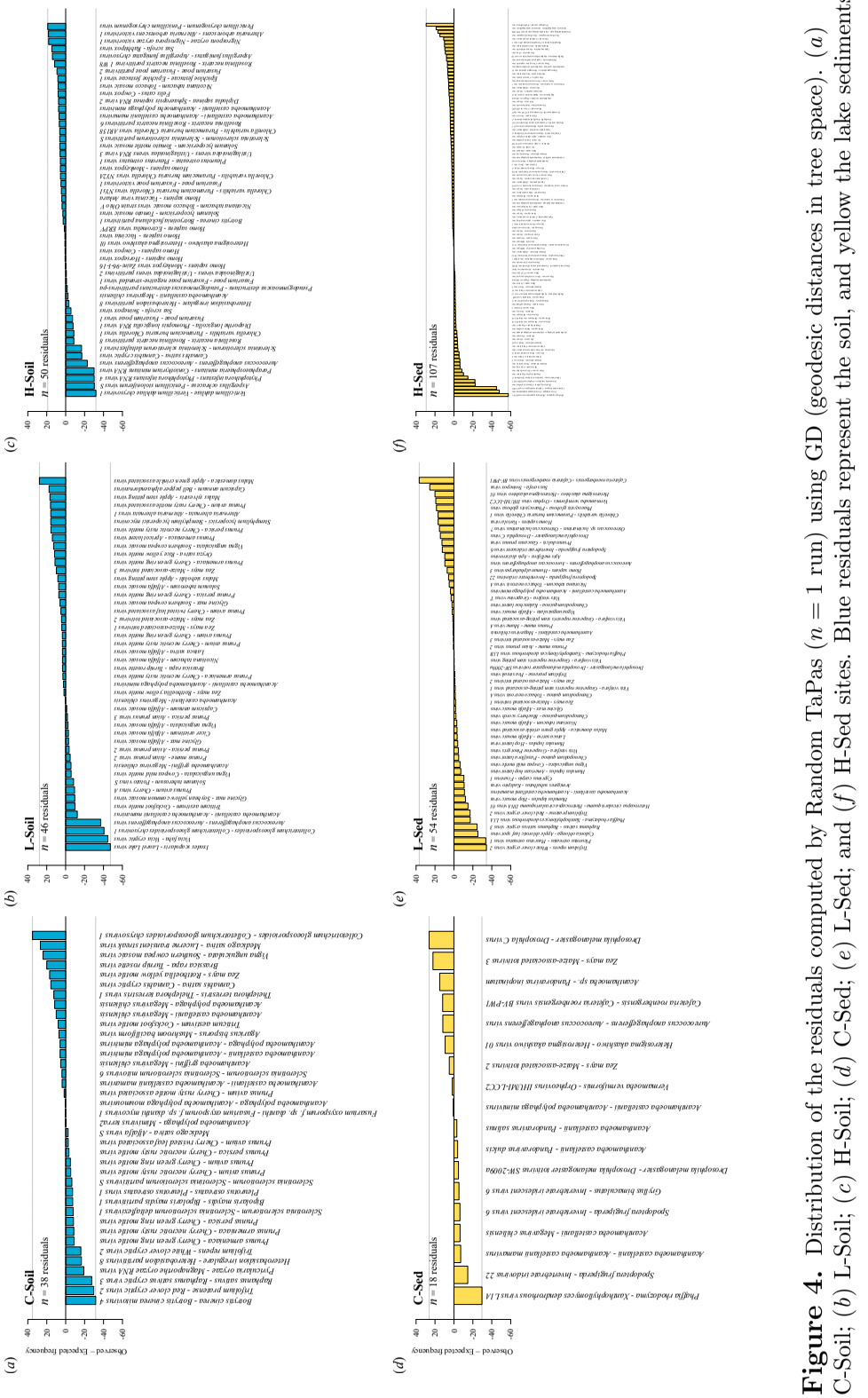

Figure 4. Distribution of the residuals computed by Random TaPas ($n = 1$ run) using GD (geodesic distances in tree space). (*a*) C-Soil; (*b*) L-Soil; (*c*) H-Soil; (*d*) C-Sed; (*e*) L-Sed; and (*f*) H-Sed sites. Blue residuals represent the soil, and yellow the lake sediments.