PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Lemieux A, Colby GA, Poulain AJ, Aris-Brosou S. 2022 Viral spillover risk increases with climate change in High Arctic lake sediments. *Proc. R. Soc. B* **289**: 20221073.
https://doi.org/10.1098/rspb.2022.1073

Received: 2 June 2022
Accepted: 27 September 2022

Subject Category:
Ecology

Subject Areas:
ecology, genomics

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

Author for correspondence:
Stéphane Aris-Brosou
e-mail: sarisbro@uottawa.ca

†Present addresses: Centre de Recherche du CHUM, Montréal, Quebec, Canada; Département de Microbiologie, Infectiologie et Immunologie, Université de Montréal, Montréal, Quebec, Canada.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6238499.

THE ROYAL SOCIETY PUBLISHING

Viral spillover risk increases with climate change in High Arctic lake sediments

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

[1]Department of Biology, and [2]Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada

AL, 0000-0002-5150-8010; SA-B, 0000-0003-4987-0296

The host spectrum of viruses is quite diverse, as they can sustainedly infect a few species to several phyla. When confronted with a new host, a virus may even infect it and transmit sustainably in this new host, a process called 'viral spillover'. However, the risk of such events is 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 a 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 in both its sediments and soils, as well as its range of eukaryotic hosts. We then estimated the spillover risk by measuring the congruence between the viral and the eukaryotic host phylogenetic trees, and show that spillover risk increases 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 replicating entities 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 that, following others [8,9], we define as 'viral spillover'. Indeed, in the past years, many viruses such as the Influenza A [10], Ebola [11], and SARS-CoV-2 [12] 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 a versatile host specificity is difficult to assess without resorting to an expert opinion [14], and gauging the probability that a virus may spill over from one host species to another, i.e. its spillover risk, is hard to quantify.

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 colonise 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]. RNA viruses are not only more likely to change host, but may do so in novel host species that have different ecological niches as well [21].

In this context of spillover, the High Arctic is of special interest as it is particularly affected by climate change, warming faster than the rest of the world [22–25]. Indeed, a warming climate and rapid transitions of the environment may both increase spillover risk by varying the global distribution and dynamics of viruses, as well as that of their reservoirs and vectors [26,27], as shown for arboviruses [28] and the Hendra virus [29]. 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 [30]. Thus, it has now become critical to be more proactive in preventing such events [31], but also to be able 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 [32–36]. 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 combining current knowledge about host range and phylogenetic incongruency can be used to quantify 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 both the virosphere and its range of hosts [3], focusing on eukaryotes, which are critically affected by viral spillovers [37]. 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 [22,23]. We show here that spillover risk increases with warming climate in lake sediments only, and suggest potential mechanisms explaining these differences.

2. Material and methods

(a) Data acquisition

An overview of data acquisition and analytical pipeline is shown in figure 1. Between 10 May and 10 June 2017, sediment 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 [23]. Sampling took place as the lake was still completely covered in ice (electronic supplementary material, table S1), as previously described [22]. 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 (electronic supplementary material, figure S1; C-Soil, L-Soil and H-Soil) in the seasonally dried riverbeds 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 (electronic supplementary material, figure S1; C-Sed, L-Sed and H-Sed). Specifically, the C (for *Control*) sites were both far from the direct influence of glacial inflows, while the L sites were at a variable distance from Blister Creek, a small glacial inflow. 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 and 21 m, and the overlying water depth for site C-Sed was 50 m.

Before sample collection, all equipment was sterilized with both 10% bleach and 90% ethanol, and non-powdered latex gloves were worn to minimize contamination. Three cores of approximately 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.

We resorted to shotgun metagenomics and metatranscriptomics to sequence the entire DNA and RNA content of each sample, and thus the genes of all the organisms present in our environments. To do so, 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: AGATCGGAAGAGCACACGTCT GAACTCCAGTCAC, and backward: AGATCGGAAGAGCGTC GTGTAGGGAAAGAGTGT) in a paired-end 125 bp configuration. Each library was replicated ($n = 2$ for DNA, $n = 3$ for RNA) for each sample. DNA and RNA yields following extractions can be, respectively, found in previous work [22] and in electronic supplementary material, table S2.

(b) Data preprocessing and taxonomic assignments

A first quality assessment of the raw sequencing data was made using FastQC v. 0.11.8 [38]. Trimmomatic v. 0.36 [39] 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 v. 3.13.1 [40] and metaSPAdes v. 3.13.1 [41] for DNA, and with Trinity v. 2.9.0 [42], rnaSPAdes v. 3.13.1 [43] and metaSPAdes for RNA. In all cases, the pipelines were used with their default settings. We chose metaSPAdes and rnaSPAdes for the DNA and RNA data, respectively, based on (i) the number of contigs generated, (ii) the taxonomic annotations, (iii) the time of assembly and (iv) the contig lengths (electronic supplementary material).

Once assembled, a high-level (superkingdom) taxonomic assignment was determined based on BLASTn v. 2.10.0 [44] searches. Those were performed at a stringent 10^{-19} *E*-value threshold against the partially non-redundant nucleotide (nr/nt) database from NCBI v. 5 [45] (ftp.ncbi.nlm.nih.gov/blast/db/nt*tar.gz; downloaded on 17 June 2020). This threshold

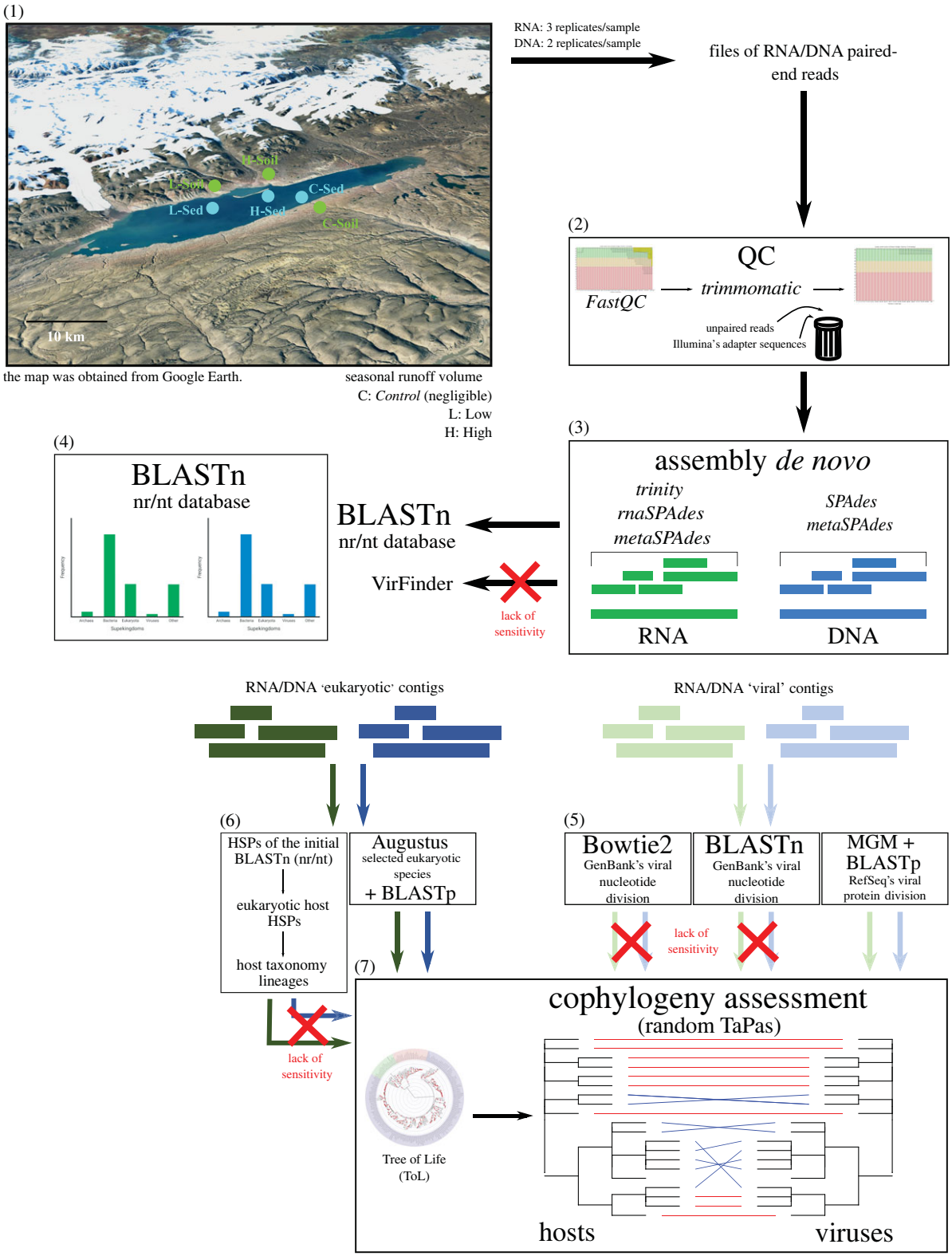

Figure 1. Overview of our analytical pipeline. The following steps are represented: (1) sampling; (2) quality control (QC); (3) assembly *de novo*; (4) contig classification; (5) refining of the viral taxonomic annotations; (6) determination of the eukaryotic hosts from the eukaryotic contigs; and (7) cophylogeny assessment. DNA and RNA are colour-coded in blue and green, respectively. (Online version in colour.)

was chosen 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 then 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.' We also tested VirFinder and its trained model for predicting prokaryotic and eukaryotic viruses, but did not use it for downstream analyses as BLASTn alignments were found to be more sensitive (electronic supplementary material).

To refine the viral taxonomic assignments, we further assessed three alternative approaches. First, GenBank's viral nucleotide sequences v238.0 [46] were retrieved (ftp.ncbi.nlm.nih.gov/genbank/gbvrl*seq.gz; downloaded on 23 July 2020), concatenated, converted to FASTA and formatted for BLAST with the `makeblastdb` command. BLASTn searches against this viral database were then conducted on the viral contigs, at

the same stringent 10^{-19} *E*-value threshold. Second, the same viral contigs were mapped against the same database, but with Bowtie2 v2.3.5.1 [47], using default settings. Third, MetaGeneMark v3.38 [48,49], which implements a hidden Markov model (HMM) used to predict genes of prokaryotic and eukaryotic viruses [50–52], was used to predict protein-coding regions in the viral contigs, followed by BLASTp v2.12.0 searches against RefSeq's v210 [53] viral protein database (ftp.ncbi.nlm.nih.gov/refseq/release/viral/*faa.gz; downloaded on 15 February 2022). As we found that the MGM + BLASTp approach was the most sensitive (a 4× to 15× increase; electronic supplementary material), this is what results from this point on are based on.

To refine the eukaryotic taxonomic assignments, we followed a similar approach. As GeneMark only implements five eukaryotic models, we employed Augustus v. 3.4.0 [54] to predict the eukaryotic protein-coding regions. This HMM implements 109 models available, including 105 eukaryotic species, spanning 76 genera. To avoid using similar models multiple times, we randomly chose one species per genus, attempting to cover the entire superkingdom. Data were combined with EVidenceModeler v1.1.1 [55], predictions were filtered at a *p*-value threshold of 0.05, and the eukaryotic model with the smallest *p*-value was then kept. In case of ties, the corresponding contigs were removed from the analyses. BLASTp searches were conducted as described above, on the following six RefSeq's protein divisions (downloaded on 15 February 2022): fungi, invertebrate, plant, protozoa, vertebrate ('mammalian') and vertebrate ('other').

For each sampling location, after combining the genes predicted from both the DNA and RNA contigs classified as 'viral' and 'eukaryotic,' the first 12 high-scoring segment pairs (HSPs) of each gene were kept. Those were then filtered with an *E*-value threshold of 10^{-50}. The accession numbers of these hits were used to retrieve their corresponding taxonomy identifiers and their full taxonomic lineages with the R packages `rentrez` v. 1.2.3 [56] and `taxonomizr` v. 0.5.3 [57]. Bacteriophages and virophages data were filtered out.

We retrieved the phylogenetic placements of all taxonomic assignments from the Tree of Life (ToL) (tolweb.org), hence generating two trees: one for the viruses and one for the eukaryotes that we identified, down to the species level. For this, we used the `classification` and `class2tree` functions from the R package `taxize` v0.9.99 [58,59]. At each site, vertices of the viral and eukaryotic trees were then put in relation with each other according to the Virus-Host DB (downloaded on 24 March 2022) [60]. These relations were saved as a binary association matrix (0: no infection; 1: infection), one for each site (see electronic supplementary material, figures S5, S9–S10). The β-diversity, or variation in species composition, between sites S_1 and S_2 was computed as $(n_{S_1} - n_c) + (n_{S_2} - n_c)$, where n_{S_i} is the total number of species in site S_i, and n_c, the number of species that the two sites have in common.

(c) Spillover quantification

To quantify viral spillovers based both on the viruses and eukaryotes identified and their known associations, we employed the Random Tanglegram Partitions algorithm (Random TaPas) [61]. This algorithm computes the cophylogenetic signal or congruence between two phylogenetic trees, the viral and the eukaryotic host trees computed as detailed above, with the normalized Gini coefficient ($G^\star$)—a quantitative proxy for spillover risk. Indeed, 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) [62], or a Procrustes distance (Procrustes Approach to Cophylogeny: PACo) [63], 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 unique virus/hosts associations 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 the normalized Gini coefficient $G^\star$ taking 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 $n = 1000$ times, each replicate being a random resolution of the multifurcating virus/host trees of life into a fully bifurcating tree. The stability of these results was assessed by running this algorithm $n = 10$ times, and by combining the results $G^\star$. Post-hoc comparisons were based on the Dunn test, controlling the False Discovery Rates with the Benjamini–Hochberg (BH) correction.

3. Results and discussion

(a) Plant and fungal viruses are overrepresented

Based on our most sensitive annotation pipeline (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, greater than or equal to 89.2% and less than or equal to 6.4%, in 11 out of 12 samples; electronic supplementary material). These results could be due to our extraction process, which might have been biased towards microbial nucleic acids, or simply to the fact that bacteria are the major members of soil and sediment samples [64].

Bacteriophages, eukaryotic viruses and even one virophage species were found among the viral HSPs (electronic supplementary material). Focusing on eukaryotic viruses, those were found to be unevenly distributed between RNA and DNA genomes, with the former (i.e. dsRNA +ssRNA and − ssRNA viruses) representing between 73.9% and 100% of all hits against eukaryotic viruses (table 1 and figure 2). This is not unexpected, as (i) fungal biomass for instance surpasses that of bacteria in Arctic environments by 1–2 orders of magnitude [69] and (ii) eukaryotes are known to be the main targets of RNA viruses [2,5–7].

All genomes and samples confounded, the majority of eukaryotic viruses was mainly found to be targeting plants and fungi, with proportions ranging from 62.1% to 92.4% (binomial tests, $p < 5.93 \times 10^{-3}$; 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]. It is also possible that, in an under-explored environment, virus/host associations not seen in better characterised environments exist, leading us

Table 1. Distribution of the viral families of the viral high-scoring segment pairs (HSPs). The host range information was obtained from the ViralZone [65], the International Committee on Taxonomy of Viruses (ICTV) [66] and the Virus-Host [60] databases. *Lavidaviridae* and *Picobirnaviridae* were neither bacteriophages nor eukaryotic viruses, as the former was shown to be virophages [67] and the latter was recently proposed to be both bacteriophages and eukaryotic viruses [68]. Viruses with no or unknown information at the family rank were classified as 'unassigned viruses'. Each viral species was kept once.

			count of HSPs by site					
			control		low runoff		high runoff	
viral family	genome	known eukaryotic host range	C-Soil	C-Sed	L-Soil	L-Sed	H-Soil	H-Sed
bacteriophages								
Ackermannviridae	dsDNA	—				1		2
Autographiviridae		—		12	1	7		34
Bicaudaviridae		—						1
Herelleviridae		—		2				2
Lipothrixviridae		—				2		
Myoviridae		—		111	12	158	5	154
Podoviridae		—		17	14	29	8	105
Schitoviridae		—						41
Siphoviridae		—	1	123	71	73	100	178
Zobellviridae		—		1		4		4
Microviridae	ssDNA	—	16	16	16	1	16	
Fiersviridae	ssRNA (+)	—	5		5			3
Leviviridae		—	5		7			5
Steitzviridae		—	1		1		1	
eukaryotic viruses								
Alloherpesviridae	dsDNA	fish, frogs						3
Ascoviridae		insects: mainly noctuids (owlet moths)						1
Baculoviridae		insects: Lepidoptera (butterflies and moths)			2			5
Herpesviridae		vertebrates				6		
Hytrosaviridae		insects: Diptera (flies)						1
Iridoviridae		insects, fish, amphibians, crustaceans		1		2		16
Marseilleviridae		amoebae						8
Mimiviridae		amoebae, protists		3	3	4		5
Nudiviridae		insects, marine crustaceans				1		1
Phycodnaviridae		algae		12		9		27
Pithoviridae		amoebae			1			
Poxviridae		humans, vertebrates, arthropods			7	1		22
Cruciviridae	ssDNA	*Solenopsis invicta* (red fire ant)						1
Chrysoviridae	dsRNA	fungi, plants	12		9		12	15
Endornaviridae		plants, fungi, oomycetes	14	23	19	17		28
Partitiviridae		fungi, plants	15	30	49	23	26	56

(*Continued*.)

Table 1. (*Continued.*)

			count of HSPs by site					
			control		low runoff		high runoff	
viral family	genome	known eukaryotic host range	C-Soil	C-Sed	L-Soil	L-Sed	H-Soil	H-Sed
Reoviridae		vertebrates, invertebrates, plants, fungi				1		3
Totiviridae		fungi, plants, insects		18	9	20	12	26
Aspiviridae	ssRNA (—)	plants			4			
Lispiviridae		arthropods					6	
Mymonaviridae		fungi	1				3	
Nyamiviridae		ticks, birds					2	
Phenuiviridae		mosquitoes, ruminants, camels, humans, plants			13		12	
Alphaflexiviridae	ssRNA (+)	plants, fungi			6		1	
Betaflexiviridae		plants			24	70		79
Botourmiaviridae		plants, fungi	2		2		2	
Bromoviridae		plants			1			
Dicistroviridae		arthropods		7		8		1
Hypoviridae		fungi			8			
Iflaviridae		arthropods	6	6		6		5
Marnaviridae		protists		13		15		11
Mayoviridae		plants						1
Mitoviridae		fungi, plants	12				22	7
Narnaviridae		fungi, protists					2	1
Nodaviridae		insects, fish				5		5
Picornaviridae		vertebrates		2		2		2
Solemoviridae		plants	14		10			
Tombusviridae		plants		1	5	8	11	6
Tymoviridae		plants					1	
Virgaviridae		plants						5
other								
Lavidaviridae	dsDNA	—				1		3
Picobirnaviridae	dsRNA	vertebrates, invertebrates, potentially bacteria				3		3
unassigned viruses	—	—	14	145	25	190	29	171
total			118	543	324	667	271	1047

to wrongly associate a virus to a given host, just because unknown associations cannot be learned from what is known. However, there is no reason to believe that such a knowledge gap would tip the bias towards plants and fungi hosts. 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 so as warming continues.

(b) Spillover risk increases with glacier runoff in lake sediments

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

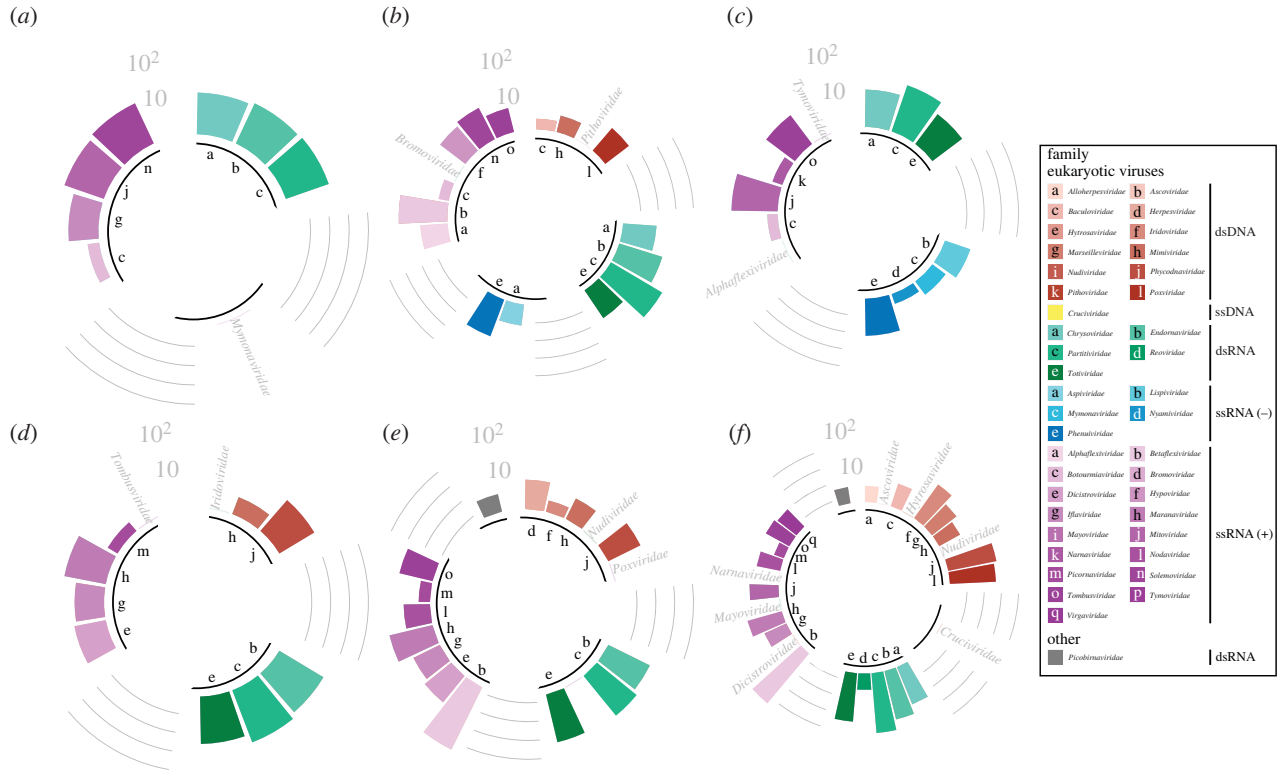

Figure 2. Distribution of the high-scoring segment pairs (HSPs) of the eukaryotic viruses, at the family rank. (*a*) C-Soil, (*b*) L-Soil, (*c*) H-Soil, (*d*) C-Sed, (*e*) L-Sed and (*f*) H-Sed sites. Each viral species was kept once. Viruses of the family *Picobirnaviridae* were neither bacteriophages nor eukaryotic viruses, as they were recently proposed to be both bacteriophages and eukaryotic viruses [68]. Counts were $\log_{10}$-transformed. The data used for this figure can be found in table 1, and electronic supplementary material, figure S7 depicts the distribution of all viral families, including bacteriophages. (Online version in colour.)

and the eukaryotic host trees, with large and weakly congruent topologies indicating low (small normalized Gini coefficient $G^\star$) and high (large $G^\star$) spillover risk, respectively.

Different patterns of spillover risks were observed in soil and in lake sediments, with both GD (figure 3) and PACo (electronic supplementary material, figure S8) exhibiting consistent results. In soil, when runoff volume is negligible (C-Soil; figure 3; electronic supplementary material, figure S8), spillover risk is low, as median $G^\star$ ranges between 0.60 and 0.61, thus below the two-thirds threshold. When runoff increases from low (L-Soil) to high (H-Soil), the median $G^\star$ increases to 0.75, but then decreases to 0.72 for GD (figure 3), with similar values for PACo (electronic supplementary material, figure S8). As such, our results show that as runoff volume increases from low to high, spillover risk remains high, but declines in soil samples (Dunn test, BH correction, $p \ll 0.001$). A possible explanation is that as glacial runoff increases, so does the erosion force of the glacier, which transports the content of the riverbed and riverbanks into the lake. This erosion would hence remove organisms from the topsoil of this environment, and hence curb the chance of interactions between viruses and hosts, that is, limit spillover risk.

Similarly, in lake sediments, spillover risk's median $G^\star \in [0.62, 0.67]$ for C-Sed (figure 3; electronic supplementary material, figure S8), again representing low to medium spillover risk. However, as runoff increases from low to high, $G^\star$ increases for both GD (L-Sed: 0.71, H-Sed: 0.75; figure 3) and PACo (L-Sed 0.75, H-Sed: 0.76; electronic supplementary material, figure S8), although the medians remains high for both sites. Altogether, and in contrast to soil samples, these results show that an increase in runoff volume in lake sediments significantly increases spillover risk (Dunn test, BH correction, $p \ll 0.001$).

This last 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 [23]. 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 (leading up to varves in undisturbed lakes) 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, as shown by an increasing β-diversity that, from the C to the H pairs of sites, goes from 21, 28 and 40 at the level of viral families (electronic supplementary material, figure S7). In turn, should this diversification be combined with 'bridge vectors' (such as mosquitoes in terrestrial systems) and/or invasive reservoir species, a corresponding increase in spillover risk could be expected [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,

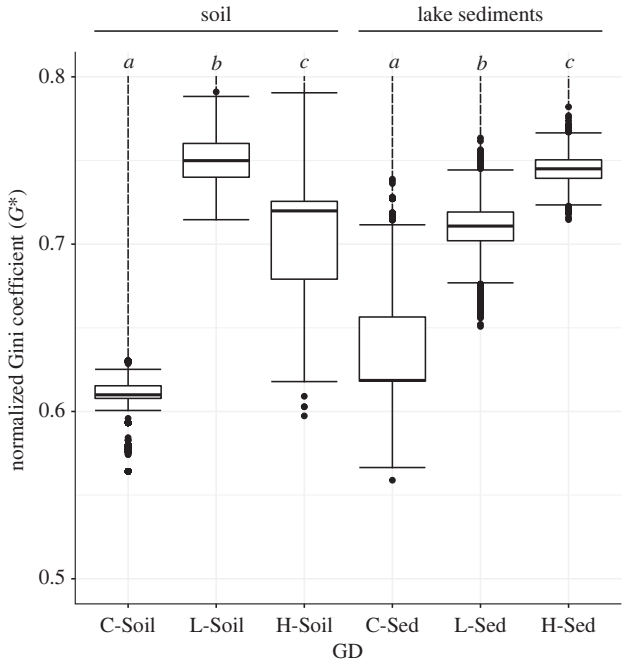

Figure 3. Normalized Gini coefficients ($G^\star$) obtained with Random TaPas ($n = 10$ runs). The values are separated by runoff volume: control (C sites), low (L sites) and high runoff (H sites). The global-fit model used was GD (geodesic distances in tree space). Significant results (Dunn test, BH correction) within each site ($n = 10 \times 1000_{\text{replicates}}$ for each site) are marked with letters from *a* to *c* ($\alpha = 0.01$). Outliers are shown as dots. Similar results were found with the PACo model (electronic supplementary material, figure S8). (Online version in colour.)

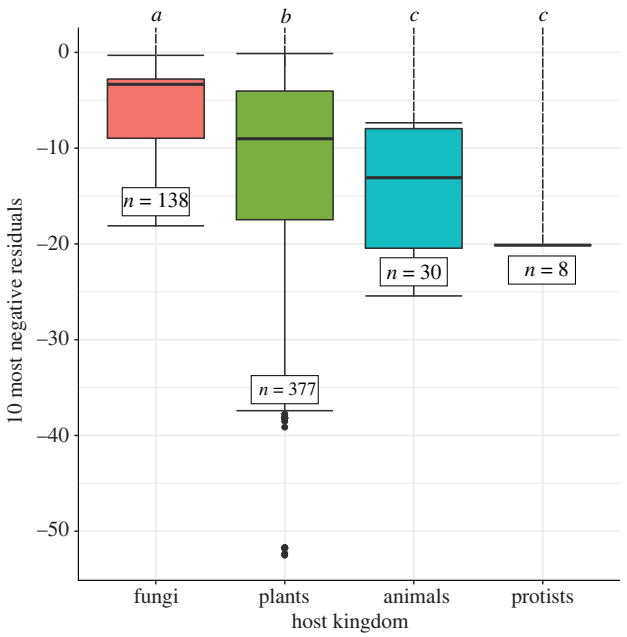

Figure 4. Eukaryotic hosts most susceptible to viral spillovers. The residuals were computed by Random TaPas ($n = 10$ runs) using GD (geodesic distances in tree space), all samples ($n = 6$) confounded, from which the 10 most negative virus/host associations were analysed (total sample size: $n = 10 \times 6 \times 10_{\text{MostNeg}} - \text{PosRes} = 553$). Significant results (Dunn test, BH correction) are marked with letters from *a* to *c* ($\alpha = 0.01$). Outliers are shown as dots. Electronic supplementary material, figure S11 further shows these results to be robust to the distance used to compare trees. (Online version in colour.)

in lake sediments (figure 3; electronic supplementary material, figure S8).

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 environment, leading us to show that increased glacier runoff, a direct consequence of climate change, is expected to increase viral spillover risk in the sediments of Lake Hazen. However, as this is the first study applying the Random TaPas algorithm, it is difficult to gauge how large $G^\star$ needs to be to reflect large spillover risks. Alternatives algorithms that estimate virus/host cophylogenies such as Coala [76] generally reconcile both trees by minimizing a number of events (e.g. codiversification, duplication, loss or host switch), and hence may not offer a direct measure of spillover risk. Furthermore, our study was limited by its number of replicates, as we only had five metagenomic libraries per sample ($n = 2$ and $n = 3$ for DNA and RNA, respectively). Additional sampling throughout the High Arctic would also be necessary to further reinforce our results, and to calibrate the 'true' risk of viral spillovers. We finally note that, because our approach relies on *known* virus/host associations, it is impossible to anticipate spillovers into novel hosts, as it was the case for viruses causing HIV [77], or through intermediate hosts as with SARS [78], and probably COVID-19 [79]. As a result, our quantification of spillover risk is likely to represent a lower bound of the actual risk.

(c) Spillovers might already be happening

To go one step further and identify the viruses and host kingdoms that are most at risk of spillovers, 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 virus/host 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 virus/host species that contribute little to the cophylogenetic signal, because they tend to create incongruent phylogenies, a signature of spillover risk (electronic supplementary material, figures S9 and S10).

When examining the virus/host associations of the ten most negative residuals of each Random TaPas run ($n = 10$), all samples confounded ($n = 6$), we found that animals and protists are the most susceptible to spillover, as they exhibit the lowest median residuals for both GD (−13.08 and −20.13; figure 4) and PACo (−14.37 and −24.52; electronic supplementary material, figure S11). On the other hand, plants and fungi showed a lower susceptibility to spillovers, as their median residuals were significantly higher (plants: −9.02 [GD] and −3.49 [PACo]; fungi: −3.33 [GD] and −3.95 [PACo]; all $p \ll 0.01$)—even though plant and fungal viruses were overrepresented in our samples (figure 4; electronic supplementary material, figure S11). Taking a more stringent threshold with the five most negative residuals led to similar results (electronic supplementary material, figure S12).

To further assess the robustness of this general finding, that animals and protists are most susceptible to viral spillovers in the High Arctic, we next examined the associations with the most negative residuals in each environment separately—as

we showed above that spillover risk increases in lake sediments only. We found that for both GD (electronic supplementary material, figure S13*a–c*) and PACo (electronic supplementary material, figure S13*b–d*), the most negative residuals were associated with animals and/or protists. Some of these hosts include the relatives of known disease vectors such as the yellow fever mosquito (*Aedes aegypti*) or the deer tick (*Ixodes scapularis*), and even pathogens such as *Pseudogymnoascus destructans*, which causes white-nose syndrome in bats, or *Fusarium poae*, involved in plant wilting (electronic supplementary material, tables S7 and S8)—hence suggesting that spillovers might already be happening.

Altogether, we provided here a novel approach to assessing spillover risk. This is not the same as predicting spillovers or even pandemics, first because we rely on *known* virus/host associations, and also because as long as viruses and their 'bridge vectors' are not simultaneously present in the environment [74], the likelihood of dramatic events probably remains low. But as climate change leads to shifts in species ranges and distributions, new associations can emerge [80], bringing in vectors that can mediate viral spillovers [81], as simulations recently highlight [82]. This twofold effect of climate change, both increasing spillover risk and leading to a northward shift in species ranges [83], 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 www.github.com/sarisbro/data/, in the archive Lemieux.etal.tar.bz2.

The data are provided in electronic supplementary material [84].

Authors' contributions. A.L.: formal analysis, investigation, methodology, software, visualization, writing—original draft; G.A.C.: data curation, methodology, resources, writing—review and editing; A.J.P.: funding acquisition, project administration, supervision, writing—review and editing; S.A.-B.: conceptualization, funding acquisition, methodology, project administration, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

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

Acknowledgements. We thank Frances Pick for her helpful comments on an early version of this work, as well as two reviewers who provided highly constructive comments.

References

1. Dávila-Ramos S *et al.* 2019 A review on viral metagenomics in extreme environments. *Front. Microbiol.* **10**, 2403. (doi:10.3389/fmicb.2019.02403)
2. Koonin EV, Dolja VV, 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 YZ, Shi M, Holmes EC. 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, UK: Oxford University Press.
5. Malik SS, Azem-e Zahra S, Kim KM, Caetano-Anollés G, Nasir A. 2017 Do viruses exchange genes across superkingdoms of life? *Front. Microbiol.* **8**, 2110. (doi:10.3389/fmicb.2017.02110)
6. Nasir A, Forterre P, Kim KM, 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)
7. Nasir A, Kim KM, Caetano-Anollés G. 2017 Long-term evolution of viruses: a Janus-faced balance. *BioEssays* **39**, 1700026. (doi:10.1002/bies.201700026)
8. Alexander KA, Carlson CJ, Lewis BL, Getz WM, Marathe MV, Eubank SG, Sanderson CE, Blackburn JK. 2018. The ecology of pathogen spillover and disease emergence at the human–wildlife–environment interface. In *The connections between ecology and infectious disease*, pp. 267–298. Berlin, Germany: Springer International Publishing.
9. Becker DJ, Washburne AD, Faust CL, Pulliam JRC, Mordecai EA, Lloyd-Smith JO, Plowright RK. 2019 Dynamic and integrative approaches to understanding pathogen spillover. *Phil. Trans. R. Soc. B* **374**, 20190014. (doi:10.1098/rstb.2019.0014)
10. Long JS, Mistry B, Haslam SM, Barclay WS. 2019 Host and viral determinants of Influenza A virus species specificity. *Nat. Rev. Microbiol.* **17**, 67–81. (doi:10.1038/s41579-018-0115-z)
11. Rewar S, Mirdha D. 2014 Transmission of Ebola virus disease: an overview. *Ann. 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 İA, Özcan O, Ilter-Akulke AZ, Scully ED, Özgen A. 2018 Invertebrate iridoviruses: a glance over the last decade. *Viruses* **10**, 161. (doi:10.3390/v10040161)
14. Grange ZL *et al.* 2021 Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proc. Natl Acad. Sci. USA* **118**, e2002324118. (doi:10.1073/pnas.2002324118)
15. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. 2014 The evolution and genetics of virus host shifts. *PLoS Pathog.* **10**, 1–8. (doi:10.1371/journal.ppat.1004395)
16. Maginnis MS. 2018 Virus–receptor interactions: the key to cellular invasion. *J. Mol. Biol.* **430**, 2590–2611. (doi:10.1016/j.jmb.2018.06.024)
17. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. 2008 Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* **72**, 457–470. (doi:10.1128/MMBR.00004-08)
18. Woolhouse MEJ, Haydon DT, Antia R. 2005 Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* **20**, 238–244. (doi:10.1016/j.tree.2005.02.009)
19. Longdon B, Day JP, Alves JM, Smith SCL, Houslay TM, McGonigle JE, Tagliaferri L, Jiggins FM. 2018 Host shifts result in parallel genetic changes when viruses evolve in closely related species. *PLoS Pathog.* **14**, 1–14. (doi:10.1371/journal.ppat.1006951)
20. Geoghegan JL, Duchêne S, Holmes EC. 2017 Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.* **13**, 1–17. (doi:10.1371/journal.ppat.1006215)
21. Wells K, Morand S, Wardeh M, Baylis M. 2020 Distinct spread of DNA and RNA viruses among mammals amid prominent role of domestic species. *Global Ecol. Biogeogr.* **29**, 470–481. (doi:10.1111/geb.13045)
22. Colby GA, Ruuskanen MO, Poulain AJ, 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)
23. Lehnherr I *et al.* 2018 The world's largest High Arctic lake responds rapidly to climate warming. *Nat. Commun.* **9**, 1290. (doi:10.1038/s41467-018-03685-z)
24. Ruuskanen MO, Colby G, Aris-Brosou S, Poulain AJ. 2020 Microbial genomes retrieved from High Arctic lake sediments encode for adaptation to cold and oligotrophic environments. *Limnol. Oceanogr.* **65**, S233–S247. (doi:10.1002/lno.11334)

25. Meredith M *et al.* 2019 Polar regions. In *IPCC special report on the ocean and cryosphere in a changing climate* (ed. H-O Pörtner *et al.*), pp. 203–320. Cambridge, UK: Cambridge University Press.
26. Parkinson AJ *et al.* 2014 Climate change and infectious diseases in the Arctic: establishment of a circumpolar working group. *Int. J. Circumpolar Health* **73**, 25163. (doi:10.3402/ijch.v73.25163)
27. 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)
28. Ciota AT, Keyel AC. 2019 The role of temperature in transmission of zoonotic arboviruses. *Viruses* **11**, 1013. (doi:10.3390/v11111013)
29. Martin G, Yanez-Arenas C, Chen C, Plowright RK, Webb RJ, Skerratt LF. 2018 Climate change could increase the geographic extent of Hendra Virus spillover risk. *Ecohealth* **15**, 509–525. (doi:10.1007/s10393-018-1322-9)
30. Messan KS, Jones RM, Doherty SJ, Foley K, Douglas TA, Barbato RA. 2020 The role of changing temperature in microbial metabolic processes during permafrost thaw. *PLoS ONE* **15**, e0232169. (doi:10.1371/journal.pone.0232169)
31. Bernstein AS *et al.* 2022 The costs and benefits of primary prevention of zoonotic pandemics. *Sci. Adv.* **8**, eabl4183. (doi:10.1126/sciadv.abl4183)
32. 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**, 59. (doi:10.1186/1471-2148-14-59)
33. Bennett AJ, Paskey AC, Kuhn JH, Bishop-Lilly KA, Goldberg TL. 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**, 750. (doi:10.3390/microorganisms8050750)
34. Jackson AP, Charleston MA. 2004 A cophylogenetic perspective of RNA–virus evolution. *Mol. Biol. Evol.* **21**, 45–57. (doi:10.1093/molbev/msg232)
35. Madinda NF *et al.* 2016 Assessing host-virus codivergence for close relatives of Merkel cell Polyomavirus infecting African great apes. *J. Virol.* **90**, 8531–8541. (doi:10.1128/JVI.00247-16)
36. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. 2017 Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650. (doi:10.1038/nature22975)
37. 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**, 343–358. (doi:10.1093/ilar/ilx026)
38. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. 2010 *FastQC: a quality control tool for high throughput sequence data*. Cambridge, UK: Babraham Institute.
39. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
40. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020 Using SPAdes de novo assembler. *Curr. Protocols Bioinform.* **70**, e102. (doi:10.1002/cpbi.102)
41. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017 metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834. (doi:10.1101/gr.213959.116)
42. Grabherr MG *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/nbt.1883)
43. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. 2019 rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience* **8**, 1–13. (doi:10.1093/gigascience/giz100)
44. Zhang Z, Schwartz S, Wagner L, Miller W. 2000 A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214. (doi:10.1089/10665270050081478)
45. National Center for Biotechnology Information. 1988 Nucleotide. 1988 See https://www.ncbi.nlm.nih.gov/nucleotide/ (cited 29 July 2021).
46. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2019 GenBank. *Nucleic Acids Res.* **47**, D94–D99. (doi:10.1093/nar/gky989)
47. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
48. Besemer J, Borodovsky M. 1999 Heuristic approach to deriving models for gene finding. *Nucleic Acids Res.* **27**, 3911–3920. (doi:10.1093/nar/27.19.3911)
49. Zhu W, Lomsadze A, Borodovsky M. 2010 Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132. (doi:10.1093/nar/gkq275)
50. Nishimura Y *et al.* 2017 Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* **2**, e00359-16. (doi:10.1128/mSphere.00359-16)
51. Laffy PW, Wood-Charlson EM, Turaev D, Weynberg KD, Botté ES, van Oppen MJH, Webster NS, Rattei T. 2016 HoloVir: a workflow for investigating the diversity and function of viruses in invertebrate holobionts. *Front. Microbiol.* **7**, 822. (doi:10.3389/fmicb.2016.00822)
52. Gupta A, Kumar S, Prasoodanan VPK, Harish K, Sharma AK, Sharma VK. 2016 Reconstruction of bacterial and viral genomes from multiple metagenomes. *Front. Microbiol.* **7**, 469. (doi:10.3389/fmicb.2016.00469)
53. O'Leary NA *et al.* 2015 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745. (doi:10.1093/nar/gkv1189)
54. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008 Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644. (doi:10.1093/bioinformatics/btn013)
55. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7. (doi:10.1186/gb-2008-9-1-r7)
56. Winter DJ. 2017 rentrez: an R package for the NCBI eUtils API. *R J.* **9**, 520–526. (doi:10.32614/RJ-2017-058)
57. Sherrill-Mix S. 2019 taxonomizr: functions to work with NCBI accessions and taxonomy. See https://CRAN.R-project.org/package=taxonomizr.
58. Chamberlain S *et al.* 2020 taxize: Taxonomic information from around the web. See https://github.com/ropensci/taxize.
59. Chamberlain SA, Szocs E. 2013 taxize: taxonomic search and retrieval in R. *F1000Research* **2**, 191. (doi:10.12688/f1000research.2-191.v2)
60. 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)
61. Balbuena JA, Blasco-Costa I. 2020 Random tanglegram partitions (Random TaPas): an alexandrian approach to the cophylogenetic Gordian knot. *Syst. Biol.* **69**, 1212–1230. (doi:10.1093/sysbio/syaa033)
62. Schardl CL, Craven KD, 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. Biol.* **57**, 483–498. (doi:10.1080/10635150802172184)
63. Balbuena JA, 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)
64. Suttner B *et al.* 2020 Metagenomics as a public health risk assessment tool in a study of natural creek sediments influenced by agricultural and livestock runoff: Potential and limitations. *Appl. Environ. Microbiol.* **86**, e02525–e02519. (doi:10.1128/AEM.02525-19)
65. 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**, D576–D582. (doi:10.1093/nar/gkq901)
66. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. 2018 Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717. (doi:10.1093/nar/gkx932)
67. Fischer MG. 2021 The Virophage family lavidaviridae. *Curr. Issues Mol. Biol.* **40**, 1–24. (doi:10.21775/cimb.040.001)
68. Ghosh S, Malik YS. 2021 The True Host/s of picobirnaviruses. *Front. Veterinary Sci.* **7**, 615293. (doi:10.3389/fvets.2020.615293)
69. Schmidt N, Bölter M. 2002 Fungal and bacterial biomass in tundra soils along an arctic transect from Taimyr Peninsula, central Siberia. *Polar Biol.* **25**, 871–877. (doi:10.1007/s00300-002-0422-7)

70. Kozieł E, Otulak-Kozieł K, Bujarski JJ. 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)
71. 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)
72. Aris-Brosou S, Parent L, Ibeh N. 2019 Viral long-term evolutionary strategies favor stability over proliferation. *Viruses* **11**, 677. (doi:10.3390/v11080677)
73. Hudson JMG, Henry GHR. 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 TS, Oaks JR. 2019 The coevolution effect as a driver of spillover. *Trends Parasitol.* **35**, 399–408. (doi:10.1016/j.pt.2019.03.010)
75. St. Pierre KA *et al.* 2019 Contemporary limnology of the rapidly changing glacierized watershed of the world's largest High Arctic lake. *Sci. Rep.* **9**, 4447. (doi:10.1038/s41598-019-39918-4)
76. Baudet C, Donati B, Sinaimeri B, Crescenzi P, Gautier C, Matias C, Sagot M-F. 2014 Cophylogeny reconstruction via an approximate Bayesian computation. *Syst. Biol.* **64**, 416–431. (doi:10.1093/sysbio/syu129)
77. Sharp PM, Hahn BH. 2011 Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspect. Med.* **1**, a006841. (doi:10.1126/sciadv.1400127)
78. Hu B *et al.* 2017 Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **13**, 1–27. (doi:10.1371/journal.ppat.1006698)
79. Platto S, Wang Y, Zhou J, Carafoli E. 2021 History of the COVID-19 pandemic: origin, explosion, worldwide spreading. *Biochem. Biophys. Res. Commun.* **538**, 14–23. (doi:10.1016/j.bbrc.2020.10.087)
80. Wallingford PD *et al.* 2020 Adjusting the lens of invasion biology to focus on the impacts of climate-driven range shifts. *Nat. Clim. Change* **10**, 398–405. (doi:10.1038/s41558-020-0768-2)
81. Rocklöv J, Dubrow R. 2020 Climate change: an enduring challenge for vector-borne disease prevention and control. *Nat. Immunol.* **21**, 479–483. (doi:10.1038/s41590-020-0648-y)
82. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM, Eskew EA, Olival KJ, Ross N, Bansal S. 2022 Climate change increases cross-species viral transmission risk. *Nature* **607**, 555–562. (doi:10.1038/s41586-022-04788-w)
83. Osland MJ *et al.* 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**, 3009–3034. (doi:10.1111/gcb.15563)
84. Lemieux A, Colby GA, Poulain AJ, Aris-Brosou S. 2022 Viral spillover risk increases with climate change in High Arctic lake sediments. Figshare. (doi:10.6084/m9.figshare.c.6238499)