nature geoscience

Article

https://doi.org/10.1038/s41561-024-01508-z

Glacier-preserved Tibetan Plateau viral community probably linked to warm–cold climate variations

Received: 2 January 2024

Accepted: 16 July 2024

Published online: 26 August 2024

Check for updates

Zhi-Ping Zhong[1,2,3] ✉, Olivier Zablocki[2,3], Yueh-Fen Li[2,3], James L. Van Etten[4], Ellen Mosley-Thompson[1,5], Virginia I. Rich[1,2,3], Lonnie G. Thompson[1,6] ✉ & Matthew B. Sullivan[1,2,3,7] ✉

Glaciers archive time-structured information on climates and ecosystems, including microorganisms. However, the long-term ecogenomic dynamics or biogeography of the preserved viruses and their palaeoclimatic connections remain uninvestigated. Here we use metagenomes to reconstruct viral genomes from nine time horizons, spanning three cold-to-warm cycles over the past >41,000 years, preserved in an ice core from Guliya Glacier, Tibetan Plateau. We recover genomes of 1,705 approximately species-level viral operational taxonomic units. Viral communities significantly differ during cold and warm climatic conditions, with the most distinct community observed ~11,500 years ago during the major climate transition from the Last Glacial Stage to the Holocene. In silico analyses of virus–host interactions reveal persistently high viral pressure on *Flavobacterium* (a common dominant glacier lineage) and historical enrichment in the metabolism of cofactors and vitamins that can contribute to host adaptation and virus fitness under extreme conditions. Biogeographic analyses show that approximately one-fourth of Guliya viral operational taxonomic units overlap with the global dataset, primarily with the Tibetan Plateau metagenomes, suggesting regional associations of a subset of the Guliya-preserved viruses over time. We posit that the cold-to-warm variations in viral communities might be attributed to distinct virus sources and/or environmental selections under different temperature regimes.

Earth has entered the Anthropocene whereby the human species is driving unprecedented climate and ecosystem changes[1]. Critically, differentiating naturally occurring variations from human-caused climate change requires long-term records of Earth's past. Glacier ice consists of layers of ancient atmospheric deposits that preserve records of past climatic conditions (for example, atmospheric gases, temperature and precipitation[2–4]) as well as cells and DNA that can illuminate ancient ecosystems and climates (for example, refs. 5,6).

Among these glacier archives are viruses that potentially played key ecological roles in the past before freezing. Viruses impact the modern marine planktonic habitats via host cell lysis, gene transfer between hosts and host metabolism reprogramming[7–9]. In cryoconite, snow and young ice at the glacier surface, viruses can be abundant (for example, 10^4–10^5 virus-like particles ml^{-1} on the Svalbard glacier) and diverse (for example, 10,840 approximately species-level viral taxonomic units were recently found in supraglacial materials) and

A full list of affiliations appears at the end of the paper. ✉e-mail: zhong.393@osu.edu; thompson.3@osu.edu; sullivan.948@osu.edu

https://doi.org/10.1038/s41561-024-01508-z

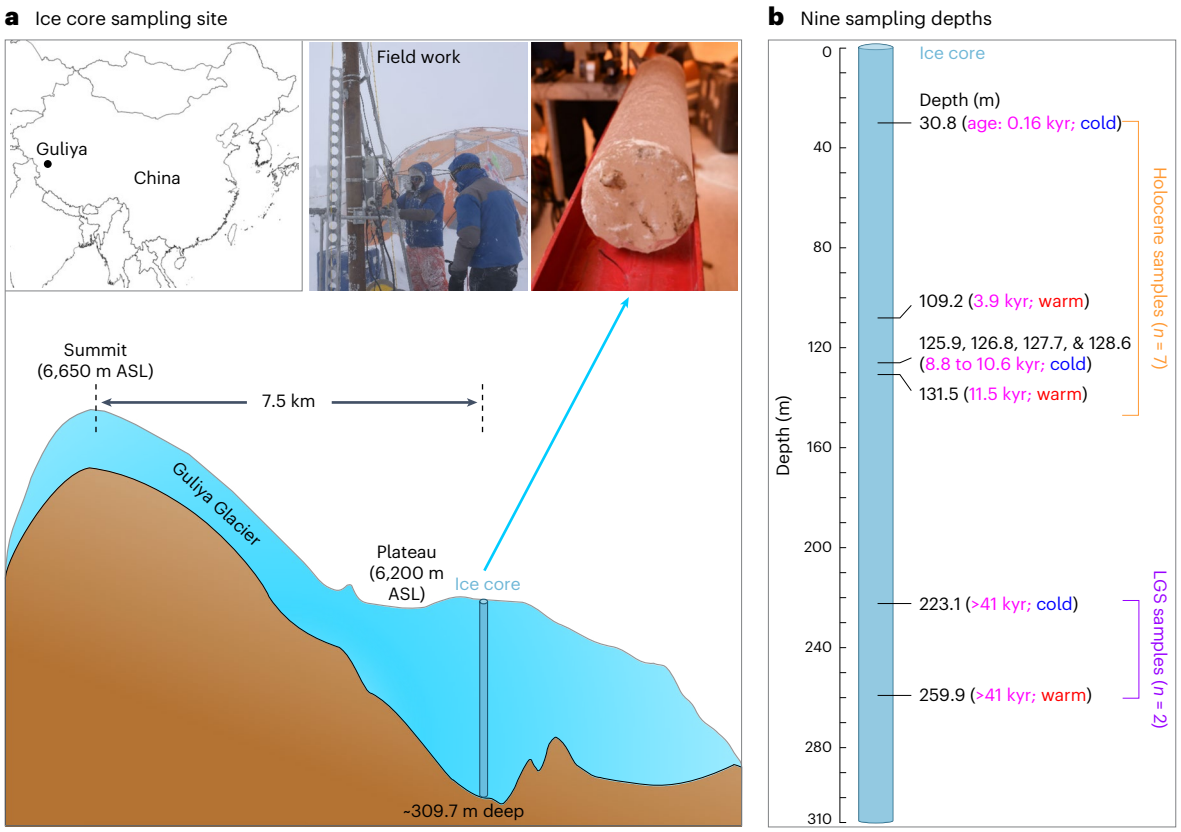


Fig. 1 | Glacier ice sampling. a, Drilling occurred on the plateau of Guliya Glacier in western China (inset), at 6,200 m ASL elevation. Drilling on the glacier surface (first inset photo) resulted in a 309.7 m core (an ~1 m section of which is shown in the second inset photo). For comparison, our previous study of viruses[15] derived from an ice core drilled on the summit (6,650 m ASL). The inset map was plotted by the package rnaturalearth (https://github.com/ropensci/rnaturalearth). **b**, Nine depths were sampled from the ice core, spanning two major climate epochs: the Holocene and LGS. The black numbers denote depth in metres from the core surface, the magenta text provides the corresponding reconstructed age and the blue/red text indicates cold/warm palaeoclimate conditions (Methods). See Supplementary Data 1 for detailed sample physicochemistry. kyr, thousand years.

can potentially modulate hosts' metabolisms and mediate horizontal gene transfers to impact supraglacial ecosystems[10–13]. However, the characterization of glacier-preserved ancient viruses (GPAVs) has thus far been limited to single gene amplification- or microscopy-based documentations[5,14], until our recent study of the first GPAV genomes (33 approximately species-level viral operational taxonomic units (vOTUs)) and communities from two time intervals of a ~15,000-year-old ice core[15]. While this latter, genome-resolved study highlighted potential viral impacts on their hosts' nutrient acquisition before freezing, there remain large gaps in our knowledge of GPAVs. Specifically, no study has documented how viruses changed across cold-to-warm cycles, examined them ecogenomically or biogeographically, or assessed their relationship with the long-term co-archived palaeoclimate conditions.

Here, we applied our previously established low-biomass glacier ice DNA extraction and metagenomic methods[15] to analyse viruses infecting prokaryotes from a ~310-m-long ice core drilled from the Tibetan Plateau (TP)'s Guliya Glacier—the highest (~6,700 m above sea level, ASL), thickest (~310 m) and second largest (>200 km^2) site among all non-polar glaciers[16]. This ice core has constructed palaeoclimate history covering the entire Holocene and Last Glacial Stage (LGS) as well as preceding periods (>115,000 years old; the dating of bottom ice is still underway)[2,16–18] and recovered a limited number of bacterial isolates[19]. Our current work samples ice from nine time intervals representing three cold-to-warm cycles over >41,000 years to investigate virus ecogenomic and biogeographic dynamics and their palaeoclimatic connections.

Increased recovery of GPAVs

We first sought to expand our previous recovery of 33 GPAV vOTUs (obtained from two time intervals representing ~355- and ~14,400-year-old ice[15]) by deeper sequencing (~2.5 times the sequencing depth per sample) and more time intervals (nine instead of two) to capture a temporal coverage window ranging from ~160 to >41,000 years ago (Fig. 1 and Supplementary Data 1 and 2). Combined with an optimized assembly method for low-biomass glacier metagenomes by assessing 18 pipelines (by this study; Supplementary Discussion and Extended Data Fig. 1) and new methods for identifying short viral contigs (previously established; Methods), these methodological improvements allowed us to recover 1,949 viral contigs belonging to 1,705 ≥5 kb vOTUs (average 26.2 kb; hereafter referred to Guliya vOTUs or Guliya viruses), including 567 'long' (≥10 kb) vOTUs. This updated catalogue of 1,705 vOTUs represents a >50-fold increase in the number of ≥5 kb GPAVs reported[15] (Fig. 2a), providing critical resources for making discoveries from the rapidly disappearing glacial records of Earth's history.

The 567 'long' vOTUs were compared with known culturable phage genomes using a gene-sharing network approach that requires ≥10 kb genomes (Methods). Approximately 97% of Guliya vOTUs could not be taxonomically classified into a known viral genus, indicating a high degree of novelty among our expanded GPAV catalogue that was largely understudied. The ~3% classified viruses were assigned to 12 viral clusters (VC, approximately genus-level viral taxonomy) belonging to the same class, Caudoviricetes. Of the 12 VCs, 9 were assigned to the genera *Carjivirus*, *Nickievirus*, *Myxoctovirus*, *Cimpunavirus*, *Phikmvvirus*, *Rauchvirus*, *Borockvirus*, *Samunavirus* and *Akihdevirus*,

https://doi.org/10.1038/s41561-024-01508-z

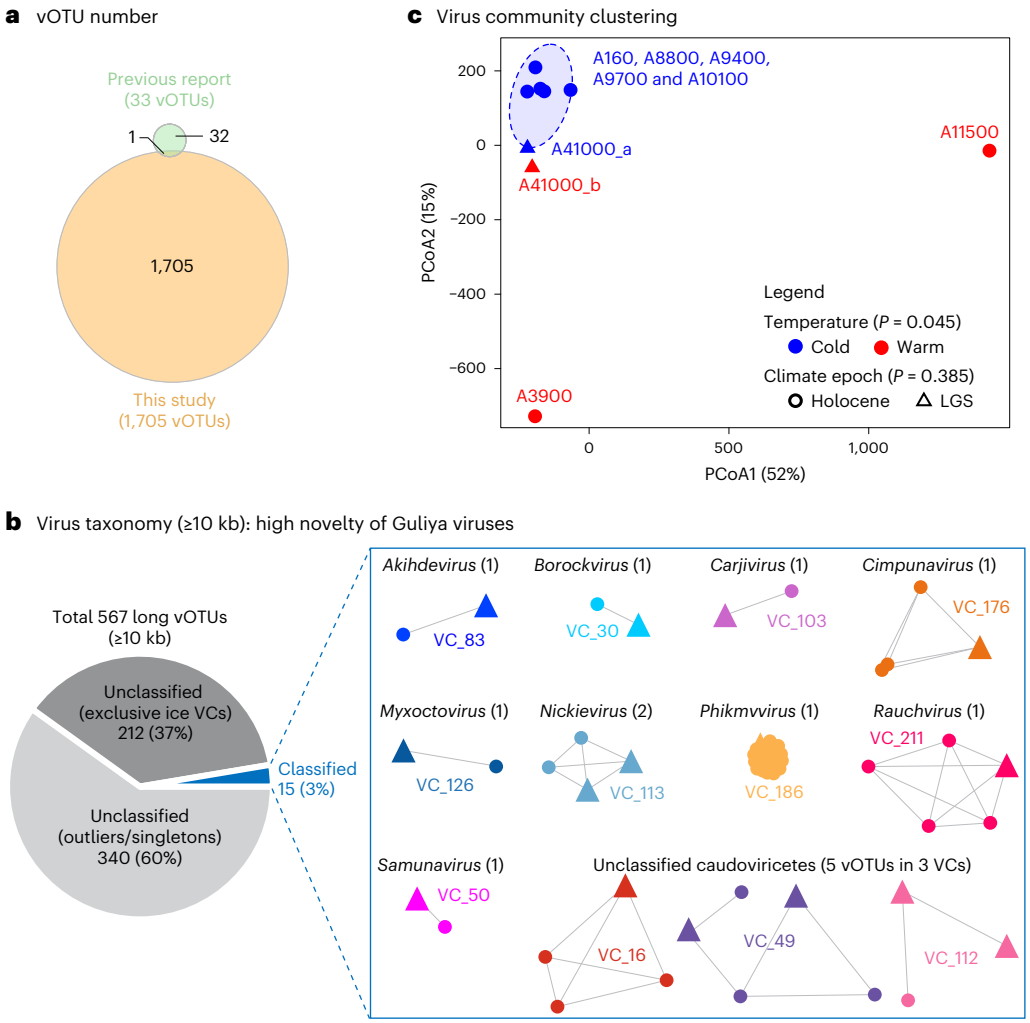


Fig. 2 | GPAV number, taxonomy and community clustering. a, This study's GPAVs relative to the sole previous report[15]. **b**, Taxonomy of this study's GPAVs ≥10 kb (see Supplementary Data 3 and Supplementary Fig. 1 for the full taxonomy results and network). The 3% classified (all within class Caudoviricetes) belonged to 12 VCs (approximately genus-level taxonomy; numbered by distinct colours) as visualized by the adjacent networks with genus names indicated; the triangles denote GPAVs, and the circles denote NCBI RefSeq viral genomes. **c**, Similarity among GPAV communities (PCoA, by Euclidean distance from vOTU coverages) in this study; sample names are indicated ('A' for 'age', and number for reconstructed ice age) and coloured by temperature condition (blue, cold; red, warm); the ellipse denotes a 95% confidence interval for cold samples. Permutational multivariate analysis of variance indicated a significant ($P = 0.045$) difference between cold and warm communities.

and the remaining three belonged to unclassified genera in Caudoviricetes (Fig. 2b, Supplementary Data 3 and Supplementary Fig. 1).

Communities reflect the co-archived climate conditions

Sampling across nine time intervals, encompassing three cold-to-warm cycles spanning >41,000 years in the Holocene and LGS (Fig. 1b), enabled us to examine the 'history' of viral communities and their relationship with the co-archived climate conditions. Ordination analysis, using Guliya vOTUs' relative abundances (Supplementary Data 4), revealed significantly different communities during cold and warm conditions of Earth history (Fig. 2c). These differences are presumably due to unique sources (for example, the released viruses from the source environments could be distinct between cold and warm conditions, and the stronger winds in cold conditions lead to a larger catchment area for glacier depositions[20]) and/or ecological selection (for example, unique selection processes during air transportation and on glacier surfaces before freezing) occurring between cold and warm conditions. The community of sample A11500 (~11,500 years old), collected from the beginning of Holocene, was most distinct from communities of other samples (Fig. 2c), which was potentially due to the sharp change in the climate transition from the LGS to Holocene (~11,700 years ago[21]).

We then assessed the ecological drivers of virus communities. Mg^{2+} (an important metal ion for maintaining phage structures and functions[22]) and $\delta^{18}O$ (a temperature proxy at the time of deposition on or near Guliya surface[16,23]) showed highest impacts (lowest P values) among the tested factors (Supplementary Data 5 and Supplementary Fig. 2). However, the findings were not statistically significant ($P > 0.05$), indicating that untested factors may have been more important, including perhaps biological impacts such as virus–host interactions (documented below).

We retrieved only one deep core from the plateau of Guliya Glacier owing to the challenges of remote and high-altitude field work. With limited core materials, only one sample per depth interval was available for microbiological and palaeoclimatic research. However, our data provide a first window into the historical changes of viral communities over tens of thousands of years and the virus–palaeoclimate linkages archived in a glacier, which will undoubtedly benefit from analyses using larger datasets in the future.

Ecological impacts of viruses over time

Our previous study revealed that GPAVs played key ecological roles in unique, relatively short-term time intervals[15]. Thus, we next sought to ask how viruses might impact ecosystems over a long-term record (>41,000 years). To assess this, we evaluated virus–host linkages associated with the dominant microbial residents and explored virus-encoded auxiliary metabolic genes (AMGs). Analyses of the lineage-specific virus–host ratios (VHRs) allow assessment of how infection dynamics of a specific lineage varies over time[24], whereas AMGs offer a proxy indicating which host metabolisms viruses might historically modulate to improve host and virus fitness[8,25].

Towards the former, we used an integrated approach and TP glacier microbial genomes to establish host predictions (Methods). This analysis linked ~67% Guliya vOTUs (1,149 of 1,705) to a host (Supplementary Data 6), which is high but comparable to the host prediction rates in other ecosystems (for example, rumen and lake sediments[26–28]) using the same method. Abundant viruses were predicted to historically infect the three most dominant bacterial genera *Flavobacterium*, *Cryobacterium* and *Polaromonas* (Fig. 3a) that are commonly dominant in glacier communities with many members having cold-adapted genomic features and the potential to impact carbon cycling[15,29–33]. These results suggest that Guliya viruses infecting dominant bacterial genera were probably important factors, over a >41,000 year record, in shaping the ancient ecosystems preserved. The VHRs varied among the genera, with VHRs for *Flavobacterium* persistently higher than those for *Cryobacterium* and *Polaromonas* (Fig. 3b), presumably reflecting higher viral pressure on *Flavobacterium* over the entire record (see Supplementary Discussion for additional ecological arguments of lineage-specific virus–host interactions).

At the phylum level, we found particularly high VHRs for Patescibacteria over time (Supplementary Fig. 4a), consistent with a previous report of high viral infection of Patescibacteria in Arctic soil with subfreezing temperature and low nutrient availability[34]. We note that Patescibacteria's VHRs could be overestimated due to the primer bias-based underestimation of Patescibacteria's 16S ribosomal RNA genes[35,36]. Alternatively, if the 16S data accurately reflect actual Patescibacteria abundances, we posit that high viral infection could result from fewer clustered regularly interspaced short palindromic repeats (CRISPRs) and spacers for defending viruses (Supplementary Discussion and Supplementary Fig. 4b–e).

Towards the latter consideration, we annotated viral genes (Supplementary Data 7) and rigorously searched for AMGs (Methods). These analyses identified 50 putative AMGs from 82 viral contigs (Supplementary Fig. 5 and Supplementary Data 8), which should be an underestimate since short-read assemblies lead to fragmentation of viral genomes particularly in AMG-enriched areas. Of the AMGs observed, they were enriched in functions of cofactor and vitamin, amino acid and carbohydrate metabolism pathways (Fig. 3c and Supplementary Data 9). Notably, AMGs of 'metabolism of cofactors and vitamins' were enriched through all nine times studied (Fig. 3c), probably due to these pathways, also found previously as enriched in supraglacial microorganisms, enabling microbial hosts to cope with stressful conditions on glacier surfaces[33] and thus, in turn, improving virus fitness.

Among these AMGs, we further analysed a cobaltochelatase gene (*cobT*) and a phosphoadenosine phosphosulfate reductase gene (*cysH*) because they are ecologically important (details in Supplementary Discussion) but have not been reported in glacier viruses. The *cobT* gene belongs to the historically enriched AMG category 'metabolism of cofactors and vitamins' and encodes a component of cobaltochelatase to catalyse cobalt insertion in the corrin ring during the biosynthesis of vitamin B_{12} (ref. 37). The *cysH* gene encodes a phosphoadenosine phosphosulfate reductase to impact the sulfur cycle via assimilatory sulfate reduction. We explored their viral genomic composition (Fig. 3d,e), gene phylogeny, conserved motifs and signatures of evolutionary pressure (details in Methods and Supplementary Discussion). Briefly, phylogenetic analyses revealed that viral *cobT* was probably transferred from microorganisms of the genus *Ralstonia* (Supplementary Fig. 6a), some members of which can participate in vitamin B_{12} metabolism[38], while viral *cysH* appeared to have been transferred from microorganisms of the genus *Janthinobacterium* (Supplementary Fig. 6b), a very common glacier lineage with cold-adaptive molecular capability[32,39]. Both AMGs were under purification selection (Supplementary Data 10 and 11) and contained the conserved functional domains of these enzymes (Supplementary Fig. 7). Taken together, these results suggest that both AMGs might functionally impact glacier microorganisms, presumably by impacting stress resistances via synthesizing vitamin B_{12} and the sulfur cycle via reducing sulfate (Supplementary Discussion).

Overall, our findings suggest that Guliya ice archives a record of ancient virus ecosystems, which can go back >41,000 years. We propose that Guliya viruses potentially prey upon dominant microorganisms and encode AMGs (including those relevant to stress responses) to influence their hosts under the extreme conditions of the ancient ecosystems (for example, on glacier surfaces) before freezing. These data support our previous findings about viral impacts on hosts[15] but provide additional insights into the long-term changes of virus–host interactions over tens of thousands of years, with particular highlights on the persistently high viral pressure on the dominant genus *Flavobacterium* (a common glacier lineage) and the historical enrichment in 'metabolism of cofactors and vitamins' that can contribute to host adaptation under extreme conditions.

Regional association of some Guliya viruses

One of the key goals in glacier research is to understand glacier's sources, which help link the preserved information to source locations[40]. Guliya Glacier is located in Northwestern TP influenced by multiple air masses (Southeast and Southwest Asian monsoons, Indian monsoon and the continental westerlies), complicating the interpretation of its sources[16,17]. Previous analyses show that Guliya experiences a broad range of air impacts (spanning Central Asia, Northwestern India, the Middle East, Northern Africa, and Eastern and Western Europe; mainly impacted by its nearby region)[41] and that Guliya's dust particles partly originated from the Taklimakan desert[42]. However, Guliya's biological source areas, including the main regions, remain to be verified[16,17]. This can be explored now using Guliya virus genomes as signatures against global virus biogeography from known sequencing datasets. To accomplish this, we assessed the distribution of the 1,705 Guliya vOTUs in 733 selected metagenomes from non-glacier environments (for example, air, China soils, deserts and seawater, which could contain viruses deposited on glaciers) and in the 187 globally available glacier metagenomes (published by July 2022) from Arctic, Asia, Europe, North and South America, and Antarctica (Fig. 4a and Supplementary Data 12). This assessment found that ~73% (1,242 of 1,705) vOTUs were unique to Guliya, while ~27% (452 of 1,705) were found in ≥1 of the 187 glacier metagenomes and the remaining <1% (11 of 1,705) were present in non-glacier environments (Fig. 4b). Notably all the later 11 vOTUs were from air samples above the Red Sea (Fig. 4a), suggesting that Guliya-preserved sources might partly have originated from the nearby regions, such as the Middle East and/or North Africa, as also indicated by air influences[41].

Of the 452 vOTUs found in the 187 globally available glacier metagenomes, virtually all (~99%; 447 of 452) were detected from TP glaciers (Fig. 4b), although they accounted for <50% (n = 92) of the 187 glacier metagenomes and ~60% of the total glacier metagenomic sequencing (1.7 of 2.8×10^{12} bp). In contrast, only ~12% (54 of 452) vOTUs were found in the 83 glaciers outside of Asia (Arctic, Europe, North and South America, and Antarctic) and almost all of them (53 of 54) were shared with TP glaciers, although these glaciers accounted for a large portion (44%; 83 of 187) of the available glacier metagenomes. The remaining 12 glacier metagenomes from Asia, but outside of the TP,

https://doi.org/10.1038/s41561-024-01508-z

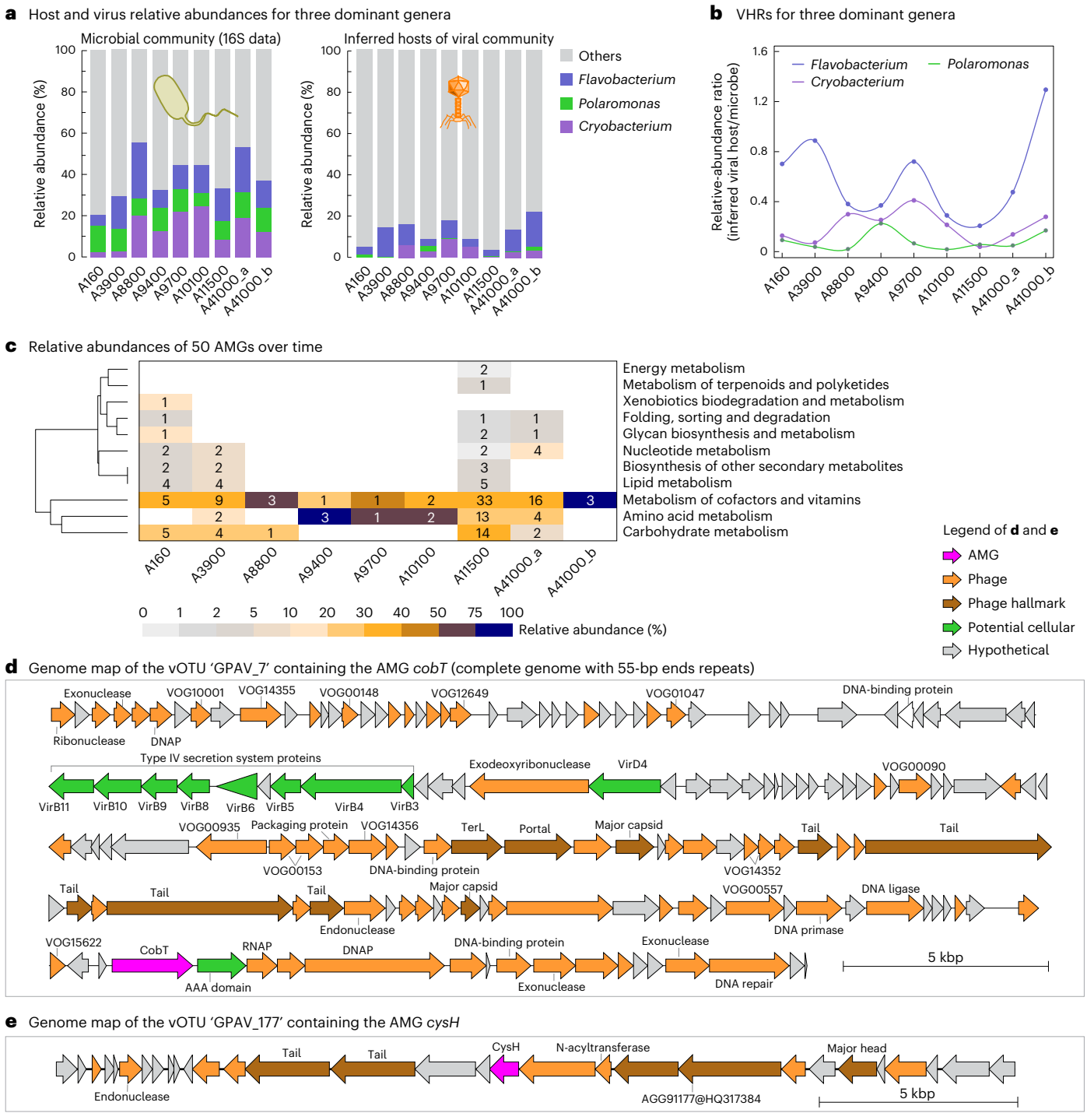


Fig. 3 | Historical changes of virus–host interactions. a,**b**, Relative abundances over time of the three most dominant genera (via 16S rRNA gene amplicons) and their linked viruses (via vOTU coverages; see Methods for in silico virus–host linkages) (**a**) and the resultant VHRs (**b**). Sample names reflect age as in Fig. 2c. **c**, Relative abundances (gradient colours; via vOTU coverages) and metabolic categories of virus-encoded AMGs ($n = 50$) over time. The numbers within the cells are the total number of AMGs present in each category and sample. **d**,**e**, The genome context of the AMGs *cobT* (**d**) and *cysH* (**e**) in the two associated viral contigs. Open reading frames (arrows) are coloured by annotation types: AMGs (magenta), phage genes (orange), phage hallmark genes (dark orange), potential cellular genes (green) and hypothetical protein genes (grey).

contained 124 Guliya vOTUs, with 120 of them shared with TP glaciers (Fig. 4b). Similarly, assessing the historical change of Guliya virus biogeography (that is, analysing the virus distribution for each of the nine Guliya samples individually rather than combining viruses of all nine samples as documented above) revealed that ~99% of Guliya vOTUs (of those presenting in ≥1 of the 187 glacier metagenomes) were detected in TP glaciers over the entire record studied (Fig. 4c).

These biogeographic assessments revealed that ~73% of the Guliya vOTUs were absent in the global dataset analysed; thus, their source information was unknown. The other ~27% were shared with the global dataset, primarily with the TP metagenomes, indicating a likely regional TP association and, thus, a potential regional origin for these Guliya viruses. This result is consistent with a previous finding (via back-trajectory frequency analyses) that the primary source of the

https://doi.org/10.1038/s41561-024-01508-z

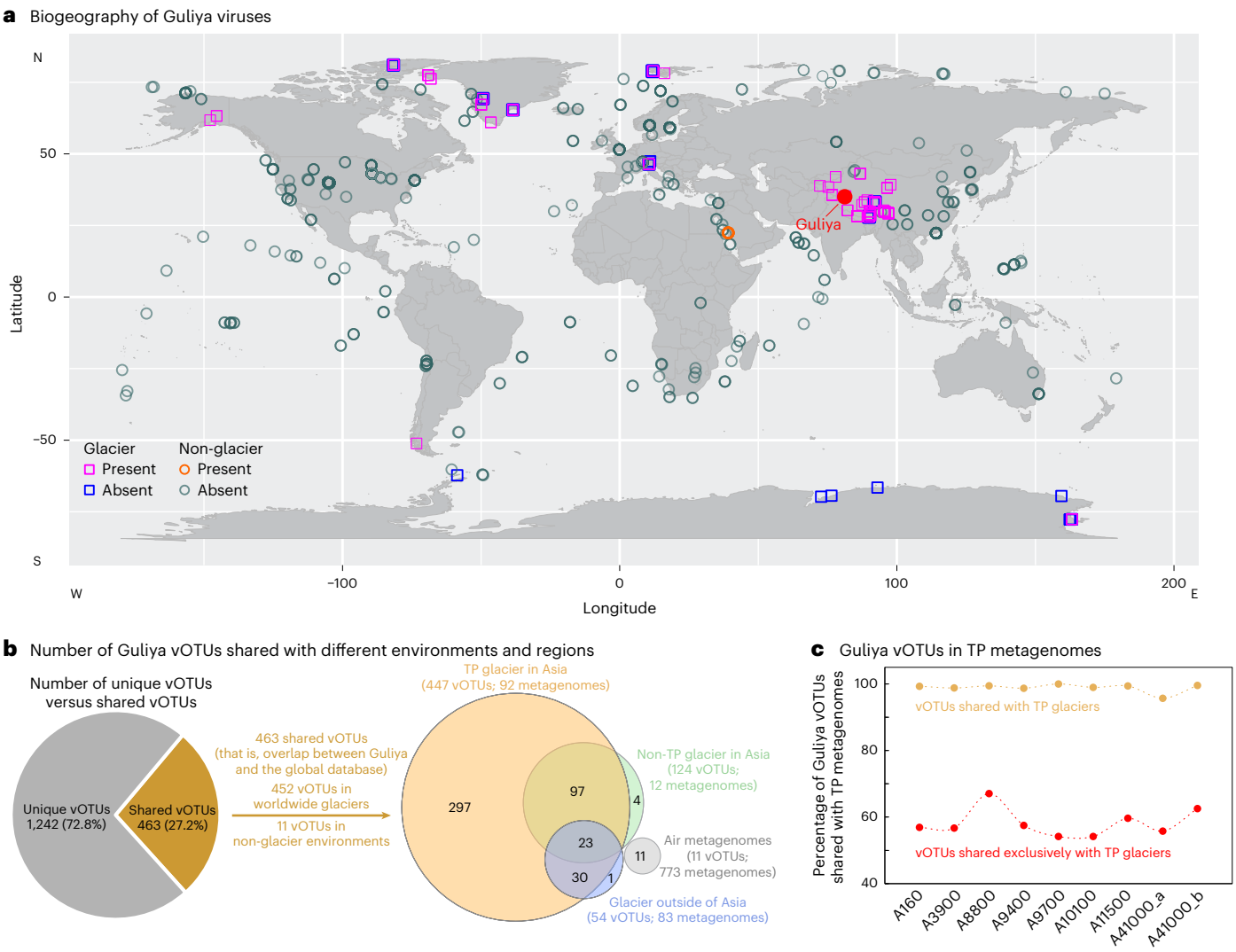


Fig. 4 | Biogeography of Guliya viruses. a, Detection in 920 globally available metagenomes. The red-filled circle indicates Guliya's location; the squares and circles denote glacier (n = 187) and non-glacier (n = 733) metagenomes, respectively; the bold circles indicate ≥2 metagenomes; the magenta squares and orange circles indicate metagenomes containing approximately species-level Guliya viruses (n = 155 and 3, respectively; the latter overlap, thus appearing as 1), while the blue squares and teal circles represent metagenomes without Guliya viruses. **b**, The summarized quantitative distribution of Guliya virus from **a**, grouped by source type; these vOTUs reflect ~27% of the total Guliya viruses (463 of 1,705, with 452 and 11 present in glacier and non-glacier/air metagenomes, respectively), with the remaining ~73% (n = 1,242) unique to Guliya. Of the 452 vOTUs shared with glacier metagenomes, nearly all (n = 447; 99%) were present and 66% (n = 297) were exclusively present in other TP glacier metagenomes; only 1% (n = 5) were exclusively present outside of the TP. **c**, The temporal variation in the biogeography summarized in **b** for the percentage of each sample's glacier-shared vOTUs present in TP glaciers.

air arriving over Guliya Glacier is from the nearby region[41]. We note that the low overlap (~27%) of viruses between Guliya and public dataset might be attributed to the limited size of the comparator datasets analysed and that the source information for most Guliya viruses remains unknown. To improve our understanding of Guliya's biological sources, future studies will benefit from increasing the viral datasets, as well as from including other biological materials (for example, prokaryotes[43]), in the biogeographic investigations.

Finally, we considered whether the ancient microorganisms/viruses could have in situ activity within the frozen ice. Indeed, there are hints of microbial activity in some basal glacier ice, along with the detection of excess gases (CO_2, CH_4 and N_2O) at certain depths that might have been produced by microorganisms[44–46] thought to occur either in cold and salty liquids along three-grain boundaries[47] or on the surface of clay grains[48]. These putative habitats are thought to occur in Antarctic basal ice with extreme conditions including no sunlight, no oxygen, high pressure, low temperature, high acidity (pH ~2.5) and high salinity[47–49], which would select microorganisms that are able to live under these conditions. However, they would preclude the enrichment of *Flavobacterium*, *Polaromonas* and *Cryobacterium*, which were historically enriched in all the nine of the Guliya ice samples studied (Fig. 3a) and are commonly dominant lineages in glacier-preserved communities and on the modern glacier surfaces[15,29–33]. Further, if in situ activity does occur in frozen ice, it is thought to probably be at only an extremely low level of metabolism to mitigate genetic and cellular damage, and microorganisms would be unlikely to proliferate to alter the original community and nucleotide compositions[50,51]. Thus, we interpret the viruses archived in Guliya Glacier as probably representing glacier surface taxa before freezing and faithfully reflecting ancient ecosystems and palaeoclimate conditions.

Conclusion

Glaciers archive palaeo-information including climates and microbial ecosystems over millennia and longer. However, GPAVs have been

largely ignored due to extensive methodological challenges such that no study has investigated the long-term historical dynamics, ecogenomics, biogeography or palaeoclimatic connections. Here, we examined the co-archived palaeoclimate and viruses in TP's Guliya Glacier over nine time intervals that spanned >41,000 years in the Holocene and LGS and covered three cold-to-warm cycles. This expanded known GPAVs >50-fold and provided foundational advances in our understanding of ancient virus roles in rapidly disappearing glacial records of Earth's history, including insights into how virus communities change across cold-to-warm transitions, as well as long-term changes in virus–host interactions and genomic-tracking data to assess the potential source materials for Guliya archives.

Future GPAV studies will benefit from diverse advances and technologies including (1) higher-resolution records to assess finer-scale ecological dynamics and connections to palaeoclimates; (2) expanding datasets to include single-strand DNA[52] and RNA[53] viruses; (3) expanding datasets to include higher per-genome coverage to evaluate population genetics and evolutionary hypotheses, (4) low- and high-throughput cultivation approaches to allow model system 'virocell' experimental testing of genome-derived hypotheses; and (5) experimental virus–host linkage measurements to evaluate the in silico host predictions[25,54]. Just as such approaches have transformed the study of marine viruses[9], similar applications will transform our understanding of GPAVs, including eukaryotic viruses, and their roles in the story of the Earth system—hopefully before the Anthropocene warming compromises all the glacial ice essential to tell such stories.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41561-024-01508-z.

References

1. IPCC *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).
2. Thompson, L. G. et al. Tropical climate instability: the last glacial cycle from a Qinghai-Tibetan ice core. *Science* **276**, 1821–1825 (1997).
3. Rosen, J. L. et al. An ice core record of near-synchronous global climate changes at the Bølling transition. *Nat. Geosci.* **7**, 459–463 (2014).
4. Bauska, T. K., Marcott, S. A. & Brook, E. J. Abrupt changes in the global carbon cycle during the last glacial period. *Nat. Geosci.* **14**, 91–96 (2021).
5. Priscu, J. C., Christner, B. C., Foreman, C. M. & Royston-Bishop, G. in *Encyclopedia of Quaternary Science* Vol. 2 (ed. Elias, S. A.) 1156–1166 (Elsevier, 2006).
6. Karl, D. M. et al. Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* **286**, 2144–2147 (1999).
7. Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
8. Roux, S. et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
9. Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13**, 147–159 (2015).
10. Anesio, A. M., Mindl, B., Laybourn-Parry, J., Hodson, A. J. & Sattler, B. Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *J. Geophys. Res.* **112**, G04S31 (2007).
11. Bellas, C. M., Anesio, A. M. & Barker, G. Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. *Front. Microbiol.* **6**, 656 (2015).
12. Liu, Y. et al. Diversity and function of mountain and polar supraglacial DNA viruses. *Sci. Bull.* **68**, 2418–2433 (2023).
13. Bellas, C. M., Schroeder, D. C., Edwards, A., Barker, G. & Anesio, A. M. Flexible genes establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems. *Nat. Commun.* **11**, 4403 (2020).
14. Castello, J. D. et al. Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biol.* **22**, 207–212 (1999).
15. Zhong, Z. P. et al. Glacier ice archives nearly 15,000-year-old microbes and phages. *Microbiome* **9**, 160 (2021).
16. Thompson, L. G. et al. Ice core records of climate variability on the Third Pole with emphasis on the Guliya ice cap, western Kunlun Mountains. *Quat. Sci. Rev.* **188**, 1–14 (2018).
17. Thompson, L. G. et al. Ice core evidence for an orbital-scale climate transition on the Northwest Tibetan Plateau. *Quat. Sci. Rev.* **324**, 108443 (2024).
18. Thompson, L. G. et al. Use of $\delta^{18}O_{atm}$ in dating a Tibetan ice core record of Holocene/Late Glacial climate. *Proc. Natl Acad. Sci. USA* **119**, e2205545119 (2022).
19. Christner, B. C. *Detection, Recovery, Isolation, and Characterization of Bacteria in Glacial Ice and Lake Vostok Accretion Ice.* Doctoral thesis, Ohio State Univ. (2002).
20. Werner, M. et al. Seasonal and interannual variability of the mineral dust cycle under present and glacial climate conditions. *J. Geophys. Res.* **107**, 4744 (2002).
21. Summerhayes, C. & Charman, D. Introduction to Holocene climate change: new perspectives. *J. Geol. Soc.* **172**, 251–253 (2015).
22. Chaturvedi, U. C. & Shrivastava, R. Interaction of viral proteins with metal ions: role in maintaining the structure and functions of viruses. *FEMS Immunol. Med. Microbiol.* **43**, 105–114 (2005).
23. Yao, T. et al. A review of climatic controls on $\delta^{18}O$ in precipitation over the Tibetan Plateau: observations and simulations. *Rev. Geophys.* **51**, 525–548 (2013).
24. Emerson, J. B. et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870–880 (2018).
25. Howard-Varona, C. et al. Phage-specific metabolic reprogramming of virocells. *ISME J.* **14**, 881–895 (2020).
26. Bolduc, B. et al. iVirus 2.0: cyberinfrastructure-supported tools and data to power DNA virus ecology. *ISME Commun.* **1**, 77 (2021).
27. Gregory, A. C. et al. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* **28**, 724–740.e8 (2020).
28. Zhong, Z. P. et al. Viral potential to modulate microbial methane metabolism varies by habitat. *Nat. Commun.* **15**, 1857 (2024).
29. Liu, Q., Liu, H. C., Zhou, Y. G. & Xin, Y. H. Genetic diversity of glacier-inhabiting *Cryobacterium* bacteria in China and description of *Cryobacterium zongtaii* sp. nov. and *Arthrobacter glacialis* sp. nov. *Syst. Appl. Microbiol.* **42**, 168–177 (2019).
30. Liu, Q., Liu, H. C., Zhou, Y. G. & Xin, Y. H. Microevolution and adaptive strategy of psychrophilic species *Flavobacterium bomense* sp. nov. isolated from glaciers. *Front. Microbiol.* **10**, 1069 (2019).
31. Liu, Y. et al. Genomic insights of *Cryobacterium* isolated from ice core reveal genome dynamics for adaptation in glacier. *Front. Microbiol.* **11**, 1530 (2020).
32. Zhong, Z. P. et al. Clean low-biomass procedures and their application to ancient ice core microorganisms. *Front. Microbiol.* **9**, 1094 (2018).
33. Liu, Y. et al. A genome and gene catalog of glacier microbiomes. *Nat. Biotechnol.* **40**, 1341–1348 (2022).
34. Trubl, G. et al. Active virus–host interactions at sub-freezing temperatures in Arctic peat soil. *Microbiome* **9**, 208 (2021).
35. Brown, C. T. et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).

36. Zhao, R., Farag, I. F., Jorgensen, S. L. & Biddle, J. F. Occurrence, diversity, and genomes of '*Candidatus* Patescibacteria' along the early diagenesis of marine sediments. *Appl. Environ. Microbiol.* **88**, e0140922 (2022).
37. Debussche, L. et al. Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a,c-diamide during coenzyme B_{12} biosynthesis in *Pseudomonas denitrificans*. *J. Bacteriol.* **174**, 7445–7451 (1992).
38. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. *J. Biol. Chem.* **278**, 41148–41159 (2003).
39. Dieser, M., Smith, H. J., Ramaraj, T. & Foreman, C. M. *Janthinobacterium* CG23_2: comparative genome analysis reveals enhanced environmental sensing and transcriptional regulation for adaptation to life in an Antarctic supraglacial stream. *Microorganisms* **7**, 454 (2019).
40. Cuffey, K. M. in *The Physics of Glaciers* (ed. Paterson, W. S. B.) 611–674 (Butterworth-Heinemann/Elsevier, 2010).
41. Sierra-Hernández, M. R., Beaudon, E., Gabrielli, P. & Thompson, L. 21st-century Asian air pollution impacts glacier in northwestern Tibet. *Atmos. Chem. Phys.* **19**, 15533–15544 (2019).
42. Beaudon, E. et al. Aeolian dust preserved in the Guliya ice cap (Northwestern Tibet): a promising paleo-environmental messenger. *Geosciences* **12**, 366 (2022).
43. Liu, Y. et al. Bacterial responses to environmental change on the Tibetan Plateau over the past half century. *Environ. Microbiol.* **18**, 1930–1941 (2016).
44. Tung, H. C., Bramall, N. E. & Price, P. B. Microbial origin of excess methane in glacial ice and implications for life on Mars. *Proc. Natl Acad. Sci. USA* **102**, 18292–18296 (2005).
45. Miteva, V., Sowers, T., Schupbach, S., Fischer, H. & Brenchley, J. Geochemical and microbiological studies of nitrous oxide variations within the new NEEM Greenland ice core during the last glacial period. *Geomicrobiol. J.* **33**, 647–660 (2016).
46. Campen, R., Sowers, T. & Alley, R. Evidence of microbial consortia metabolizing within a low-latitude mountain glacier. *Geology* **31**, 231–234 (2003).
47. Price, P. B. A habitat for psychrophiles in deep Antarctic ice. *Proc. Natl Acad. Sci. USA* **97**, 1247–1251 (2000).
48. Tung, H. C., Price, P. B., Bramall, N. E. & Vrdoljak, G. Microorganisms metabolizing on clay grains in 3-km-deep Greenland basal ice. *Astrobiology* **6**, 69–86 (2006).
49. Barletta, R. E., Jones, W. L., Mader, H. M., Priscu, J. C. & Roe, C. H. Chemical analysis of ice vein microenvironments: II. Analysis of glacial samples from Greenland and Antarctica. *J. Glaciol.* **58**, 1109–1118 (2017).
50. Price, P. B. & Sowers, T. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc. Natl Acad. Sci. USA* **101**, 4631–4636 (2004).
51. Doyle, S. M., Montross, S. N., Skidmore, M. L. & Christner, B. C. Characterizing microbial diversity and the potential for metabolic function at −15 °C in the basal ice of Taylor Glacier, Antarctica. *Biology* **2**, 1034–1053 (2013).
52. Kazlauskas, D., Varsani, A., Koonin, E. V. & Krupovic, M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat. Commun.* **10**, 3425 (2019).
53. Dominguez-Huerta, G. et al. Diversity and ecological footprint of Global Ocean RNA viruses. *Science* **376**, 1202–1208 (2022).
54. Jang, H. B. et al. Viral tag and grow: a scalable approach to capture and characterize infectious virus–host pairs. *ISME Commun.* **2**, 12 (2022).

[1]Byrd Polar and Climate Research Center, Ohio State University, Columbus, OH, USA. [2]Department of Microbiology, Ohio State University, Columbus, OH, USA. [3]Center of Microbiome Science, Ohio State University, Columbus, OH, USA. [4]Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska–Lincoln, Lincoln, NE, USA. [5]Department of Geography, Ohio State University, Columbus, OH, USA. [6]School of Earth Sciences, Ohio State University, Columbus, OH, USA. [7]Department of Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, OH, USA. ✉e-mail: zhong.393@osu.edu; thompson.3@osu.edu; sullivan.948@osu.edu

Methods

Guliya ice core sampling and physicochemical conditions

The ice core (309.7 m deep drilled to the glacier bedrock; location: 35° 13′ 58.8′ N, 81° 28′ 5.7′ E; 6,200 m ASL) was drilled on the plateau of Guliya Glacier (named 'Guliya Plateau core', GP core) in 2015 (Fig. 1a). The drilling site had an air temperature of −21.2 °C; the ice temperatures were −11.6 °C and −2.1 °C at 10 m and 309.7 m (bedrock) depth, respectively[16]. The core was 10 cm in diameter, which was cut into ~1 m sections. Core sections were sealed in plastic tubing, placed in cardboard tubes covered with aluminium and transferred at −20 °C by a freezer truck from the drill site to freezers in Lhasa, by airplane to freezers in Beijing, by airplane to Chicago and then by freezer truck to the Byrd Polar and Climate Research Center at The Ohio State University in Columbus where they have been stored at −34 °C.

Nine samples were collected from the GP core at depths of 30.80–30.96 (sample name A160, ~160 years old), 109.23–109.42 (A3900, ~3,900 years old), 125.93–126.89 (A8800, ~8,800 years old), 126.89–127.75 (A9400, ~9,400 years old), 127.75–128.61 (A9700, ~9,700 years old), 128.61–129.38 (A10100, ~10,100 years old), 131.57–131.77 (A11500, ~11,500 years old), 223.15–223.29 (A41000_a, >41,000 years old) and 259.95–260.16 (A41000_b, >41,000 years old) metres, respectively (Fig. 1b and Supplementary Data 1). The putative ages and physicochemical parameters of each sample are summarized in Supplementary Data 1. Specifically, these samples represented ice that was ~160 to >41,000 years old[2,17,18,42], with seven and two samples from the two major climate epochs Holocene and LGS, respectively. In addition to the time gradient, these samples were intentionally selected from the peak of cold or warm conditions based on GP core's $\delta^{18}O$ curve[2,17,18,42], which is the temperature proxy at the time of deposition on or near the surface of Guliya Glacier[23,55], for investigating the potential impacts of temperature conditions on communities. This designation generated samples covering three cold-to-warm cycles of Earth history (Fig. 1b and Supplementary Data 1).

These ice samples were collected according to the clean low-biomass procedures we established previously[15]. Briefly, ice core sections were transferred from −34 °C to the sampling temperature of −5 °C overnight to reduce the possibility of fracturing during surface decontamination procedures, which consisted of three steps to remove the ice core's surface ice and, finally, sample the inner ice for further analyses[15]. The sampling steps were conducted in a cold room laboratory (−5 °C), which was exposed to ultraviolet light for more than 12 h before ice core processing to kill microorganisms in the air and on the surface of the instruments. All the biological work in this study after the ice sampling in the cold room laboratory was performed in a hood within a small (~2 m^2 in area) clean room laboratory that is reserved for microbial experiments with low-biomass samples. The hood was exposed to ultraviolet light for more than 1 h before experiments.

Concentrations of insoluble dust, major ions and oxygen isotopes of glacier ice were analysed via a Coulter Counter (TAII), Thermo Scientific Dionex ion chromatograph (ICS-5000) and Picarro cavity ring-down spectrometer (L2120-i), respectively, with methodological details described in our previous work[17]. The chronology of the GP core, from which the nine ice samples were collected, was established by ^{14}C accelerator mass spectrometry dating of englacial plant fragments and ^{36}Cl age, as well as by matching the oxygen isotopic data between the GP core and another ice core that was drilled and dated in 1992 near the GP core from Guliya Glacier[2,17,18,56]. Until now, we had confident dating for GP core's ice from the surface to 187.4 m deep, which represents materials from the current time (that is, 2015 CE, at the time of core drilling) to ~41,000 years ago[16,17]. Seven of the nine samples in the current study were collected above 187.4 m deep of the GP core and had robust dating (Fig. 1b and Supplementary Data 1), while the other two samples were collected from ice deeper than 187.4 m (that is, 223.15–223.29 and 259.95–260.16 m deep) and were only listed as older than 41,000 years.

Background controls

Four controls were used to trace possible sources of background contaminations during ice processing as established previously[32]. First, we assessed what microorganisms inhabited the air of the cold and clean room laboratories, respectively, in which we sampled the ice and conducted the biological work (sample names of these two air controls: AirColdRoom and AirCleanRoom, respectively). Microorganisms from about 17.3 m^3 of air were collected over 4 days of continuous sampling in the cold and clean rooms, respectively, with an air sampler (SKC) using the established methods[32], during which the ice core samples were processed at the same time. These two controls provided evaluations of the background contamination due to ice exposure to air during the processing. Second, a sterile artificial ice core was processed in parallel with the authentic ice core samples through the entire analysis (sample name: ArtificialIce). This control allowed the evaluation of contamination from the instruments used to process the ice. Third, a blank control was established by filtering and extracting DNA directly from 50 ml of sterile water in parallel with the ice samples (sample name: Blank). This control allowed the evaluation of contamination downstream of the ice processing, including the water filtering and molecular procedures (for example, DNA extraction, polymerase chain reaction (PCR), library preparation and sequencing).

Genomic DNA isolation

Viruses, together with other biological particles (for example, cells), in GP core samples and two liquid controls (that is, ArtificialIce and Blank) were concentrated to 0.8 ml using 100 kDa Amicon Ultra Concentrators (EMD Millipore) and preserved at 4 °C until DNA extraction (within 2 h). The bulk genomic DNA, which theoretically captures both intracellular and extracellular viruses, from the above concentrates and the filters of two air controls (that is, AirColdRoom and AirCleanRoom) was isolated with a DNeasy PowerSoil Kit (cat. no. 12888-100, Qiagen) according to the manufacturer's instructions, with an additional step of bead beating to disrupt bacterial spores and Gram-positive cells before cell lysis. The bead beating was conducted using the MiniBeadBeater-16 (Model 607, BioSpec Products) by homogenization at 3,400 rpm for 1 min with 100 mg of autoclaved (121 °C for 30 min) 0.1-mm-diameter glass beads (cat. no. 13118-400, Qiagen). The isolated DNA was preserved at −20 °C.

Metagenomic sequencing

After DNA extraction, metagenomic sequencing for all samples was performed at the Joint Genome Institute (JGI), Department of Energy, USA. Briefly, the DNA libraries were prepared using the Nextera XT Library Prep Kit (cat. no. 15032354, Illumina) with 9–20 cycles of PCR amplification to increase template concentrations in each library. All libraries were sequenced by Illumina NovaSeq platform (2 × 150 bp).

Metagenomic read processing and viral identification

All metagenomic data analyses were supported by the Ohio Supercomputer Center[57], except that the sequencing reads were filtered for quality by JGI using the previously established standard pipeline[58]. Briefly, BBDuk (BBMap v38.26 (ref. 59)) was used to remove potential contaminants that might be introduced during sampling processing and trim reads with adapters or low-quality bases[58]. The first and last ten bases were trimmed off each read. In addition, a read was removed if it contained four or more 'N' bases, had an average quality score <10 across the read, or had a minimum length ≤51 bp or 33% of the full read length. After quality filtering, the metagenomic sequence data were assembled to contigs using an optimized pipeline for assembling low-biomass glacier ice metagenomes (see details in 'Assembly method optimization' section).

The assembled contigs (length ≥5 kb or circular contigs with length 1.5–5.0 kb) were then used to identify viral contigs according to the previously established methods[60]. Briefly, this study focused on viruses infecting prokaryotes and used three tools VirSorter v1.1.0

(ref. 61), DeepVirFinder v1.0 (ref. 62) and MARVEL v0.2 (ref. 63) for predicting viruses. Contigs were classified as viruses if they met one of the following four criteria: (1) categories 1, 2, 4 or 5 of VirSorter; (2) DeepVirFinder score ≥0.9 and $P < 0.05$; (3) MARVEL probability score ≥90%; or (4) DeepVirFinder score ≥0.7 and $P < 0.05$ and MARVEL probability score ≥70%. Viruses identified by all methods were combined and dereplicated for further analyses.

The viral contigs were first checked for contaminants by comparing them with viral genomes considered as putative laboratory contaminants (phages cultivated in our laboratory: *Synechococcus*, *Cellulophaga* and *Pseudoalteromonas* phages) using Blastn v2.10.0+. Two Guliya viral contigs (GPAV_392 and GPAV_677) had hits to the *Cellulophaga* phage genomes of lab contaminants (Supplementary Data 13). However, the alignment coverages (≤1%) and identities (88–91%) were low. Thus, no Guliya viruses were lab contaminants. The remaining contigs were clustered into vOTUs if they shared ≥95% nucleotide identity across 80% of their lengths[64]. The longest contig within each vOTU was selected as the seed sequence to represent that vOTU. A coverage table of each vOTU was generated using iVirus' methods (tools: BowtieBatch v1.0.0 and Read2RefMapper v1.1.0) by mapping quality-controlled reads to vOTUs (read identity of ≥95% and vOTU coverage of ≥70%), and the resulting coverage depths, which were weighted by genome length, were further normalized by library size to 'coverage per gigabase of virome'[26,65]. Any vOTUs that were identified from both glacier ice and controls were interpreted as potential contaminants and removed in further analyses.

Assembly method optimization

Technical advances have enabled metagenomic sequencing of low-biomass communities with nanogram to subnanogram amounts of DNA[66,67]. In addition, we have previously used mock viral communities to optimize the in silico pipeline for investigating viral communities, including the assessment of the rate of chimerism in the contigs assembled by the assemblers SPAdes and MEGAHIT[68]. We further used 169 PCR-amplified metagenomes to specifically optimize the de novo assembly pipeline in terms of the size of assembled contigs with length ≥10 kb and the error rates to propose the best-suited assembly method for PCR-amplified low-biomass metagenomes[69]. Here, we adapted the recommended assembly method established previously[68,69] but further used three of the nine glacier ice metagenomes from this study to optimize the assembly method in terms of the size of shorter assembled contigs and viral contigs with length ≥5 kb. We tested a suit of combinations of assembly pipelines (Extended Data Fig. 1): (1) two types of read selection (that is, read deduplication by BBMap's (v38.26) clumpify.sh[59] versus no read deduplication before assembly); (2) two types of read correction (that is, read error correction versus no read error correction with option --only-assembler in SPAdes v3.11.1 assemblies); (3) two modes of the SPAdes assembler (that is, metaSPAdes with option -meta versus single-cellSPAdes with option -sc); and (4) two sets of *k*-mers in SPAdes assembler (that is, -k 21,33,55 versus -k 21,33,55,77,99,127). These combinations generated a total of 16 assembly tests (Extended Data Fig. 1), which were then compared with two additional tests with two sets of *k*-mers (that is, -k 21,33,55 versus the default *k*-mer parameters -k 21,29,39,59,79,99,119,141) using the tool MEGAHIT (v1.1.2) that has a better computational efficiency than SPAdes[70].

Taxonomy and ecology analyses

Because viruses lack any single universally shared gene, we established taxonomy using gene-sharing network analysis from viral sequences ≥10 kb in length using vConTACT2 v2 (ref. 71). Briefly, this analysis compared the vOTUs in this study with viral genomes in the National Center for Biotechnology Information (NCBI) RefSeq database (release v201) and generated VCs approximately equivalent to known viral genera[8,71,72].

Principal coordinate analysis (PCoA) was performed using the Euclidean distance matrix based on the coverage of each vOTU. Associations between the viral community compositions and environmental parameters were evaluated through both distance-based redundancy analysis and two-tailed Mantel tests by comparing their Euclidean distance matrices. Environmental parameters were compiled and tested for normality (one-sample Kolmogorov–Smirnov test; $P > 0.05$) using SPSS (release 18). Parameters found to have a non-normal distribution were transformed as close to normality as possible. These statistical analyses provide estimates of the independent contribution from each parameter, and both methods generated consistent results that none of the tested parameters had a significant impact on the glacier viral communities across the nine samples (Supplementary Data 5), indicating that there are untested variables that are important in shaping glacier viral communities, such as the interactions between viruses and their hosts.

Viral host prediction

The putative virus–host linkages were predicted in silico using the iVirus tool VirMatcher v0.3.3 (ref. 26), which aggregates four different methods to provide a statistical confidence score for each host prediction and these methods are based on (1) tRNA match, (2) nucleotide sequence composition, (3) nucleotide sequence similarity and (4) CRISPR spacer match. Viral host prediction benefits from the database that contains microbial genomes from the same ecosystems as viruses[24]. Therefore, we used the 2,358 microbial MAGs that were recovered from TP glaciers[33] as the microbial database for linking the Guliya viruses to their hosts. A summary of the predictions is available in Supplementary Data 6.

Once the viral hosts were predicted, we assessed lineage-specific virus/host abundance ratios at both phylum and genus levels according to the relative abundances of microbial lineages and their viruses[24], which were obtained on the basis of the microbial (bacterial/archaeal) 16S rRNA gene amplicon reads and each vOTU's coverage, respectively. At the phylum level, the virus/host abundance ratio of Patescibacteria was significantly (~20 times; $P = 0.007$) higher than that of all other phyla (Supplementary Fig. 4a). Thus, we compared the genomic features of the 2,358 TP glacier MAGs belonging to Patescibacteria and other phyla. The taxonomy, genome size and GC content were summarized on the basis of the original report of these MAGs[33]; while we extracted the CRISPRs and spacers (≥3 direct repeats) from these MAGs using MinCED (v0.4.2; mining CRISPRs in environmental datasets[73]).

Virus-encoded AMGs

The putative AMGs were identified and evaluated according to our previously established methods[74]. Specifically, all vOTUs were processed with VIBRANT v1.2.1 (ref. 75) to obtain gene functional annotations and identify putative AMGs. To obtain high-quality AMGs and rule out false-positive AMGs from microbial contamination, CheckV v1.1.10 (with default parameters) and manual inspections were then used to identify putative cellular genes, assess host–virus boundaries and remove the potential host fraction of the viral contigs[76]. Only AMGs that were surrounded by phage genes and did not contain transposon regions were included for further analyses. Metabolism categories of AMGs were summarized on the basis of Kyoto Encyclopedia of Genes and Genomes annotations and the pathway modules[77]. Visualization of the genome maps for viruses containing AMGs was performed using Easyfig v2.2.5 (ref. 78). Phage genes and hallmark genes were identified by VirSorter[61,79]. The relative abundances of AMGs were assessed on the basis of the coverages of vOTUs containing the AMGs.

Two AMGs of particular interest (that is, the *cobT* gene involved in vitamin B_{12} synthesis and *cysH* gene involved in sulfur metabolism) were subjected to further analyses to infer their evolutionary histories. DIAMOND BLASTP (v2.10.0+)[80] was used to query an AMG's amino acid sequence against NCBI RefSeq database (release v214) in a sensitive mode with default settings, to obtain the top 40 hits as the reference sequences. In addition, microorganism-encoded *cobT* and *cysH* genes

were extracted from the nine glacier metagenomes to investigate possible gene transfers between viruses and their microbial hosts. Multiple sequence alignment was performed using MAFFT (v.7.017)[81] with the E-INS-I strategy for 1,000 iterations. The aligned sequences were then trimmed using TrimAl v1.2 (ref. 82) with the flag gappyout. The substitution model was selected by ModelFinder[83] for accurate phylogenetic analysis. Phylogenies were generated using IQ-TREE v1.6.11 (ref. 84) with 1,000 bootstrap replicates and then visualized in iTOL v5 (ref. 85). Potential recombination among genes was evaluated using nine programs: RDP[86], GENECONV[87], BootScan[88], MaxChi[89], Chimaera[90], SiScan[91], LARD[92], Phylpro[93] and 3Seq[94] within RDP5 (ref. 95). A Bonferroni correction with a *P* value cut-off of 0.05 was applied in each of the tests. A sequence was considered as a true recombinant if supported by at least four of the nine programs. Branch and site selection pressure (d*N*/d*S*) analysis across lineages was carried out using codon models with maximum likelihood estimated with the CODEML package in PAML v4.9 (ref. 96) (see details in Supplementary Data 10 and 11).

Biogeographic assessment

Considering the low temperatures of Guliya (described above), the preserved viruses should reflect the species signal from the deposition sources[50,51]. To explore the biogeography of glacier viruses identified, the genome fragments of 1,705 Guliya vOTUs were used as baits to recruit reads (≥95% identity) from 733 non-glacier metagenomes and 187 globally available glacier metagenomes (published by July 2022), based on the methods described above using iVirus' BowtieBatch and Read2RefMapper tools[26,65]. A Guliya vOTU was considered to be present in a metagenome if ≥70% of its genomic content was covered by the recruited reads. The details, including environments, locations and citations of the above metagenomes, are summarized in Supplementary Data 12. The global map was plotted by the packages tidyverse[97] and ggplot2 (ref. 98) in R v3.6.1.

Data availability

Metagenomic data of Guliya samples are available to the public via both the Integrated Microbial Genomes (IMG) system (https://genome.jgi.doe.gov/portal/) and NCBI Sequence Read Archive (SRA) database, with an individual accession number for each sample summarized in Supplementary Data 2. The Guliya virus contigs are available via figshare at https://doi.org/10.6084/m9.figshare.24523849 (ref. 99). The accession information of 920 published metagenomes, including the 187 globally available glacier metagenomes, is provided in Supplementary Data 12. Source data are provided with this paper.

Code availability

The custom scripts used for analysing data are available via GitHub at https://github.com/zhiping393/GPAV.

References

55. Tian, L. et al. Oxygen-18 concentrations in recent precipitation and ice cores on the Tibetan Plateau. *J. Geophys. Res.* **108**, 4293 (2003).
56. Thompson, L. G. et al. A 1000 year ice-core climate record from the Guliya ice cap, China: its relationship to global climate variability. *Ann. Glaciol.* **21**, 175–181 (1995).
57. Ohio Supercomputer Center. Ohio Supercomputer Center. *Ohio Supercomputer Center* http://osc.edu/ark:/19495/f5s1ph73 (1987).
58. Clum, A. et al. DOE JGI metagenome workflow. *mSystems* **6**, e00804-20 (2021).
59. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner* (Lawrence Berkeley National Laboratory, 2014).
60. Zhong, Z. P. et al. Lower viral evolutionary pressure under stable versus fluctuating conditions in subzero Arctic brines. *Microbiome* **11**, 174 (2023).
61. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
62. Ren, J. et al. Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77 (2020).
63. Amgarten, D., Braga, L. P. P., da Silva, A. M. & Setubal, J. C. MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Front. Genet.* **9**, 304 (2018).
64. Gregory, A. C. et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123.e14 (2019).
65. Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B. L. & Sullivan, M. B. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J.* **11**, 7–14 (2016).
66. Rinke, C. et al. Validation of picogram- and femtogram-input DNA libraries for microscale metagenomics. *PeerJ* **4**, e2486 (2016).
67. Duhaime, M. B., Deng, L., Poulos, B. T. & Sullivan, M. B. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* **14**, 2526–2537 (2012).
68. Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A. & Sullivan, M. B. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817 (2017).
69. Roux, S. et al. Optimizing de novo genome assembly from PCR-amplified metagenomes. *PeerJ* **7**, e6902 (2019).
70. van der Walt, A. J. et al. Assembling metagenomes, one community at a time. *BMC Genomics* **18**, 521 (2017).
71. Jang, H. B. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
72. Lima-Mendez, G., Van Helden, J., Toussaint, A. & Leplae, R. Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol. Biol. Evol.* **25**, 762–777 (2008).
73. Skennerton, C. T., Soranzo, N. & Angly, F. MinCED—mining CRISPRs in environmental datasets. *GitHub* https://github.com/ctSkennerton/minced (2019).
74. Pratama, A. A. et al. Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. *PeerJ* **9**, e11447 (2021).
75. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).
76. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2020).
77. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
78. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
79. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
80. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
81. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
82. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

83. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
84. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
85. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
86. Martin, D. & Rybicki, E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563 (2000).
87. Padidam, M., Sawyer, S. & Fauquet, C. M. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218–225 (1999).
88. Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retrovir.* **11**, 1423–1425 (1995).
89. Smith, J. M. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**, 126–129 (1992).
90. Posada, D. & Crandall, K. A. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl Acad. Sci. USA* **98**, 13757–13762 (2001).
91. Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573–582 (2000).
92. Holmes, E. C., Worobey, M. & Rambaut, A. Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16**, 405–409 (1999).
93. Weiller, G. F. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Mol. Biol. Evol.* **15**, 326–335 (1998).
94. Lam, H. M., Ratmann, O. & Boni, M. F. Improved algorithmic complexity for the 3SEQ recombination detection algorithm. *Mol. Biol. Evol.* **35**, 247–251 (2018).
95. Martin, D. P. et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **7**, veaa087 (2021).
96. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
97. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
98. Villanueva, R. A. M. & Chen, Z. J. ggplot2: elegant graphics for data analysis (2nd ed.). *Meas.-Interdiscip. Res.* **17**, 160–167 (2019).
99. Zhong, Z. P. Genomes of glacier-preserved ancient viruses. *figshare* https://doi.org/10.6084/m9.figshare.24523849 (2024).

Acknowledgements

This work was supported by a collaborative programme for ice core drilling and analyses between The Ohio State University's Byrd Polar and Climate Research Center and the Institute of Tibetan Plateau Research of the Chinese Academy of Sciences, funded by the National Science Foundation's Paleoclimate Program award no. 1502919 and the Chinese Academy of Sciences, respectively, to L.G.T. Partial support was provided by a Gordon and Betty Moore Foundation Investigator Award no. 3790 to M.B.S., the Byrd Postdoc Fellowship to Z.-P.Z. and the Heising-Simons Foundation award no. 2022-4014 to L.G.T., Z.-P.Z., V.I.R. and E.M.-T. A portion of this research was funded by the US Department of Energy Joint Genome Institute CSP project no. 503428 to M.B.S. and was performed under the JGI-EMSL Collaborative Science Initiative and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under contract nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). We greatly appreciate the help by M. E. Davis for providing dating and environmental data of the ice, by E. Beaudon, M. R. Sierra-Hernández, D. V. Kenny and P.-N. Lin with ice core sampling, by J. Wainaina, A. Gregory, J. Guo, K. Gerhardt and B. Christner for helpful discussions, by Y. Zhou with figure modifications, by N. E. Solonenko with shipment staff for metagenomic sequencing, by S. Roux with metagenomic processing at JGI and by A. Jermy with manuscript commenting and revising. Bioinformatics were supported by the Ohio Supercomputer Center.

Author contributions

Z.-P.Z., J.L.V.E., E.M.-T., V.I.R., L.G.T. and M.B.S. conceived and designed the research. E.M.-T., V.I.R., L.G.T. and M.B.S. supervised this work. E.M.-T., L.G.T. and Z.-P.Z. coordinated the sampling efforts. Y.-F.L. extracted the DNA. Z.-P.Z. performed biological tests before DNA extraction and analysed sequencing data. Z.-P.Z. wrote and O.Z., E.M.-T., V.I.R., L.G.T. and M.B.S. critically revised the manuscript. All authors revised and approved the final manuscript to be published.

Competing interests

The authors declare no competing interests.

Additional information

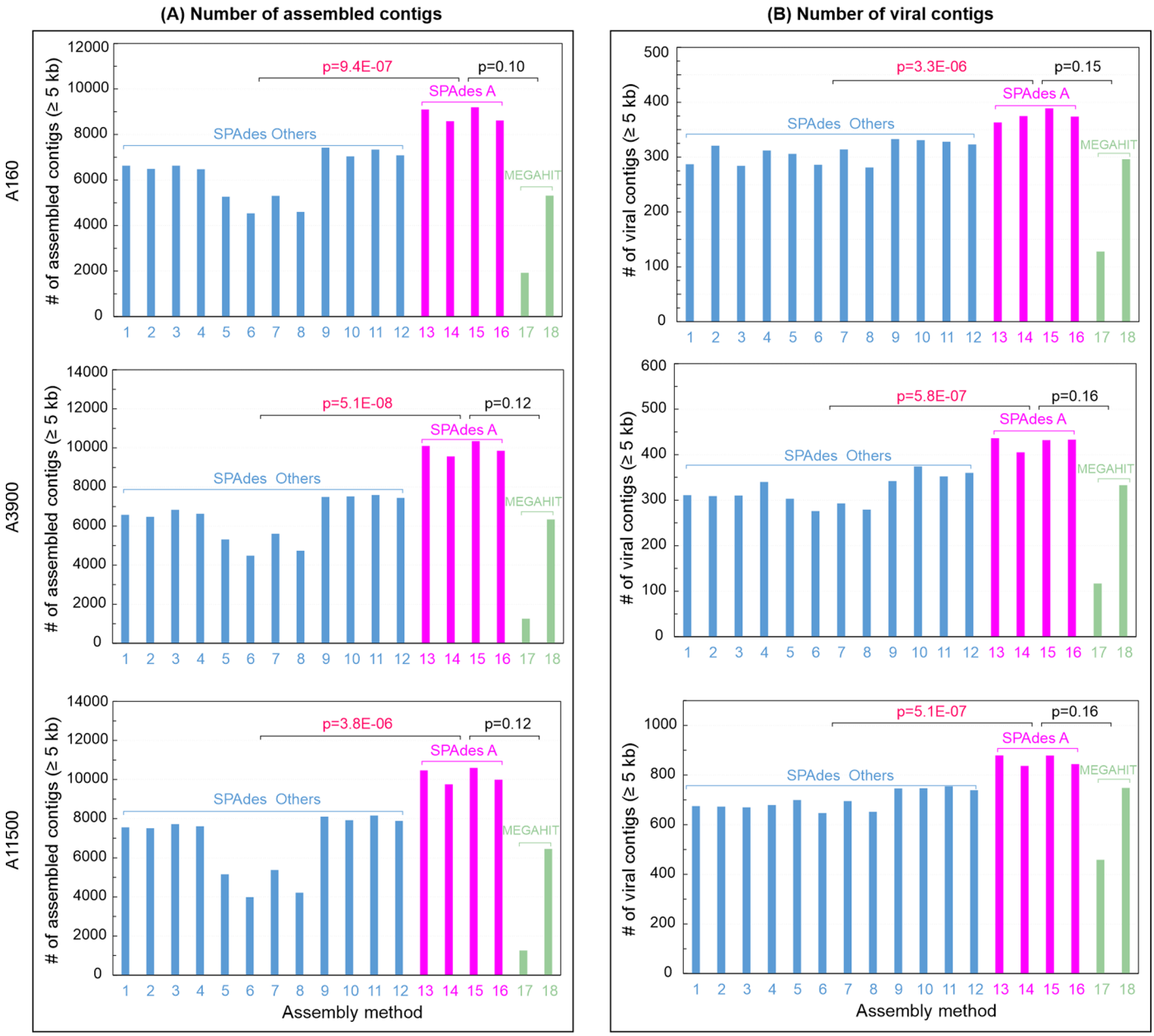
Extended data is available for this paper at https://doi.org/10.1038/s41561-024-01508-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41561-024-01508-z.

Correspondence and requests for materials should be addressed to Zhi-Ping Zhong, Lonnie G. Thompson or Matthew B. Sullivan.

Peer review information *Nature Geoscience* thanks Matthieu Legendre, Wei Li and Ruonan Wu for their contribution to the peer review of this work. Primary Handling Editor: James Super, in collaboration with the *Nature Geoscience* team.

https://doi.org/10.1038/s41561-024-01508-z



Legend for assembly methods

Parameter	SPAdes Others												SPAdes A				MEGAHIT	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Reads deduplication			✓	✓			✓	✓			✓	✓			✓	✓		
Reads error correction	✓		✓		✓		✓		✓		✓		✓		✓		✓	✓
metaSPAdes	✓	✓	✓	✓	✓	✓	✓	✓										
single-cell SPAdes									✓	✓	✓	✓	✓	✓	✓	✓		
MEGAHIT																	✓	✓
kmers: 21,33,55	✓	✓	✓	✓					✓	✓	✓	✓					✓	
kmers: 21,33,55,77,99,127					✓	✓	✓	✓					✓	✓	✓	✓		
kmers: 21,29,39,59,79,99,119,141																		✓

Extended Data Fig. 1 | Optimization of assembly pipelines. The number of assembled contigs (**a**) and viral contigs (**b**) ≥5 kb was compared across 18 assembly pipelines using metagenomes of three glacier-ice samples: A160, A3900, and A11500. Parameters of each pipeline are summarized in legends. These pipelines comprised three groups of comparisons: (i) SPAdes A (in magenta), this group used 'single-cellSPAdes + six different kmer frequencies' for assemblies and generated best assemblies; (ii) SPAdes Others (in light blue), this group used metaSPAdes or 'single-cellSPAdes + three or six different kmer frequencies' for assemblies and generated significantly fewer contigs than the group SPAdes A; and (iii) MEGAHIT (in green), this group used the tool MEGAHIT for assembly and generated substantially fewer contigs than the group SPAdes A. The two-sided p values are provided for comparisons between groups, with significant difference indicated in red.