nature portfolio

Peer Review File

Elevation-structured viral ecological strategies along glacier-fed rivers on the Qinghai-Tibet Plateau

Corresponding Author: Professor Jinren Ni

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript of Li., et al presents a large-scale metagenomic study of DNA viral communities spanning glaciers, plateau rivers, and downstream plains across the Qinghai–Tibet Plateau (QTP). Leveraging 682 metagenomes (597 river samples from the Yangtze, Yellow, Lancang, and Yarlung Tsangpo rivers, and 85 glacial metagenomes from the TG2G project) the authors generate the Glacier-to-River Virome Catalogue (GtRVC) comprising 36,358 non-redundant vOTUs and 897,250 viral protein clusters. They integrate viral identification, host prediction, genomic characterization, and functional annotation to explore how environmental gradients shape viral diversity, genomic traits, infection strategies, virus–host interactions, and distributions of auxiliary metabolic genes (AMGs).
The study reports a clear biogeographic structuring of viral communities, with viral diversity increasing downstream and with strong regional. A major finding is the systematic shift from temperate to lytic viral lifestyles along the elevation gradient, accompanied by changes in virus–host relative abundance patterns, inferred co-evolutionary dynamics, and defense/anti-defense gene repertoires. Viral hosts differ markedly across regions and include broad bacterial and archaeal lineages.
The authors further identify extensive viral functional potential across nitrogen, sulfur, phosphorus, and methane cycling, showing that AMGs related to nutrient scavenging are enriched in glaciers, whereas those supporting denitrification and anaerobic ammonium oxidation increase downstream. They also examine pathogen-associated MAGs and report a decline in virulence and antibiotic-resistance–carrying hosts from glacier to plain rivers, alongside evidence of intensified virus–pathogen interactions in glacial environments. Finally, microdiversity analyses suggest distinct evolutionary pressures on viral generalists and specialists across habitats.
Overall, the manuscript aims to position viruses as key ecological regulators in Asia's major headwater systems and to provide a foundational resource for understanding viral adaptation and ecosystem function in rapidly changing cryosphere-influenced environments.

However, several issues overshadow the true relevance of their work. One of the major issues is AMGs and the functional annotation of viral genes. Despite applying different annotation pipelines, the authors fail to provide a comprehensive annotation of relevant viral functions. The current broad classification of genes transforms the dataset into just another catalog among those that exist today. It is crucial that authors delve deeply into the genes within the broad pathways described in the main text to provide clear evidence of the relevant viral functions. A good example of this type of good practice is the following work, although not about viruses, it provides a good description of what readers look for in functional annotations https://doi.org/10.1038/s41564-023-01322-0.

Another important issue is the sections that suggest viruses from glacials preferentially infect cold-adapted or pathogenic hosts, as
the manuscript fails to provide relevant evidence for this claim. In general, it is unclear what proportion of the total cold-adapted microbes are infected by viruses, or even whether this proportion is statistically significant. The same problem occurs with pathogenic microbes: what are the ratios, and are they significant?

Specific comments are below.

RESULTS

Line 96-97: “clustered at 95% ANI, 90% coverage.” Why did you choose this threshold? A gold standard is to use 95% average nucleotide identity over 85% alignment fraction. https://doi.org/10.1038/nbt.4306. Please include a rationale sentence for your threshold choice in the main text.

Line 99: “Protein clustering (80% coverage, 60% AAI)”: There is no reference for these thresholds.
Lines 127-141: The relative VHR section depends on a small portion of the total viruses (~10%). A cautionary note regarding the methodological limitations of these results should be included in the main text. Also, tone down these findings.
Lines 141-151: These analyses based on the host can be easily misinterpreted. For example, “suggesting viral targeting of cold-tolerant microbes.” Indeed, your analysis suggests that glaciers harbor cold-tolerant microbes that can be infected by viruses, as any other microorganism. I think this section does not contribute to the main focus of this work and can be easily removed.

Line 169-172: “A considerable proportion of viral ORFs were annotated as
“Energy production and conversion (C)”, “Carbohydrate transport and metabolism (G)”, “Lipid transport and metabolism (I)” and “Inorganic ion transport and metabolism (P)”, indicating the potential viral involvement in host respiration, nutrient utilization, and facilitate viral manipulation of host membrane composition.”
It is difficult to establish this kind of conclusion without specifying the particular functions within those broad classifications. So, what specific functions support viral involvement in host respiration, nutrient utilization, and facilitate viral manipulation of host membrane composition?

Lines 180-188: Again, these broad classifications do not contribute to the understanding of viral auxiliary metabolism. Please list some relevant genes for each mentioned category in the main text.

I suggest unifying the two sections of AMGs, despite the different annotation pipelines being used, the final purpose is the same.

DISCUSION

Lines 294-298: In extreme glacial and plateau river environments, temperate viruses are dominant. In contrast, downstream nutrient-rich environments favor lytic infection.
Line 305-307: This suggests that glacial systems may align with the Kill-the-Winner (KtW) model, whereas downstream communities follow Piggyback-the-Winner (PtW) dynamics.
This is a clear contradiction because KtW indicates a dominance of lytic viruses, while PtW indicates a dominance of lysogenic viruses. Therefore, an explanation is necessary.

Line 329-331: “In addition, viruses in cold environments preferentially infected hosts carrying cold-adaptation genes, supporting co-adaptation to extreme conditions”
How does this support co-adaptation?

Line 354-361: Why not consider superinfection exclusion as another beneficial option for lysogens?

Reviewer #2

(Remarks to the Author)
Manuscript review for:

‘Viral lifestyle shift regulates eco-dynamics along glacier-to-river gradients’
by Li et al.

Summary:

This study by Li et al. aims at investigating the previously underexplored viral communities in the Qinghai-Tibet Plateau, by in depth computational analysis of metagenomes that the authors obtained from the Yangtze, Yellow, Lancang, and Yarlung Tsangpo rivers and publicly available data sets from glacial metagenomes. The overarching goal of the study is to reveal novel features related to the diversity, distribution, relative abundance, and lifestyle of viruses in this important geographical location that encompasses glacier, plateau river and plain river habitats as a downstream gradient. Secondary objectives are the characterization of potential host species, co-evolutionary hos-virus dynamics, functional characteristics of viruses (including the presence of auxiliary metabolic genes and their type), characterization of specialist vs generalist viruses, and detection of virulence factors and antimicrobial genes. As such this is a very important study, because very few studies have looked at viruses and their hosts on such a continental, land-based scale. And although the study is purely computational in nature, it does reveal many novel insights on viruses and their hosts in this important part of the world, which is prone to anthropogenically driven ecosystem changes.

General Comments:

Although I found the manuscript very interesting and insightful, I also found it very disjoined and a bit all over the place due to the huge amount of data that the authors are showing and the many different analysis/directions that they explore through these datasets. It’s an extremely ambitious body of work, and I don’t think that all that is currently in the manuscript should be

there. The manuscript contains 5 major figures and 15 supp. figures. In addition, very few figures are explained with the level of detail that is required to fully understand them. This is specially important for readers that may not be bioinformaticians. I believe that a lot of these issues stem predominantly from the fact that the authors try to explore too many different themes within this one manuscript.

For example, the current themes that I identified are:

1. Virus distribution and composition
2. Virus and host phylogeny
3. Virus-host interactions
4. Virus genomic features and function
5. Specific focus on AMGs
6. Pathogens and associated viruses
7. Micro-diversity and evolutionary dynamics of generalist and specialist viruses

These are too many different themes to cover in one manuscript, so the result of this is that none of these themes are actually explored in a clear way with the required detail that each and one of them warrants. They are all important and the bioinformatic skills of the researchers can definitely address most of them but not in the way they are all currently presented. It's a classical example in which a "less is more" approach should be taken.

Specific Comments:

Here I provide specific comments on the various section of the manuscript. Many more additional comments can be found on the tracked-changes document of the manuscript which I have submitted together with this review.

Introduction:

One question that came up frequently during my review was "Why is the Qinghai-Tibet Plateau (QTP) considered an extreme environment?" I do not doubt that it is – but I think that adding something that explains in what way this area can be considered extreme would be helpful in putting the subsequent results into context. What specifically in its extremity could be influencing viral populations or host populations? Likewise, I imagine that there is a difference in the environments of the glaciers and the plains. Some more context as to what those are would be helpful.

Line 68 – "In extreme environments, they [viruses] may drive host adaptation through auxiliary metabolic genes (AMGs) or serve as reservoirs for antibiotic resistance and virulence factors". This is accurate but leads to the suggestion that such things do not happen in non-extreme environments, which is not the case. Clarification would be appreciated here.

Results:

Line 100 – The authors report that 99.3% of vOTUs had no hits to IMG/VR. Given the cutoffs that were used this makes sense but I'd be interested in hearing also what the ones that did hit were.

Line 104 – How many viral clusters are there in total? Even a percentage of total would make this more meaningful to the reader.

Line 110 – BACPHLIP operates on the following assumptions: it assumes you're inputting phage data, it assumes the phage genome is complete, and it assumes that all inputs are virulent until proven otherwise. The authors here inputted genomic data that was not complete, as only 9159 sequences were complete (noted in line 97), and included 663 sequences that were later identified as NCLDV or virophage (noted in line 116). I would go through to remove any sequences identified as temperate that were from NCLDV or virophage sequences, and would caution against including sequences that were of medium quality, as they might not be classified properly.

Line 142 – Extreme conditions in QTP were never explained.

Line 152 – This entire section is only referring to material in the supplemental. Is that appropriate? I feel like there's too much material in the supplementary and some should be brought back into a main figure.

Line 177 – The status of many of these as AMGs is arguable, specifically replication as that is a core viral process. See https://www.nature.com/articles/s41564-025-02095-4 A proper discussion on this is warranted.

Line 198 – Based on Figure 4a, it appears like there was only 1 nrfA gene detected, which means saying it was "most common in glaciers" a bit strange.

Line 222 – In my view classifying something as a pathogen simply because it has a virulence gene seems like an oversimplification.

Line 239 – Another section where all material referenced is in the supplemental. I would suggest to think about if any of this information should be promoted to main figure status.

Discussion:

Line 319 – I think it is an overstatement to say AMGs were “extensively involved” in nitrogen and methane cycling, when (based on Figure 4) it appears no gene involved in these two pathways cleared more than 10.

Methods:

Line 393 – The authors only filtering through 0.22 µm filters which means that they would lose a ton of viruses this way. The fact that all the data refers to only cell associated viruses and not those that are free passing though this filter cutoff should be noted.

Line 425 – Something that may be a bit of a red flag here is the use of six identifiers? This is a very wide net to get viral sequences, and it means that there are likely a lot of false positives that could be cropping up. There is, of course, something to be said about the fact that each identifier has its blind spots and a combination can get you more than if you had just used a single identifier, but this needs to be emphasized more,

Line 437 – See previous notes about BACPHLIP. Here the authors say that they use default parameters, and documentation for BACPHLIP clearly state the assumptions it uses, so they haven’t changed anything with how BACPHLIP functions.

Figures:

Overall – The colors of Plateau and Plains are just similar enough to make them hard to distinguish in certain graphs. For example, Figure 1b, 3a, or Figure S6b. Would recommend they make them easier to distinguish. Already stated a few times but there’s a lot of stuff in supplemental. Might be better to be less picky about what goes in the main figure section.

Figure 1 – A map of China with the location of the QTP would be nice. I can’t tell based on the sampling location where in Asia this is.

Figure 3 – The colors are egregious here – I can’t tell gray from green from blue very well.

Figure 4 – I find A and B to be particularly hard to parse. B in particular – Saying average relative abundance and then having a scale that only goes up to 6 is a bit strange. They also don’t follow the order that they do in panel A of that figure which makes it more challenging too.

Reviewer #3

(Remarks to the Author)
The authors have used a considerable number of metagenomes (n=682) across a spatial (Tibetan plateau rivers) and temporally gradient – samples from 2014 to 2018 to produce a database of virus data – of which they have also accessed ecologically. The authors applied several programmes and methods to carry out each analyses, for example several programmes to identify viral contigs/MAGs – producing a high confidence dataset. The authors identify interesting spatial patterns, such as differences in genome size, GC content, and gene functional annotations, generalists vs specialists, etc. – showcasing ecological advantages within specific environments. HGT, alongside selection typing (evolution) is also explored in association with lytic and lysogenic lifestyles to further profile these viruses. Ultimately, the authors showcase the viral diversity and ecology of viruses along this gradient, and highlight their implied impacts on the system. These methods are clearly reproducible, the methods are the standards expected in the field (and above), and these data/conclusions will be interesting to other researchers in the future.

Comments:

The filtration methods do have a bias against large DNA viruses (Nucleocytoviricota) in that they are mostly excluded from the pore size used – which is fine. However, it may be worth commenting that these patterns are reflected in these specific viral groups, it is not entirely clear that giant viruses would have the exact patterns shown – although I imagine some of the genes with specific ecological advantages would overlap in environment. RNA viruses are also excluded as well (unless the DNA stages of some were captured). I would wonder if the GC content and genome size phenomena would also be observed w/ giant viruses. This could be included late in the discussion, but also in certain cases bacteriophages are stated to be the most prevalent DNA viruses (line 280), which very well may be the case but giant viruses (also DNA) were not explored accurately.

The discussion of VHR (lines 301 and beyond) might need to consider that viral DNA presence does not always indicate successful viruses, many viral progeny cannot successfully infect hosts. How do the authors think this could impact VHR? Potentially glacier environments a higher production of viruses for successful infection. VHR discussion should be accompanied by discussions of viral success as well. This also cannot be fully solved via flow cytometry/viral particle counts. Potentially gene expression could help (metatranscriptomes) using infection markers.

The filtration methods do have a bias against large DNA viruses (Nucleocytoviricota) in that they are mostly excluded from

the pore size used – which is fine. However, it may be worth commenting that these patterns are reflected in these specific viral groups, it is not entirely clear that giant viruses would have the exact patterns shown – although I imagine some of the genes with specific ecological advantages would overlap in environment. RNA viruses are also excluded as well (unless the DNA stages of some were captured). I would wonder if the GC content and genome size phenomena would also be observed w/ giant viruses. This could be included late in the discussion, but also in certain cases bacteriophages are stated to be the most prevalent DNA viruses (line 280), which very well may be the case but giant viruses (also DNA) were not explored accurately.

The discussion of VHR (lines 301 and beyond) might need to consider that viral DNA presence does not always indicate successful viruses, many viral progeny cannot successfully infect hosts. How do the authors think this could impact VHR? Potentially glacier environments a higher production of viruses for successful infection. VHR discussion should be accompanied by discussions of viral success as well. This also cannot be fully solved via flow cytometry/viral particle counts. Potentially gene expression could help (metatranscriptomes) using infection markers.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have made several revisions in response to the reviewers' comments, resulting in notable improvements to the manuscript. While many concerns have been addressed, some issues from the initial round of review remain only partially resolved, particularly those related to the interpretation of AMG's results and cold-adaptations. The following outlines these outstanding points for further consideration.

Lines 215-221: Please clarify how your findings compare to those in https://doi.org/10.1128/msystems.00396-21. Given that your cold-adaptation section is succinct, provide a discussion highlighting similarities or differences to this reference.

Lines 222-325: This section remains lengthy and somewhat disorganized. Most annotated viral proteins are common in viral genomes and may not be directly relevant to your model system. For example, detecting "soluble lytic murein transglycosylases (related phage lysozyme – muramidase was also detected) and N acetylmuramoyl-L-alanine amidase," is expected, as viruses typically encode lysozymes to lyse their hosts. Thus, annotation of these is routine in phage genomes. To improve focus and clarity, consider emphasizing the genes discussed in lines 436-488.

Reviewer #3

(Remarks to the Author)
I want to thank the authors for their thorough responses. I am content with the modifications and justifications.

Response to the Reviewers' comments

Response to Reviewer #1

We thank the Reviewer for the thoughtful and constructive feedback on our manuscript. We have carefully considered all comments and have revised the manuscript accordingly. Our point-by-point responses are provided below.

General Comment: The manuscript of Li., et al presents a large-scale metagenomic study of DNA viral communities spanning glaciers, plateau rivers, and downstream plains across the Qinghai–Tibet Plateau (QTP). Leveraging 682 metagenomes (597 river samples from the Yangtze, Yellow, Lancang, and Yarlung Tsangpo rivers, and 85 glacial metagenomes from the TG2G project) the authors generate the Glacier-to-River Virome Catalogue (GtRVC) comprising 36,358 non-redundant vOTUs and 897,250 viral protein clusters. They integrate viral identification, host prediction, genomic characterization, and functional annotation to explore how environmental gradients shape viral diversity, genomic traits, infection strategies, virus–host interactions, and distributions of auxiliary metabolic genes (AMGs).

The study reports a clear biogeographic structuring of viral communities, with viral diversity increasing downstream and with strong regional. A major finding is the systematic shift from temperate to lytic viral lifestyles along the elevation gradient, accompanied by changes in virus–host relative abundance patterns, inferred co-evolutionary dynamics, and defense/anti-defense gene repertoires. Viral hosts differ markedly across regions and include broad bacterial and archaeal lineages.

The authors further identify extensive viral functional potential across nitrogen, sulfur, phosphorus, and methane cycling, showing that AMGs related to nutrient scavenging

are enriched in glaciers, whereas those supporting denitrification and anaerobic ammonium oxidation increase downstream. They also examine pathogen-associated MAGs and report a decline in virulence and antibiotic-resistance–carrying hosts from glacier to plain rivers, alongside evidence of intensified virus–pathogen interactions in glacial environments. Finally, microdiversity analyses suggest distinct evolutionary pressures on viral generalists and specialists across habitats.

Overall, the manuscript aims to position viruses as key ecological regulators in Asia's major headwater systems and to provide a foundational resource for understanding viral adaptation and ecosystem function in rapidly changing cryosphere-influenced environments.

However, several issues overshadow the true relevance of their work. One of the major issues is AMGs and the functional annotation of viral genes. Despite applying different annotation pipelines, the authors fail to provide a comprehensive annotation of relevant viral functions. The current broad classification of genes transforms the dataset into just another catalog among those that exist today. It is crucial that authors delve deeply into the genes within the broad pathways described in the main text to provide clear evidence of the relevant viral functions. A good example of this type of good practice is the following work, although not about viruses, it provides a good description of what readers look for in functional annotations https://doi.org/10.1038/s41564-023-01322-0.

Another important issue is the sections that suggest viruses from glacials preferentially infect cold-adapted or pathogenic hosts, as the manuscript fails to provide relevant evidence for this claim. In general, it is unclear what proportion of the total cold-adapted microbes are infected by viruses, or even whether this proportion is statistically significant. The same problem occurs with pathogenic microbes: what are the ratios, and are they significant?

Response: We thank the Reviewer for the positive assessment of our study's scope and aims, and for highlighting key areas for improvement. We agree that a more detailed

functional annotation of viral genes and stronger evidence for host infection patterns are crucial. In response:

(1) Regarding AMGs and the functional annotation of viral genes: We have significantly expanded our functional analysis beyond broad pathway-level categorization. For each metabolic pathway discussed, we now explicitly report and compare the key viral genes with the highest counts or differential abundance across the glacier-to-river gradient. Furthermore, we interpret viral gene functions within the comprehensive "auxiliary viral gene" (AVG) framework (as proposed in Martin et al., *Nature Microbiology*, 2025[1]), distinguishing between auxiliary metabolic genes (AMGs), auxiliary physiology genes (APGs), and auxiliary regulatory genes (AReGs). We also referred to the study you provided[2] and have strengthened our functional interpretations accordingly. These provide a more nuanced and evidence-based interpretation of viral functional potential.

(2) Regarding evidence for infection of cold-adapted and pathogenic hosts: We have revised our analyses and interpretations to address this concern.

For cold adaptation, we acknowledge that our initial wording implying viral "targeting" or "preference" was an overinterpretation. The higher density of cold-adaptation genes in hosts associated with plateau viruses likely reflects the genomic background of the resident microbial community, not an active viral strategy. We have therefore removed claims of "preferential infection" and refocused the narrative on the environmental filtering of both hosts and their viruses. New analysis shows viruses themselves carry a higher density of cold-adaptation-related genes in glacial/plateau environments, suggesting direct environmental selection on viral genomic content.

For pathogens, we provide quantitative infection statistics. Based on our host prediction results, the proportion of infected genomes was consistently higher in pathogens (including those with/without ARGs, over 80%) than for non-pathogens (~ 60%) across all regions. Statistical tests on the number of viruses per host and the Virus-to-Host Ratio (VHR) further indicate that pathogen-associated genomes, particularly

ARG-carrying ones, tend to be associated with a greater number and density of viruses. We have toned down causal language and present this as an observed pattern.

Below are our detailed responses to each specific comment.

Comment No.1: Line 96-97: "clustered at 95% ANI, 90% coverage." Why did you choose this threshold? A gold standard is to use 95% average nucleotide identity over 85% alignment fraction. https://doi.org/10.1038/nbt.4306. Please include a rationale sentence for your threshold choice in the main text.

Response: We thank the reviewer for catching this discrepancy. The "90% coverage" in the Results was a typographical error. We indeed used the standard species-level threshold of 95% ANI over 85% alignment fraction for vOTU clustering, as correctly stated in the original Methods. We have corrected the Results text and added a brief rationale in the Methods citing the suggested reference.

Comment No.2: Line 99: "Protein clustering (80% coverage, 60% AAI)": There is no reference for these thresholds.

Response: We have added clarification in the Methods. While vOTU clustering follows a standardized species-level threshold, protein clustering parameters are more flexible. We used 60% AAI over 80% coverage, a threshold commonly employed in large-scale viral protein clustering studies[3–5] to group homologous sequences while preserving functional diversity. Corresponding references are now cited, and the added brief rationale in the Material and Methods section (Lines 645-648) reads:

"These thresholds are widely used in large-scale viral comparative genomics, with the vOTU-level cutoff reflecting species-level delineation[96] and the protein-level cutoff designed to identify homologous proteins across diverse viral lineages while preserving functional diversity[63,97,98]."

Comment No.3: Lines 127-141: The relative VHR section depends on a small portion of the total viruses (~10%). A cautionary note regarding the methodological limitations of these results should be included in the main text. Also, tone down these findings.

Response: We agree. Virus-host prediction is inherently conservative, and our high-confidence assignments cover ~10% of vOTUs but ~70% of MAGs, a trade-off between precision and recall common in metagenomic studies[6–8]. In this study, virus–host associations were inferred using three alignment-based approaches that rely on distinct virus-host genomic features, including CRISPR spacer matching, tRNA sequence matching, and nucleotide sequence homology. These tools are widely used in previous studies[9–13]. The trade-off between accuracy and recall represents a core limitation of current virus–host prediction strategies in metagenomic studies[14]. To address this concern, we have added a cautionary note in the Discussion, clarifying that VHR trends are based on this host-linked subset and may not represent the entire community. We have toned down related conclusions accordingly.

The added descriptions for methodological limitations in **Discussion** (Lines 424-429 in the new version) read:

"It's noticeable that host prediction results included only a subset of viral populations, therefore VHR-related results may not reflect patterns across entire viral communities. However, the hosts predicted to be infected account for nearly 70% of recovered MAGs, suggesting that the observed VHR trends still provide meaningful insights into virus–host abundance relationships along the glacier-to-river gradient."

Comment No.4: Lines 141-151: These analyses based on the host can be easily misinterpreted. For example, "suggesting viral targeting of cold-tolerant microbes." Indeed, your analysis suggests that glaciers harbor cold-tolerant microbes that can be infected by viruses, as any other microorganism. I think this section does not contribute to the main focus of this work and can be easily removed.

Response: We agree with this careful interpretation. The association between viruses and hosts with more cold-adaptation genes is likely a reflection of the host community's genomic background in cold environments, not evidence of active viral targeting. We have therefore removed the claim of "viral targeting" from the Abstract and Results. We have retained and reframed the analysis of cold-adaptation genes encoded by viruses themselves (new analysis), which shows a direct environmental signal in viral genomes, and integrated this into the broader discussion of environmental adaptation.

Comment No.5: Line 169-172: "A considerable proportion of viral ORFs were annotated as "Energy production and conversion (C)", "Carbohydrate transport and metabolism (G)", "Lipid transport and metabolism (I)" and "Inorganic ion transport and metabolism (P)", indicating the potential viral involvement in host respiration, nutrient utilization, and facilitate viral manipulation of host membrane composition." It is difficult to establish this kind of conclusion without specifying the particular functions within those broad classifications. So, what specific functions support viral involvement in host respiration, nutrient utilization, and facilitate viral manipulation of host membrane composition?

Response: We agree that broad COG categories are insufficient for specific functional claims. We have revised this section. Instead of making broad inferences, we now describe the most frequent or characteristic genes within these COG categories (e.g., DNA replication, recombination and repair, cell wall, posttranslational modification) and discuss their potential roles based on established literature, providing a more precise and conservative interpretation.

The revised contents in Results section (Lines 229-253 in the new version) read:

"Each category was characterized by its most represented functions, for example, …"

Comment No.6: Lines 180-188: Again, these broad classifications do not contribute to the understanding of viral auxiliary metabolism. Please list some relevant genes for each mentioned category in the main text.

Response: As per this comment and Comment No.5, we have revised the manuscript to include gene-level details for key functional categories. We now list representative genes (e.g., specific AMGs for defense system like restricted modification, bacterial motility) to substantiate the functional discussion and avoid over-reliance on broad classifications.

The revised contents in Results Section (Lines 257-291 in the new version) read:

" Among viruses enriched in different regions, genes associated with DNA replication, prokaryotic defense systems, and transcriptional regulation exhibited higher occurrence ratios than other functional categories, particularly in glacier-enriched temperate viruses, where their occurrence ratios exceeded 10% …"

Comment No.7: I suggest unifying the two sections of AMGs, despite the different annotation pipelines being used, the final purpose is the same.

Response: Thank you for this helpful suggestion. We have merged the two AMG-related sections into a cohesive narrative. In the revised Discussion, we interpret these findings under the integrated "auxiliary viral gene" (AVG)[1] framework, which provides a clearer conceptual structure.

Comment No.8: Lines 294-298: In extreme glacial and plateau river environments, temperate viruses are dominant. In contrast, downstream nutrient-rich environments favor lytic infection. Line 305-307: This suggests that glacial systems may align with the Kill-the-Winner (KtW) model, whereas downstream communities follow Piggyback-the-Winner (PtW) dynamics. This is a clear contradiction because KtW

indicates a dominance of lytic viruses, while PtW indicates a dominance of lysogenic viruses. Therefore, an explanation is necessary.

Response: We thank the reviewer for highlighting this apparent contradiction. Our re-analysis using a more stringent lifestyle prediction strategy (considering Reviewer 2's Comment 5) confirmed the dominant lifestyle shift and consistent decrease of VHR with increasing host relative abundance. We recognize that our initial interpretation linking KtW/PtW to absolute dominance of lytic/lysogenic viruses was imprecise. These models describe ecological strategies or tendencies along a continuum influenced by host density and environmental conditions, not strict predictions about viral population counts.

An extensive review of the literature suggests that the pattern observed in our study occurred across many environmental systems[15]. In fact, among 11 previously surveyed ecosystems, 8 reported a decline in virus-to-microbe ratio (VMR) with increasing host abundance. This study consistently observed a peak VMR at approximately ~10^6 microbes mL^{-1} or g^{-1}, followed by a decline as host density increased further. This pattern has been interpreted as a dynamic continuum of virus–host interaction strategies[15,16], encompassing Piggyback-the-Loser (PtL), Kill-the-Winner (KtW)[17–19], and Piggyback-the-Winner (PtW)[20] dynamics as follows. Within this integrated framework, PtL, KtW, and PtW do not represent contradictory models but rather different segments of a continuous virus–host interaction spectrum along host density gradients.

(1) Piggyback-the-Loser (PtL): Under oligotrophic or otherwise extreme conditions, host densities are low and microbial growth rates are slow. Energy supply and nutrient limitation suppresses the expression of lytic genes and prolongs the decision window for lysogenic commitment, thereby increasing the likelihood of coinfection and subsequent genome integration. Viruses tend to adopt lysogenic

strategies, integrating into host genomes rather than killing hosts and host cells effectively serve as a long-term refugium for viral persistence.

(2) Kill-the-Winner (KtW): At intermediate host densities, microbial growth rates and metabolic efficiency are elevated. Under these conditions, ATP-dependent host proteases become more active and can cleave phage lytic repressors[21], while the expression efficiency of lytic genes increases. In addition, prophages accumulated during PtL or PtW phases may be induced into the lytic cycle. Viruses preferentially infect and lyse the most abundant and fastest-growing host populations, resulting in frequency-dependent control of dominant hosts.

(3) Piggyback-the-Winner (PtW): At very high host densities, such as in eutrophic aquatic systems or host-associated environments (e.g., gut microbiomes), energy availability is high but microbial metabolic efficiency may decline. High host densities increase the frequency of multiple infections, leading to elevated phage genome copy numbers within cells and increased production of lytic cycle repressors[21,22]. Under these conditions, viruses again tend to favor lysogenic lifestyles rather than lytic.

Applying this framework to our study system, glacial and plateau rivers are characterized by cold temperatures, oligotrophic conditions, high radiation and relatively low host metabolic activity compared to downstream plain rivers. These conditions are consistent with the PtL regime, in which viruses preferentially adopt lysogenic strategies and persist within host cells to ensure long-term survival. In contrast, downstream plain rivers receive greater external nutrient inputs and experience more favorable temperatures, resulting in higher host densities. Reported host densities in lake and river systems[15] often exceed ~10^6 microbes mL^{-1} or g^{-1}, comparable to conditions in our plain river sites, placing these environments within the PtW regime. In additions, higher host diversity and more heterogeneous host communities in plain rivers may favor viruses with broader host ranges[23], which can

reduce per-host infection efficiency and lead to lower VHR at the level of individual host taxa as host abundance increases.

We note that the PtW, KtW, and PtL frameworks proposed in the literature do not explicitly define quantitative expectations for the relative abundance or density ratios between temperate and lytic viruses. These models describe strategic preferences in viral infection modes as host density and environmental conditions change. In the original article of the PtW model[15], enhanced lysogeny was inferred from multiple lines of evidence, including a decline in virus-to-microbe ratios (VMR) based on virus-like particle (VLP, lytic and free viruses) counts, increased abundance of lysogeny-associated markers (e.g., integrase, excisionase, and prophage reads), and experimental observations indicating reduced lytic activity at high host density. Importantly, these observations were interpreted as a shift in viral strategy preference, rather than a strict prediction that temperate viruses must numerically dominate lytic viruses. From this perspective, our observation of decreasing temperate virus abundance and proportion along the glacier-to-river gradient does not inherently conflict with these ecological frameworks, as the models describe relative tendencies in infection strategies rather than direct abundance relationships between temperate and lytic viral populations.

Nevertheless, inferences based on metagenomic data alone are inherently limited. Future studies integrating direct measurements, such as virus-like particle (VLP) and microbial cell counts, flow cytometry, or other experimental approaches, will be essential to quantify virus-host abundance and density relationships in natural environments, thereby providing stronger empirical constraints for virus–host ecological theory.

We have revised this section extensively. We now interpret our observations (decreasing VHR with increasing host abundance) within an integrated framework where Piggyback-the-Loser (PtL) dynamics may prevail in oligotrophic glacial settings (favoring lysogeny for persistence), while Piggyback-the-Winner (PtW) dynamics may

become more relevant in downstream, higher-density communities (where frequent coinfection can also promote lysogeny). We clarify that the observed numerical dominance of virulent viruses downstream is not inherently contradictory to a PtW strategy, as the model describes a shift in infection mode preference, not a requirement for temperate viruses to outnumber lytic ones. We have toned down the model assignments and focused on describing the observed patterns and their plausible ecological context.

The revised text in Results section (Lines 185-187 in the new version) reads:

"However, the virus-host relative abundance ratio (VHR) decreased with increasing host relative abundance within the glacier-to-river gradient regardless of lifestyle (SI Fig. 6b)"

The revised text in Discussion section (Lines 415-424 in the new version) correspondingly reads:

"Previous studies indicate that virus–host interactions follow a continuum of ecological strategies[114,115]. Virus-to-microbe ratios often peak at intermediate host densities and decline as host abundance increases or decreases, reflecting continuum of strategies among Piggyback-the-Loser (PtL)[115], Kill-the-Winner (KtW)[116], and Piggyback-the-Winner (PtW)[114] dynamics. Under cold, oligotrophic and high radiation conditions characteristic of glacial and plateau rivers, low host density and slow growth favor lysogenic persistence when lytic replication is inefficient, consistent with PtL-like dynamics. In plain rivers with higher host densities, increased encounter rates and frequent coinfection can again promote lysogeny (PtW)[117], and higher host diversity favors broader viral host ranges which may further reduce per-host infection efficiency[118]."

Comment No.9: Line 329-331: "In addition, viruses in cold environments preferentially infected hosts carrying cold-adaptation genes, supporting co-adaptation to extreme conditions". How does this support co-adaptation?

Response: We agree that the observed pattern does not constitute strong evidence for active co-adaptation. As also addressed in our response to Comment No.4, it primarily reflects the background genomic signature of hosts in cold environments. We have removed the term "co-adaptation" and revised the statement to a more cautious description of the observed association, avoiding causal implications.

Comment No.10: Line 354-361: Why not consider superinfection exclusion as another beneficial option for lysogens?

Response: This is an excellent point. Indeed, superinfection exclusion is an important benefit of lysogeny[24–26] to be considered. Temperate phages can protect their hosts from subsequent infections by related phages through immunity mechanisms, potentially enhancing host survival under certain conditions. We have revised the Discussion to acknowledge that superinfection exclusion is a potential benefit of lysogeny. We now discuss the trade-offs along the environmental gradient: while lysogeny may offer benefits like superinfection exclusion in stable conditions, the metabolic cost of prophage maintenance and the risk of induction under changing conditions (e.g., in dynamic plain rivers) may favor lytic strategies. This provides a more balanced perspective.

The revised contents in Discussion section (Lines 511-517 in the new version) read:

"Although temperate viruses may confer ecological advantages to their hosts, including auxiliary gene carriage and protection against other phages infections via superinfection exclusion[105,106], lysogeny also imposes metabolic and regulatory costs associated with prophage maintenance. In plain rivers dominated by virulent viruses,

these costs, together with possibility to prophage induction under environmental perturbations[107,108], may reduce the competitive fitness of lysogenized hosts."

Lastly, we would like to express our sincere thanks to the anonymous Reviewer for the thoughtful and constructive comments and suggestions, which is of great significance for the improvement of our manuscript quality.

Response to Reviewer #2

Summary: This study by Li et al. aims at investigating the previously underexplored viral communities in the Qinghai-Tibet Plateau, by in depth computational analysis of metagenomes that the authors obtained from the Yangtze, Yellow, Lancang, and Yarlung Tsangpo rivers and publicly available data sets from glacial metagenomes. The overarching goal of the study is to reveal novel features related to the diversity, distribution, relative abundance, and lifestyle of viruses in this important geographical location that encompasses glacier, plateau river and plain river habitats as a downstream gradient. Secondary objectives are the characterization of potential host species, co-evolutionary hos-virus dynamics, functional characteristics of viruses (including the presence of auxiliary metabolic genes and their type), characterization of specialist vs generalist viruses, and detection of virulence factors and antimicrobial genes. As such this is a very important study, because very few studies have looked at viruses and their hosts on such a continental, land-based scale. And although the study is purely computational in nature, it does reveal many novel insights on viruses and their hosts in this important part of the world, which is prone to anthropogenically driven ecosystem changes.

Response: We sincerely thank the Reviewer for the thorough and constructive evaluation, and for the positive assessment of our study's importance and scope. We have carefully considered all comments and have substantially revised the manuscript to address the concerns raised. Our point-by-point responses are below.

General comment: Although I found the manuscript very interesting and insightful, I also found it very disjoined and a bit all over the place due to the huge amount of data that the authors are showing and the many different analysis/directions that they explore through these datasets. It's an extremely ambitious body of work, and I don't think that

all that is currently in the manuscript should be there. The manuscript contains 5 major figures and 15 supp. figures. In addition, very few figures are explained with the level of detail that is required to fully understand them. This is specially important for readers that may not be bioinformaticians. I believe that a lot of these issues stem predominantly from the fact that the authors try to explore too many different themes within this one manuscript.

For example, the current themes that I identified are:

1. Virus distribution and composition
2. Virus and host phylogeny
3. Virus-host interactions
4. Virus genomic features and function
5. Specific focus on AMGs
6. Pathogens and associated viruses
7. Micro-diversity and evolutionary dynamics of generalist and specialist viruses

These are too many different themes to cover in one manuscript, so the result of this is that none of these themes are actually explored in a clear way with the required detail that each and one of them warrants. They are all important and the bioinformatic skills of the researchers can definitely address most of them but not in the way they are all currently presented. It's a classical example in which a "less is more" approach should be taken.

Response: We thank the Reviewer for the insightful summary and for the valuable critique regarding the manuscript's structure. We agree that integrating multiple analytical dimensions risked making the narrative appear fragmented.

We acknowledge that the initial presentation could have better connected these threads. In revision, we have:

(1) **Reorganized the narrative flow** to strengthen logical connections between sections. We expanded the environmental background section to better contextualize the glacier-to-river gradient and its ecological relevance.

(2) **Consolidated and streamlined analyses** (e.g., merging AMG sections, integrating key supplementary findings into main figures) to reduce redundancy.

(3) **Adopted the "auxiliary viral gene" (AVG) framework** to provide a unifying conceptual lens for discussing viral functional traits.

(4) **Added more detailed explanations** in both the main text and figure legends to enhance accessibility for readers beyond the bioinformatics community.

[**Notes for the relations among the themes**: themes 1 & 2 (Distribution/Phylogeny) establish *who* is present and *how* community structure changes. Themes 4 & 5 (Genomic features/AMGs) reveal *how* viruses are functionally adapted to different habitats. Themes 3 & 6 (Host Interactions/Pathogens) explore the ecological *consequences* of these shifts for host communities and potential risk dissemination. Theme 7 (Micro-diversity) provides an evolutionary perspective on *adaptation strategies*.]

Our overarching goal was to construct a multidimensional framework to understand how viral communities adapt and function along the glacier-to-river environmental gradient. The seven themes identified by the Reviewer are not independent explorations but complementary lenses addressing a unified ecological question: how do viral genomic features, host interactions, functional potential, and evolutionary dynamics shift systematically across this critical environmental transect?

We believe the revisions, following the reviewer's comments, have significantly improved the manuscript's structure and clarity while preserving its integrative ecological perspective.

Comment No.1: Introduction One question that came up frequently during my review was "Why is the Qinghai-Tibet Plateau (QTP) considered an extreme environment?" I do not doubt that it is - but I think that adding something that explains in what way this area can be considered extreme would be helpful in putting the subsequent results into context. What specifically in its extremity could be influencing viral populations or host

populations? Likewise, I imagine that there is a difference in the environments of the glaciers and the plains. Some more context as to what those are would be helpful.

Response: We thank the Reviewer for this essential suggestion. Providing clearer environmental context is crucial. We have expanded the Introduction to explicitly define the QTP as an extreme environment due to its high elevation, concomitant low temperatures and atmospheric pressure, intense ultraviolet radiation, and oligotrophic conditions. We also now include a new section in the Results (and a corresponding panel in Fig. 1) presenting contemporaneous environmental data (temperature, nutrients) and surface energy flux variables (solar radiation, latent/sensible heat) to quantitatively characterize the stark differences between plateau and plain river sites. This establishes a concrete physicochemical and energetic basis[27] for interpreting the observed biological patterns.

The revised contents in Introduction section (Lines 44-49, 54-73 in the new version) read:

"The QTP is widely regarded as an extreme environment due to its high elevation, low temperatures, intense ultraviolet radiation, hypoxia, and oligotrophic conditions[4-6]. At the same time, the QTP exhibits pronounced spatiotemporal heterogeneity, with strong southeast-to-northwest gradients in precipitation, air temperature, and the frequency and intensity of extreme climatic events[7-9]."

"The extreme environment of the Qinghai–Tibet Plateau imposes strong selective pressures on biological adaptation. Studies on macroscopic organisms have shown widespread tolerance to cold, hypoxia, and intense ultraviolet radiation at high altitudes. For example, low- and mid-altitude fish species are largely unable to migrate onto the plateau[17], plateau frogs exhibit reduced metabolic rates and metabolic reorganization under lower temperature and dissolved oxygen conditions[18] and plants form adaptation at morphological, physiological, and molecular levels through enhanced photosynthetic efficiency, strengthened antioxidant systems, and other molecular regulatory

mechanisms[19,20]. Host-associated microbial communities also contribute substantially to adaptation to the harsh plateau environment. Rapid and persistent shifts in gut microbiota have been observed in humans[21,22] and other plateau animals[23,24] following high-altitude exposure, with enriched microbial taxa linked to altered energy metabolism, antioxidant capacity, and physiological regulation under hypoxic and high-radiation conditions. Environmental microorganisms also show strong adaptation to plateau conditions, with total microbial biomass generally declining along elevational gradients, while functional investment in resource acquisition becomes increasingly pronounced[25,26]. Soil microbial communities on the plateau exhibit elevation-dependent reorganization toward facultatively anaerobic, stress-tolerant, and saprotrophy-dominated strategies[27]. Diverse plateau lakes harbor microbial communities exhibiting multifaceted functional adaptations related to carbon cycling, energy metabolism, and stress tolerance[28,29].”

The revised contents in Results section (Lines 112-134 in the new version) read:

“To provide a clearer environmental context for viral communities, we first compiled environmental and surface energy flux data for a subset of sampling sites to characterize environmental differences. Due to the lack of environmental metadata for glacier samples obtained from public databases, comparisons of environmental and heat flux variables were restricted to monitoring sites along plateau and plain rivers, with the exception of elevation (Fig. 1b). The mean elevations of glacier, plateau river, and non-plateau river sites were 5,168.85 m, 3,341.18 m, and 448.51 m, respectively, revealing a pronounced elevational gradient (Fig. 1b, Table S9). Plateau rivers exhibited significantly lower mean water temperatures (7.73 °C) compared to Plain rivers (14.60 °C). Similarly, several nutrient-related parameters were significantly lower in plateau rivers (mean values: DO = 7.73 mg/L, NH_4^+ = 2.23 mg/L, TP = 89.10 mg/L) than in Plain rivers (mean values: DO = 8.24 mg/L, NH_4^+ = 8.87 mg/L, TP = 432.12 mg/L). In contrast, surface energy flux variables showed the opposite pattern (Fig. 1b,

Table S9), with plateau rivers characterized by significantly higher values of surface solar radiation (srad = 84.61 w/m^2), surface latent heat flux (LE = 7.48 w/m^2), and surface sensible heat flux (Heat = 3.90 w/m^2) compared to plain rivers (srad = 63.36 w/m^2, LE = −24.73 w/m^2, Heat = −25.73 w/m^2). ”

The revised text in Discussion section (Lines 369-379 in the new version) reads:

“**Elevational gradients in physicochemical conditions and surface energy fluxes**

The observed elevational gradient is associated with clear differences in physicochemical conditions and surface energy fluxes between QTP and plain (Fig. 1b, Table S9). Plateau rivers exhibit lower temperatures and reduced nutrient concentrations, consistent with previous studies[102,103]. Differences in dissolved oxygen likely reflect the combined influence of temperature and reduced atmospheric pressure in QTP[104]. In contrast, surface energy fluxes, including solar radiation as well as latent and sensible heat fluxes, are elevated in plateau rivers, reflecting strong radiative forcing, large diurnal temperature variability, and persistent wind[105–107]. Together, these patterns characterize a distinct physicochemical and energetic flow for plateau, providing essential background context for interpreting biological patterns along the glacier-to-river gradient.”

Comment No.2: Introduction Line 68 – “In extreme environments, they [viruses] may drive host adaptation through auxiliary metabolic genes (AMGs) or serve as reservoirs for antibiotic resistance and virulence factors”. This is accurate but leads to the suggestion that such things do not happen in non-extreme environments, which is not the case. Clarification would be appreciated here.

Response: We agree and thank the Reviewer for the correction. Our intention was to emphasize the particular relevance of these processes in extreme settings, not to imply exclusivity. We have revised the sentence to (Lines 90-92 in the revised Introduction): “Viruses may drive host adaptation through auxiliary metabolic genes (AMGs) or serve

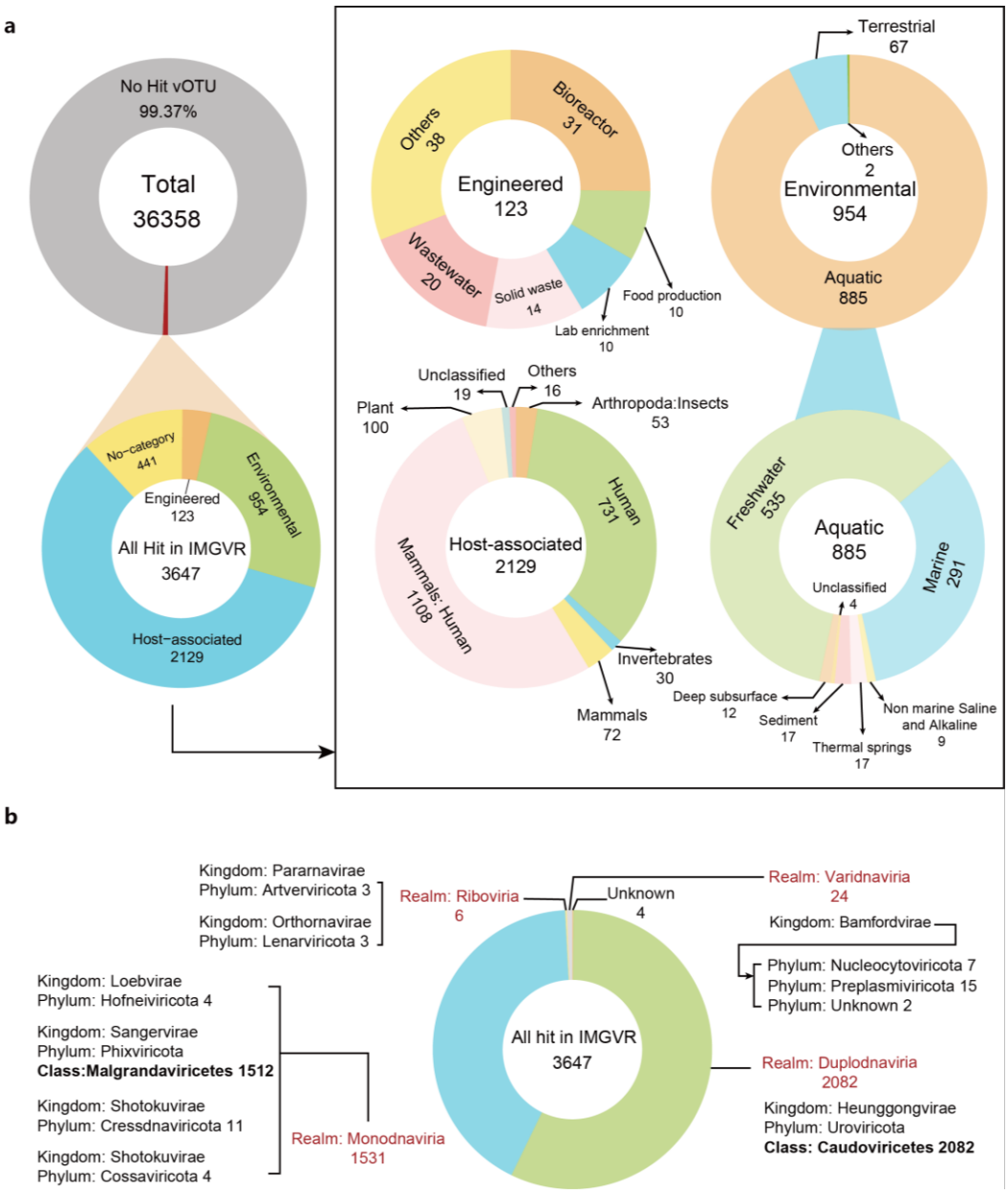
as reservoirs of antibiotic resistance and virulence factors that enhance competitive fitness[1,38–41].” This clarifies that these are general viral strategies, potentially under distinct selective pressures in extreme environments.

Comment No.3: Results Line 100 - The authors report that 99.3% of vOTUs had no hits to IMG/VR. Given the cutoffs that were used this makes sense but I’d be interested in hearing also what the ones that did hit were.

Response: This is a valuable point. We have now added a new supplementary figure (SI Fig. 2) visualizing the source environments and taxonomy of the GtRVC vOTUs that did have matches in IMG/VR. The matches were primarily to sequences from host-associated and environmental (freshwater/marine) sources. Taxonomically, over 98.5% belonged to the realms *Duplodnaviria* (class *Caudoviricetes*) and *Monodnaviria* (class *Malgrandaviricetes*). This analysis confirms the novelty of our catalogue while contextualizing the small fraction of known relatives.

The revised contents (Lines 144-150 in the new version) read:

“The majority of GtRVC viral sequences showing significant BLAST matches to the IMG/VR database were derived from host-associated and environmental sources. Within the environmental category, most matched viruses were associated with freshwater and marine habitats (SI Fig. 2a). Taxonomically, more than 98.5% of these matched viruses belonged to the realm *Duplodnaviria*, primarily represented by the class *Caudoviricetes*, and to the realm *Monodnaviria*, represented by the class *Malgrandaviricetes* (SI Fig. 2b).”



SI-Figure 2: BLAST-based comparison of GtRVC viruses with the IMG/VR database. (a) Distribution of GtRVC viral sequences that show significant BLAST matches to the IMG/VR database across different sample categories. Numbers indicate the number of viral sequences corresponding to each sample category. (b) Taxonomic composition of GtRVC viral sequences with BLAST matches to IMG/VR. Viral realms are highlighted in red, with numbers indicating the total number of matched viral sequences within each realm. Annotations at the kingdom and phylum levels are also shown within each realm to illustrate finer taxonomic resolution.

Comment No.4: Results Line 104 - How many viral clusters are there in total? Even a percentage of total would make this more meaningful to the reader.

Response: Thank you. We have added this information (Supplementary Table 8). The gene-sharing network analysis formed a total of 5,318 viral clusters (VCs). We now explicitly report that 58.8% of VCs were composed exclusively of plain river viruses, while only 0.45% contained sequences from all four source types (glacier, plateau, plain, RefSeq). These percentages provide a clearer sense of the endemicity and sharedness of viral populations across the gradient.

The revised text (Lines 151-157 in the new version) reads:

"The gene-sharing network analysis formed a total of 5,318 viral clusters (Table S8). Among these, only 0.45% of viral clusters contained sequences originating from glacier, plateau, plain, and RefSeq datasets simultaneously. 1.65% of viral clusters comprised viral sequences shared among glacier, plateau, and plain environments, while 18.1% of viral clusters were shared between plateau and plain rivers. Notably, viral clusters composed exclusively of viral sequences assembled in plain rivers accounted for 58.8% of all clusters (SI Fig. 1d)."

Comment No.5: Results Line 110 - BACPHLIP operates on the following assumptions: it assumes you're inputting phage data, it assumes the phage genome is complete, and it assumes that all inputs are virulent until proven otherwise. The authors here inputted genomic data that was not complete, as only 9159 sequences were complete (noted in line 97), and included 663 sequences that were later identified as NCLDV or virophage (noted in line 116). I would go through to remove any sequences identified as temperate that were from NCLDV or virophage sequences, and would caution against including sequences that were of medium quality, as they might not be classified properly.

Response: We thank the Reviewer for this critical methodological point. BACPHLIP is designed for bacteriophage genomes and performs optimally on complete phage

sequences, with the detection of certain conserved lysogeny-associated protein domains[28]. We have followed your recommendation rigorously. First, we excluded all NCLDV and virophage sequences from lifestyle prediction. Second, we restricted the analysis to high-quality and complete viral genomes only, as suggested. We confirm that the core finding, a significant decrease in the proportion of temperate viruses from glacier to plain, remains robust under this stricter filtering (see the Figure below). We discuss this choice and its potential implications (e.g., the underrepresentation of temperate phages in incomplete genomes is a known bias of domain-based tools like BACPHLIP) in the revised Methods and Discussion.

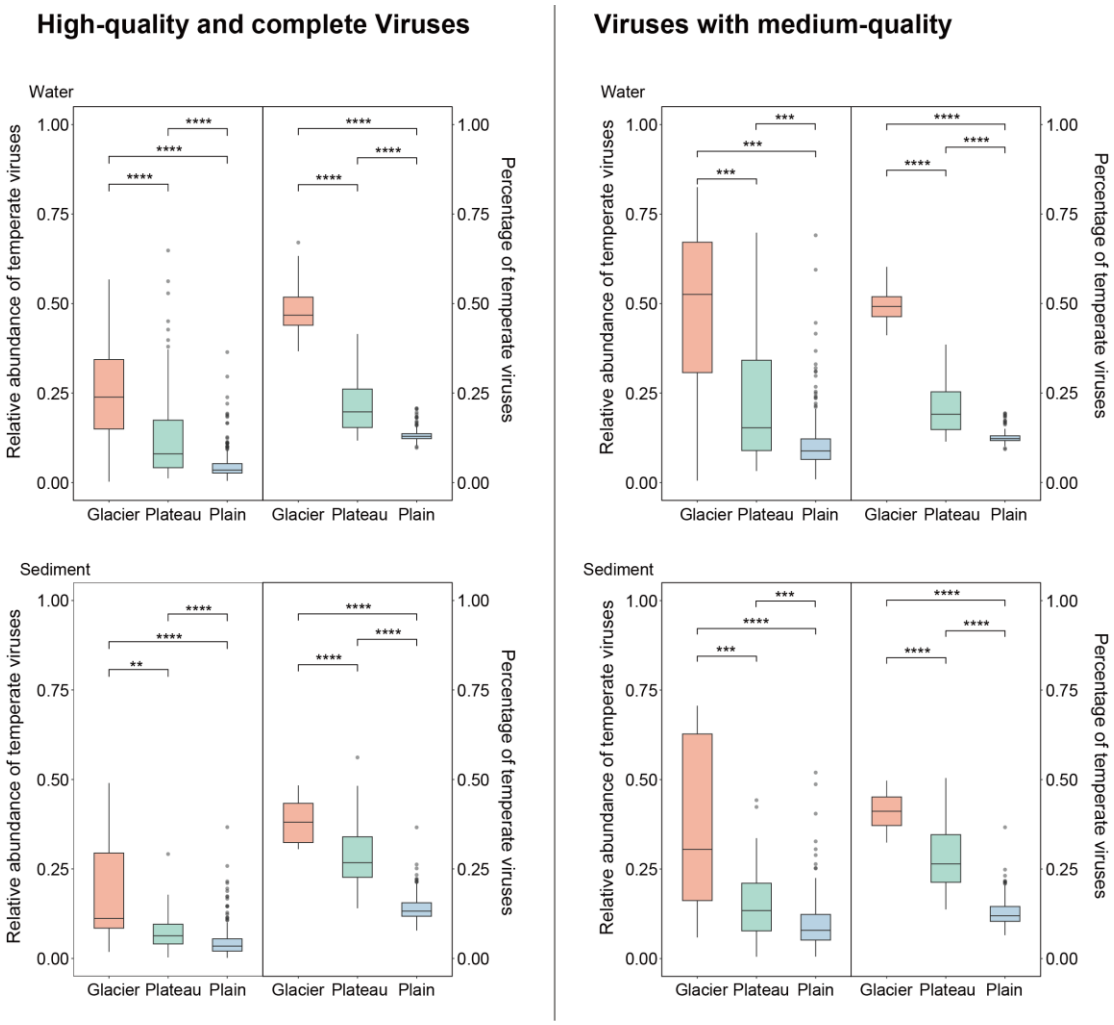


Figure: Changes in temperate virus ratios along the glacier-to-river gradient under different virus selection strategies. The left panels show the relative abundance

and proportion of temperate viruses based on high-quality and complete viral genomes from water and sediment samples, whereas the right panels include medium-quality, high-quality and complete viral genomes, illustrating temperate virus ratios in terms of abundance and counts across the same gradient. All comparisons exclude sequences classified as NCLDVs and virophages. Significance levels were estimated using the Wilcoxon rank-sum test, with asterisks denoting significance (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

As shown in the figure, the numerical proportion of temperate viruses remains largely consistent regardless of whether medium-quality genomes are included, indicating that BACPHLIP predicts broadly similar lifestyle ratios across viruses of different quality levels. In contrast, a more pronounced decrease is observed in abundance-based temperate ratios when restricting the analysis to high-quality and complete genomes. This pattern suggests that virulent viruses tend to be more abundant among high-quality and complete genomes, which is biologically plausible, as viruses that can be assembled into near-complete genomes are often those with higher representation in the community, potentially reflecting active lytic replication and release at the time of sampling.

However, we consider that, for community-level lifestyle patterns, restricting the analysis exclusively to complete viral genomes would provide an incomplete representation of the viral community, even though the overall trend and the purpose we analysis lifestyle remains robust across different virus selection strategies. Medium-quality vOTUs account for approximately 49% of the GtRVC dataset and therefore constitute a substantial component of the overall viral community. Notably, all medium-quality genomes were larger than 10kbp and 62% of these medium-quality genomes have estimated completeness exceeding 60%, providing not too short sequence length to recover lysogeny-associated conserved protein domains. As noted by the authors of BACPHLIP, the method exhibits an inherent classification asymmetry, whereby

temperate viruses are more likely to be misclassified as virulent due to the absence of detectable lysogeny-related domains, given the initial assumption that all input sequences are virulent. Consequently, temperate lifestyle predictions are generally more reliable than virulent assignments.

In the revised manuscript, we replaced the lifestyle related analyses and figures with results based exclusively on high-quality and complete viral genomes. Correspondingly, we added a discussion addressing the implications of this choice and the observed patterns in the Discussion section.

The corresponded revised contents in **Discussion** (Lines 405-410 in the new version) read:

"In addition, viral lifestyle inference is influenced by genome quality and tool-specific assumptions. The lifestyle analyses in this study only include high-quality and complete sequences. Future work may be needed to validate potential biases arising from restricting analyses to complete genomes to satisfy some tools' requirements, given the substantial contribution of incomplete sequences to the whole community in metagenome analyses."

Comment No.6: Results Line 142 - Extreme conditions in QTP were never explained.

Response: We have addressed this in our response to Comment No.1. The term "extreme conditions" is now explicitly defined in the Introduction and quantified in the new environmental context section of the Results, referring to low temperature, high UV radiation, hypoxia, and oligotrophy.

The revised contents (Lines 200 in the new version) read:

"Given the low temperatures and intense ultraviolet radiation characteristics of the QTP, …"

Comment No.7: Results Line 152 - This entire section is only referring to material in the supplemental. Is that appropriate? I feel like there's too much material in the supplementary and some should be brought back into a main figure.

Response: We agree. In line with this comment and Reviewer #1's suggestion to unify AMG sections, we have integrated key results on viral functional genes (including AMGs involved in N, S, P, and methane cycles) into the revised main figures (now Fig. 3). The corresponding narrative in the Results has been consolidated and strengthened, moving these central findings from the supplement to the main text for greater impact and clarity.

First, we have integrated key results previously shown only in the Supplementary Information into the revised Fig. 3, bringing the relevant findings into the main figures. Second, we have merged this section with "Genomic features and potential functions of viruses" to provide a more coherent presentation of viral functional characteristics. The implications of these results are now discussed in a unified manner in the Discussion section.

Comment No.8: Results Line 177 - The status of many of these as AMGs is arguable, specifically replication as that is a core viral process. See https://www.nature.com/articles/s41564-025-02095-4 A proper discussion on this is warranted.

Response: We thank the Reviewer for highlighting this important conceptual paper[1]. We fully agree that classifying genes involved in core viral processes (e.g., replication, nucleotide metabolism) as AMGs is contentious. In our original analysis, we used the term "potential functions" cautiously. We have now revised the manuscript to adopt the more rigorous "auxiliary viral gene" (AVG) framework proposed in the cited perspective. We distinguish between core viral genes (essential for viral replication) and AVGs (which include auxiliary metabolic [AMGs], physiological [APGs], and

regulatory [AReGs] genes). We have re-evaluated our annotations accordingly and discuss this distinction explicitly in the Discussion, providing a more precise and conservative interpretation of viral functional potential.

As highlighted by Martin et al.[1], many genes annotated as "metabolic" based on database homology may in fact participate in essential viral functions or have multifunctional roles, and therefore should not be considered AMGs in the strict sense without additional contextual or experimental evidence.

In fact, in the original version of the manuscript, we intentionally titled this section "Genomic features and potential functions of viruses", emphasizing functional potential rather than definitive assignment of viral auxiliary metabolic genes (AMGs). In addition, the AMGs prediction implemented in DRAMv[29] explicitly considers local genomic context, and only genes with auxiliary scores ≤ 3 -corresponding to genes flanked by viral or viral-like genes—were retained. This criterion is consistent with the principles for in silico identification of putative AMGs which is flanked by viral/viral like genes outlined by Martin et al.[1]. Nevertheless, we acknowledge that further manual curation and validation based on experiments are still required to confidently resolve the functional roles of these genes. To ensure a more conservative interpretation, we have avoided referring to DNA replication, nucleotide metabolism, ribosomal proteins and other essential viral processes-related genes as AMGs in the revised manuscript and instead describe them more generally as viral genes or potential functional genes.

We have also explicitly discussed this distinction and its implications in the **Discussion**.

The revised contents in **Discussion** (Lines 437-488 in the new version) read:

"Viruses play essential roles in ecological processes and ecosystem stability[100] by shaping host population dynamics and cellular processes, while also depend on host to buffer against environmental stress. Following the framework proposed in recent perspective[101], we interpret viral genes as part of a broader functional continuum rather

than AMGs. Genes involved in DNA replication, nucleotide metabolism, ribosomal proteins and other essential viral processes are treated as core viral functions that primarily support viral propagation. In contrast, genes not directly required for viral core processes are discussed in terms of 'auxiliary viral genes' (AVGs)[101], including auxiliary metabolic genes (AMGs), auxiliary physiology genes (APGs) and auxiliary regulatory genes (AReGs). Although viruses don't directly participant in biogeochemical cycling, they influence the pools of cellular- and virus-derived materials via host lysis and modulate host metabolism through AVGs[40] …"

Comment No.9: Results Line 198 - Based on Figure 4a, it appears like there was only 1 nrfA gene detected, which means saying it was "most common in glaciers" a bit strange.

Response: This is an excellent point. The wording "most common" was imprecise. We have revised the text to accurately reflect the data: while *nrfA* was detected at very low absolute frequency, the relative abundance of *nrfA*-carrying viruses was orders of magnitude higher in glaciers than downstream.

The revised contents in Results (Lines 297-301 in the new version) read:

"Viruses carrying denitrification genes (nirS, nirK) and anaerobic ammonium oxidation genes (HZS subunit), while were relatively more abundant in plain rivers (2.63) than in plateau rivers (0.115) and glaciers (0.0025). Notably, *nrfA* peaked in glaciers (9.11) but remained at very low levels ($<10^{-3}$) in both plateau and plain rivers."

Comment No.10: Results Line 222 - In my view classifying something as a pathogen simply because it has a virulence gene seems like an oversimplification.

Response: We agree completely. Presence of a virulence factor gene (VFG) does not equate to confirmed pathogenicity, which requires clinical or experimental evidence. While viruses or hosts carrying virulence factor genes (VFGs) may pose potential

pathogenic risks[30–32], the presence of such genes alone does not constitute direct evidence of pathogenicity. Classifying a host as a pathogen based solely on the presence of a single virulence-associated gene would be an oversimplification, though the potential pathogenicity of bacteria carrying virulence factor genes (VFGs) and antibiotic resistance genes (ARGs) has been mentioned in previous studies[31,33–35].

Our aim was to track the distribution of potential risk-associated genes (VFGs and ARGs). We have revised the terminology throughout to avoid overstatement. We now refer to "hosts carrying VFGs," "hosts with potential pathogenicity," or "potential pathogens," and frame the discussion around the distribution and potential dispersal risk of these genetic elements, not the definitive classification of pathogens.

Comment No.11: Results Line 239 - Another section where all material referenced is in the supplemental. I would suggest to think about if any of this information should be promoted to main figure status.

Response: We thank the reviewer for this suggestion. We have reorganized this section and promoted the key results on the microdiversity of generalist/specialist viruses and their host evolutionary dynamics to a new main figure (now Fig. 5). This revision ensures that core findings from each Results subsection are supported by a corresponding main figure, improving narrative flow and visibility.

Comment No.12: Discussion Line 319 - I think it is an overstatement to say AMGs were “extensively involved” in nitrogen and methane cycling, when (based on Figure 4) it appears no gene involved in these two pathways cleared more than 10.

Response: Thank you very much. The phrase “extensively” indeed overstates the degree of involvement given the relatively low absolute counts of nitrogen- and methane-related genes detected. We have therefore deleted the term “extensively” and, in line with the suggestions raised in Comment No.8, revised the discussion of section

"Functional adaptation strategies of viruses in QTP" to adopt a more conservative and integrative interpretation.

The corresponded revised contents in **Discussion** (Lines 465-468 in the new version) read:

"Viral AMGs were linked to methane, nitrogen, phosphorus, and sulfur cycling, exhibiting distinct spatial patterns along the glacier-to-river gradient … (Fig. 3)"

Comment No.13: Methods Line 393 - The authors only filtering through 0.22 μm filters which means that they would lose a ton of viruses this way. The fact that all the data refers to only cell associated viruses and not those that are free passing though this filter cutoff should be noted.

Response: Thank you for raising this important methodological point. Accumulating evidences[36,37] indicate that filtrations through different pore sizes have bias on both viral and microbial communities, thereby shaping the recovered assemblages and downstream interpretations. The filtration through 0.22 μm size cutoffs inevitably excludes a substantial fraction of free viral particles (viral fraction, <0.2μm), particularly smaller virions, and therefore biases the dataset toward cell-fraction viruses, including intracellular, particle-attached, and actively infecting viruses[38].

Here, our sampling strategy was designed to capture viral communities that are closely linked to microbial hosts, which are particularly relevant for analyses of virus-host interactions. Importantly, cell-fraction viruses recovered from metagenomics using 0.22 μm filtration is widely adopted in microbial and environmental virology[6,39,40]. However, more comprehensive investigations integrating data across multiple size fractions will be necessary to achieve a more complete view of environmental viral diversity and dynamics[38].

In this study, water samples were **sequentially filtered through 3.0 μm and 0.22 μm membranes, and both size fractions were subjected to metagenomic sequencing**.

The inclusion of the 3.0 μm size fraction was motivated by the high suspended sediment loads present in several rivers. To distinguish microorganisms associated with abundant particles from those in the water column, the 3.0 μm filtration step was used to retain particle-associated microbial communities, which are reported as the "particle" fraction in the **Supplementary Table 1** (as referenced in Comments 66 and 67).

We have now explicitly acknowledged this limitation in the **Materials and Methods** and **Discussion** sections, clarifying that our results primarily reflect cell-fraction viral communities and should not be interpreted as representing the entire free viral assemblage present in these environments.

The corresponding revised contents in **Discussion** (Lines 392-397 in the new version) read:

"It is important to note that water and particle samples in the GtRVC is primarily derived from metagenomes of size-fractionated microbial communities (>0.22 μm). As such, the recovered viral genomes predominantly represent viruses that are closely associated with microbial hosts, including intracellular and actively replicating viruses, rather than the full assemblage of free viral particles. This host-linked viral fraction is particularly informative for understanding virus–host coupling along environmental gradients[113]."

The revised contents in **Materials and Methods** (Lines 554-556 in the new version) read:

"…was filtered again through 0.22 μm polycarbonate membranes (Millipore, USA) to capture free-living microbes and cellular fraction viruses…"

Comment No.14: Methods Line 425 - Something that may be a bit of a red flag here is the use of six identifiers? This is a very wide net to get viral sequences, and it means that there are likely a lot of false positives that could be cropping up. There is, of course, something to be said about the fact that each identifier has its blind spots and a

combination can get you more than if you had just used a single identifier, but this needs to be emphasized more.

Response: Thank you very much. We agree that applying multiple viral identification tools constitutes a broad detection strategy and, if used without caution, could increase the risk of false positives.

The six viral identification tools applied in this study represent complementary methodological strategies with distinct strengths and blind spots. VIBRANT, ViralVerify, and VirSorter2 all rely primarily on protein annotation and conserved functional signals for viral identification, but they differ substantially in their underlying classification frameworks and decision strategies. VIBRANT performs viral detection by combining protein annotation–derived features with a neural network–based classifier and a v-score metric. VirSorter2 extends viral detection beyond Caudovirales by integrating multiple classifier modules tailored to diverse viral groups, including large dsDNA and ssDNA viruses. ViralVerify classifies contigs as viral, bacterial, or uncertain using a Naive Bayesian classifier trained on Pfam HMM profiles. In contrast, DeepVirFinder and PPR-Meta adopt reference-free deep-learning approaches that capture discriminative nucleotide- and codon-level sequence patterns, enabling the recovery of highly divergent or novel viral sequences. The Earth Virome Pipeline prioritizes specificity through a curated viral protein family database, offering conservative yet robust identification of viruses. The combination of six complementary viral identification approaches was designed to maximize viral recovery across diverse viral groups during the initial virus identification step. While this strategy increases recall, it inevitably introduces a higher risk of false positives, as expected when combining multiple tools with different underlying assumptions.

To mitigate the accumulation of false-positive viral sequences resulting from this multi-tool approach, we implemented a stringent downstream quality-control procedure. Specifically, all predicted viral contigs were assessed using CheckV[41] to evaluate

genome quality and completeness. Low-quality sequences were completely excluded to reduce false positives arising from short contigs, severe contamination, or spurious viral predictions. Considering that retaining only high-quality and complete viral genomes would exclude nearly half (49%) of the medium-quality nonredundant viral sequences, this filter would also introduce substantial bias into assessments of overall viral diversity in the environment. Finally, medium-quality sequences were retained only if their lengths exceeded 10 kb, together with high-quality and complete viral genomes, and were kept for downstream analyses. Furthermore, the modification of analyses related to viral lifestyle prediction (addressed in **Comments 5 and 15**), which require high-quality and complete viral genomes also minimize the potential impact of medium-quality sequences on downstream analyses.

The description for six methods used in virus identification steps.

Tool name	Description
VIBRANT[42]	A protein annotation–driven viral identification framework that combines **neural network–based** classification with the **v-score metric** to capture conserved functional patterns across viral protein families, enabling accurate and high-confidence recovery of viral genomes, especially integrated prophages, with low false-positive rates..
Virsorter2[43]	A modular, **multi-classifier** viral detection framework that integrates protein annotation features with group-specific machine-learning models to detect diverse DNA (dsDNA, ssDNA, NCLDV, *Lavidaviridae*) and RNA viruses from contigs. VirSorter2 provides a framework to go beyond the detection of viruses in the order *Caudovirales*.
DeepVirFinder[44]	A reference-free, alignment-free viral identification tool that uses **convolutional neural networks** to automatically learn discriminative nucleotide-level sequence patterns for distinguishing viral from metagenomes.

Earth Virome Pipeline[6,45]	A viral discovery approach that identifies viruses directly from metagenomic assemblies using an expanded, manually curated set of viral protein families, prioritizing high specificity over sensitivity to robustly recover dsDNA viruses from microbial metagenomes.
ViralVerify[46]	A module that classifies assembled contigs as viral, bacterial, or uncertain using a Naive Bayesian classifier trained on Pfam HMM profiles.
PPR-Meta[47]	A tool identifies viral contigs from metagenomic assemblies using a deep-learning framework that leverages base-level and codon-level sequence representations within a bi-path convolutional neural network.

To address this concern more explicitly, we have revised the Materials and Methods section to clarify the rationale behind combining multiple viral identifiers, emphasizing their complementary nature and the role of downstream curation in ensuring dataset reliability.

The corresponded revised contents in **Materials and Methods** (Lines 602-608 in the new version) read:

"These tools rely on different principles, including protein annotation–based classification, machine-learning models, and reference-free deep-learning approaches, thereby compensating for individual methodological limitations. To minimize false positives introduced by this broad identification strategy, all predicted viral contigs were subjected to stringent downstream quality control. …"

Comment No.15: Methods Line 437 - See previous notes about BACPHLIP. Here the authors say that they use default parameters, and documentation for BACPHLIP clearly state the assumptions it uses, so they haven't changed anything with how BACPHLIP functions.

Response: Thank you. As detailed in our response to Comment No.5, we have addressed this concern by: (1) Excluding NCLDV and virophage sequences from

lifestyle analysis, and (2) Restricting the analysis to high-quality and complete viral genomes as input for BACPHLIP. All lifestyle-related results and figures in the revised manuscript are based on this stricter, more conservative dataset. We acknowledge the tool's assumptions in the revised manuscript.

Comment No.16: Figures Overall - The colors of Plateau and Plains are just similar enough to make them hard to distinguish in certain graphs. For example, Figure 1b, 3a, or Figure S6b. Would recommend they make them easier to distinguish. Already stated a few times but there's a lot of stuff in supplemental. Might be better to be less picky about what goes in the main figure section.

Figure 1 - A map of China with the location of the QTP would be nice. I can't tell based on the sampling location where in Asia this is.

Figure 3 - The colors are egregious here – I can't tell gray from green from blue very well.

Figure 4 - I find A and B to be particularly hard to parse. B in particular – Saying average relative abundance and then having a scale that only goes up to 6 is a bit strange. They also don't follow the order that they do in panel A of that figure which makes it more challenging too.

Response: We sincerely thank the reviewer for the careful evaluation of the figures and the constructive suggestions regarding color differentiation, layout clarity, and figure organization. We have comprehensively revised the relevant figures to improve readability and visual coherence.

Overall Figure Adjustments

To improve group discrimination, we replaced several light-toned colors that were previously difficult to distinguish with more contrasting and visually distinct color schemes. In addition, we reorganized the figure content and moved some information

from the supplementary materials to the main figures. This adjustment ensures that each major section in the manuscript is directly supported by corresponding main figures, thereby improving narrative clarity and reducing unnecessary reliance on supplementary materials.

For Figure 1

A map of China highlighting the QTP region has been added.

For Figure 3 (now Fig.2)

We replaced the original light green and blue color tones with more contrasting colors to enhance differentiation among glacier, plateau and plain samples.

For Figure 4 (now Fig.3)

We reorganized the layout so that original panels A and B now follow a consistent and matching order, improving logical alignment and interpretability.

The upper limit of "6" in panel B resulted from log-transformation of relative abundance values for heatmap visualization. To avoid confusion, we have now explicitly indicated in the figure legend that the heatmap values are log-transformed.

Additional Comments in attached marked-up files:

Comment No.17: Title Line 1 - The title does not reflect the findings of the work very well. I have not seen direct evidence of viral regulation related to viral life-tyle shift.

Response: We thank the reviewer for this comment. We have revised the title to more reasonably reflect the main findings of this study. The new title is slightly changed to "Viral lifestyle shift drives eco-dynamics along glacier-to-river gradients".

Comment No.18: Abstract Line 21 - Not clear what this (TG2G) refers to. It's an acronym but of what?

Response: Thank you very much for pointing this out. TG2G is the acronym for the Tibetan Glacier Genome and Gene catalog, a publicly available glacier metagenome dataset.[48]. We have now explicitly defined this acronym at its first occurrence in the Abstract, consistent with the revision made in response to Comment No. 22, to ensure clarity for readers.

Comment No.19: Abstract Line 32 - What does microbial resilience mean in this context? Line 33-34 - Not clear what water towers refer to. Correct word?

Response: Thank you. We have replaced the term "microbial resilience" with a more specific description to avoid ambiguity. The revised text now reads (Lines 32 in the new version):

"These findings position viruses as pivotal regulators of microbial community structure and functional dynamics to glacier-to-river gradient change and biogeochemistry in the QTP, …"

Comment No.20: Abstract Line 33-34 - Not clear what *Asian water towers* refer to. Correct word?

Response: "Asian Water Tower" is an established term in hydrology and Earth sciences referring to the QTP's role as a crucial freshwater source[49,50]. For clarity, we have added a brief explanation in the Introduction (Lines 40-44 in the new version):

"The Qinghai–Tibet Plateau (QTP), the highest and one of the most vulnerable ecosystems on Earth, functions as Asia's primary high-mountain headwater system, often refers to as the Asian Water Tower, by storing water in glaciers, snow, permafrost, and lakes and regulating its release to sustain major river systems such as the Yangtze, Yellow, Lancang, and Indus Rivers[2,3]."

Comment No.21: Introduction Line 85 - Not clear what this means (environmental filtering). Is this related to filtering samples from the environment? Does it relate to methodology?

Response: We apologize for the lack of clarity. "Environmental filtering" is an ecological concept referring to the process by which abiotic conditions (e.g., temperature, nutrients) select for organisms with traits suited to those conditions, thereby shaping community assembly. It is not related to sample filtration. This concept is widely used in soil and freshwater microbial studies[51,52]. In the present study, "environmental filtering" is applied to viral communities, where glacier-to-river environmental gradients impose ecological constraints that shape viral genomic features, infection strategies, functional gene repertoires, and virus–host associations. To avoid ambiguity, we have revised the text to clarify this term: "We examine how environmental gradients shape viral genomic features..." (Lines 105-108 in the new version)

Comment No.22: Results Line 94-101 – Define on first use for *TG2G*, *ANI*, *AAI*, *vPCs*.

Response: Thank you very much. We have defined all acronyms at their first use in the text: Tibetan Glacier Genome and Gene catalog (TG2G), average nucleotide identity (ANI), average amino acid identity (AAI), and viral protein clusters (vPCs).

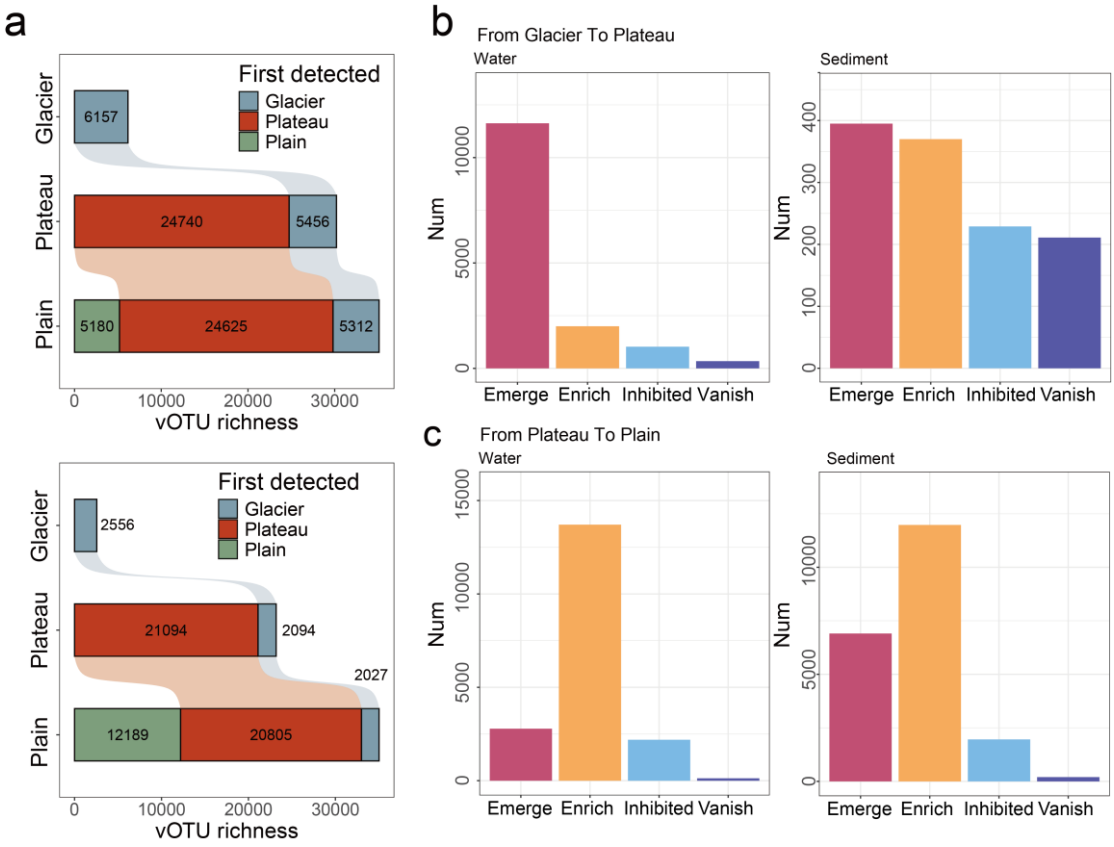
Comment No.23: Results Line 101 – Please check this number (99.3%). I can see in Fig 1e hat the % of classified ones is around 3% with 16.9% being as other than Caudoviricetes. Unless I am missing something? Please revise accordingly and/or explain.

Response: Thank you for identifying this potential confusion. The 99.3% refers to the proportion of vOTUs with no significant match to the IMG/VR database, indicating novelty. The ~3% and 16.9% values in the figure describe the taxonomic breakdown of the small subset that could be classified. We have revised the figure (now Fig. 1e) and

its legend to clearly distinguish between these two metrics (novelty vs. taxonomy of classified fraction) and added a supplementary figure (SI Fig. 2) detailing the IMG/VR matches to avoid ambiguity.

Comment No.24: Results Line 108 – "and most vOTUs present in glaciers or plateau rivers displayed significantly higher relative abundance downstream." Fig? Not clear where I can see this.

Response: We have added a new supplementary figure (SI Fig. 4) and revised the text to clarify this analysis. We categorized vOTUs based on differential abundance testing across the gradient (Emerge, Enrich, Inhibit, Vanish). Quantitatively, vOTUs in the "Emerge" and "Enrich" categories (i.e., those increasing downstream) constituted 63.5%–90.9% of vOTUs in the transitions, supporting the statement.



SI Figure 4: Changes of abundance pattern along the glacier-to-river gradient. (a) Distribution of vOTUs from glaciers, plateau rivers, and plain rivers. Bars represent the number of vOTUs, with colors indicating the regions where vOTUs were first detected along the glacier–plateau–plain gradient. The bar plot shows the number of viral species exhibiting different abundance-change patterns during transitions from glaciers to plateau rivers **(b)** and from plateau to plain rivers **(c)**. Viral species were classified into four categories based on their abundance change patterns: Emerge: viruses absent upstream but detected downstream; Enrich: viruses present upstream and showing significantly increased abundance downstream; Inhibited: viruses present upstream but significantly decreased downstream; and Vanish: viruses present upstream but undetected downstream.

Comment No.25: Results Line 127 – "… and plain river viruses targeted more diverse phyla (SI Fig. 5b)". This should probably be a main figure rather than supp.

Response: Thank you very much. Following this suggestion, this result has been integrated into the revised main Figure 2.

Comment No.26: Results Line 137 – "… Glacier hosts were infected by more viruses …" Not clear if this refers to more lytic or lysogenic infecting viruses. "… more viruses on average (8.3) than those in plateau (4.2) …". When providing an average is there something more that can be added like SD? Other types of stats?

Response: We have provided stratified and detailed statistics.

Specifically, the number of virulent viruses infecting each host (mean ± SD) was 1.81 ± 1.10 in glaciers, 1.54 ± 1.01 in plateau rivers, and 2.31 ± 4.80 in plain rivers. In contrast, temperate viruses showed higher infection numbers in glaciers (3.33 ± 3.01) compared with plateau (2.53 ± 1.81) and plain rivers (2.35 ± 1.89). These values have

now been added to the revised Results to provide a clearer and more quantitative description of viral infection patterns across regions.

The text now reads (Results: Lines 193-195 in the new version): "Hosts in glaciers were infected by 1.81 ± 1.10 (mean ± SD) virulent and 3.33 ± 3.01 temperate viruses, compared with 1.54 ± 1.01 and 2.53 ± 1.81 in plateau rivers, and 2.31 ± 4.80 and 2.35 ± 1.89 in plain rivers, respectively."

Comment No.27: Results Line 145 - "Hosts of glacier and plateau river viruses carried more cold adaptation genes, …" Is this novel? It would be expected that microbes at extreme habitats would be adapted to their environment. This has been shown previously in Arctic and Antarctic regions. Line 145-146 - Any numbers to support this? What is more? More than? Some quantitative info here would be useful. Also please specify if you are talking about Water or Sediment; Line 146 "… suggesting viral targeting of cold-tolerant microbes". This is part of a discussion and should not be in the results. Isn't it expected that cold adapted hosts will be targeted given that these are the ones present there and adapted to those environments?

Response: We agree this host trait is environmentally expected and not novel. As also addressed in response to Reviewer #1 (Comment No.4), we have significantly downplayed this point in the Results, removed the interpretive "suggesting viral targeting" language, and refocused on virus-encoded genes. We now specify gene densities (per kbp) for viruses and their hosts and note patterns were consistent in water and sediment (SI Fig. 8, see revised text in response to Comment No.29).

Comment No.28: Results Line 147 – "… also had more ROS resistance genes". As per prior comment, need some numbers here.

Response: Many thanks. We have added explicit quantitative data (Results: Lines 207-210 in the new version): "ROS resistance genes... peaking in plateau river water and sediment (0.2137 and 0.1917 per kbp), compared with glaciers and plain rivers."

Comment No.29: Results Line 148-152 – "Light utilization genes were more prevalent in plain river water hosts…". Makes sense no? Due to photosynthesis? Isn't this expected? "… varied by phylum and region …" Fig?

Response: We acknowledge this trend may be expected. We have revised the text to a more descriptive and cautious tone.

Previous studies have shown that short-term exposure of bacteria to high radiation and low temperature can downregulate the expression of photosynthesis-related genes[53,54]. In addition, photosystem II genes are highly susceptible to UV-induced damage, which requires neo-synthesis of PSII subunits for repair[54]. Therefore, it remains possible that photosynthesis-related genes exhibit higher transcriptional activity in plain river environments than in plateau regions, given the more favorable light and temperature conditions downstream. However, because our study is based on metagenomic data and lacks corresponding transcriptomic evidence, we cannot directly evaluate transcriptional activity. To avoid over-interpretation, we did not further speculate on light utilization genes in the manuscript. Whether long-term exposure to plateau-like environments leads to genomic loss or retention of light utilization genes remains poorly constrained. Addressing this question would require long-term evolutionary observations or experimental validation, which are beyond the scope of the current study and the focus of this work. We have therefore revised the text to adopt a more cautious interpretation.

Regarding the statement that patterns "varied by phylum and region," this variation is shown in SI Fig. 9b, where a heatmap presents the normalized counts of light utilization genes across different phyla and regions. We have now added the corresponding figure reference and numbers in the Results to improve clarity.

The revised contents in **Results** for cold-adaptation, ROS resistance and light-related genes (Lines 201-213 in the new version) read:

"Cold-adaptation–related genes were widespread across hosts, particularly within Pseudomonadota, Actinomycetota, and Bacteroidota (SI Fig. 8a). Their densities were similar between water and sediment, with the highest values in plateau rivers (0.0321 and 0.0313 per kbp), exceeding those in glaciers (0.0295 and 0.0299) and plain rivers (0.0307 and 0.0303) (SI Fig. 8b). Viruses themselves encoded the highest density of cold-adaptation genes in glaciers (>0.33 COGs per kbp), exceeding those observed in downstream (SI Fig. 8c). ROS resistance genes in hosts (e.g., SOD FeMn[65], catalase–peroxidase[66], glutathione peroxidase[67]) showed a consistent pattern, peaking in plateau river water and sediment (0.2137 and 0.1917), compared with glaciers and plain rivers (SI Fig. 9a). In contrast, light utilization genes (e.g., photosynthetic reaction centers[68], proteorhodopsin[69,70], chlorophyll-binding proteins) displayed an increasing trend from glaciers to plateau and plain rivers, although not all differences were statistically significant (SI Fig. 9b)."

Comment No.30: Results Line 165-169 – Why is the focus on those? What about the other categories of which there are many? "A considerable proportion of viral ORFs …" This is vague.

Response: Thank you very much. In the revised manuscript, we have clarified this rationale and added explicit gene counts and proportions. In addition, we have replaced the vague phrasing with specific numbers: "Category L (78,191 genes), M (36,184 genes), K (25,163 genes), and O (17,747 genes) were the four most abundant categories, together accounting for 55.4% of all functionally annotated genes (Fig. 3a)." (Results: Lines 227-229 in the revised version)

Comment No.31: Results Line 175 – This should be written as only AMGs because AMGs are known to be a very specific feature unique to viruses. So when we talk about AMGs we know that we talk about viral genes so no need for the V before AMG.

Response: Thank you very much. We agree and have removed the redundant "v" prefix, using "AMGs" throughout the revised manuscript. We initially adopted this terminology because similar usage has appeared in previous studies[55].

Comment No.32: Results Line 177 – The authors need to explain what they mean exactly by enriched when talking about viruses. Enrichment can have very different meanings depending on the context.

Response: We thank the reviewer for raising this point. In this study, when referring to "enriched" viruses, we specifically mean vOTUs whose relative abundances are statistically significantly higher in a given region, based on formal statistical testing. The detailed criteria and analytical procedures are described in the section "Identification of dominant, enriched and generalist/specialist viruses" in **Materials and Methods**.

To avoid ambiguity, we have added a clarification at the first use: "Based on relative abundance and statistical testing (Materials and Methods), we identified viruses significantly enriched in each region." (Lines 179-180 in the new version) We also revised other uses of "enriched" that did not explicitly refer to vOTUs, replacing them with clearer descriptions.

Comment No.33: Results Line 184, 186 – "… were higher …" How much higher? Numbers?

Response: Thank you very much. We have systematically revised these sections to add specific quantitative data (counts, percentages) to replace qualitative terms like "higher." These revisions clarify how much higher the reported numbers are and better reflect the contribution of genes across regions.

Comment No.34: Results Line 186-189 – This separation of these different types of genes and their distribution in temperate vs lytic viruses is known. Is there something new here? And again, what is meant by preferential in this context?

Response: Many thanks. Indeed, similar functional differences between temperate and lytic viruses have been reported previously[56]. We acknowledge the known functional differences between temperate and lytic viruses and have streamlined the narrative accordingly. We have also removed the term "preferential" and instead describe observed distributions based on proportion and relative abundance.

Comment No.35: Results Line 195 – "(Fig. 4b)" The numbering of figures and their appearance in the text should follow some sort of a logical chronological order. We are now starting straight with Figure 4b before Figure 4a was even mentioned. Also as it relates to this figure specifically, I don't see black arrows on the figure that indicate viral AMG involvement yet the black arrows are described as such at the bottom of panel C.

Response: We have reordered the figure citations in the text to follow a logical sequence. We have also revised the figure (now Fig. 3) to ensure all arrows and pathways are clearly and consistently labeled and explained in the legend.

Comment No.36: Results Line 194 – "··· viruses encoding nrtB, nrtC, and nrtD were more abundant ···" Numbers? How much more? Line 198 – "… more prevalent …" Numbers needed. How much more?

Response: Many thanks. As with other comments regarding insufficient quantitative detail, we have conducted a systematic revision throughout the manuscript, adding explicit numerical data to replace qualitative descriptions and improve overall clarity.

Comment No.37: Results Line 158-159 – This is for discussion and should not be here; Line 192-193 – Belongs in discussion. Line 195-196 – "These genes, which are

involved in nitrate/nitrite transport, may enhance host nitrogen acquisition under nutrient-limited conditions" Belongs in discussion. Line 198-200 – "… confers an advantage in nitrate-limited environments due to high substrate affinity[56], was most common in glaciers." Maybe more for the discussion part? Line 208-209 – "… suggesting viral enhancement of host phosphorus uptake in oligotrophic environments." This is discussion.

Response: Thank you very much. We have moved this content from the Results section to the Discussion section and integrated it with the related discussion.

Comment No.38: Results Line 212 – "… glaciers and plateau rivers." Fig X?

Response: Many thanks. The relative abundance of genes across different regions are shown in Fig. 3c. To improve clarity and facilitate comparison, we have now added the corresponding figure references immediately after the relevant sentences describing gene relative abundance, allowing readers to more easily link the textual descriptions with the figure.

Comment No.39: Results Line 214-215 – "… prevalent in plateau rivers, … was abundant in glaciers." Prevalent and abundant, what do they mean exactly? These are not quantitative terms?

Response: In this context, both terms were intended to indicate a higher relative abundance. To improve clarity and precision, we have revised the text to explicitly report the quantitative values of relative abundance, based on the data shown in the corresponding figure, thereby clearly indicating that these genes occur at higher relative abundance in the respective environments.

The revised contents in **Results** (Lines 310-320 in the new version) read:

"Viruses with sulfate transport gene *cysA* showed the highest relative abundance in glaciers, being 3.7-fold and 17.2-fold higher than in plateau and plain rivers,

respectively (Fig. 3c). Sulfur oxidation genes *soxX* and *soxD* which can mediate thiosulfate oxidation under low-oxygen conditions, exhibited highest relative abundances in plateau rivers, being 126.9-fold and 13.2-fold higher than those observed in glaciers and plain rivers (Fig. 3c). In contrast, viruses in glaciers encoding *tsdA* involved in thiosulfate-to-tetrathionate conversion and previously reported in deep-sea systems[96], exceeded those in plateau and plain rivers by approximately four orders of magnitude. Meanwhile, viruses carrying assimilatory sulfate reduction genes (*cysC*, *PAPSS*, *cysNC*, and *cysH*[97]) were most abundant in plain rivers (590.66), compared with plateau rivers (504.67) and glaciers (195.68) (Fig. 3c, d)."

Comment No.40: Results Line 232 – "… is relatively higher compared with the others." Difficult to see if this is significant of ARG carrying pathogens vs the others based on Figure 5 A. Are there any further statistics to back this up?

Response: We have improved the visualization and analysis. The bar chart in the original Fig. 5a has been replaced with a box plot in the revised Fig. 4a to better show distributions and statistical comparisons. We now report median VHR values and the results of statistical tests (Wilcoxon rank-sum) showing ARG-carrying pathogens have significantly higher VHR than other groups. We also present complementary data on the mean number of viruses per host category, which supports the same conclusion.

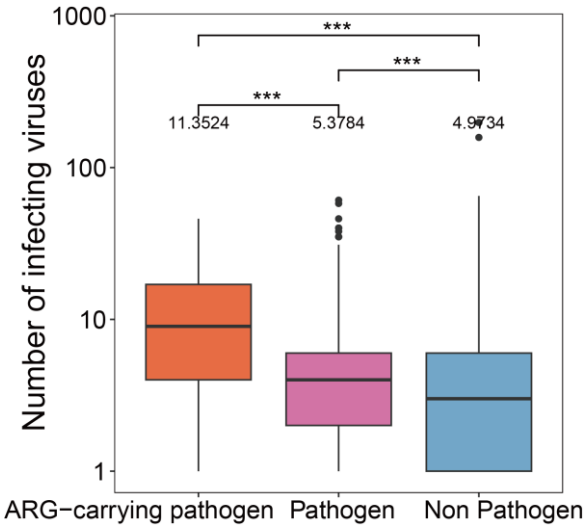


Figure: the number of predicted viruses infecting different host categories. Significance levels were estimated using the Wilcoxon rank-sum test, with asterisks denoting significance (*p < 0.05; **p < 0.01; ***p < 0.001).

The revised contents in **Results** (Lines 335-337 in the new version) read:

"ARG-carrying pathogens exhibited the highest VHR across the glacier-to-river gradient (Fig. 4a), exceeding 100 in glaciers and consistently higher than that of other pathogens and non-pathogens."

Comment No.41: Results Line 236 – "… VFG …" Define on first use.

Response: We have now defined VFG (virulence factor genes) at its first occurrence in the revised manuscript to improve clarity for readers.

Comment No.42: Results Line 240 – This entire section relies on supp Figures. Is this part of the research less important than the previous parts? If not then some main figures for this section should be part of the main body of text. "… generalist and specialist." For viruses? Please specify.

Response: We agree this section was underrepresented in the main figures. We have moved key results on viral microdiversity and host evolutionary dynamics into a new main figure (Fig. 5). We also clarify that "generalist" and "specialist" refer to viruses

based on their niche breadth (widespread vs. habitat-restricted), as defined in the Methods. The revised section title (Line 346-347 in the new version) now reads: "Microdiversity of generalist and specialist viruses and evolutionary dynamics of their hosts"

Comment No.43: Results Line 245-246 – "… generalist viruses exhibited much higher speciation rates than those infected by specialists …" It is not clear what criteria were used to define a virus generalist and another specialist. What is the context for these two definitions?

Response: Many thanks. The criteria used to define generalist and specialist viruses are described in detail in the **Materials and Methods section** entitled "Identification of dominant, enriched and generalist/specialist viruses." Briefly, these definitions are based on ecological niche breadth.

To improve clarity for readers, we have also added a brief explanation in the Results section where these terms are first introduced, explicitly stating the context and criteria used to distinguish generalist and specialist viruses.

The added contents in **Results** (Lines 348-349 in the new version) read:

"In ecological systems, organisms can adopt generalist strategies by occupying broad niches or specialist strategies by exploiting narrower subsets of available resources or habitats[100,101]."

Comment No.44: Line 247-252 – This is discussion. Line 273-277 – This is repetitive from intro; Line 277 – 279 – Repetition from results; Line 280-281 – This is results; Line 283-285 – This is results; Line 290 – Results.

Response: Thank you very much. We have thoroughly revised the Discussion to remove repetitive content from the Introduction and Results, and relocated result-

oriented statements to the appropriate sections, ensuring a clear distinction and improving narrative flow.

Comment No.45: Discussion Line 291 – "... geographic isolation and nutrient availability" What about soil vs water? Also glacial? Geographic distribution would not be the only factor here. As far as I can tell there is no data in the manuscript to support correlations with nutrients or any other metadata.

Response: We agree this statement was not well-supported. As this point was not central to the main argument, we have removed it from the revised Discussion to maintain focus. We have, however, added a new section in the Results (as per Comment 1) presenting environmental metadata (nutrients, temperature, energy fluxes) to provide concrete context for the discussed gradients.

Comment No.46: Discussion Line 291-292 – "Viral diversity peaked in downstream plain rivers …" It is not clear why this here talks about viral diversity, given that the title of this section is supposed to focus on viral life styles, e.g. lysogeny vs lytic infection.

Response: This sentence has been moved to the earlier Discussion section titled "Diverse viral communities in rivers originating from the Qinghai–Tibet Plateau," where it logically fits the theme of biogeographic patterns.

Comment No.47: Discussion Line 295-296 – "… extreme glacial and plateau river environments …" I think that the term extreme needs to be defined. What makes one environment extreme compared to another? This is poorly defined.

Response: We have explicitly defined "extreme" in the revised Introduction and Results, referring to the combined stresses of low temperature, high UV radiation, hypoxia, and oligotrophy characteristic of the QTP, supported by the newly added environmental data. As summarized in previous studies on microbial life in extreme

environment[57], extremeness is not defined by a single factor, but rather by the presence of multiple physicochemical stresses that impose strong constraints on biological entities.

In the context of this study, Qinghai-Tibetan Plateau (QTP) containing glaciers and plateau rivers are considered extreme primarily due to the combined effects of low temperature, strong nutrient limitation, high radiation exposure, and reduced oxygen, rather than in comparison to downstream rivers based on any single variable alone. These conditions are widely recognized as characteristic stressors shaping microbial community structure and function in cold and high-altitude ecosystems. In addition, as detailed in our response to Comment No. 1, we collected and incorporated multiple environmental parameters to better contextualize the environmental background of the study regions, thereby supporting our definition of environmental extremeness.

Comment No.48: Discussion Line 297 – "… through auxiliary genes." Do you mean AMGs?

Response: Thank you. We have revised the text to use the precise term "auxiliary viral genes (AVGs)" as per the conceptual framework adopted, which includes AMGs as a subset. This ensures consistency with the conceptual framework adopted in previous article[1].

Comment No.49: Discussion Line 298 – "… nutrient-rich …" It would be useful to have some more metadata in support of the fact that the downstream environments were nutrient rich. I have no doubt that they were more nutrient rich but its very hypothetical with lack of data or at least references to prior works that show the nutrient status of these water bodies.

Response: We have added quantitative environmental data (e.g., TP, NH_4^+ concentrations) in the new Results section (Fig. 1b, Table S9) showing significantly

higher nutrient levels in plain rivers, replacing the hypothetical description with evidence-based statements.

Comment No.50: Discussion Line 299 – "… similar to what is observed in soil environments …" This is a very bold claim and generalizes viral life style in soil environments. Soil environments are not all the same and can be very diverse. Some more rich than others in nutrients. It is usually a more context-dependent viral activity, with actually lysogeny often being hypothesized as prevalent in many soil environments due to nutrient scarcity and host patchiness. Of course, there is a lot of variability so I don't think that this claim can be made by the authors.

Response: We agree this was an overgeneralization. We have removed the comparison to soil environments from the revised text. Our intention was only to refer to observations from some nutrient-rich soil environments, where increased lytic strategy has been reported[58], rather than to generalize across all soil ecosystems.

Comment No.51: Discussion Line 301 – "… pathogen loads …" Not clear why we are talking about pathogens here.

Response: We have clarified the context: the discussion of pathogen-associated hosts is presented as a potential ecological consequence of the observed viral lifestyle shift, suggesting lytic activity downstream may influence the abundance of hosts carrying virulence and antibiotic resistance genes.

Comment No.52: Discussion Line 302 – "… (VHR) …" Throughout this work it was not clear to me if this VHR takes into account only lytic viruses or if it also includes temperate ones.

Response: The primary VHR results presented in the main text are based on all detected viruses. This approach is consistent with previous environmental metagenomic studies[9,10,59], in which VHR is typically computed using the entire viral community

rather than being restricted to a specific viral lifestyle. In response to this comment, we have performed an additional analysis calculating VHR separately for temperate and virulent viruses (Comment 63). The trend (highest VHR in glaciers) holds for both subsets, supporting our main conclusion. This clarification has been added to the Methods and Results.

Comment No.53: Discussion Line 302 – "… due to higher host diversity" The authors are confusing here diversity with relative abundance. Host diversity is one thing, virus to host ratio based on relative abundances is another.

Response: Thank you very much. We apologize for the lack of clarity. We have revised the text to distinguish between the concepts clearly. Our hypothesis is that higher host diversity downstream may promote broader viral host ranges, which can be associated with a trade-off of reduced per-host infection efficiency, potentially contributing to the observed lower VHR. We present this as a plausible explanation, not a direct equivalence.

Comment No.54: Discussion Line 303-304 – "… where viruses exhibit broader host ranges but reduced per-host infection efficiency". But what about the fact that those downstream environments have more lytic activity? A high lytic activity will surely mean a higher VHR?

Response: This is an important point. We clarify that a higher proportion of lytic viruses does not directly equate to higher lytic activity or VHR, as many lytic progeny may not successfully infect new hosts. Accurately quantifying lytic activity would require additional approaches, such as metatranscriptomic analyses or direct measurements of infection dynamics, which are beyond the scope of this study. In our manuscript, we therefore did not intend to make absolute claims about lytic activity. Instead, our discussion focused on the relative proportions of viral lifestyles, highlighting that

glaciers are dominated by temperate virus infections, whereas downstream environments show a higher proportion of lytic viruses.

Moreover, the virus-to-host relative abundance ratio (VHR) is not determined solely by viral lysis. Even if lytic activity is high and large numbers of viral progeny are produced, these viruses may fail to establish successful infections. In addition, rapid host population growth could offset host mortality caused by viral lysis, resulting in no net increase in VHR. The broader host range/reduced efficiency trade-off is presented as one factor potentially influencing the net VHR in complex communities. We have refined this discussion to avoid overinterpretation.

The inference in this section is therefore based on an ecological trade-off framework: downstream environments harbor more diverse and abundant host communities, providing viruses with a larger pool of potential hosts. As viruses expand their host range under such conditions, this may come at the cost of reduced infection efficiency per host, leading to lower individual host VHR despite a higher proportion of lytic viruses. Expanding host range and reducing infection efficiency at the level of individual hosts represent an inherent trade-off[23]. These two factors are expected to jointly influence effective lytic activity, rather than acting independently. From this perspective, our interpretation does not conflict with the observation of a higher proportion of lytic viruses in downstream environments, but instead reflects different ecological strategies shaping virus–host interactions.

Comment No.55: Discussion Line 306-308 – Inferring to these two models from the data that the authors presented in this manuscript is an overstretch and premature. They don't explain the two models wither. It also doesn't make sense given that piggyback the winner is more aligned with a temperate like mode of infection which the authors claim anyway is not the case downstream where more lytic activity is seen. If anything the opposite may be true with KtW seen downstream and PtW in glacial.

Response: Thank you. We agree with the Reviewers that our initial interpretation of the KtW/PtW models was premature. We have significantly revised this section. We now present our observations (decreasing VHR with increasing host abundance) within a broader, non-mutually-exclusive framework of virus-host interaction continuums (PtL-KtW-PtW) discussed in recent literature. We emphasize these as potential contributing dynamics rather than definitive assignments, and note the limitations of metagenomic data for testing these models directly.

A synthesis of previous studies indicates that the pattern observed here is not unique to our system but has been reported across a wide range of environments[15]. Among 11 ecosystems surveyed in earlier work, 8 exhibited a decline in virus-to-microbe ratio (VMR) with increasing host abundance, with VMR typically peaking at around ~10^6 microbes mL^{-1} or g^{-1} before decreasing at higher host densities[15]. This recurring pattern has been interpreted as reflecting a continuum of virus–host interaction strategies[16], spanning Piggyback-the-Loser (PtL), Kill-the-Winner (KtW), and Piggyback-the-Winner (PtW) dynamics.

Importantly, PtL, KtW, and PtW frameworks do not make explicit quantitative predictions regarding the relative abundances of temperate versus lytic viruses. In the original development of the Piggyback-the-Winner (PtW) model[15], reduced lytic activity and increased lysogeny was inferred from converging evidence, including declining VMR with increased microbe density based on virus-like particles (VLP, these free viral particles may originate from newly released progeny lytic viruses following recent lytic events) counts, increase of lysogeny-associated markers, and other experimental observation, rather than from a requirement that temperate viruses numerically dominate lytic viruses. From this perspective, the observed decrease in temperate virus abundance and proportion along the glacier-to-river gradient does not conflict with these models.

Within this conceptual framework, PtL is thought to dominate under oligotrophic or otherwise extreme conditions where host densities and growth rates are low. Under such conditions, nutrient limitation suppresses lytic gene expression and extends the window for lysogenic commitment, increasing opportunities for coinfection and genome integration, thereby favoring viral persistence through lysogeny. At intermediate host densities, elevated microbial growth and metabolic activity enhance lytic gene expression and prophage induction, leading to frequency-dependent control of fast-growing dominant hosts, consistent with KtW dynamics. At high host densities, such as in eutrophic aquatic systems or host-associated environments, frequent coinfection and increased intracellular phage copy numbers can again favor lysogeny, corresponding to PtW dynamics.

When applied to our study system, glacial and plateau rivers are characterized by cold temperatures, oligotrophic conditions, high radiation, and reduced host metabolic activity relative to downstream plain rivers. These features are broadly consistent with conditions favoring PtL-like strategies, whereas plain rivers experience higher nutrient inputs, warmer temperatures, and higher host densities, placing them within the host-density range commonly associated with PtW dynamics as the microbe density for temperate lake/river in figure 2 of previous articles[15]. In addition, greater host diversity and community heterogeneity downstream may promote viruses with broader host ranges, reducing per-host infection efficiency[23] and contributing to lower VHR as host relative abundance increases.

Nevertheless, interpretations based solely on metagenomic data remain limited. Future integration of metagenomics with direct quantification approaches, such as virus-like particle and microbial cell counts, flow cytometry, metatranscriptome[38] and experimental validation, will be necessary to more precisely resolve virus–host abundance, viral lifestyle and density relationships in natural ecosystems.

The revised contents in **Discussion** (Lines 415-424, 432-435 in the new version) read:

"Previous studies indicate that virus–host interactions follow a continuum of ecological strategies[98,99]. Virus-to-microbe ratios often peak at intermediate host densities and decline as host abundance increases or decreases, reflecting continuum of strategies among Piggyback-the-Loser (PtL)[115], Kill-the-Winner (KtW)[116], and Piggyback-the-Winner (PtW)[114] dynamics. Under cold, oligotrophic and high radiation conditions characteristic of glacial and plateau rivers, low host density and slow growth favor lysogenic persistence when lytic replication is inefficient, consistent with PtL-like dynamics. In plain rivers with higher host densities, increased encounter rates and frequent coinfection can again promote lysogeny (PtW)[117], and higher host diversity favors broader viral host ranges which may further reduce per-host infection efficiency[118]."

"Furthermore, absolute quantification using virus-like particle and microbial cell counts, flow cytometry, metatranscriptome[90] (targeting successful infection-related markers[98,103]), and experimental validation is required to distinguish viral DNA presence from successful infection and replication[104]."

Comment No.56: Discussion Line 308-309 – "… absolute quantification via methods such as flow cytometry is needed to confirm these patterns …" Exactly!

Response: Thank you. We fully agree and have strengthened this point in the revised Discussion, explicitly stating that future work integrating methods like flow cytometry, VLP counts, and metatranscriptomics is essential to move beyond relative abundance patterns.

At present, most purely omics-based studies lack absolute quantification, and interpretations based solely on relative abundance cannot fully capture true virus–host numerical and density relationships. We therefore emphasize that absolute quantification approaches, such as flow cytometry, virus-like particle (VLP) enumeration, and direct microbe–virus detections, are essential to validate and refine

the patterns inferred from metagenomic data. We have retained and strengthened this statement in the Discussion and other related comments. We will continue to improve sampling designs and experimental strategies in future studies to better resolve virus-host abundance dynamics.

Comment No.57: Discussion Line 312 – "… nutrient release …" It's not only nutrients that are being released but also other types of dissolved organic and inorganic materials.

Response: Corrected. The text now reads: "Although viruses don't directly participant in biogeochemical cycling, they influence the pools of cellular- and virus-derived materials via host lysis and modulate host metabolism through AVGs[40]." (**Discussion:** Lines 445-447 in the revised version).

Comment No.58: Discussion Line 323 – "…energy scarcity in extreme environments …" This concept needs to be explained.

Response: We have replaced the vague term "energy scarcity" with a more precise description: "the combined constraints of nutrient limitation and cold-induced metabolic depression in glacial environments."

Comment No.59: Discussion Line 330 – "… antimicrobial compounds" This is an interesting idea but why would glacial microbes need to be more resistant to antimicrobials? One would have expected that those microbes downstream that are closer to urban areas will be the ones having more of these systems.

Response: We have expanded the explanation: antimicrobial resistance in glacial environments may be driven by biotic competition (e.g., antibiotic production as a competitive strategy under stress) rather than just anthropogenic pressure. We cite relevant studies showing high ARG abundance in pristine, extreme environments to support this ecological perspective.

In addition to anthropogenic inputs, biotic interactions within microbial communities may also play a critical role in shaping ARG distributions[60]. Under environmental stresses such as nutrient limitation and low temperatures, microbes may secrete antibiotics or antimicrobial compounds as competitive strategies to inhibit or eliminate rivals[61]. Such interactions can, in turn, impose selective pressure favoring antibiotic-resistant bacteria, even in environments with limited direct human influence. Consistent with this ecological perspective, previous studies have reported the highest ARG abundance and diversity in high-altitude or high-latitude environments in soil[62] and phyllosphere[61] ecosystems, suggesting that extreme environmental conditions can promote antibiotic resistance traits. This pattern aligns with our observations along the glacier-to-river gradient, where glaciers exhibited elevated proportions of virus-associated antimicrobial resistance genes.

Comment No.60: Discussion Line 331 – "… preferentially …" 'Preferentially' implies that the viruses there can chose who to infect and they chose hosts that have cold-adaptation genes. I did not see data to support this "choice" made by the viruses there.

Response: Many thanks. As also addressed in response to Reviewer #1 (Comments No.4 & No.9), we have removed the term "preferentially" and all language implying viral choice.

Comment No.61: Discussion Line 339 – "… generalist …" The context of generalist and specialist needs to be explained in order for this section to make sense. These two aspects usually relate to host range, so is a virus able to infect only few hosts in which case it is a specialist, or is it able to infect many hosts in which case it may be generalist. But my understanding from reading this manuscript is that the two concepts are discussed more in the context of niche and environment.

Response: Thank you. We have clarified in the Results that in this study, "generalist" and "specialist" refer to viral distribution breadth across sampling sites (niche width),

not experimentally validated host range. This is based on the analysis detailed in the Methods.

These terms are also widely used in ecological research[63–65] to describe taxa that occupy broad ecological niches, occurring across a wide range of environmental conditions and/or habitats (generalists), versus taxa with narrow ecological niches that are restricted to specific environmental conditions or habitats (specialists).

In this study, we adopt the latter ecological definition. Specifically, we use these terms to indicate whether viruses are broadly distributed across multiple sampling sites (generalists) or are associated with specific sites (specialists). This classification is based on niche width using the *EcolUtils* package (See **Materials & Methods**), rather than on experimentally validated host-range breadth.

The revised contents in **Results** (Lines 350-351 in the new version) read:

"We identified the generalists and specialists in each region based on ecological niche width (Table S6) and …"

Comment No.62: Discussion Line 345 – "… infection rates …" I have not seen any data in the manuscript to support a discussion on infection rates. This is something that is normally evaluated experimentally over time, i.e. a rate.

Response: Thanks. We have replaced "infection rates" with more accurate terms based on our data: "Hosts showed higher virus-to-host relative abundance ratios (VHR)…"

Comment No.63: Discussion Line 350 – "… viral infection …" Virulent or temperate?

Response: We have specified that the pattern of "stronger" viral association (higher number of viruses per host and higher VHR) in glaciers was observed for both virulent and temperate viruses when analyzed separately.

First, hosts in glacial environments harbored a higher average number of viral infections, reaching up to 8.3 infections per host, which is nearly twice that observed in plateau

(4.2) and plain rivers (4.7). Second, the virus-to-host relative abundance ratio (VHR) was significantly higher in glaciers than in plateau and plain river systems (Fig. 2c).

To further disentangle the effect of viral lifestyle, we separately calculated VHR values for virulent and temperate viruses. Both lifestyles showed patterns consistent with the overall trend, with the highest VHR observed in glacial environments (log_{10} VHR - temperate: 2.68; virulent: 1.89). These results indicate that the stronger viral infection in glaciers is not for a single lifestyle but is shared by both virulent and temperate viruses.

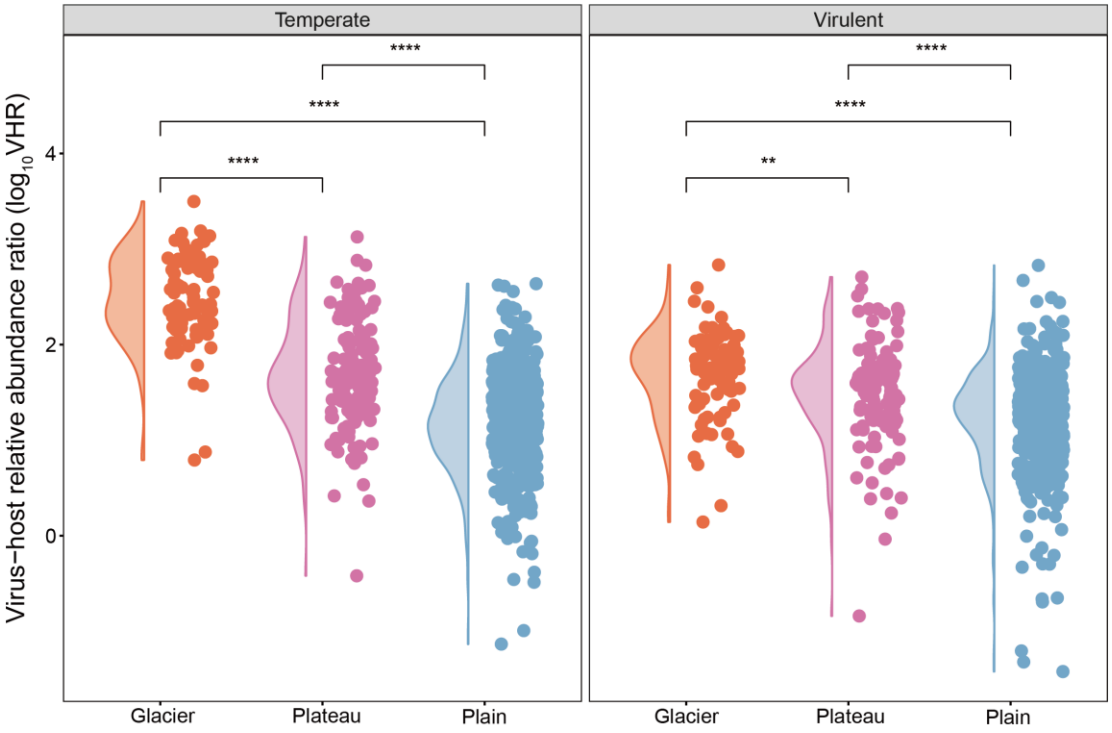


Figure: Virus–host relative abundance ratios across the three regions and different lifestyles.

Comment No.64: Materials and Methods Line 387 – "Table S1" Not clear what particle samples are.

Response: We have clarified that "particle samples" refer to the >3.0 μm size fraction, capturing particle-associated microbes, as distinguished from the 0.22–3.0 μm free-living fraction.

As noted in our response to Comment No.13, "particle samples" refer to larger suspended particulate matters present in the water column, which primarily captures microorganisms and viruses attached to suspended particles or aggregates in river water. These samples were collected to distinguish particle-associated communities from the free-living microbes cellular fraction (0.22–3.0 μm)[66]. However, the primary aim of this study was not to explicitly compare microbial communities across different physical media. In subsequent analyses, we focused primarily on the differences along the glacier-to-river gradient.

Comment No.65: Materials and Methods Line 389 – "… 10 L of river water…" Surface? Integrated? From certain depth? Not clear.

Response: Thank you. All water samples were collected at approximately 0.5 m below the water surface. We have revised the Materials and Methods section to explicitly state the sampling depth to improve clarity and reproducibility.

The revised contents in **Materials and Methods** (Lines 548-550 in the new version) read:

"At each sampling site, approximately 10 L of river water was collected from ~0.5 m below the water surface and transferred into two sterile polyethylene terephthalate (PET) bottles for downstream processing"

Comment No.66: Materials and Methods Line 394-395 – "… 0.22 μm polycarbonate membranes (Millipore, USA) to capture free-living microbes …" This will capture most microbes including large viruses but will let through many phages. So now that I am reading this section after I read the manuscript it seems like only cell associated

viruses are analyzed in this study. So this means that a lot of the ones that are free floating and are a result of lytic events are not captured. So in other words the authors need to make it clear that only actively infecting/cell associated viruses are analyzed rather than the entire viral community.

Response: Thank you very much. This crucial point is now explicitly stated in the Methods and Discussion: our protocol enriches for cell-associated viruses, and thus our analysis primarily reflects this fraction, not the free viruses.

As noted in our response to Comment No.13, the 0.22 μm filtration step for water used in this study captures **cellular fraction viruses** , rather than free viral particles[38]. As a result, viral sequences recovered from these samples originate predominantly from cell-associated viruses, including intracellular viruses, actively replicating viruses, and viruses tightly linked to microbial hosts, while free viral particles smaller than 0.22 μm, including many phages released during lytic events, were not comprehensively targeted by our sampling strategy.

The revised text in **Discussion** (Lines 402-405 in the new version) reads:

"It should be noted that our analyses for water samples primarily reflect host-associated (cell fraction) viruses, as free viral particles smaller than 0.22 μm released from lytic events were not comprehensively captured (see Materials and Methods)"

Comment No.67: Materials and Methods Line 379, 400 – This is for water samples. What about sediment?

Response: We have added a dedicated subsection detailing sediment sample collection, processing, and DNA extraction protocols, ensuring parallel description with water samples.

The revised text in **Materials and Methods** (Lines 562-566, 568-572 in the new version) reads:

“Surface sediment samples for each river were sampled at sites with a water depth of approximately 0.5 m, transferred into sterile 50-mL polypropylene tubes, and then kept on dry ice during transport to the laboratory, just like procedures described in our previous studies[105,106]. All samples were subsequently preserved at −80 °C until downstream DNA extraction”

“For water and particle samples, two filter membranes were transferred into sterile centrifuge tubes, flash-frozen in liquid nitrogen, and mechanically disrupted prior to DNA extraction. For sediment samples, frozen material was retrieved from −80 °C storage and allowed to thaw. After brief centrifugation to remove water, approximately 0.5–1.0 g of sediment was subsampled into sterile centrifuge tubes for subsequent DNA extraction.”

Comment No.68: Materials and Methods Line 409-410 – More info is needed about the sequencing. Was it shotgun? Amplicon? Barcode?

Response: The sequencing performed in this study was shotgun metagenomic sequencing. DNA extracted from each sample was used to construct whole-genome libraries and then sequenced on the Illumina HiSeq platform to generate pair-end (2x150 bp) metagenomic reads. We have now explicitly clarified the sequencing strategy in the Materials and Methods section to avoid any ambiguity.

The revised text in **Materials and Methods** (Lines 579-582 in the new version) reads:

“DNA (DNA amount > 1 μg and concentration > 10 ng μL−1) was used for shotgun metagenomic libraries construction, and then was sequenced on an Illumina’s HiSeq platform (Majorbio Company, Shanghai, China) using a paired-end strategy (2 × 150 bp with an insert size of 300 bp).”

Comment No.69: Materials and Methods Line 431 – Result of CheckV should be reported somewhere.

Response: The CheckV results are reported in the manuscript. Specifically, SI Fig. 1b shows the number of vOTUs across different quality categories as well as their corresponding length distributions. In addition, the CheckV quality and MIUViG quality for each vOTU are provided in Table S2, allowing full traceability at the individual vOTU level. The original manuscript also reported the numbers of vOTUs across different CheckV quality categorie, "…, including 9,351 high-quality, 17,848 medium-quality (≥10 kb), and 9,159 complete genomes".

Comment No.70: Materials and Methods Line 463 – "… high-confidence genome …" What is the criteria for high-confidence genomes?

Response: This term is from the IMG/VR database metadata[67]. For consistency, we further filtered these to only include sequences also classified as "high-quality" by MIUViG standards, matching our own vOTU quality criteria.

We have added clarification to the revised manuscript to improve transparency and to help readers better understand the criteria used for genome selection.

The revised text in **Materials and Methods** (Lines 639-641 in the new version) reads:

"Based on the metadata provided by the IMG/VR database[57], the sequences annotated as high-confidence and high-quality from aquatic ecosystems were selected for comparison."

Comment No.71: Data Availability Line 562-563 – There is a fundamental issue here in comparing data from what the authors obtained from the rivers compared to what was previously obtained and deposited in the Tibetan Glacier Genome and Gene Catalogue and that it is not clear how data was normalized between the two types of datasets. Specifically, was all sampling across the two datasets conduced in the same way? The authors used a 0.2 um cutoff for most of the river samples filtering 10 L. Was the same done in the Glacial data set? If so how? Also, was the quality of the raw data

from each dataset the same? More information is needed in this manuscript on the glacier data set because many of the differences between the 3 habitats may stem from differences in sampling and downstream processing. It is very risky to make major conclusions using such datasets from various researchers, if things have not been normalised in some way or standardised across datasets/samples.

Response: Thank you. This is a critical point. We have added substantial detail to the Methods to address it.

(1) Sampling consistency: Both the TG2G glacier samples and our river samples used 0.22 μm filtration for microbial biomass collection. Regarding sampling and filtration, the glacial dataset used in this study was derived from the Tibetan Glacier Genome and Gene Catalogue (TG2G)[48]. As described in the original TG2G publication, ice and snow samples were melted and filtered through 0.22 μm polycarbonate membranes, which is consistent with the 0.22 μm filtration applied to our river water samples. In addition, both datasets were sequenced using Illumina HiSeq platforms with paired-end 150 bp reads, ensuring comparable sequencing strategies.

(2) Processing uniformity: We downloaded raw TG2G reads and processed them through our identical bioinformatics pipeline (QC, assembly, binning, annotation), ensuring methodological consistency from raw data onward. To minimize potential biases introduced by differences in downstream processing, we did not directly use assembled contigs, MAGs, or annotations provided by the TG2G study. Instead, to ensure full methodological consistency, we retrieved the raw read files (FASTQ) from the TG2G dataset and processed them together with our newly generated river metagenomes using an identical pipeline, including quality control, assembly, binning, and downstream analyses. This approach ensured that all samples were treated uniformly starting from raw reads.

(3) Analysis robustness: Key findings (e.g., temperate virus proportion trend) are based on within-sample ratios/proportions, which are less sensitive to absolute

differences in sequencing depth or biomass. Many of the key comparisons in this study are based on within-sample proportions rather than direct comparisons of absolute values across datasets. For example, our central conclusions regarding the gradual decrease in the relative abundance and number proportion of temperate viruses along the glacier-to-river gradient are derived from ratios calculated within each sample. Similarly, comparisons of auxiliary metabolic genes (AMGs) are primarily based on the proportion of gene-carrying viral populations or the fraction of genes within each sample, rather than raw abundance of genes across samples, which substantially reduce the impact of differences in sequencing depth or sampling intensity.

Finally, comparative analyses across datasets generated by different studies are widely used in metagenomic research, including in the TG2G study itself, where glacial samples were compared with 45 non–Plateau datasets retrieved from public repositories[48].

We acknowledge that perfect standardization across independent studies is challenging, but believe our approach minimizes technical biases. We have added and explicitly emphasized these aspects of sampling consistency and data processing in the revised manuscript to declare the consistency in sampling and the unified processing pipeline applied across all datasets.

The added text in **Materials and Methods** (Lines 543-546, 560-562 in the new version) reads:

"Raw sequencing reads from the glacial samples were downloaded and processed using the same pipeline as the river samples, starting from read quality control, followed by assembly, binning, gene annotation, relative abundance profiling and all subsequent downstream analyses."

"In the TG2G[5], ice and snow samples were also melted and filtered through 0.22 μm polycarbonate membranes prior to DNA extraction."

Again, we sincerely thank the anonymous Reviewer for the thorough and constructive comments, which are of great importance to the improvement of our manuscript quality.

Response to Reviewer #3

We greatly thank the Reviewer for the positive assessment of our work and for the constructive and thoughtful comments. We have carefully considered both points and have revised the manuscript accordingly.

General Comment:

The authors have used a considerable number of metagenomes (n=682) across a spatial (Tibetan plateau rivers) and temporally gradient – samples from 2014 to 2018 to produce a database of virus data – of which they have also accessed ecologically. The authors applied several programmes and methods to carry out each analyses, for example several programmes to identify viral contigs/MAGs – producing a high confidence dataset. The authors identify interesting spatial patterns, such as differences in genome size, GC content, and gene functional annotations, generalists vs specialists, etc. – showcasing ecological advantages within specific environments. HGT, alongside selection typing (evolution) is also explored in association with lytic and lysogenic lifestyles to further profile these viruses. Ultimately, the authors showcase the viral diversity and ecology of viruses along this gradient, and highlight their implied impacts on the system. These methods are clearly reproducible, the methods are the standards expected in the field (and above), and these data/conclusions will be interesting to other researchers in the future.

Response: We thank the Reviewer for their encouraging feedback on the scale of the dataset, the methodological rigor, and the ecological insights of our study. Our aim was indeed to move beyond cataloging to place viral diversity within an integrated ecological and evolutionary framework across the glacier-to-river gradient. We acknowledge that a single study cannot address all facets of viral ecology in this system. Future work will benefit from targeted investigations of specific viral groups (e.g., NCLDVs, RNA viruses) and the integration of complementary approaches like

metatranscriptomics to provide more direct insights into viral activity. We view the present work as a foundational step toward a more comprehensive understanding of viral roles in cryosphere-influenced freshwater systems.

Comment No.1: The filtration methods do have a bias against large DNA viruses (Nucleocytoviricota) in that they are mostly excluded from the pore size used – which is fine. However, it may be worth commenting that these patterns are reflected in these specific viral groups, it is not entirely clear that giant viruses would have the exact patterns shown – although I imagine some of the genes with specific ecological advantages would overlap in environment. RNA viruses are also excluded as well (unless the DNA stages of some were captured). I would wonder if the GC content and genome size phenomena would also be observed w/ giant viruses. This could be included late in the discussion, but also in certain cases bacteriophages are stated to be the most prevalent DNA viruses (line 280), which very well may be the case but giant viruses (also DNA) were not explored accurately.

Response: We thank the Reviewer for this important methodological point. The filtration strategy would introduce a systematic bias against viral and microbial communities[36,37]. The use of 0.22 μm pore-size filtration is currently common in environmental microbiology and virology studies[6,39,40]. In most workflows, the material retained on the filter is subjected to shotgun metagenomic sequencing, whereas the filtrate is further processed for viral particle enrichment and viromic sequencing. We consider that the ecological bias introduced by analyzing the >0.22 μm fraction varies across different viral groups. For large DNA viruses, particle sizes span a wide range from <0.22 μm to up to ~2 μm[68]; consequently, a proportion of these viruses may be retained on the filter, while others pass through. Previous studies have shown that many NCLDVs are preferentially detected in size fractions >0.45 μm[37]. In contrast, most other dsDNA viruses typically range from ~100–200 nm in diameter, meaning that 0.22 μm filtration allows the majority of free viral particles to pass through, thereby

enriching the retained fraction in cell-associated viruses, including intracellular, particle-attached, and actively infecting viruses[38]. As for RNA viruses, the approach employed in this study is not effective for capturing their genomes. As a result, the viral patterns described in this manuscript primarily reflect cell-associated and bacteriophage-dominated DNA viral communities, rather than the full spectrum of environmental viruses[38].

We therefore caution against extrapolating these patterns to contigs identified as belonging to the phylum *Nucleocytoviricota*. To minimize potential biases introduced by genome incompleteness, we restricted the comparison to high-quality and complete NCLDV contigs. However, the relatively small number of recovered NCLDV genomes limited statistical power (the number of high-quality and complete NCLDV contigs in glaciers is only one), and the observed differences did not reach statistical significance. Although median and mean values for all three genomic features consistently showed higher values in plateau samples than in plain river samples, these trends should be interpreted with caution and require further confirmation. From a metagenomic perspective, future studies could apply targeted binning strategies specifically optimized for NCLDVs to improve genome recovery and enable more robust comparative analyses of NCLDV genomic characteristics.

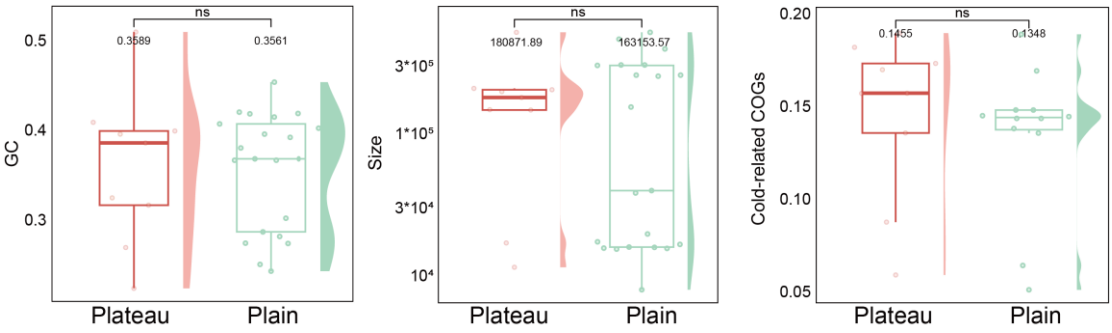


Figure: Comparative analyses of high-quality NCLDV genome features. GC content (guanine–cytosine proportion), genome size, and the number of cold-adaptation–related COGs.

We have revised the relevant text to explicitly acknowledge this limitation and to avoid overgeneralization when stating that bacteriophages are the most prevalent DNA viruses in our dataset. Future studies integrating multiple size fractions, virus-like particle enrichments, and RNA-targeted approaches will be required to robustly evaluate whether similar genomic and ecological patterns extend to entire DNA and RNA viral lineages.

Comment No.2: The discussion of VHR (lines 301 and beyond) might need to consider that viral DNA presence does not always indicate successful viruses, many viral progeny cannot successfully infect hosts. How do the authors think this could impact VHR? Potentially glacier environments a higher production of viruses for successful infection. VHR discussion should be accompanied by discussions of viral success as well. This also cannot be fully solved via flow cytometry/viral particle counts. Potentially gene expression could help (metatranscriptomes) using infection markers.

Response: Thank you for this insightful comment. We fully agree that the presence of viral DNA in metagenomes does not equate to successful infection or productive viral replication, as a significant proportion of viral sequences may originate from defective particles or non-infectious stages. This inherent limitation means that VHR, calculated from metagenomic relative abundance, reflects a genomic signal rather than a direct measure of infection success or viral fitness.

In our study, VHR reflects the relative abundance relationship between viral genomes and their predicted hosts rather than a direct measure of infection success or viral fitness. As such, VHR may overestimate effective virus–host interactions in environments where a large proportion of viral particles are non-infective. This effect could be particularly relevant in glacial environments, where a higher numerical or relative abundance of temperate viruses may reduce the fraction of successful infections through superinfection, as lysogenic hosts are often protected against secondary viral invasion by prophage-encoded defense mechanisms[24–26] (e.g., superinfection exclusion

systems). The slow growth rates characteristic of microbial communities in glaciers can prolong the lysogenic commitment time, increasing the window for multiple viral encounters and coinfections[16], which in turn favors prophage integration rather than productive lytic infection. These potentially inflating VHR estimates without a proportional increase in effective host infection.

We explicitly acknowledge in the Discussion that VHR captures genome-level ratios rather than infection efficiency, and that differences in viral "success" across environments may decouple viral relative abundance of metagenomes from effective infection outcomes. We also agree that this limitation cannot be fully resolved by virus-like particle counts or flow cytometry alone, as these approaches similarly do not distinguish infective from non-infective particles. Integrating metatranscriptomic[38] data targeting infection-related markers (e.g., integrase, excisionase[15], capsid, replication and others lytic-related genes[69]) would provide a more direct assessment of lytic virus activity[70] and viral success. We highlight this as an important direction for future work and have revised the Discussion to clarify these points.

The revised text in **Discussion** (Lines 429-435 in the new version) reads:

"Crucially, current metagenomic approaches cannot discriminate between productive and unsuccessful infections, which may lead to an overestimation of VHR if a substantial fraction of detected viral genomes didn't result in effective host infection or progeny production. Furthermore, absolute quantification using virus-like particle and microbial cell counts, flow cytometry, metatranscriptome[113] (targeting successful infection-related markers[114,119]), and experimental validation is required to distinguish viral DNA presence from successful infection and replication[120]."

Indeed, we would like to express our gratefulness to the anonymous Reviewer for the important comments and suggestions, which are of great help to improve the manuscript quality.

References

1. Martin, C., Emerson, J. B., Roux, S. & Anantharaman, K. A call for caution in the biological interpretation of viral auxiliary metabolic genes. *Nat Microbiol* **10**, 2122–2129 (2025).
2. Lappan, R. *et al.* Molecular hydrogen in seawater supports growth of diverse marine bacteria. *Nat Microbiol* **8**, 581–595 (2023).
3. Brum, J. R. *et al.* Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).
4. Elbehery, A. H. A., Feichtmayer, J., Singh, D., Griebler, C. & Deng, L. The Human Virome Protein Cluster Database (HVPC): A Human Viral Metagenomic Database for Diversity and Function Annotation. *Front. Microbiol.* **9**, 1110 (2018).
5. Hurwitz, B. L. & Sullivan, M. B. The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *PLoS ONE* **8**, e57355 (2013).
6. Paez-Espino, D. *et al.* Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).
7. Cheng, R. *et al.* Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome* **10**, 235 (2022).
8. Yu, M. *et al.* Diversity and potential host-interactions of viruses inhabiting deep-sea seamount sediments. *Nat Commun* **15**, 3228 (2024).
9. Emerson, J. B. *et al.* Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* **3**, 870–880 (2018).

10. Li, Z. *et al.* Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *The ISME Journal* **15**, 2366–2378 (2021).

11. Chen, T. *et al.* Temporal dynamics, microdiversity, and ecological functions of viral communities during cyanobacterial blooms in Lake Taihu. *npj Biofilms Microbiomes* **11**, 178 (2025).

12. Zhang, Q. *et al.* Eutrophication impacts the distribution and functional traits of viral communities in lakes. *Science of The Total Environment* **946**, 174339 (2024).

13. Roux, S. *et al.* iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol* **21**, e3002083 (2023).

14. Cisneros-Martínez, A. M. *et al.* Comparative evaluation of bioinformatic tools for virus-host prediction and their application to a highly diverse community in the Cuatro Ciénegas Basin, Mexico. *PLoS ONE* **19**, e0291402 (2024).

15. Knowles, B. *et al.* Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).

16. Silveira, C. B., Luque, A. & Rohwer, F. The landscape of lysogeny across microbial community density, diversity and energetics. *Environmental Microbiology* **23**, 4098–4111 (2021).

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

18. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology & Oceanography* **45**, 1320–1328 (2000).

19. Winter, C., Bouvier, T., Weinbauer, M. G. & Thingstad, T. F. Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the "Killing the Winner" Hypothesis Revisited. *Microbiol Mol Biol Rev* **74**, 42–57 (2010).

20. Silveira, C. B. & Rohwer, F. L. Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms Microbiomes* **2**, 16010 (2016).

21. Cheng, H. H., Muhlrad, P. J., Hoyt, M. A. & Echols, H. Cleavage of the cII protein of phage lambda by purified HflA protease: control of the switch between lysis and lysogeny. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7882–7886 (1988).

22. Herskowitz, I. & Hagen, D. THE LYSIS-LYSOGENY DECISION OF PHAGE λ: EXPLICIT PROGRAMMING AND RESPONSIVENESS. *Annu. Rev. Genet.* **14**, 399–445 (1980).

23. Sant, D. G., Woods, L. C., Barr, J. J. & McDonald, M. J. Host diversity slows bacteriophage adaptation by selecting generalists over specialists. *Nat Ecol Evol* **5**, 350–359 (2021).

24. Dedrick, R. M. *et al.* Prophage-mediated defence against viral attack and viral counter-defence. *Nat Microbiol* **2**, 16251 (2017).

25. Mavrich, T. N. & Hatfull, G. F. Evolution of Superinfection Immunity in Cluster A Mycobacteriophages. *mBio* **10**, e00971-19 (2019).

26. Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* **10**, 2854–2866 (2016).

27. Copernicus Climate Change Service. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/CDS.E2161BAC (2019).

28. Hockenberry, A. J. & Wilke, C. O. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ* **9**, e11396 (2021).

29. Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research* **48**, 8883–8900 (2020).

30. Méthot, P.-O. & Alizon, S. What is a pathogen? Toward a process view of host-parasite interactions. *Virulence* **5**, 775–785 (2014).

31. Forsberg, K. J. *et al.* The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science* **337**, 1107–1111 (2012).

32. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol Mol Biol Rev* **68**, 560–602 (2004).

33. Haley, B. J., Kim, S. W., Salaheen, S., Hovingh, E. & Van Kessel, J. A. S. Virulome and genome analyses identify associations between antimicrobial resistance genes and virulence factors in highly drug-resistant Escherichia coli isolated from veal calves. *PLoS ONE* **17**, e0265445 (2022).

34. Liu, Z.-T. *et al.* Organic fertilization co-selects genetically linked antibiotic and metal(loid) resistance genes in global soil microbiome. *Nat Commun* **15**, 5168 (2024).

35. Pan, Y. *et al.* Coexistence of Antibiotic Resistance Genes and Virulence Factors Deciphered by Large-Scale Complete Genome Analysis. *mSystems* **5**, 10.1128/msystems.00821-19 (2020).

36. Gao, F.-Z. *et al.* Unveiling the overlooked small-sized microbiome in river ecosystems. *Water Research* **265**, 122302 (2024).

37. Palermo, C. N., Shea, D. W. & Short, S. M. Analysis of Different Size Fractions Provides a More Complete Perspective of Viral Diversity in a Freshwater Embayment. *Appl Environ Microbiol* **87**, e00197-21 (2021).

38. Roux, S. & Coclet, C. Viromics approaches for the study of viral diversity and ecology in microbiomes. *Nat Rev Genet* **27**, 32–46 (2026).

39. Gios, E., Mosley, O. E., Hoggard, M. & Handley, K. M. High niche specificity and host genetic diversity of groundwater viruses. *The ISME Journal* **18**, wrae035 (2024).

40. Worp, N. *et al.* Unveiling the global urban virome through wastewater metagenomics. *Nat Commun* **16**, 10707 (2025).

41. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**, 578–585 (2021).

42. 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).

43. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

44. Ren, J. *et al.* Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77 (2020).

45. Paez-Espino, D., Pavlopoulos, G. A., Ivanova, N. N. & Kyrpides, N. C. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. *Nat Protoc* **12**, 1673–1682 (2017).

46. Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. METAVIRALSPADES: assembly of viruses from metagenomic data. *Bioinformatics* **36**, 4126–4129 (2020).

47. Fang, Z. *et al.* PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience* **8**, giz066 (2019).

48. Liu, Y. *et al.* A genome and gene catalog of glacier microbiomes. *Nat Biotechnol* **40**, 1341–1348 (2022).

49. Immerzeel, W. W. *et al.* Importance and vulnerability of the world's water towers. *Nature* **577**, 364–369 (2020).

50. Zhang, F., Zeng, C., Zhang, Q. & Yao, T. Securing water quality of the Asian Water Tower. *Nat Rev Earth Environ* **3**, 611–612 (2022).

51. Li, X. *et al.* Multi-functional trait-based community stability reveals climate-driven environmental filtering on riverine phytoplankton in a transitional climate zone of China. *Water Research* **290**, 125057 (2026).

52. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).

53. Kataria, S., Jajoo, A. & Guruprasad, K. N. Impact of increasing Ultraviolet-B (UV-B) radiation on photosynthetic processes. *Journal of Photochemistry and Photobiology B: Biology* **137**, 55–66 (2014).

54. Guyet, U. *et al.* Synergic Effects of Temperature and Irradiance on the Physiology of the Marine Synechococcus Strain WH7803. *Front. Microbiol.* **11**, 1707 (2020).

55. Yuan, L. & Ju, F. Potential Auxiliary Metabolic Capabilities and Activities Reveal Biochemical Impacts of Viruses in Municipal Wastewater Treatment Plants. *Environ. Sci. Technol.* **57**, 5485–5498 (2023).

56. Luo, X. *et al.* Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* **10**, 190 (2022).

57. Shu, W. & Huang, L. Microbial diversity in extreme environments. *Nat Rev Microbiol* **20**, 219–235 (2022).

58. Huang, X. *et al.* Soil nutrient conditions alter viral lifestyle strategy and potential function in phosphorous and nitrogen metabolisms. *Soil Biology and Biochemistry* **189**, 109279 (2024).

59. Fan, X. *et al.* Global diversity and biogeography of DNA viral communities in activated sludge systems. *Microbiome* **11**, 234 (2023).

60. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).

61. Ding, Y. *et al.* Increased antibiotic resistance gene abundance linked to intensive bacterial competition in the phyllosphere across an elevational gradient. *Environ Microbiol Rep* **16**, e70042 (2024).

62. Delgado-Baquerizo, M. *et al.* The global distribution and environmental drivers of the soil antibiotic resistome. *Microbiome* **10**, 219 (2022).

63. Wang, X. *et al.* Abundant and rare fungal taxa exhibit different patterns of phylogenetic niche conservatism and community assembly across a geographical and environmental gradient. *Soil Biology and Biochemistry* **186**, 109167 (2023).

64. Hernandez, D. J., Kiesewetter, K. N., Almeida, B. K., Revillini, D. & Afkhami, M. E. Multidimensional specialization and generalization are pervasive in soil prokaryotes. *Nat Ecol Evol* **7**, 1408–1418 (2023).

65. Gweon, H. S. *et al.* Contrasting community assembly processes structure lotic bacteria metacommunities along the river continuum. *Environmental Microbiology* **23**, 484–498 (2021).

66. Gao, Y. *et al.* The mystery of rich human gut antibiotic resistome in the Yellow River with hyper-concentrated sediment-laden flow. *Water Research* **258**, 121763 (2024).

67. Camargo, A. P. *et al.* IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Research* **51**, D733–D743 (2023).

68. Schulz, F., Abergel, C. & Woyke, T. Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat Rev Microbiol* **20**, 721–736 (2022).

69. Yang, T. *et al.* Diversity and evolution of prokaryotic viral lytic proteins. *The ISME Journal* **19**, wraf200 (2025).

70. Coclet, C. *et al.* Virus diversity and activity is driven by snowmelt and host dynamics in a high-altitude watershed soil ecosystem. *Microbiome* **11**, 237 (2023).

Response to the Reviewers' comments

Response to Reviewer #1

Many thanks. We have further revised the manuscript by adding a comparison with the Southern Ocean viral study, refining the interpretation of cold-adaptation-associated signals, and shortening the AMG-related sections. Our point-by-point responses are provided below.

Comment No.1: Lines 215-221: Please clarify how your findings compare to those in https://doi.org/10.1128/msystems.00396-21. Given that your cold-adaptation section is succinct, provide a discussion highlighting similarities or differences to this reference.

Response: We thank the reviewer for this helpful suggestion. We carefully compared our results with the Southern Ocean study[1] recommended by the reviewer. That study reported viral cold-adaptation at both the functional level, including genes related to nucleic acid metabolism, protein folding, and quorum-sensing-associated lysogenic–lytic regulation, and the protein level, including lower hydrophobicity, lower molecular weight, and altered amino acid composition.

In the revised manuscript, we have added a concise comparison highlighting that viruses from glaciers and plateau rivers in our study also encoded lower-molecular-weight proteins and carried cold-adaptation-associated COGs, including GroEL/GroES chaperonins, CspA cold-shock proteins, rare antifreeze proteins, and genes related to nucleic acid metabolism such as DNA-binding proteins, DnaB helicases, and DNA methylases. These shared patterns are broadly consistent with Southern Ocean viromes and suggest that viruses from distinct cold environments may develop partially conserved functional and molecular responses to low-temperature selective pressures.

The added description for comparation in Discussion (Lines 483-488 in the new version) reads:

"In addition, viruses from glaciers and plateau rivers showed signatures of cold adaptation, including proteins with lower molecular weight and a higher density of cold-adaptation-associated genes (Supplementary Fig. S8c), encoding GroEL/GroES chaperonins, Csp family protein and few antifreeze proteins. These patterns are consistent with Southern Ocean viromes[124] and suggesting shared viral responses to low-temperature stress."

Comment No.2: Lines 222-325: This section remains lengthy and somewhat disorganized. Most annotated viral proteins are common in viral genomes and may not be directly relevant to your model system. For example, detecting "soluble lytic murein transglycosylases (related phage lysozyme – muramidase was also detected) and N acetylmuramoyl-L-alanine amidase," is expected, as viruses typically encode lysozymes to lyse their hosts. Thus, annotation of these is routine in phage genomes. To improve focus and clarity, consider emphasizing the genes discussed in lines 436-488.

Response: We thank the reviewer for this important suggestion. We agree with the reviewer that many annotated viral proteins are common in phage genomes and should not be overinterpreted as ecosystem-specific functions. Accordingly, we have substantially shortened this section and removed the detailed discussion of routine lysis-related genes, including lysozyme-, muramidase- and amidase-related proteins.

At the same time, we retained some common but environmentally relevant genes, because they provide useful context for interpreting viral adaptation along the glacier-to-river gradient. For example, GroEL/GroES chaperonins, which were also highlighted in the Southern Ocean viral study[1], may be associated with protein folding under cold stress. Genes related to exopolysaccharide and lipid biosynthesis were also retained because they may be linked to host physiological buffering under cold,

oligotrophic or low-oxygen conditions. Similarly, although replication proteins and DNA repair/recombination genes are common in viruses and bacteria, we retained representative examples because their regional patterns may help readers understand the potential role of genome maintenance and stress tolerance in glacier-enriched viruses. These retained descriptions also provide necessary context for the revised Discussion, where we interpret these genes within the broader framework of auxiliary viral genes (AVGs), including AMGs, APGs and AReGs, rather than treating all annotations as ecosystem-specific AMGs.

Response to Reviewer #3

We are grateful to the Reviewer for the encouraging assessment of our work and the constructive feedback.

General Comment:

I want to thank the authors for their thorough responses. I am content with the modifications and justifications.

Response: We thank the reviewer for the positive assessment and are pleased that the revisions and justifications have addressed the concerns.

References

1. Alarcón-Schumacher, T., Guajardo-Leiva, S., Martinez-Garcia, M. & Díez, B. Ecogenomics and Adaptation Strategies of Southern Ocean Viral Communities. *mSystems* **6**, e00396-21 (2021).