Supplementary Information for

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

Yinghao Li[1,2], Tianyi Chen[1,3], Pengwei Li[1,2], Guohua Zhang[1], Yucheng Tian[1,3], Jiawen Wang[1], Jinren Ni[1,4]*

[1]Environmental Microbiome and Innovative Genomics Laboratory, College of Environmental Sciences and Engineering, Peking University, Beijing 100871, P. R. China

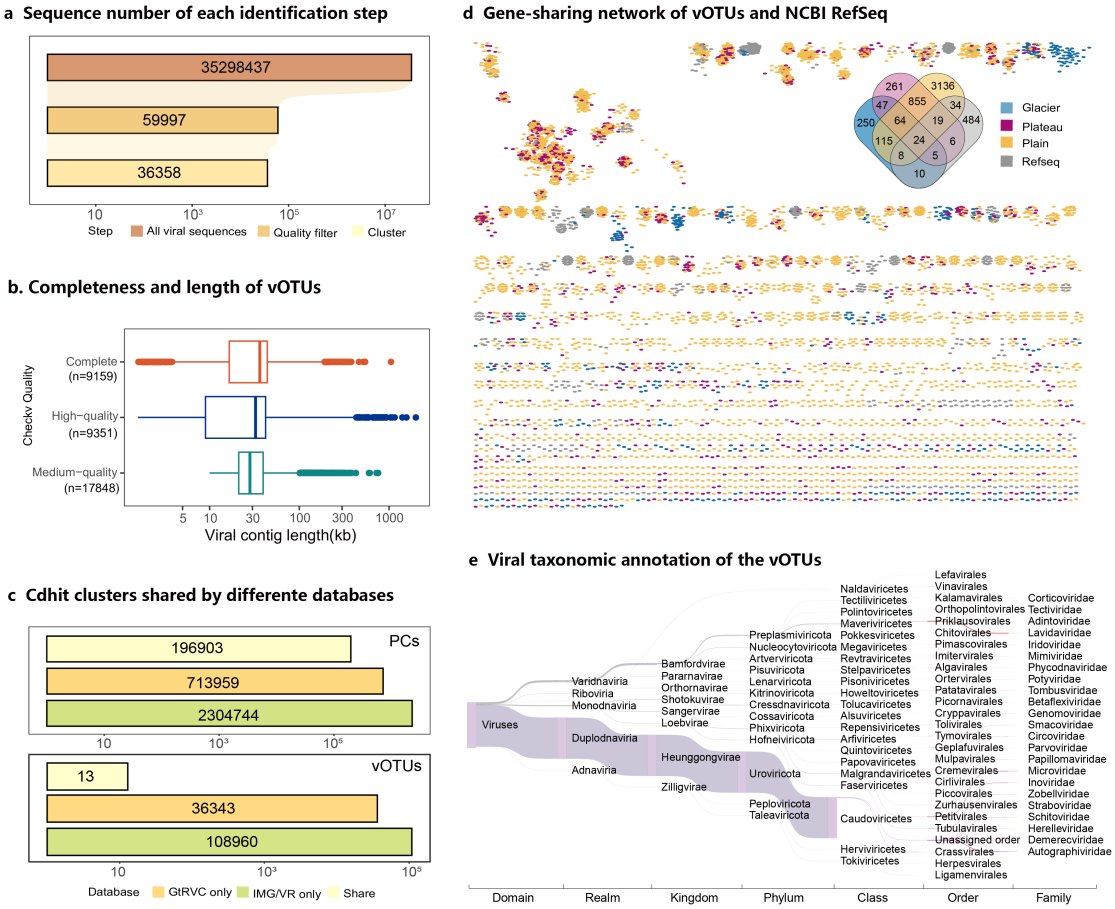
[2]Key Laboratory of All Material Fluxes in River Ecosystems, Ministry of Ecology and Environment, Beijing 100871, P. R. China

[3]Eco-environment and Resource Efficiency Research Laboratory, School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen 518055, P. R. China

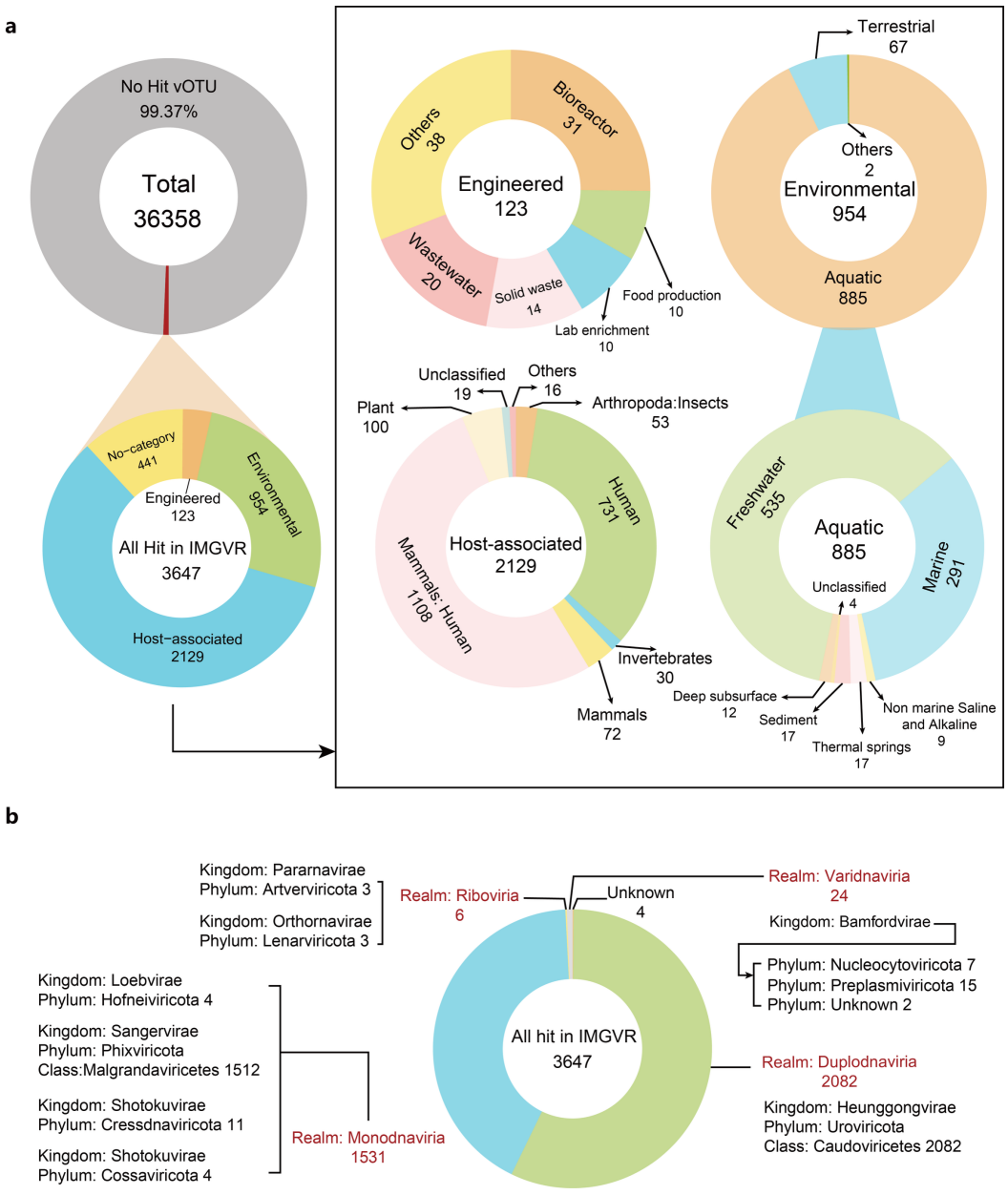
[4]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, P. R. China

***Correspondence:** Jinren Ni, College of Environmental Sciences and Engineering, Peking University, Beijing 100871, P. R. China. E-mail: jinrenni@pku.edu.cn

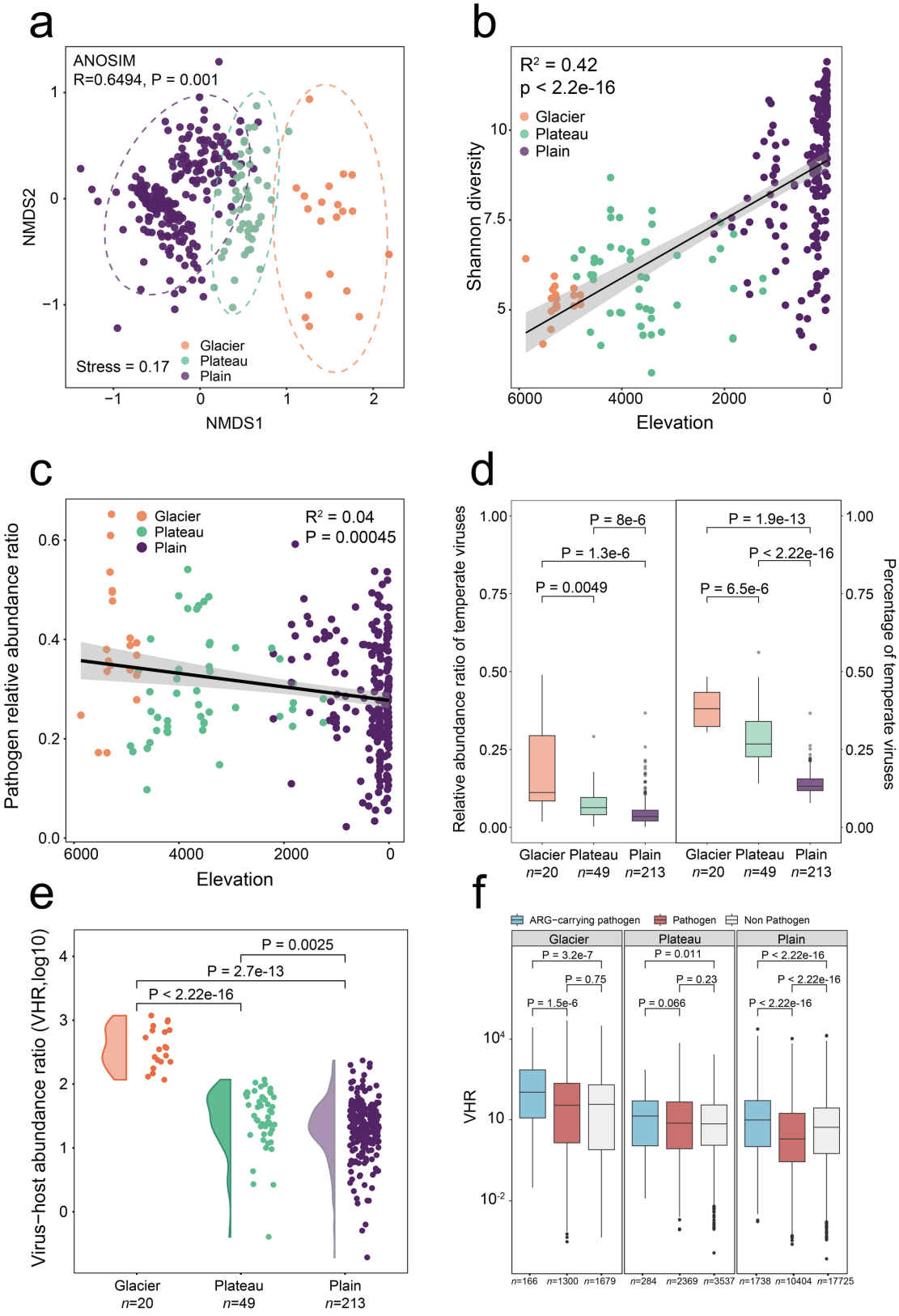
Supplementary Figures



Supplementary Figure 1: **The identification, clustering, quality, database comparison, and taxonomic annotation of viral sequences. (a)** Number of sequences at each processing step. “All viral sequences” represents potential viral sequences merged from all identification tools (see Materials and Methods). “Quality filter” refers to filtering steps based on CheckV results and sequence length. “Cluster” refers to viral sequences clustered using CD-HIT. **(b)** Viral contig numbers and length for each Checkv quality category. **(c)** Comparison with the IMG/VR4 database. The bar plot shows the number of clusters generated by CD-HIT. “GtRVC only” denotes vOTUs or protein clusters derived exclusively from this study, “IMG/VR only” indicates vOTUs or protein clusters sourced solely from the IMG/VR public database, and “Shared” refers to vOTUs or protein clusters present in both this study and the IMG/VR database. **(d)** Gene-sharing network of viral sequences in GtRVC and NCBI RefSeq. Nodes represent viral genomes, and edges indicate similarity among viruses based on shared protein clusters. The Venn diagram shows the overlap among viral clusters containing vOTUs in different regions within the network. **(e)** Detailed viral taxonomic annotation of all vOTUs. Source data are provided as a Source Data file.

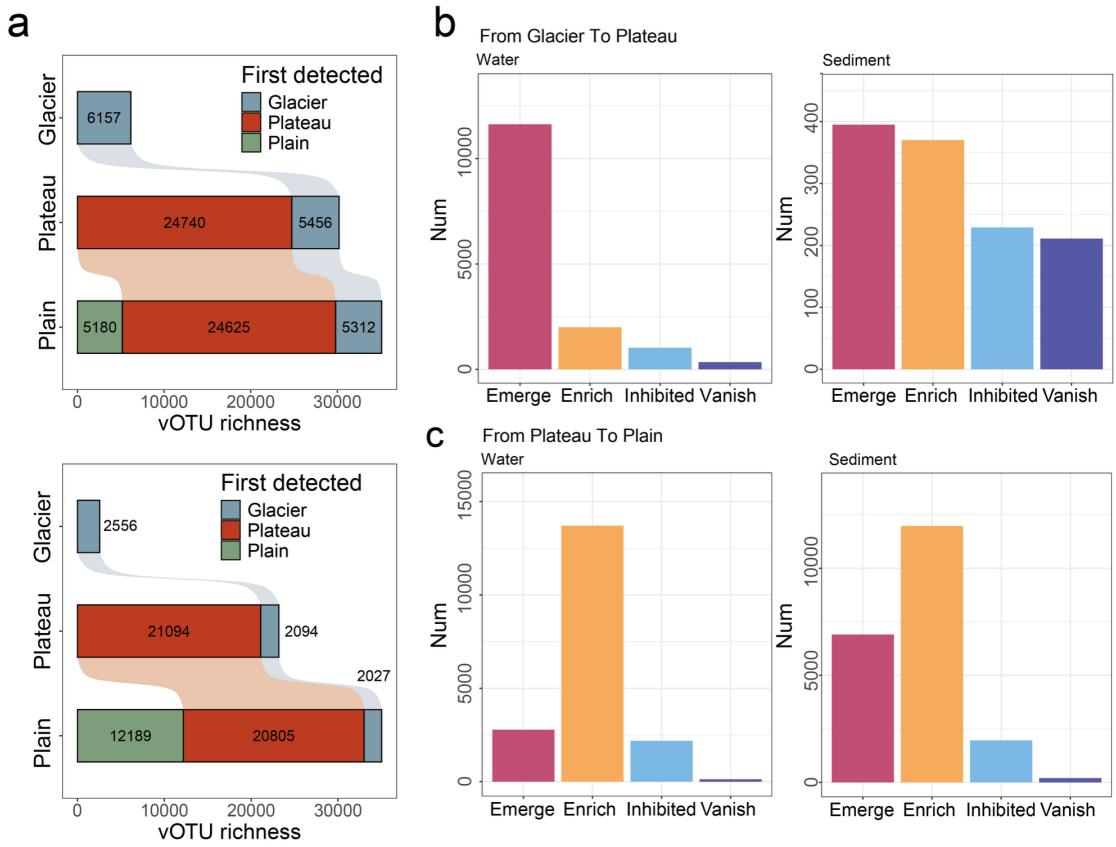


Supplementary Figure 2: BLAST-based comparison of GtRVC viruses with the IMG/VR database. (a) Distribution of GtRVC viral sequences that show significant BLAST matches to the IMG/VR database across different sample categories. Numbers indicate the number of viral sequences corresponding to each sample 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. Source data are provided as a Source Data file.

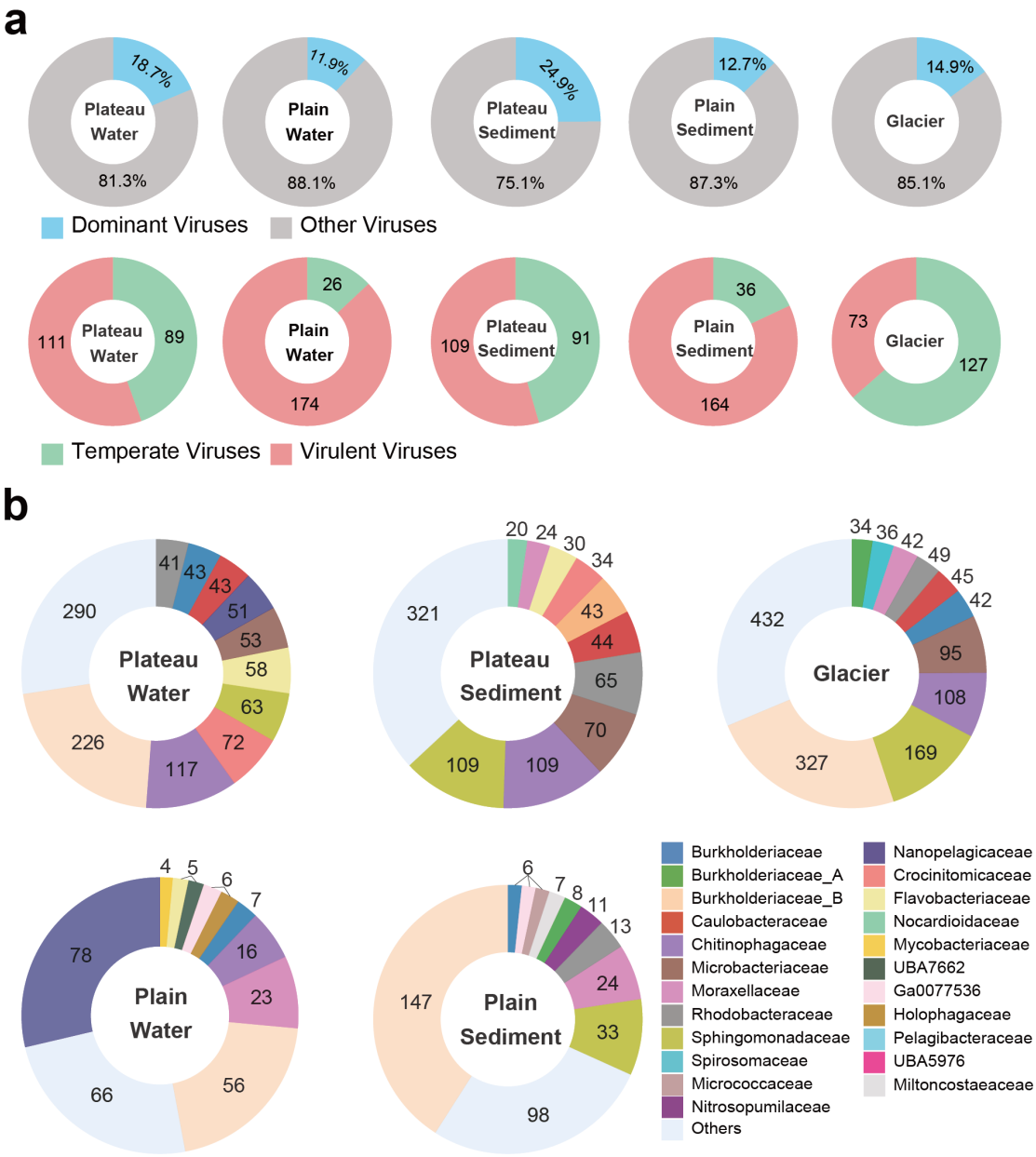


Supplementary Figure 3: **Viral community diversity, structure, and pathogen dynamics in sediments across the three regions. (a)** Non-metric multidimensional scaling (NMDS) analysis of

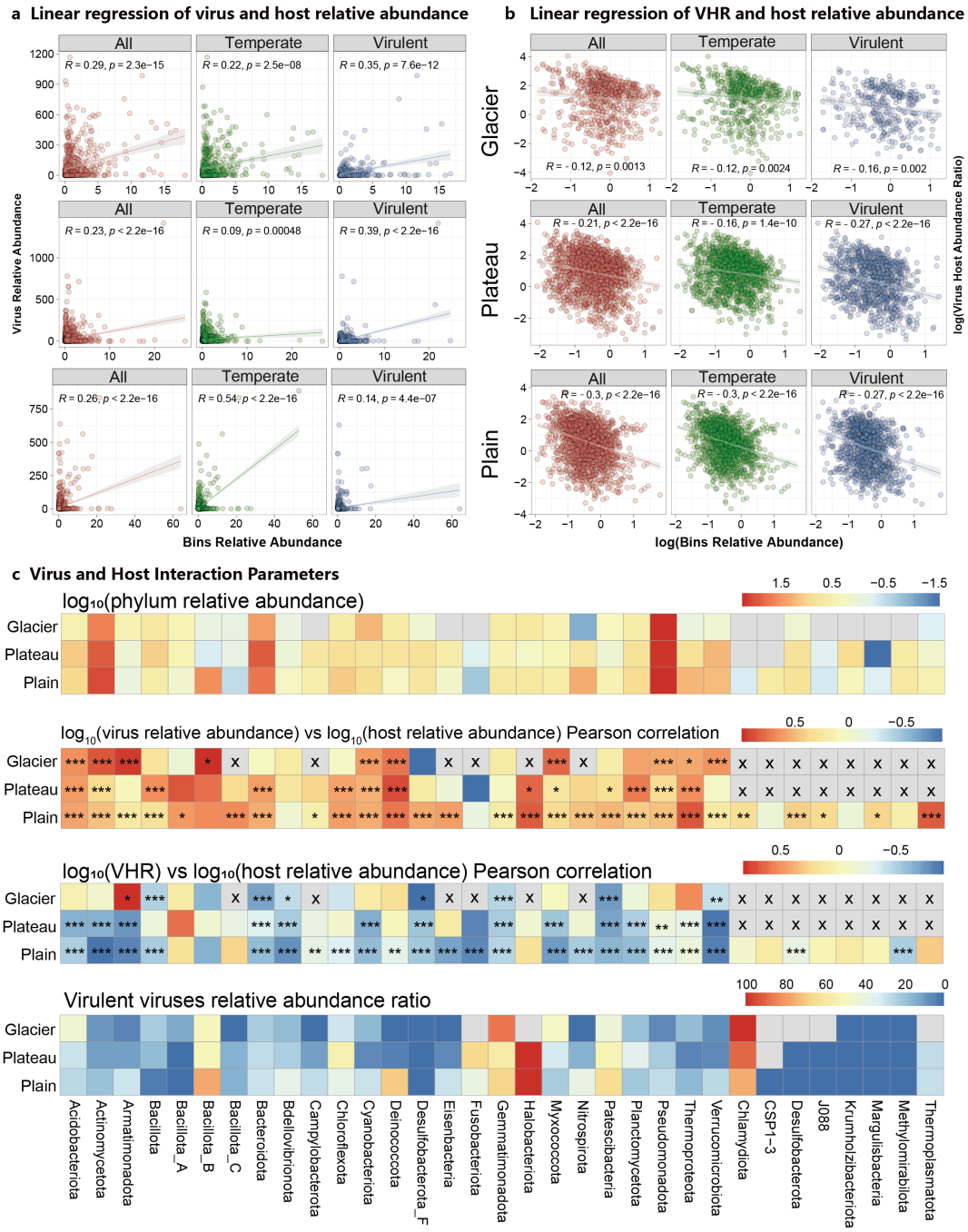
viral communities across the three regions. Analysis of similarities (ANOSIM) was used to evaluate community composition differences among different groups (999 permutations). ANOSIM R statistic and P values are shown in the figure. **(b)** Linear regression between Shannon diversity index and altitude. **(c)** Linear regression between pathogen relative abundance and elevation. The solid line in **(b)**, **(c)** indicates the fitted linear regression, with the 95% confidence interval denoted by gray shading. Adjusted R^2 and associated P value was derived from "lm" function in R. **(d)** Relative abundance and proportion of temperate viruses in each region. **(e)** Virus–host relative abundance ratios across the three regions. **(f)** Virus–host ratios (VHRs) of different hosts in the three regions. MAGs were classified into pathogens with antibiotic resistance genes (ARGs), pathogens without ARGs, and non-pathogens. Statistical significance in **(d)**, **(e)**, **(f)** was evaluated using the two-sided Wilcoxon rank-sum test with exact P values provided. The box in **(d)**, **(f)** indicates the 25th–75th percentiles, the inner horizontal line marks the median, and whiskers extend to 1.5× the interquartile range. The number of data points analyzed is shown as n values. Source data are provided as a Source Data file.



Supplementary 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. Source data are provided as a Source Data file.

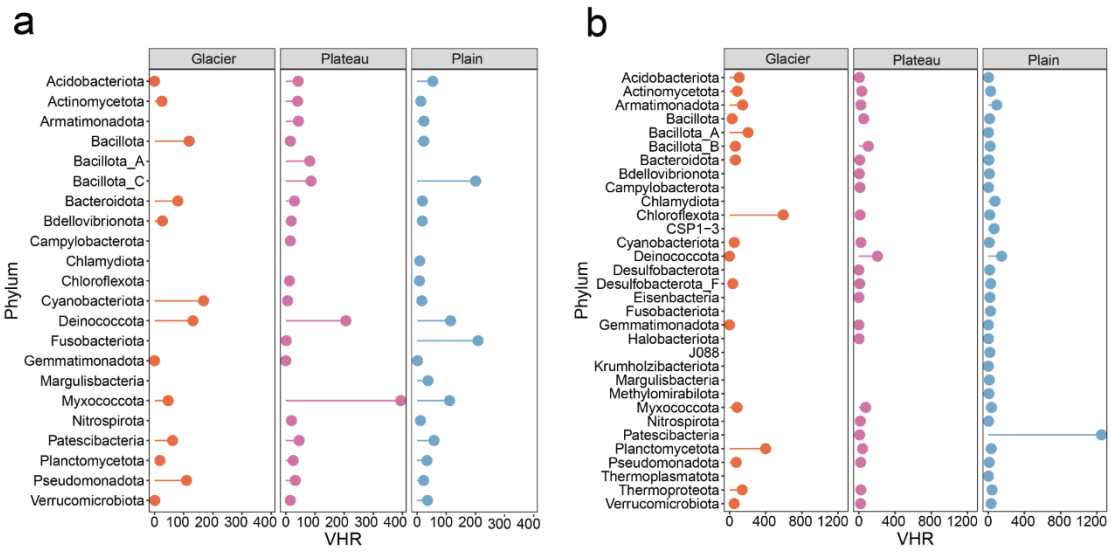


Supplementary Figure 5: **Composition of dominant viruses and their hosts in each region. (a)** Relative abundance ratio of dominant viruses (top 200 vOTUs ranked by relative abundance) (top) and counts of different lifestyles in dominant viruses (bottom). **(b)** Host phyla infected by dominant viruses in different regions. The donut plots show the number of hosts in each family infected by dominant viruses in the corresponding regions. Source data are provided as a Source Data file.

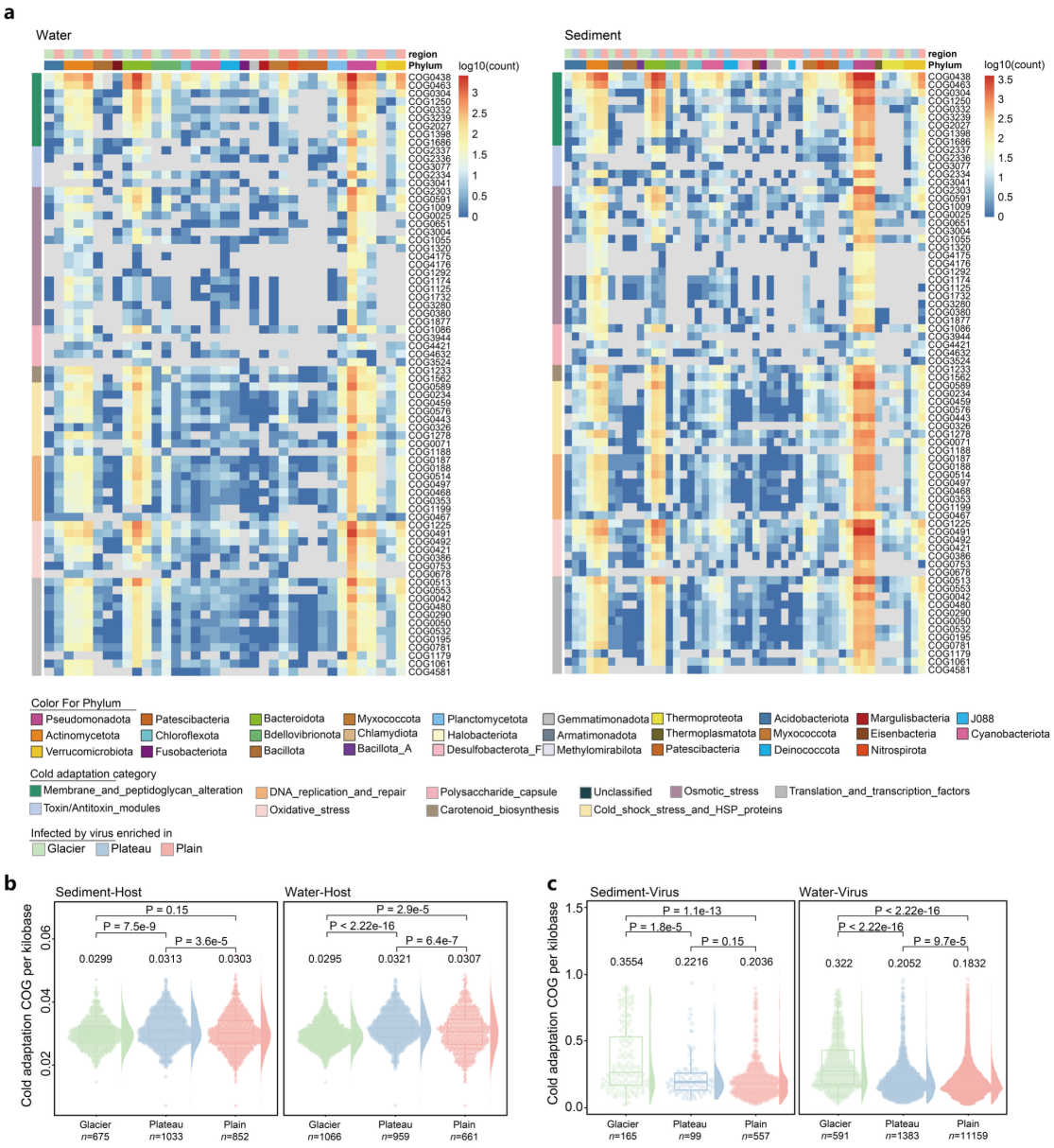


Supplementary Figure 6: **Parameters of virus–host interactions across the three regions. (a)** Linear regression of viral relative abundance with host relative abundance. **(b)** Linear regression of virus–host relative abundance ratio (VHR) with host relative abundance. Two-sided pearson correlation coefficients R and P values are indicated in each panel of **(a)**, **(b)** using "stat_cor" method in ggpubr package. **(c)** $\log_{10}$(Host relative abundance), Pearson correlation coefficient of $\log_{10}$ (viral relative abundance) versus $\log_{10}$ (host relative abundance), Pearson correlation coefficient of $\log_{10}$ (VHR) versus $\log_{10}$ (host relative abundance) for each phylum and virulent virus relative abundance ratio from top to bottom. Color scales in the heatmaps of $\log_{10}$ (host relative abundance) and virulent virus relative abundance ratio indicate the respective exact values, whereas

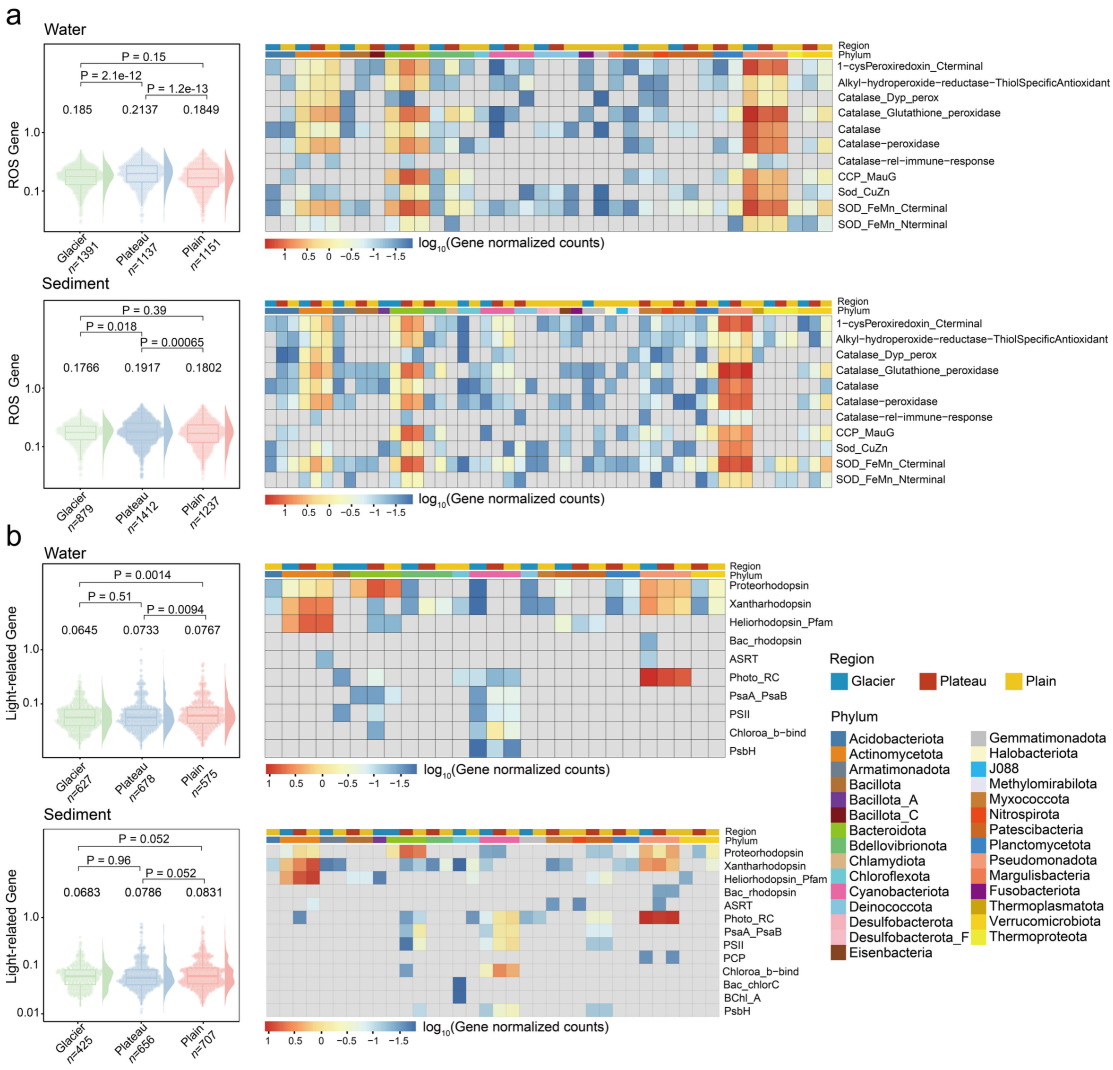
those in the other two heatmaps indicate Pearson correlation coefficients. P values of Pearson correlation coefficients are indicated with asterisks (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; X: Relative abundance for host is 0, corresponding values were omitted). Source data are provided as a Source Data file.



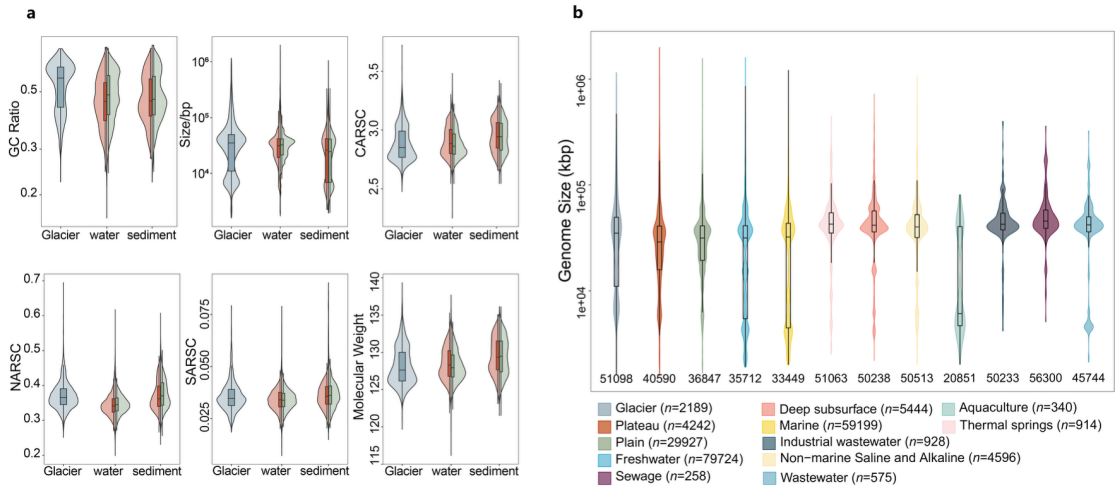
Supplementary Figure 7: **Virus–host relative abundance ratios of different phyla in water and sediment. (a)** Virus–host ratios (VHRs) of different phyla in water samples, including glacier ice/snow, plateau river water, and plain river water. **(b)** Virus–host ratios (VHRs) of different phyla in sediment samples, including glacier cryoconite, plateau river sediment, and plain river sediment. Source data are provided as a Source Data file.



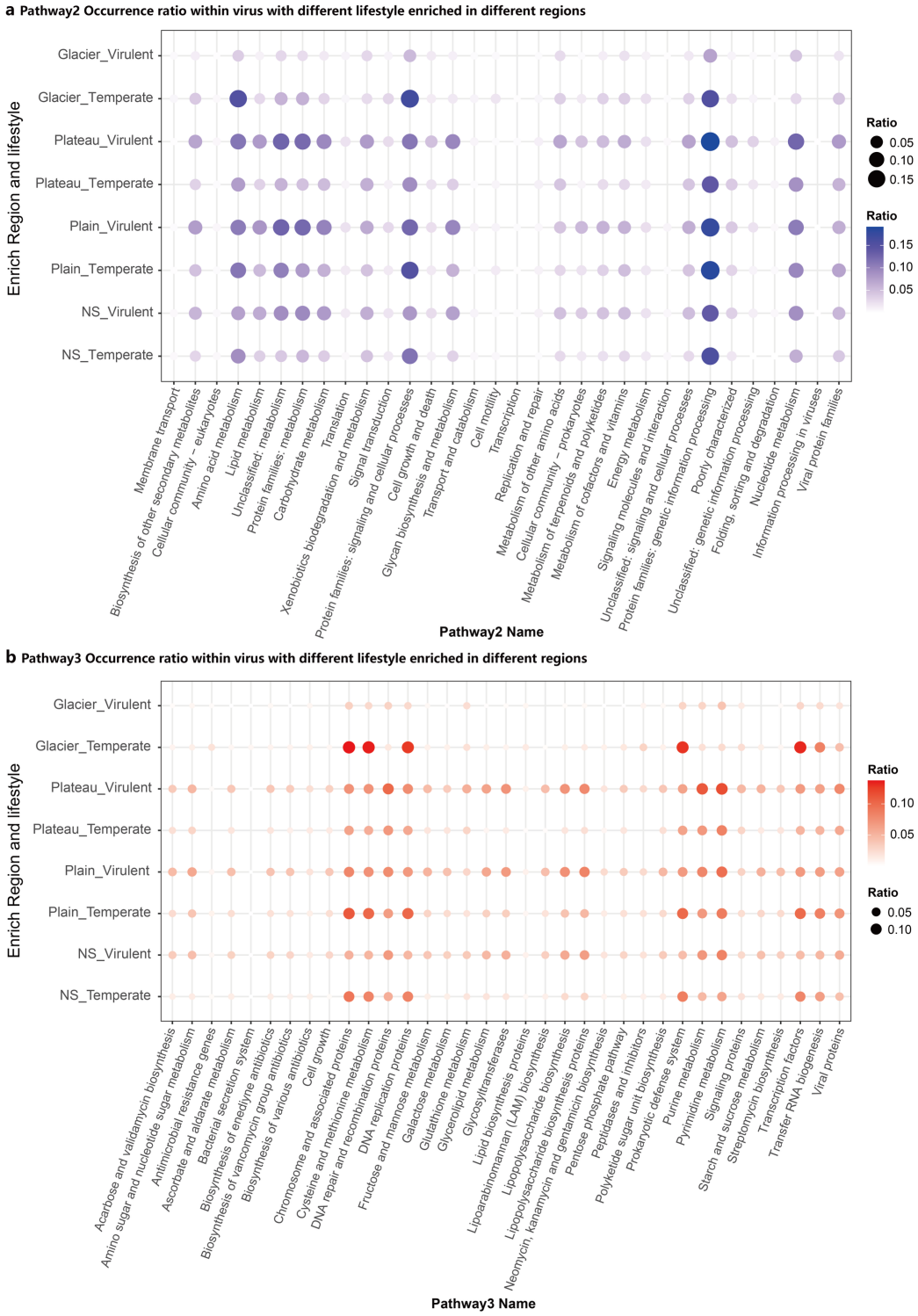
Supplementary Figure 8: **Cold-adaptation genes in viruses and their hosts. (a)** Heatmap showing the counts of cold-adaptation genes. Rows represent clusters of orthologous groups (COGs) associated with adaptive genes, and columns represent hosts from different phyla infected by enriched viruses in different regions. Phylum and region information is indicated above the columns. **(b)** Boxplots showing the number of cold-adaptation COGs per kilobase in host infected by enriched viruses across regions. **(c)** Boxplots showing the number of cold-adaptation COGs per kilobase in viruses across regions. Statistical significance in **(b), (c)** was evaluated using the two-sided Wilcoxon rank-sum test with exact P values provided. The box in **(b), (c)** indicates the 25th–75th percentiles, the inner horizontal line marks the median, and whiskers extend to 1.5× the interquartile range. The value displayed above each box corresponds to the mean of the data in that group. The number of data points analyzed is shown as n values. Source data are provided as a Source Data file.



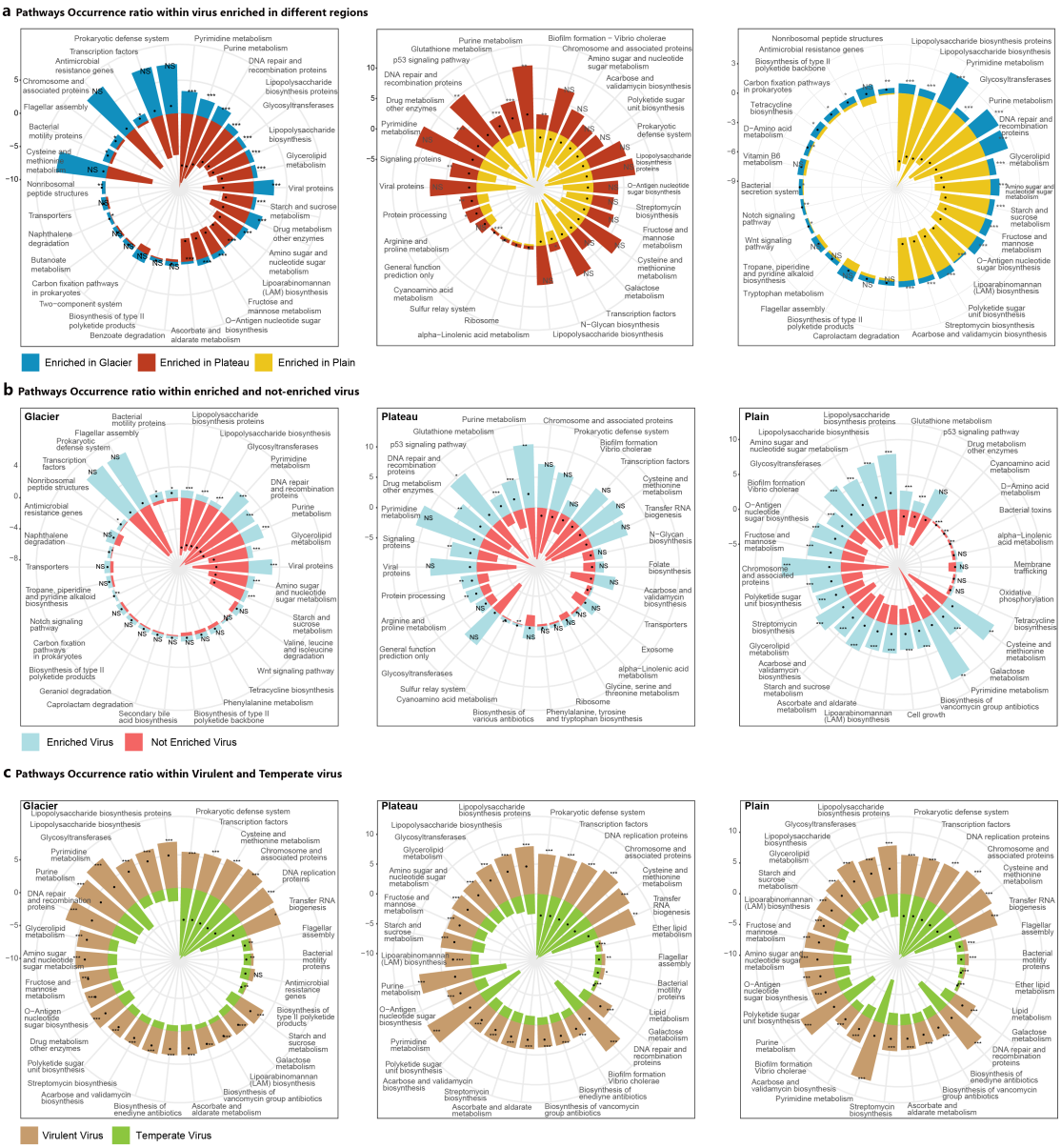
Supplementary Figure 9: **Distribution of adaptation-related genes in hosts infected by enriched viruses. (a)** Reactive oxygen species (ROS)-related genes. Boxplots show the normalized counts of these genes in each host infected by enriched viruses across regions. Heatmap showing the sum of normalized counts of ROS-related genes. **(b)** Photosystem-, rhodopsin-, and light-production-related genes. Boxplots show the normalized counts of these genes in each host infected by enriched viruses across regions. Heatmap showing the sum of normalized counts of genes related to photosystem, rhodopsin, and light production in different regions and phyla. For both heatmaps, rows represent groups of adaptive genes, and columns represent hosts from different phyla infected by enriched viruses in different regions. Phylum and region information is indicated above the columns. Statistical significance in **(a)**, **(b)** was evaluated using the two-sided Wilcoxon rank-sum test with exact P values provided. The box in **(a)**, **(b)** indicates the 25th–75th percentiles, the inner horizontal line marks the median, and whiskers extend to 1.5× the interquartile range. The value displayed above each box corresponds to the mean of the data in that group. The number of data points analyzed is shown as n values. Source data are provided as a Source Data file.



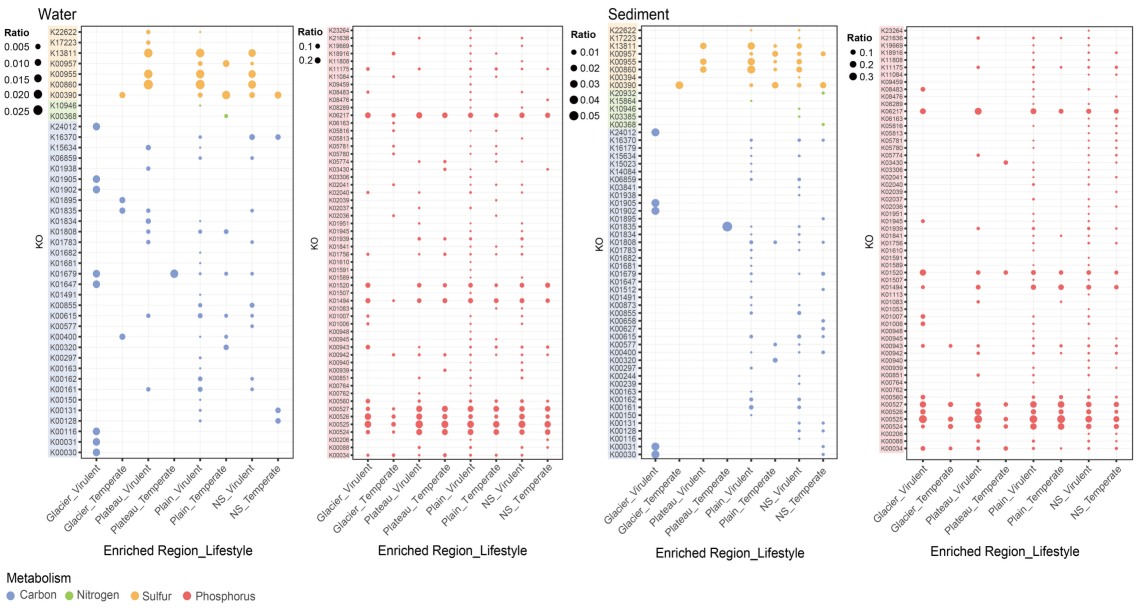
Supplementary Figure 10: Genomic characteristics of Viruses. **(a)** Genomic characteristics of MIUViG high-quality viruses, including GC content, genome size, C/N/S atoms per residue side chain (ARSC), and molecular weight, calculated and displayed across different regions and sample types. Colors are consistent with those in **(b)**. Sample sizes (n) are 2,189 (Glacier), 2,271 (Plateau water), 409 (Plateau sediment), 19,915 (Plain water) and 2,013 (Plain sediment). **(b)** Genome lengths of viral sequences from this study (glacier, plateau, and plain) and from different types in the IMG/VR4 database. The box in **(a)**, **(b)** indicates the 25th–75th percentiles, the inner horizontal line marks the median, and whiskers extend to 1.5× the interquartile range. The number below each boxplot indicates the average genome length. Source data are provided as a Source Data file.



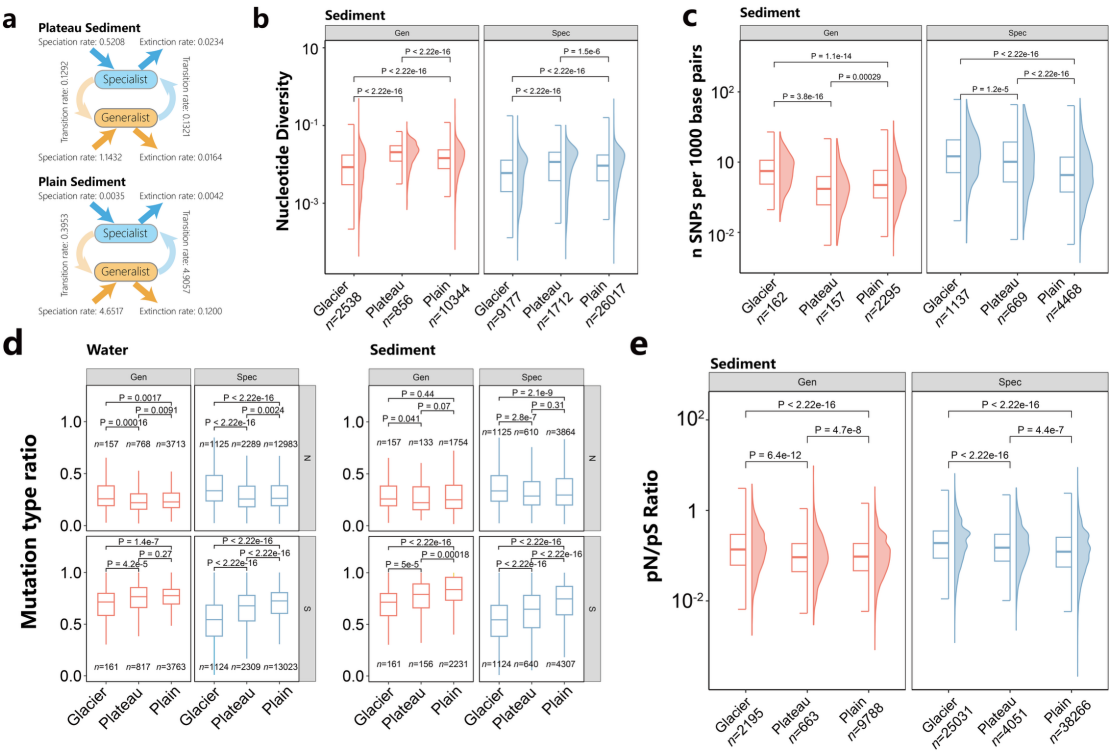
Supplementary Figure 11: **Distribution of potential auxiliary metabolic genes (AMGs) across pathways.** Proportion of viruses carrying AMGs assigned to KOs at different pathway levels: **(a)** level 2 and **(b)** level 3, in enriched viral groups. The y-axis represents enriched viruses in each region and their lifestyles; for example, "Glacier_Virulent" indicates virulent viruses enriched in glaciers. "NS" denotes no significant differences based on enrichment statistical testing (See Methods). Source data are provided as a Source Data file.



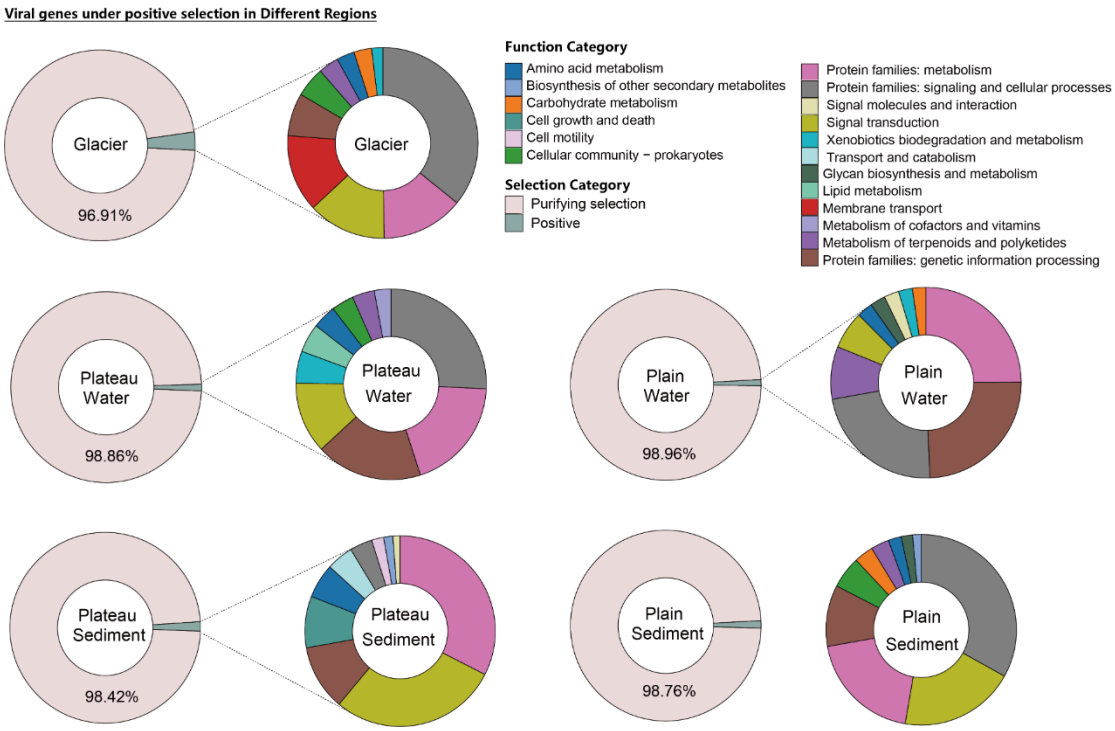
Supplementary Figure 12: **Pathway occurrence ratios in different viral groups. (a)** Enriched viruses across different regions. **(b)** Comparison of enriched and non-enriched viruses within each region. **(c)** Comparison of enriched virulent and temperate viruses within each region. Black dots on each bar represent frequency differences between the compared groups. For each comparison (**a-c**), pathways were ranked by the absolute difference in occurrence ratios between the compared groups, and the categories among the top were selected for visualization. Asterisks above the bars indicate statistical significance (Fisher's test; *P < 0.05; **P < 0.01; ***P < 0.001; NS, not significant). The exact Fisher's P values are provided in Source Data files.



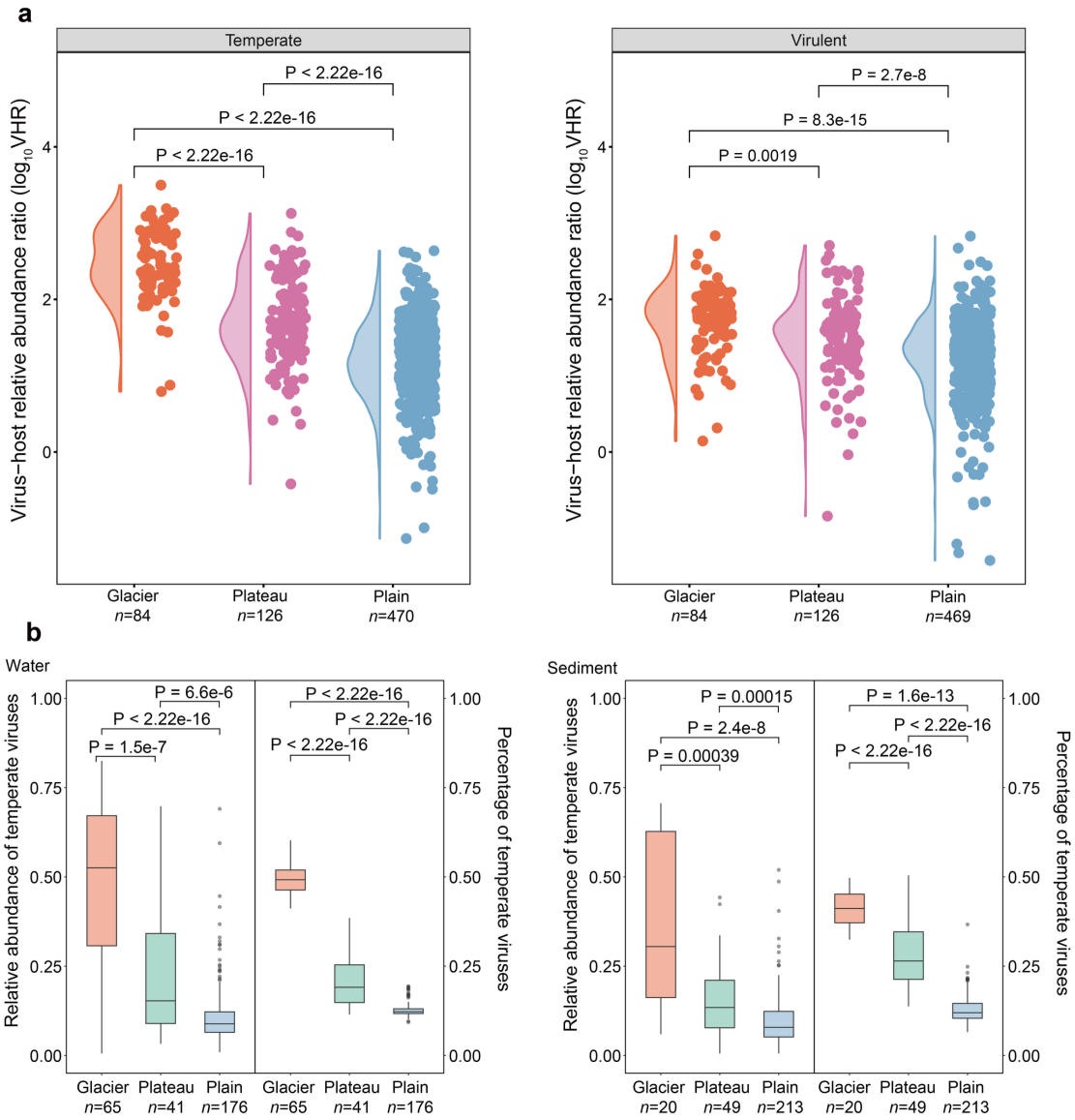
Supplementary Figure 13: **Auxiliary metabolic genes related to biogeochemical cycle of viruses identified in this study.** Proportion of viruses carrying auxiliary metabolic genes (AMGs) assigned to KOs related to carbon, nitrogen, sulfur, and phosphorus cycling in viruses enriched from water and sediment samples. The x-axis represents enriched viruses in each region and their lifestyles; for example, 'Glacier_Virulent' indicates virulent viruses enriched in glaciers. "NS" denotes no significant differences based on enrichment statistical testing (See Methods). Source data are provided as a Source Data file.



Supplementary Figure 14: **Microdiversity of generalist and specialist viruses and evolutionary characteristics of their hosts. (a)** Speciation, extinction, and transition rates of hosts infected by generalist and specialist viruses in sediments. Hosts infected exclusively by either generalist or specialist viruses were considered in the BiSSE model. **(b)** Nucleotide diversity of specialists, generalist across different regions in sediments. **(c)** SNPs per kilobase in viruses with different ecotypes across regions in sediments. **(d)** Ratios of mutation types (S, synonymous; N, non-synonymous) for each virus across different regions. **(e)** *pN/pS* ratios of viral genes carried by viruses with different ecotypes across regions in sediments. The box in (**b-e**) indicates the 25th–75th percentiles, the inner horizontal line marks the median, and whiskers extend to 1.5× the interquartile range. Statistical significance was evaluated using the two-sided Wilcoxon rank-sum test with exact P values provided. The number of data points in each boxplot is shown as n values. Source data are provided as a Source Data file.



Supplementary Figure 15: **Proportion of positively selected genes and their functional categories across regions.** Viral genes were divided into two groups based on *pN/pS* values: genes under positive selection (*pN/pS* > 1) and genes under purifying selection (*pN/pS* < 1). Source data are provided as a Source Data file.



Supplementary Figure 16: **Robustness of glacier-to-river viral patterns (a)** Virus–host relative abundance ratios across glacier, plateau river and plain river regions for different viral lifestyles**. (b)** Relative abundance- and count-based proportions of temperate viruses along the glacier-to-river gradient from water and sediment samples. Sequences classified as NCLDVs or virophages were excluded from all comparisons. The box indicates the 25th–75th percentiles, the inner horizontal line marks the median, and whiskers extend to 1.5× the interquartile range. Statistical significance was evaluated using the two-sided Wilcoxon rank-sum test with exact P values provided. The number of data points in each boxplot is shown as n values. Source data are provided as a Source Data file.