Article

Diversity and biogeography of the bacterial microbiome in glacier-fed streams

https://doi.org/10.1038/s41586-024-08313-z

Received: 15 March 2023

Accepted: 30 October 2024

Published online: 1 January 2025

Open access

Check for updates

Leïla Ezzat[1,2]✉, Hannes Peter[1], Massimo Bourquin[1], Susheel Bhanu Busi[3,4], Grégoire Michoud[1], Stilianos Fodelianakis[1], Tyler J. Kohler[1,5], Thomas Lamy[2], Aileen Geers[1], Paraskevi Pramateftaki[1], Florian Baier[1], Ramona Marasco[6], Daniele Daffonchio[6], Nicola Deluigi[1], Paul Wilmes[4], Michail Styllas[1,7], Martina Schön[1], Matteo Tolosano[1], Vincent De Staercke[1] & Tom J. Battin[1]✉

The rapid melting of mountain glaciers and the vanishing of their streams is emblematic of climate change[1,2]. Glacier-fed streams (GFSs) are cold, oligotrophic and unstable ecosystems in which life is dominated by microbial biofilms[2,3]. However, current knowledge on the GFS microbiome is scarce[4,5], precluding an understanding of its response to glacier shrinkage. Here, by leveraging metabarcoding and metagenomics, we provide a comprehensive survey of bacteria in the benthic microbiome across 152 GFSs draining the Earth's major mountain ranges. We find that the GFS bacterial microbiome is taxonomically and functionally distinct from other cryospheric microbiomes. GFS bacteria are diverse, with more than half being specific to a given mountain range, some unique to single GFSs and a few cosmopolitan and abundant. We show how geographic isolation and environmental selection shape their biogeography, which is characterized by distinct compositional patterns between mountain ranges and hemispheres. Phylogenetic analyses furthermore uncovered microdiverse clades resulting from environmental selection, probably promoting functional resilience and contributing to GFS bacterial biodiversity and biogeography. Climate-induced glacier shrinkage puts this unique microbiome at risk. Our study provides a global reference for future climate-change microbiology studies on the vanishing GFS ecosystem.

Mountain ecosystems, particularly those interfacing with the cryosphere, are highly vulnerable to climate change[6]. Glaciers, and the streams that they feed (glacier-fed streams, GFSs), are among the most iconic of these ecosystems. Whereas GFSs dominated the Earth's fluvial landscapes in the aftermath of snowball Earth and major ice ages[7], today they are largely restricted to mountain tops, the world's water towers, where they initiate the flow of water for some of the largest river systems globally[8]. In doing so, they supply freshwater resources to large human populations living in mountain regions, and to ecosystems further downstream, where they provide important services including regulation of water quality and supporting fisheries[2,8].

Climate-induced glacier shrinkage changes the structure and function of GFSs, putting their ecosystem services and biodiversity at risk[2,9]. Biofilms, the microbial communities attached to streambed sediments, dominate life in GFSs and regulate key ecosystem processes, drive biogeochemical cycles and form the basis of food webs[3]. Nevertheless, despite the critical functions fulfilled by biofilms, microbial life in GFSs remains poorly studied[5,10,11], with all work to date focusing on local to regional scales (for example, refs. 12–15). As a result, no global and systematic study on the GFS microbiome is available. This knowledge gap makes it challenging to assess how the world's GFS microbiome might respond to climate-induced glacier shrinkage. Therefore, to fill this gap, we first need to put the GFS microbiome on the map of the world's major mountain ranges and unravel the eco-evolutionary processes that generate and maintain its regional and global biodiversity.

In this article, we provide a global atlas of the bacterial microbiome in GFS benthic biofilms and infer mechanisms underpinning its biodiversity and biogeography. To this end, the Vanishing Glaciers project used standardized protocols to systematically collect benthic biofilms and environmental data related to glaciology, streamwater physicochemistry and sediment mineralogy from 152 GFSs spanning the Southern Alps of New Zealand, Nepalese Himalayas, Caucasus Mountains, Pamir and Tian Shan, European Alps, Scandinavian Mountains, Southwest Greenland, Alaska Range, the Rwenzori Mountains in Africa and both Ecuadorian and Chilean Andes (Fig. 1a and Extended Data Fig. 1a). This sampling effort covered an elevational gradient from 93 m above sea level (a.s.l.) in Greenland to 5,093 m a.s.l. in the Himalayas. We sampled benthic biofilms from triplicate sediment patches within one upstream reach, as close to the glacier snout as possible, and from one downstream reach, around the terminal moraine of the Little Ice Age (Methods), to capture the environmental heterogeneity of GFSs. Leveraging a total of 54,837 bacterial amplicon sequence variants

[1]River Ecosystems Laboratory, Alpine and Polar Environmental Research Center, Ecole Polytechnique Fédérale de Lausanne, Sion, Switzerland. [2]MARBEC, University of Montpellier, CNRS, IFREMER, IRD, Montpellier, France. [3]UK Centre for Ecology and Hydrology, Wallingford, UK. [4]Systems Ecology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [5]Department of Ecology, Faculty of Science, Charles University, Prague, Czechia. [6]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia. [7]Institut de Physique du Globe de Paris, Paris, France. ✉e-mail: leila.ezzat@gmail.com; tom.battin@epfl.ch

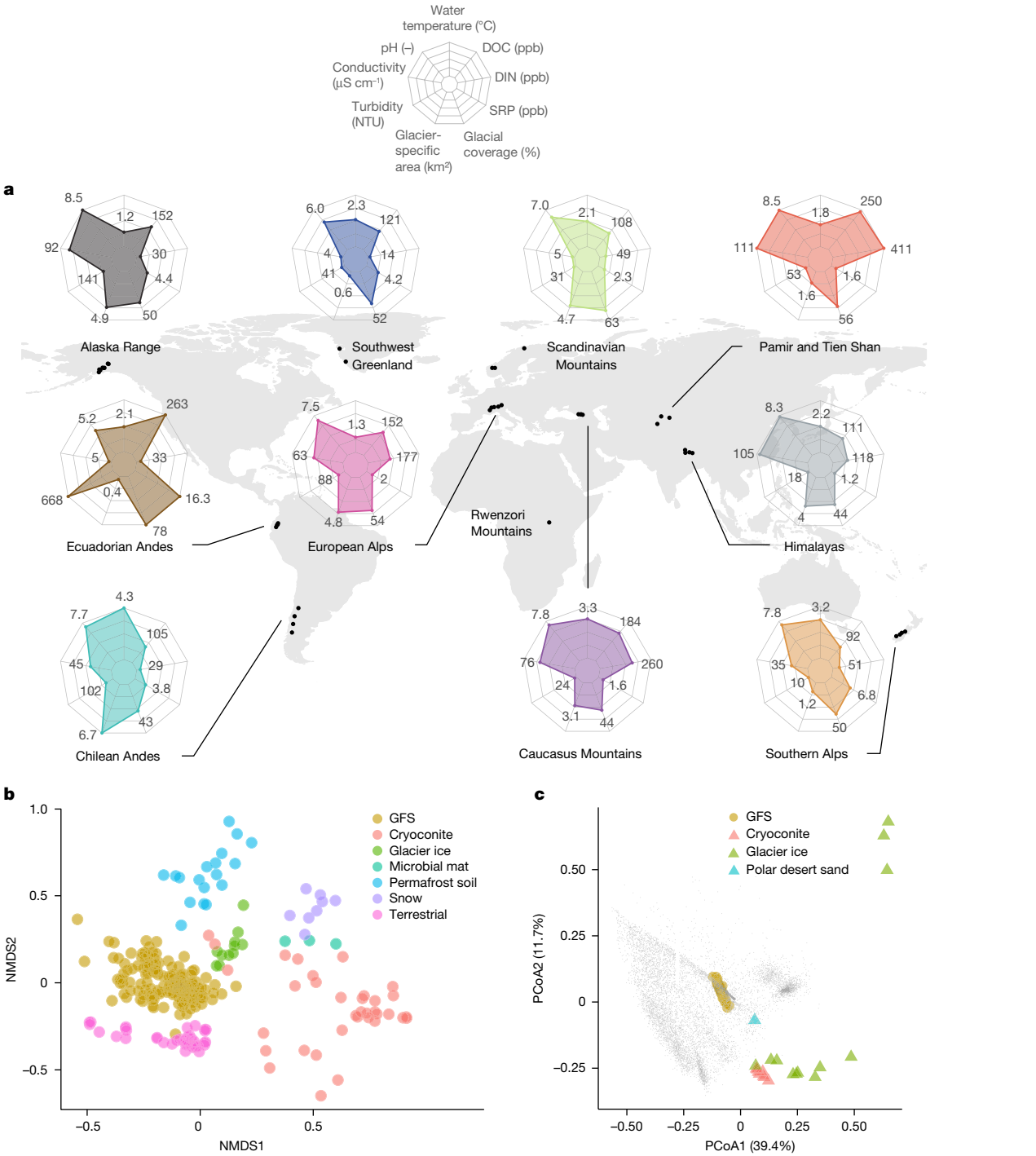


Fig. 1 | GFSs sampled by the Vanishing Glaciers project, their environmental characteristics and microbiome structure and function. a, World map showing the mountain ranges and locations (black dots) where the Vanishing Glaciers project sampled GFSs. Radar charts depict key physicochemical parameters of GFS water and glacier metrics. Shown are median values of normalized parameters. No radar chart is shown for the Rwenzori Mountains, because only one GFS was sampled. DIN, dissolved inorganic nitrogen; DOC, dissolved organic carbon; NTU, nephelometric turbidity units; ppb, parts per billion; SRP, soluble reactive phosphorus. **b**, Non-metric multidimensional scaling (NMDS) based on Bray–Curtis dissimilarity, illustrating the composition of bacterial communities of 152 GFSs and other cryospheric ecosystems[21] (n_{sample} = 268, k = 2, stress = 0.17). Taxonomic composition of the GFS microbiome differs significantly from that of other microbiomes, as shown by both permutational analysis of variance (PERMANOVA, $F_{6,264}$ = 8.6, R^2 = 0.16, P = 0.001; pairwiseAdonis, P_{adj} < 0.02 for all tests, based on 151 GFSs, excluding the Rwenzori Mountains) and multivariate generalized linear models (P_{adj} = 0.01 for all tests; Supplementary Table 2). **c**, Principal coordinate analysis (PCoA), based on 8,502 KEGG normalized abundances of KOs, shows that the functional composition of the GFS differs from that of other cryospheric microbiomes (PERMANOVA, $F_{1,111}$ = 92.5, R^2 = 0.46, P = 0.001; n_{sample} = 189; see Supplementary Table 5 for adjusted pairwise comparison P values). Grey dots indicate underlying KOs contributing to dissimilarity (Bray–Curtis dissimilarity) among GFSs and other cryospheric microbiomes. Highlighted are relevant pathways (thiamine and thienamycin biosynthesis; amino acid, carbon, pyruvate and sulfur metabolism; and bacterial motility, two-component system, hydrolases, energy metabolism and uncharacterized proteins), based on grouped KOs, contributing to the variation within and among samples and cryospheric systems, respectively.

(ASVs) and 8,518 Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologue groups (KOs), we first compared the structure and function of the GFS microbiome with other cryospheric ecosystems. We then performed analyses rooted in microbial biogeography and niche- and dispersal-based community ecology (for example, refs. 16,17) to disentangle the effects of historical contingencies (for example, dispersal limitation) and contemporary environmental selection on the biodiversity and biogeography of the global GFS microbiome. Thereby, we shed new light on an old debate in microbial biogeography, rooted in the Baas–Becking hypothesis positing that 'everything is everywhere' but that 'the environment selects'[18]. Our study unveils a distinct bacterial microbiome at the tips of some of the world's largest river systems at risk because of climate change, thereby further aggravating the already imperilled biodiversity of the Earth's running waters[19].

The GFS environment

Glacier-fed streams figure among Earth's harshest freshwater ecosystems. Due to their inextricable interface with glaciers, they share some common characteristics to which ecological communities have adapted, and that manifest independently of geography (Fig. 1a, Supplementary Table 1 and Extended Data Fig. 1b). For instance, the major contributions of glacier meltwater to GFS runoff maintain water temperatures near freezing point. Given recently deglaciated terrain at the glacier snout, GFS sediments are poorly consolidated and regularly scoured, particularly by daily meltwater peaks[19]. These mobilized sediments, along with glacial weathering products, increase streamwater turbidity during summer, thereby limiting light availability and photoautotrophic production in GFSs. Although repeated streambed reworking, and possibly also debris flows, may affect life in GFSs, such legacy effects are difficult to quantify. On an annual basis, photoautotrophic production is further reduced by prolonged darkness during winter when ice and snow cover GFSs. This, together with dilute glacial meltwater, results in overall oligotrophic conditions, depriving the microbial food web of organic carbon and inorganic nutrients[2,20]. On a regional scale (for example, Ecuadorian Andes, Greenland), catchment geology and associated sediment mineralogy can imprint on GFS streamwater geochemistry, including pH and ions (Extended Data Fig. 1b), which may be relevant in regard to microbial redox reactions.

A distinct bacterial microbiome

Glacier-fed streams share some of their fundamental environmental constraints (for example, near-freezing temperatures, oligotrophy) with other cryospheric ecosystems. Despite this, we found marked differences in the composition of benthic GFS bacteria and those contained in glacier cryoconite, proglacial lakes and permafrost soils[21] (Fig. 1b and Supplementary Table 2). To further explore these apparent differences, we investigated the GFS and other cryospheric microbiomes for patterns of functional potential as inferred from metagenomes (Methods). In line with patterns of taxonomic composition, we found significant differences in KEGG pathways segregating the GFS microbiome from other cryospheric microbiomes (Fig. 1c). Notably, pathways involved in biofilm formation (for example, secretion and two-component systems) drive this segregation. The biofilm mode of life protects bacteria from flow-induced erosion and resource scarcity in streams, offering a clear adaptive advantage in GFSs. It also enables bacteria to diversify energy acquisition and metabolic pathways (for example, degradation of amino acids, sugars and xenobiotic compounds and sulfur and nitrogen metabolism; Extended Data Fig. 2) to exploit resources when they become available (for example, pulses of algal exudates during vernal and autumnal windows of opportunities), which can vary from weeks to months in GFSs, depending on glacier influence[20].

Global biodiversity patterns

Whereas the cryospheric microbiome has gained attention in recent years, we still lack a comprehensive appreciation of the bacterial biodiversity of GFSs globally. Our survey (including Mount Stanley in the Rwenzori Mountains) unveiled bacteria spanning 44 phyla (Fig. 2a) and comprising novelty across all taxonomic ranks, including 219 ASVs unclassified at class level and 4,545 unclassified at genus level. For systematic assessment of biodiversity patterns (that is, gamma diversity per mountain range), we first explored asymptotic estimates based on incidence data (Methods). This shows that we sampled a median of 86.2% (interquartile range (IQR) 79.7–90.03%) of ASVs per mountain range (Fig. 2b and Supplementary Table 3), indicating that our sampling is representative and enabling us to draw generalizable conclusions about bacterial biodiversity in GFSs. We also acknowledge that these figures imply that there is further bacterial diversity yet to be discovered in the world's GFSs, especially given our conservative filtering and denoising approaches (Methods).

We expected relatively low bacterial biodiversity and biomass in GFSs due to overall low resource availability. However, in exploring alpha diversity, we found median observed richness values of 2,791 ASVs (IQR: 2,268–3,346 ASVs) (Fig. 2c) and Shannon H (based on Hill numbers) of 327.0 (IQR: 200.8–464.1 ASVs). These values fall between those found for other cryospheric ecosystems (Supplementary Table 4) and even for globally distributed soils[22,23], suggesting that GFS sediments provide niches that sustain a level of bacterial diversity similar to that in other cryospheric and matrix-dominated systems. The notion of viable niches in GFSs is corroborated by sediment bacterial abundances (median, 7.6×10^6 cells per gram, IQR: 3×10^6–1.8×10^7 cells per gram; Supplementary Table 1), bracketed by abundances reported from other extreme sedimentary environments (for example, refs. 24,25), and further by marked compositional differences between sediment communities and those suspended in streamwater or locked in glacier ice[14,26,27].

Rank abundance curves (Extended Data Fig. 3a) point to a few abundant ASVs structuring alpha diversity, with 11.1% of all ASVs having a relative abundance above 0.1%. Comparison with log-normal rank-abundance models (Extended Data Fig. 3b) suggests a paucity of rare ASVs[23]. Rarity, a hallmark of many microbial communities, can either reflect microbial lifestyle adaptations or arise from local colonization and extinction dynamics[23]. We attribute the limited rarity of GFS bacteria to the unstable and selective environment, which is probably unfavourable for transient and rare taxa, thereby increasing the risk of local extinction.

Biodiversity patterns in mountains follow not only latitudinal but also elevational gradients—a recurrent topic in microbial ecology[28]. However, how taxonomic and functional diversity of the GFS microbiome changes with elevation at the global scale remains unclear. Leveraging our 5,000 m elevational gradient and accounting for latitudinal effects, we found that ASV richness decreases with elevation (Extended Data Fig. 4a). Whereas such elevational patterns have been reported for GFSs from several regions[14], our findings underline their global character, consistent with findings from streams draining non-glacierized catchments. We suggest that shrinking metacommunity size with greater elevation from which local communities recruit[29], along with increasing energetic constraints[28] imposed by mountain topography, elevation and glacier influence, may collectively underlie the observed elevational gradients in ASV richness. The notion of glacier influence is indeed supported by the fact that ASV richness also decreases with increasing glacier coverage within the catchment (Extended Data Fig. 4b). Interestingly, the potential functional diversity of GFS bacteria did not follow the elevational gradient of ASV richness, as inferred from 8,518 KOs, from our metagenomes (Extended Data Fig. 4c and Supplementary Table 5). This contrasts with findings from mountain streams that are not under glacier influence, where climate-related parameters largely drive microbial functional diversity[30]. We attribute

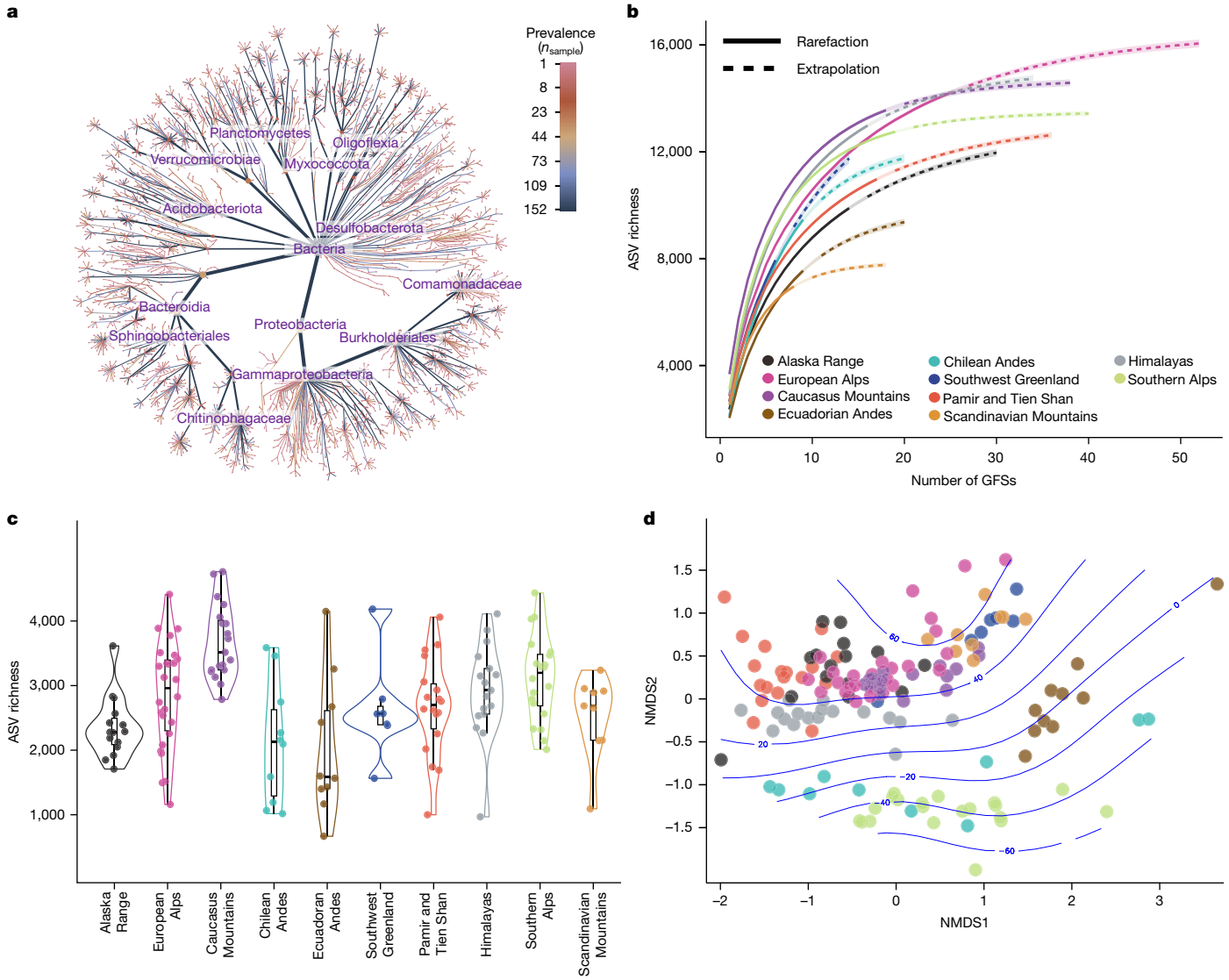


Fig. 2 | Global diversity patterns of the GFS microbiome. a, Heat tree illustrating the bacterial taxonomic structure, from domain to genus level, and highlighting the most prevalent lineages. Edge colour reflects occupancy, and node size indicates ASV frequency in a given lineage. **b**, Observed and estimated gamma diversity per mountain range based on asymptotic estimates ($n = 151$ GFSs). **c**, Violin plots illustrating ASV richness across mountain ranges. Horizontal lines represent the median, box height represents IQR and whiskers extend 1.5 times beyond IQR. Each dot represents one GFS ($n = 151$ GFSs). **d**, NMDS, based on Bray–Curtis dissimilarity, illustrating the composition of bacterial communities from 151 GFSs across ten mountain ranges (excluding the Rwenzori Mountains; $k = 2$, stress = 0.15). Colours correspond to mountain ranges shown in **b**. Also shown are latitudinal isolines. Community composition varied significantly across mountain ranges based on (1) PERMANOVA (based on 151 GFSs, excluding the Rwenzori Mountains, $F_{9,147} = 7.8$, $R^2 = 0.33$, $P_{mountain\_range} = 0.001$; pairwiseAdonis, two-sided tests, $P_{adj} = 0.045$ for all tests, except for the Scandinavian Mountains and Southwest Greenland; pairwiseAdonis, $P_{adj} = 0.08$) and (2) multivariate generalized linear models ($P_{adj} = 0.01$ for all tests; Supplementary Table 7). Community composition varied significantly between hemispheres, based on both PERMANOVA ($F_{1,147} = 15.9$, $R^2 = 0.097$, $P_{hemisphere} = 0.001$) and multivariate generalized linear models (two-sided tests; $P_{adj} = 0.01$; Supplementary Table 8).

this apparent discrepancy to the dominance of glacier influence over climatic controls on GFS bacteria, and to its functional repertoire required to thrive in an ecosystem with globally similar environmental constraints (Extended Data Fig. 1b).

Our worldwide sampling effort enabled us to resolve the spatial structure of the GFS bacterial microbiome. We noted high community dissimilarities (Bray–Curtis dissimilarity index) both within (0.68, IQR: 0.64–0.70) and between (0.89, IQR: 0.82–0.91) mountain ranges (Extended Data Fig. 5a). This pattern was mirrored by the presence–absence-based Sørensen index, indicating that ASV replacement drives beta diversity (Extended Data Fig. 5b and Supplementary Table 6). Through further analysis of beta diversity, we identified conspicuous biogeographical patterns in GFS bacterial composition, with a significant segregation between the Southern and Northern hemispheres, and with GFS bacterial composition clustering according to mountain range (Fig. 2d and Supplementary Tables 7 and 8). This biogeographical pattern was driven by 20,522 indicator ASVs (median relative abundance 21%, IQR: 15.3–37.2%; Extended Data Fig. 6a,b), as shown by dissimilarity-based multivariate analysis of variance (Methods). Indicator ASVs contributed 35.4% to global beta diversity (as total variance; Methods), emphasizing the relevance of regionality (that is, at the scale of mountain ranges) for the compositional turnover of the GFS microbiome. Pronounced dissimilarities among mountain ranges suggest that spatial processes (for example, dispersal limitation, spatial structuring of the environment) play a role in shaping GFS microbiome structure. Patterns of taxonomic composition were not reflected by spatial patterns of functional composition based on KOs (Extended Data Fig. 7a,b). Moreover, KO numbers were essentially invariant

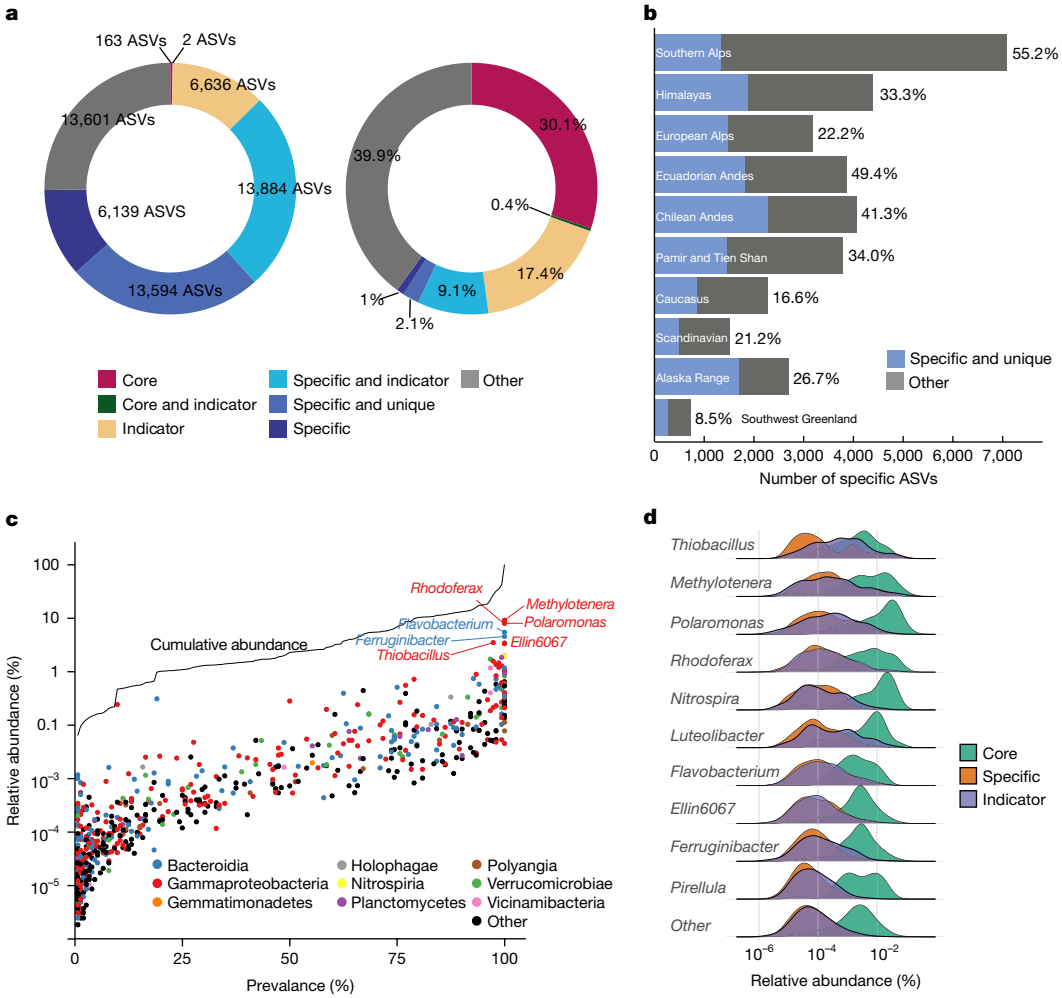


Fig. 3 | Components and taxonomy of the GFS microbiome. a, Absolute (left) and relative (right) distribution of the various components (that is, specific, core, unique and indicator ASVs) of the GFS microbiome from 151 GFSs (excluding Rwenzori Mountains). **b**, Distribution of specific ASVs across mountain ranges, shown as absolute numbers (*x* axis) and fraction (%) of ASVs per mountain range. Blue and grey correspond to unique and other ASVs, respectively. **c**, Distribution of global GFS microbiome classes according to their relative prevalence and abundance. Dot colour denotes class; line represents cumulative relative abundance as a function of prevalence; *y* axis represents the fraction of samples in which these taxa are present. **d**, Relative abundance of the ten most abundant genera for indicator, core and specific ASVs.

(ranging 7,202–7,928) across the global GFS alpha-diversity gradient (Extended Data Fig. 7c), with 75.3% present in all GFSs. Levin's niche breadth, derived from KO prevalence and abundance, suggests that common KOs explain the subtle differences in global GFS functional potential (Extended Data Fig. 7d,e). Consistent with the elevational patterns of taxonomic and functional diversity, this suggests a decoupling of the structure and functional potential of the GFS bacterial microbiome, probably due to functional redundancy among closely related community members. Indeed, functional redundancy can confer resilience to microbial communities[31], which would be particularly advantageous within fluctuating environments such as GFSs[32].

High mountain-range specificity

These marked beta-diversity patterns prompted us to further explore the GFS bacterial microbiome for ASVs that are specific to single mountain ranges (hereafter referred to as 'range-specific' ASVs) (Methods). The determination of range-specific ASVs is not trivial because of sampling effort, taxonomic resolution, spatial scales and rarity[16]. Therefore, we applied a conservative data pretreatment, benchmarked denoising approaches (that is, DADA2 (ref. 33), Deblur[34] and UNOISE[35]), and conducted analyses of mock communities and sensitivity analyses for each, to investigate the effects of rarefaction depth and the number of GFSs sampled per mountain range (Methods and Extended Data Fig. 8). As a result of these efforts, we retained Deblur which, in combination with our filtering, yielded the most conservative and consistent estimates of specific ASVs and biodiversity (Extended Data Fig. 8).

Following this methodology, we identified 62.2% of all ASVs as range specific, occurring at low relative abundances (9.0%; IQR: 3.7–22.2%; Fig. 3a). Despite low relative abundances, specific ASVs contribute 20.9% to beta diversity as estimated by total variance (Methods), underscoring their relevance in regard to overall GFSs bacterial biodiversity. Furthermore, 25.2% of all ASVs (excluding the only GFS from the Rwenzori Mountains) were repeatedly found across multiple sediment samples in a single GFS only, and are therefore considered 'unique' (Methods and Fig. 3a); by definition, unique ASVs are also range specific. Notably, 869 (6.4%) of unique ASVs were found in all sediment samples (six out of six) collected within a given GFS.

Amplicon sequence variant specificity was not uniformly distributed across mountain ranges (Fig. 3b). Instead, it was highest in the Southern Alps of New Zealand and the Ecuadorian Andes, consistent with the elevated endemicity of flora and fauna in these regions[36]. Such patterns of floral and faunal endemism align with island biogeography, with the Southern Alps and Ecuadorian Andes emerging as biogeographical

'islands' from the ocean and tropical lowlands, respectively, and also with the fact that spatial isolation, selection pressure and long-term climate stability promote speciation and endemism in the high mountains[37]. Although it cannot be unequivocally demonstrated whether specific ASVs are also endemic to mountain ranges, the GFS bacterial microbiome adds to the ongoing debate on endemism in microbes (for example, ref. 38).

Range specificity and the uniqueness of GFS bacteria underscore the relevance of spatial isolation and selective pressures for biogeographic patterns across spatial scales. Long-distance dispersal limitation between mountain ranges produces regional isolation. Furthermore, GFSs are often incised within mountainous landscapes and form the tips of dendritic river networks, which limits dispersal across headwaters even within the same network[39]. The GFS bacterial microbiome is therefore an example that refutes the idea of 'everything is everywhere' that has been so influential in microbial biogeography[18].

A small core microbiome

The taxonomy of the GFS bacterial microbiome is characterized by a few dominant phyla, including Pseudomonadota, Bacteriodota, Gemmatomonadota and Verrucomicrobiota. Within these phyla, relatively few genera (for example, *Methylotenera*, *Polaromonas*, *Rhodoferax*) prevail, whereas numerous genera occur at low abundance and prevalence (Fig. 3c). The prevalence of a few genera prompted us to further explore the GFS microbiome for core ASVs. Consistent with the high dissimilarities among and within mountain ranges, we identified only 165 (0.42%) core ASVs (based on over 0.1% of relative abundance and over 50% of prevalence thresholds (Methods) contributing to 28.3% (IQR: 21.2–38.04%) of total relative abundance. Similarly small core microbiomes were also reported from global surveys of wastewater, freshwater and soil systems (for example, refs. 40,41). Small but relatively abundant core taxa probably reflect their competitive advantage. Interestingly, core, specific and indicator ASVs broadly share taxonomies (Fig. 3d), dominated by the phyla Pseudomonadota (55.9 %), Bacteroidota (18.0%) and Verrucomicrobiota (5.5%), the families Comamonadaceae (22.2%), Methylophilaceae (8.6%) and Chitinophagaceae (5.2%) and the genera *Methylotenera* (8.2%), *Polaromonas* (8.0%) and *Rhodoferax* (5.5%). Several of these genera were also reported from other cryospheric ecosystems (Extended Data Fig. 9), underscoring their ability to thrive in these extreme environments.

Drivers of GFS microbiome composition

Biogeographical patterns result from ecological and evolutionary processes, including contemporary environmental selection and historical processes (that is, dispersal limitation, drift and past environmental selection)[42]. Exploring distance-decay patterns (DDPs) for all ASVs (that is, the entire GFS microbiome) across distances ranging from 95 m to over 18,000 km, we found that the GFS microbiome is spatially structured (Fig. 4a). Both ASV replacement and changes in relative abundance drive DDPs generated by Sørensen and Bray–Curtis dissimilarity indices, respectively (Fig. 4a). In line with the dominance of a few taxonomies across GFSs worldwide, weaker DDPs produced by phylogeny-informed analyses suggest that phylogenetically similar taxa are present even across distant GFSs (Fig. 4a).

To further quantify the processes underlying these biogeographical patterns, we computed distance-based redundancy analyses (db-RDAs) and multiscale variance partitioning (accounting for a linear gradient based on geographic coordinates and altitude, along with spatial and environmental variations) among and within mountain ranges (Methods). The model explained 54.9% of total compositional variance within the GFS bacterial microbiome. The linear gradient explained 15.9% of total variance and pure spatial processes, both among and within mountain ranges, explained 23.6%, whereas pure and spatially structured environmental processes explained 15.4% (Fig. 4c, Supplementary Table 9 and Supplementary Note 1). The dominance of spatial effects for GFS bacterial beta diversity contrasts with that of environmental effects on various terrestrial and aquatic microbial communities[16]. This finding highlights the role of geographic isolation and dispersal limitation for structuring the GFS bacterial microbiome, consistent with the numerous specific and unique bacteria dwelling in global GFSs.

In line with regional differences in some of the GFS environmental properties (Fig. 1a and Supplementary Table 10), spatially structured environmental processes among and within mountain ranges explained 9.2 and 4.2% of explained variance, respectively (Fig. 4b). Much of this can be attributed to regional differences in catchment geology and associated streamwater geochemistry, as illustrated by volcanic rocks lowering streamwater pH in Ecuador (Fig. 1a and Extended Data Fig. 1b). However, whereas differences in pH presumably reflect weathering of different bedrock types, dissolved inorganic nitrogen and precipitation are also influential (Supplementary Table 10), further indicating the importance of atmospheric processes. The interactive roles of geology and atmosphere for GFS bacteria at the scale of mountain ranges are consistent with the emerging awareness that geodiversity (that is, combined geological, geomorphological and hydrological processes) can shape biodiversity, particularly in mountains[43,44].

In contrast to the composition of GFS bacterial communities, their functional potential (as KOs) does not demonstrate strong biogeographical patterns (Fig. 4c). In fact, DDPs show that KO turnover remains largely similar across geographic distances whereas KO abundances exhibit a weak spatial trend, which we attribute to the overall low KO dissimilarity (Extended Data Fig. 7). We interpret these contrasting biogeographical patterns between GFS bacterial composition and functional potential as further supporting the notion of community-level functional redundancy. However, more studies are needed to verify such a pattern, ideally utilizing proteomics, because evolutionary changes in bacterial genomes could modulate the extent of their functional redundancy[45].

Phylogeography of the microbiome

The phylogenetic structure of ecological communities bears evolutionary and ecological signatures, which can inform on processes underlying biogeographic patterns[16]. Within this framework, we analysed the phylogeography of GFS bacteria to assess the role of deterministic (that is, environmental selection) and stochastic (that is, dispersal and drift) assembly processes. Among these processes, homogeneous environmental selection can induce phylogenetic clustering of closely related ASVs prevailing within putatively well-adapted clades, whereas dispersal limitation can be reflected by higher compositional turnover than that expected by chance[46]. Consequently, homogeneous selection may promote community similarity whereas dispersal limitation would promote dissimilarity and specificity.

Quantifying phylogenetic and compositional turnover compared with null-model expectations[47] (Methods), we found that dispersal limitation was the most common assembly process (59.4% of community pairs), followed by ecological drift (30.0%) and homogeneous selection (11.7%) (Fig. 5a). Whereas dispersal limitation shaped clades across the entire phylogenetic tree, homogeneous selection consistently prevailed among several clades. Accordingly, clades predominantly shaped by dispersal limitation accounted for 45.4% (IQR: 38.8–54.0%) of relative abundance within each mountain range. In comparison, the contribution of clades under homogeneous selection to relative abundance varied more across mountain ranges. These clades contributed 43.9% and 30.4% to relative abundance in the Caucasus Mountains and New Zealand, respectively, whereas they garnered 16.6% in the Chilean Andes and 18.8% in Southwest Greenland. These clades included the globally distributed genera *Methylotenera* and *Polaromonas*, as well

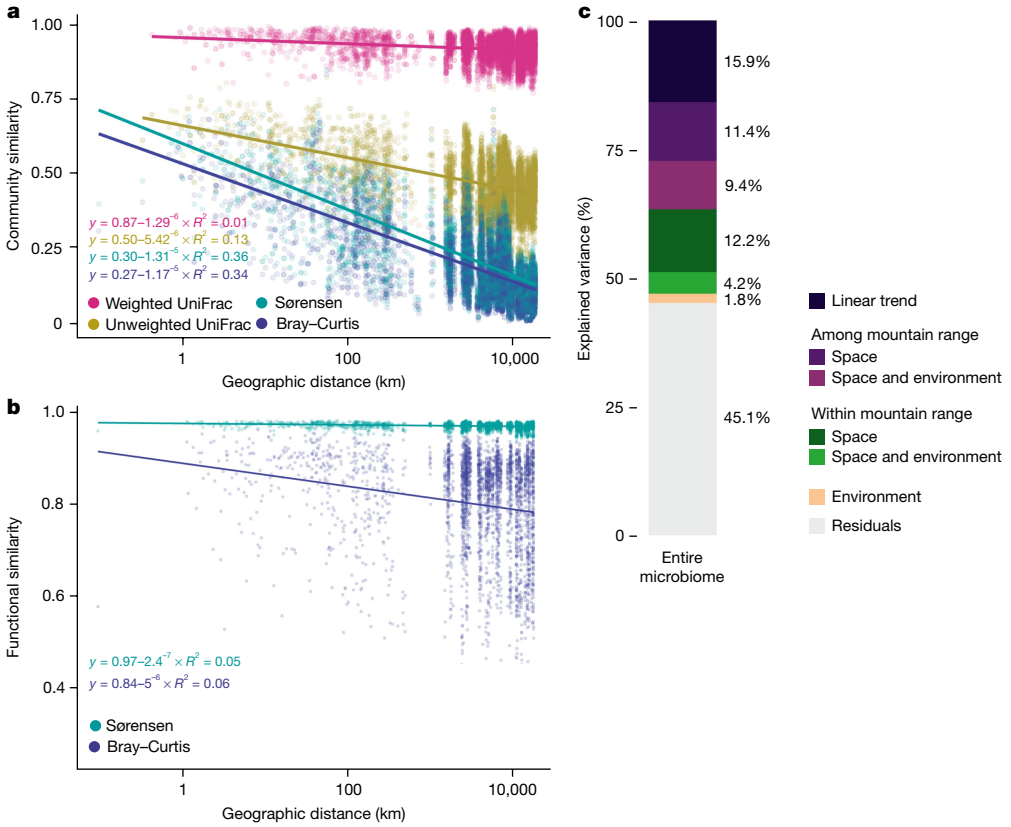


Fig. 4 | Processes driving GFS biogeographic patterns. a, DDPs for the GFS microbiome based on 151 GFSs and geographic distance, ranging from 95 m to over 18,000 km. Similarity significantly decreases across GFSs with increasing geographic distance, independently of the dissimilarity indices or distances used (Mantel tests: $r_{BC} = 0.58$, $P_{BC} = 0.001$, $r_{SOR} = 0.60$, $P_{SOR} = 0.001$, $r_{WU} = 0.11$, $P_{WU} = 1 \times 10^{-4}$, $r_{UW} = 0.36$, $P_{UW} = 8 \times 10^{-4}$; two-sided tests; where BC is Bray–Curtis, SOR is Sørensen, WU is weighted UniFrac and UW is unweighted UniFrac). Significant differences were noted between slopes of DDPs based on Sørensen and Bray–Curtis dissimilarity indices (linear model, estimate = −0.0024, s.e. = 0.0006, $t = -3.958$, $P = 7.57 \times 10^{-5}$, adj. $R^2 = 0.37$) and those based on weighted and unweighted UniFrac distances (linear model, estimate = −0.016, s.e. = 0.0004, $t = -41.73$, $P < 2 \times 10^{-16}$, adj. $R^2 = 0.9$). Because of the large file size only a subset of data is shown, whereas equations and statistics reflect the full dataset (see full figure in data repository). **b**, KOs based on 84 GFSs. Decrease in similarity was significant for both indices ($r_{BC\_KEGG} = 0.24$, $P_{BC\_KEGG} = 0.001$, $r_{SOR\_KEGG} = 0.21$, $P_{SOR\_KEGG} = 0.005$). As a result, dissimilarity increased more rapidly with geographic distance based on Bray–Curtis compared with Sørensen (permutation tests for analysis of variance (aovperm), $F_{1,13940} = 161.4$, $P = 0.0002$). **c**, Variance partitioning of beta diversity of the GFS microbiome ($n = 140$ GFSs) among spatial and environmental components within and among mountain ranges. Shown are adjusted R^2 values (%) for each component, including residuals. The model explained 54.9% of total variance (linear gradient, 15.9%; analysis of variance (ANOVA) on a db-RDA model, $F_{3,134} = 9.8$, $P = 0.001$; among mountain ranges: 20.8%, ANOVA $F_{9,131} = 5.97$, $P = 0.001$; within mountain ranges: 16.4%, ANOVA $F_{60,122} = 1.7$, $P = 0.001$; based on two-sided tests).

as less abundant genera, such as *Pirellula* (Planctomycetota) and *Ferruginibacter* (Bacteroidota). We found homogeneous environmental selection invariant over geographical distance, whereas dispersal limitation became more important with increasing distance between GFSs within the same mountain range (Fig. 5b). These phylogeographic findings further support our observations from variance partitioning, that dispersal limitation and globally consistent environmental constraints jointly shape the GFS microbiome. The apparent relevance of ecological drift also points towards the role of stochasticity in structuring these communities, and is potentially related to the environmental instability typical for these nascent ecosystems[48].

Whereas environmental constraints may imprint on the phylogeny of clades under selection, we anticipated that spatial isolation would result in the turnover of phylogenetically closely related ASVs, even among GFSs within the same mountain ranges. To test this hypothesis, we assessed how beta diversity changed as we successively agglomerated the phylogenetic tree from its tips inwards (Methods). We found that beta diversity within mountain ranges decreased continuously from 0.383 (IQR: 0.355–0.390) to 0.238 (IQR: 0.208–0.256) along the inwards gradient of the phylogenetic tree (Fig. 5c). Regional beta diversity within mountain ranges decreased by 27.5% (IQR: 24.5–28.9%) at phylogenetic distances shorter than the average phylogenetic distance among ASVs. The exponential decrease of beta diversity in all mountain ranges suggests that most compositional turnover indeed occurs among phylogenetically closely related taxa at the tips of the phylogenetic tree. Similarly, the number of unique ASVs decreases exponentially with increasing phylogenetic agglomeration, reflecting that unique and specific ASVs are phylogenetically closely related to ASVs that occur in multiple mountain ranges. We suggest that this reflects the time scales relevant for global dispersal (and dispersal limitation) in GFS.

Furthermore, we found that several of the genera under homogeneous selection are microdiverse—that is, they contain numerous ASVs with distinctly shorter phylogenetic distances compared with other genera (Fig. 5d). Among microdiverse genera are prevalent members of the GFS core microbiome, including *Polaromonas* (0.09; IQR: 0.07–0.10), *Rhodoferax* (0.083; IQR: 0.079–0.087) and *Methylotenera* (0.093; IQR: 0.085–0.098) (Fig. 5d). Whereas microdiversity was previously reported from several GFSs in New Zealand[15], our global survey shows that microdiverse genera resolve global biogeographic patterns similar to those observed for the entire GFS microbiome (Extended Data Fig. 10a). That microdiversity can contribute to biogeographic

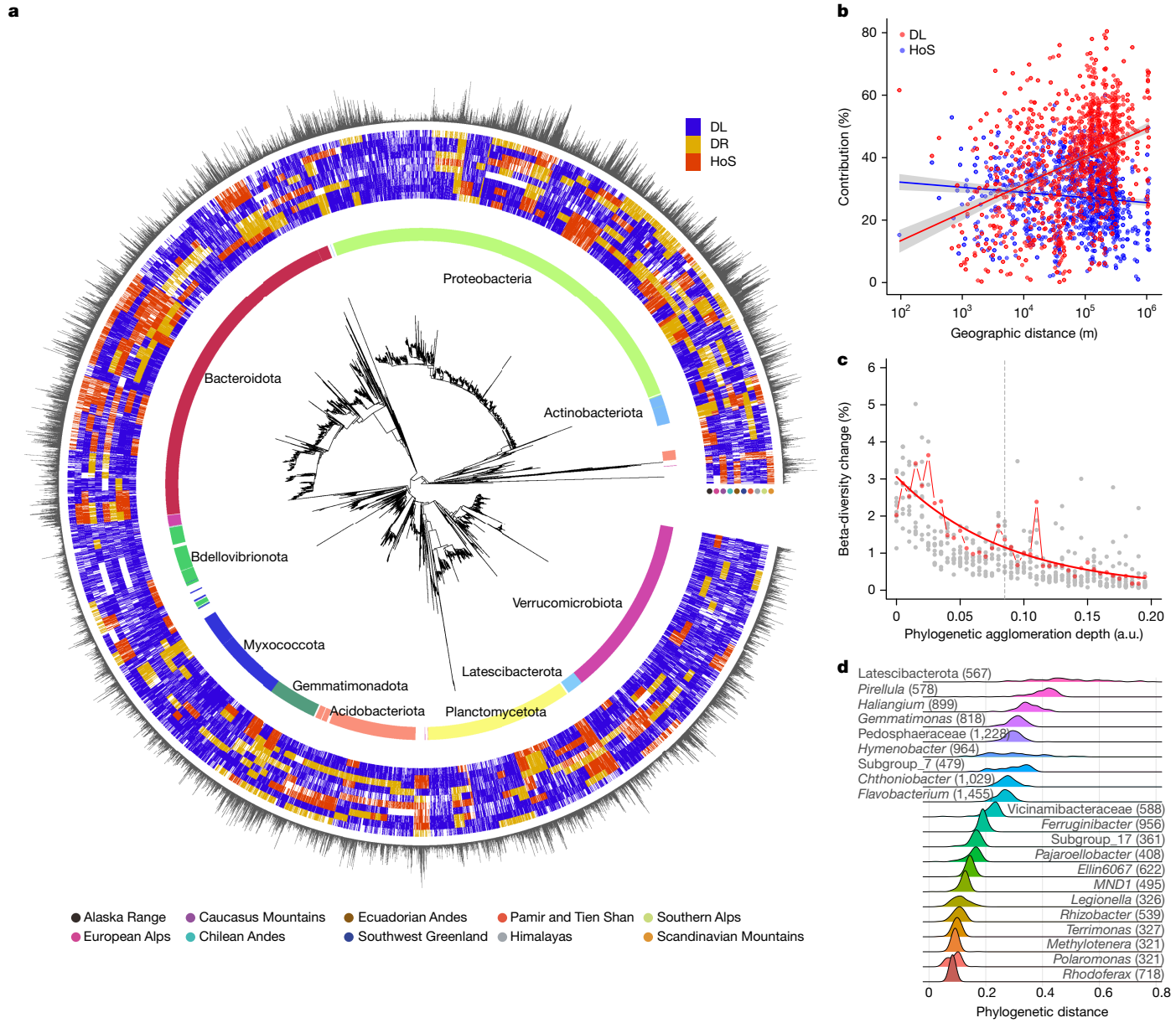


Fig. 5 | Phylogenetic structure of the GFS microbiome. a, Phylogenetic tree amended with phylum-level taxonomic information. The outer rings show bin-level dominant assembly processes for each mountain range. Homogeneous environmental selection (HoS), dispersal limitation (DL) and ecological drift (DR) dominate GFS community assembly across mountain ranges. Phylogenetic clades were often consistent across multiple regions under the influence of HoS. **b**, Whereas HoS was invariant over spatial distance among mountain ranges, the importance of dispersal limitation increased with increasing distance among GFS. Lines and shaded areas represent, respectively, linear model fits and standard error for HoS (linear model, $\text{estimate}_{\text{slope}} = -1.04 \times 10^{-5}$, s.e. $= 1.58 \times 10^{-6}$, $t = -6.58$, $P = 7.1 \times 10^{-11}$, adj. $R^2 = 0.03$) and for dispersal limitation (linear model, $\text{estimate}_{\text{slope}} = 2.39 \times 10^{-5}$, s.e. $= 2.36 \times 10^{-6}$, $t = 10.13$, $P < 2 \times 10^{-16}$, adj. $R^2 = 0.08$). **c**, Turnover among phylogenetically closely related ASVs dominated both regional (grey) and global (red) beta diversity. Agglomerative merging of phylogenetic distances (from tips towards the root) resulted in a rapid decrease in regional and global beta diversity. The greatest change in beta diversity occurred when closely related community members were considered, whereas deeper-branching phylogenies had little influence. For visual guidance, the mean nearest taxon distance is shown as a vertical dashed line, and exponential models by solid red lines. a.u., arbitrary units. **d**, Clades under the influence of HoS (**a**) include microdiverse genera. Distributions of phylogenetic distances within genera that contribute to relative abundance, diversity and beta diversity in GFS are shown. The number of ASVs per genus is given in parentheses. Genera with a high degree of phylogenetic clustering are considered microdiverse.

patterns is further corroborated by the genus *Polaromonas* which, despite being cosmopolitan, exhibits phylogeographic signatures (Extended Data Fig. 10b).

Theory posits that microdiversity can lead to the optimization of niche space[49], which is arguably constrained in GFSs due to limited resources and strong environmental selection. Our findings suggest that few clades are capable of optimizing niche space exploitation in GFSs, albeit through microdiversification rather than diversifying into deeper-branching taxa. Furthermore, our findings suggest that fine-tuned niche exploitation and dispersal limitation collectively foster microdiversification, which ultimately contributes to the global biodiversity and biogeography patterns of GFS bacteria.

In conclusion, we present a comprehensive study showing a bacterial microbiome that is taxonomically and functionally distinct from other cryospheric microbiomes, and characterized by high regional specificity and even local uniqueness. Dispersal limitation across spatial scales and strong selection imposed by the GFS environment shape a microbiome with a small core and biogeographic patterns that arise from

variation within a few dominant but microdiverse clades. We attribute this microbiome structure to the GFS environment, which maintains selective filters over time scales relevant for the eco-evolutionary dynamics of bacteria at global scales. Community-level functional redundancy, associated with microdiversity, may help GFS bacteria resist and recover from natural environmental fluctuations. However, climate-induced changes beyond such natural fluctuations may put specific, particularly unique, ASVs at risk because geographic isolation limits recovery through dispersal. Our study fills knowledge gaps in microbial community assembly and biogeography, as well as in cryospheric microbial ecology. It serves as a fundamental reference for future studies on the GFS microbiome. The consequences of global warming on mountain glaciers and downstream ecosystems are profound and, unlike the microbiome of terrestrial environments[50], the GFS bacterial microbiome cannot be restored or managed. Its genetic potential should, therefore, be explored before it is too late.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08313-z.

1. Hugonnet, R. et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature* **592**, 726–731 (2021).
2. Milner, A. M. et al. Glacier shrinkage driving global changes in downstream systems. *Proc. Natl Acad. Sci. USA* **114**, 9770–9778 (2017).
3. Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M. & Packmann, A. I. The ecology and biogeochemistry of stream biofilms. *Nat. Rev. Microbiol.* **14**, 251–263 (2016).
4. Cauvy-Fraunié, S. & Dangles, O. A global synthesis of biodiversity responses to glacier retreat. *Nat. Ecol. Evol.* **3**, 1675–1685 (2019).
5. Hotaling, S., Hood, E. & Hamilton, T. L. Microbial ecology of mountain glacier ecosystems: biodiversity, ecological connections and implications of a warming climate. *Environ. Microbiol.* **19**, 2935–2948 (2017).
6. Pörtner, H.-O. et al. (eds) *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (Cambridge Univ. Press, 2019).
7. Ménot, G. et al. Early reactivation of European rivers ruring the last deglaciation. *Science* **313**, 1623–1625 (2006).
8. Immerzeel, W. W. et al. Importance and vulnerability of the world's water towers. *Nature* **577**, 364–369 (2020).
9. Clason, C. et al. Contribution of glaciers to water, energy and food security in mountain regions: current perspectives and future priorities. *Ann. Glaciol.* **63**, 73–78 (2022).
10. Jacobsen, D., Milner, A. M., Brown, L. E. & Dangles, O. Biodiversity under threat in glacier-fed river systems. *Nat. Clim. Chang.* **2**, 361–364 (2012).
11. Hotaling, S., Finn, D. S., Joseph Giersch, J., Weisrock, D. W. & Jacobsen, D. Climate change and alpine stream biology: progress, challenges, and opportunities for the future. *Biol. Rev.* **92**, 2024–2045 (2017).
12. Vega, E., Bastidas Navarro, M., Martyniuk, N., Balseiro, E. & Modenutti, B. Glacial recession in Andean North-Patagonia (Argentina): microbial communities in benthic biofilms of glacier-fed streams. *Hydrobiologia* **850**, 3965–3979 (2023).
13. Ren, Z., Gao, H., Elser, J. J. & Zhao, Q. Microbial functional genes elucidate environmental drivers of biofilm metabolism in glacier-fed streams. *Sci. Rep.* **7**, 12668 (2017).
14. Wilhelm, L., Singer, G. A., Fasching, C., Battin, T. J. & Besemer, K. Microbial biodiversity in glacier-fed streams. *ISME J.* **7**, 1651–1660 (2013).
15. Fodelianakis, S. et al. Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. *ISME J.* **16**, 666–675 (2022).
16. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
17. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).
18. Martiny, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
19. Milner, A. M. & Petts, G. E. Glacial rivers: physical habitat and ecology. *Freshw. Biol.* **32**, 295–307 (1994).
20. Kohler, T. J. et al. Global emergent responses of stream microbial metabolism to glacier shrinkage. *Nat. Geosci.* **17**, 309–315 (2024).
21. Bourquin, M. et al. The microbiome of cryospheric ecosystems. *Nat. Commun.* **13**, 3087 (2022).
22. Bastida, F. et al. Soil microbial diversity–biomass relationships are driven by soil carbon content across global biomes. *ISME J.* **15**, 2081–2091 (2021).
23. Shoemaker, W. R., Locey, K. J. & Lennon, J. T. A macroecological theory of microbial biodiversity. *Nat. Ecol. Evol.* **1**, 107 (2017).
24. Orsi, W. D. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat. Rev. Microbiol.* **16**, 671–683 (2018).
25. Danovaro, R., Corinaldesi, C., Rastelli, E. & Anno, A. D. Towards a better quantitative assessment of the relevance of deep-sea viruses, Bacteria and Archaea in the functioning of the ocean seafloor. *Aquat. Microb. Ecol.* **75**, 81–90 (2015).
26. Hotaling, S. et al. Microbial assemblages reflect environmental heterogeneity in alpine streams. *Glob. Chang. Biol.* **25**, 2576–2590 (2019).
27. Ezzat, L. et al. Benthic biofilms in glacier-fed streams from Scandinavia to the Himalayas host distinct bacterial communities compared with the streamwater. *Appl. Environ. Microbiol.* **88**, e00421–e00422 (2022).
28. Wang, J. et al. Embracing mountain microbiome and ecosystem functions under global change. *New Phytol.* **234**, 1987–2002 (2022).
29. Bertuzzo, E. et al. Geomorphic controls on elevational gradients of species richness. *Proc. Natl Acad. Sci. USA* **113**, 1737–1742 (2016).
30. Picazo, F. et al. Climate mediates continental scale patterns of stream microbial functional diversity. *Microbiome* **8**, 92 (2020).
31. Louca, S. et al. Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**, 936–943 (2018).
32. Allison, S. D. & Martiny, J. B. H. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl Acad. Sci. USA* **105**, 11512–11519 (2008).
33. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
34. Amir, A. et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* **2**, e00191-16 (2017).
35. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/081257 (2016).
36. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
37. Rahbek, C. et al. Humboldt's enigma: what causes global patterns of mountain biodiversity? *Science* **365**, 1108–1113 (2019).
38. Souza, V., Eguiarte, L. E., Siefert, J. & Elser, J. J. Microbial endemism: does phosphorus limitation enhance speciation? *Nat. Rev. Microbiol.* **6**, 559–564 (2008).
39. Rinaldo, A., Gatto, M. & Rodríguez-Iturbe, I. *River Networks as Ecological Corridors: Species, Populations, Pathogens* (Cambridge Univ. Press, 2020).
40. Nemergut, D. R. et al. Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* **13**, 135–144 (2011).
41. Wu, L. et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat. Microbiol.* **4**, 1183–1195 (2019).
42. Clark, D. R., Underwood, G. J. C., McGenity, T. J. & Dumbrell, A. J. What drives study-dependent differences in distance–decay relationships of microbial communities? *Glob. Ecol. Biogeogr.* **30**, 811–825 (2021).
43. Rahbek, C. et al. Building mountain biodiversity: geological and evolutionary processes. *Science* **365**, 1114–1119 (2019).
44. Antonelli, A. et al. Geological and climatic influences on mountain biodiversity. *Nat. Geosci.* **11**, 718–725 (2018).
45. Shapiro, B. J. et al. Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48–51 (2012).
46. Stegen, J. C. et al. Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069–2079 (2013).
47. Ning, D. et al. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4717 (2020).
48. Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A. & Daffonchio, D. Direct quantification of ecological drift at the population level in synthetic bacterial communities. *ISME J.* **15**, 55–66 (2021).
49. Larkin, A. A. & Martiny, A. C. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ. Microbiol. Rep.* **9**, 55–70 (2017).
50. Averill, C. et al. Defending Earth's terrestrial microbiome. *Nat. Microbiol.* **7**, 1717–1725 (2022).

Methods

Study sites and sample collection

Benthic biofilms were collected from 152 GFSs encompassing the European Alps, Scandinavian Mountains, Himalayas, Pamir and Tian Shan, Ecuadorian and Chilean Andes, Southwest Greenland, Alaska Range, Caucasus, Rwenzori Mountains in Africa and Southern Alps in New Zealand, between January 2019 and July 2022. Sampling was predominantly performed in spring or autumn during hydrological 'windows of opportunity' to avoid high-flow and scouring conditions typical for summer ice melt. GFSs from heavily debris-covered and rock glaciers were avoided, as were those with proglacial lakes, debris flows or tributaries in the reaches above the sampling site. At each GFS, we sampled three independent patches (that is, ecological replicates) within approximately 10 m for each of two stream reaches. The upstream reach was sampled as close as possible to the glacier snout (median distance to glacier snout, 78 m (IQR: 29.5–319 m)), whereas the downstream reach was sampled close to the terminal moraine of the Little Ice Age (if present and/or accessible; median distance to glacier snout, 773 m (IQR: 348–1,300 m)). From each patch, we sampled sandy sediments (0.25–3.15 mm size fraction) from the benthic layer (upper 5 cm of the streambed); all sampling devices were flame-sterilized in the field. Sediment samples were transferred into sterile cryovials, immediately flash-frozen in liquid nitrogen in the field and subsequently stored at −80 °C preceding and following shipping to Switzerland for DNA extraction and biomass analyses. As required, permits for sampling and sample export were obtained from the respective authorities.

Streamwater and sediment physicochemical characteristics

In the field, we measured streamwater temperature, pH, specific conductivity and dissolved oxygen concentration (MultiLine Multi 3630 IDS, WTW), as well as turbidity (Turb 430 IR, WTW), expressed as nephelometric turbidity units (NTU). We filtered streamwater (precombusted GF/F filters, Whatman) into Nalgene HDPE bottles and froze samples within 48 h pending nutrient analyses, using a LaChat QuikChem 8500 flow-injection analyser for ammonium ($N\text{-}NH_4^+$; QuikChem method, catalogue no. 10-107-06-3-D), nitrate ($N\text{-}NO_3^-$, QuikChem method, catalogue no. 10-107-05-1-C) and soluble reactive phosphorus (P-SRP, method no. 10-115-01-1-M). We combined inorganic forms of nitrogen into dissolved inorganic nitrogen (DIN). Filtered (precombusted GF/F filters, Whatman) dissolved organic carbon (DOC) samples were collected in acid-washed, precombusted glass vials (upstream reaches only) and stored in the dark at 4 °C pending analysis. DOC concentration was measured on a Sievers M9 TOC Analyser (GE). Sediment mineralogy was determined using an X-TRA ThermoARL Diffractometer. Errors varied between 5 and 10% for phyllosilicates, and 5% for grain minerals. Raw data files were generated and transformed using WinXRD 2.0-6 (ThermoFisher). Relative abundances of the main mineral groups were estimated from raw counts of mica, chlorite, amphibole, feldspars, calcite and quartz, divided by the sum of counts.

Glacier metrics

We calculated the distance of each stream reach to the glacier snout based on georeferencing (GPSMAPR 66 s, GARMIN) of the sampling sites, as well as glacier surface area and glacierized percentage catchment based on satellite imagery (Sentinel-2, level 2a, March 2019–July 2022 from scihub.copernicus.eu), and a catchment definition derived from the ASTER Global Digital Elevation Model v.3. (NASA/Meti/Aist/Japan Spacesystems and US/Japan Aster Science Team, 2019).

Microbial biomass

Benthic microbial biomass was determined as described previously (for example, ref. 15). Briefly, chlorophyll *a*, a proxy for algal biomass, was extracted from the sediment (90% EtOH) in a hot (78 °C) water bath for 10 min and further incubated at 4 °C for 24 h. Following vortexing and centrifugation, chlorophyll *a* concentration in the supernatant was quantified using a plate reader (BioTek Synergy H1, catalogue no. EX/EM: 436/680) and a spinach standard (Sigma-Aldrich). Concentrations were normalized to the dry mass of sediment. Bacterial abundance was determined using flow cytometry (NovoCyte, ACEA Biosciences) on cells stained with SybrGreen and detached from sediments using pyrophosphate and sonication[51].

DNA extraction, library preparation and sequencing

DNA was extracted from sediment using a phenol–chloroform protocol specifically adapted for GFS sediments[52]. To avoid PCR bias, we amplified the V3–V4 hypervariable region of the bacterial 16S ribosomal RNA gene using primers 341 forward (5′-CCTACGGGNGGCWGCAG-3′) and 785 reverse (5′-GACTACHVGGGTATCTAATCC-3′)[53]. We used KAPA HiFi DNA polymerase (Roche, Hot Start and Ready Mix formulation) in a 25 µl amplification reaction containing 1× PCR master mix, 1 µM each primer, 0.48 µg μl^{-1} bovine serum albumin and 1 µl of template DNA. Amplification was performed on a biometra Trio (Biometra) instrument. Thermal conditions applied, following initial denaturation at 95 °C for 3 min, were 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s for 25 cycles, with a final extension at 72 °C for 5 min. Amplification was subsequently verified on 1.5% agarose gel. Amplicon libraries were prepared according to the MiSeq manufacturer's protocol. In short, PCR was conducted for the addition of dual indices to purified amplicon PCR products. This allowed extensive multiplexing of samples on a single sequencing lane of the MiSeq (Illumina) platform, following quantification and normalization. Samples were sequenced using a 300-base-paired-end protocol at the Bioscience Core Lab of King Abdullah University of Science and Technology, Saudi Arabia. We also sequenced blanks and three types of mock community, which included community DNA (Zymo, catalogue no. 6305), cells (containing eight bacterial species in identical cell quantities; Zymo, catalogue no. 6300) and communities created with strains isolated from alpine GFSs (Extended Data Fig. 11 and Supplementary Tables 11 and 12). Cell- and strain-based mock communities were extracted using the same DNA extraction protocol.

Metabarcoding

A total of 883 amplicon sequence libraries were produced, comprising 868 benthic sediment samples, six blanks and nine mock communities. Paired-end sequencing generated a total of 158,774,383 reads, with an average read number of 163,697 ± 48,997 for sediment samples. Raw sequences were initially trimmed of primers using the plugin cutadapt[54]. Amplicon sequences were processed using Quantitative Insights into Microbial EcologyQ 2 (QIIME2, 2020.8)[55] workflow. The plugin demux was used to visualize interactive quality plots and assess read quality. We used three different denoising approaches for the identification of ASVs, namely UNOISE3 (ref. 35), Deblur[34] and DADA2 (ref. 33 and Supplementary Methods 2 and 3). Following inspection of mock communities, we further proceeded with Deblur to perform quality control (that is, denoising, dereplication and filtering chimeras) and identify ASVs. Sequences were trimmed at 410 base pairs (bp), and included a minimum quality score of 25. These steps resulted in a total of 40,801,795 reads and 65,228 ASVs (one blank was discarded through the pipeline procedure due to an insufficient number of reads). Taxonomy was assigned against the SILVA reference database[56] (v.138) using classify-sklearn from QIIME2. ASV and taxonomy tables, along with the metadata, were transferred to the R (v.4.1) environment for subsequent statistical analyses. Eukaryote, Archaea, mitochondria, chloroplast and blank-related sequences were discarded. We then applied a prevalence threshold to our dataset such that only ASVs present in at least two sediment patches from the same GFS were retained, independent of their prevalence in other GFSs. GFSs with fewer than three sediment patches were discarded ($n = 6$). Read counts were subsequently averaged across

replicates for a given GFS and multiplied by six before rarefaction ($n_{rarefaction}$ = 135,665). The final ASV table comprised 54,837 ASVs across 152 GFSs, including the sole GFS from the Rwenzori Mountains. A phylogenetic tree was constructed using VeryFastTree[57].

Shotgun metagenomics

DNA samples obtained from benthic sediments also underwent whole-genome shotgun sequencing, as described previously[58]. Briefly, libraries were prepared using the NEBNext Ultra II FS library kit, in which 50 ng of DNA was used. Library preparation included six PCR amplification cycles following a 12.5 min enzymatical fragmentation of input DNA. On average, an insert size of 450 bp was obtained, in which libraries were quantified by Qubit (Invitrogen), coupled with quality and size estimation using a bioanalyser (Agilent). Sequencing was performed at the Functional Genomics Centre Zurich on an S4 flow cell (150 bp paired-end; NovaSeq).

In regard to metagenomic sequence data, the Integrated Meta-omic Pipeline (IMP, v.3.0, catalogue no. 9672c874; available at https://git-r3lab.uni.lu/IMP/imp3)[59] was used to process paired forward and reverse reads from 97 GFS metagenomes. The IMP workflow for GFS analyses has previously been described[59]. Briefly, IMP uses MEGAHIT (v.1.2.9)[60] for the assembly, following an adaptor- and primer-trimming step with cutadapt[53]. Subsequently, the assemblies were used for functional gene calling through Prodigal[61], yielding 24,946,385 non-redundant gene clusters following clustering with mmseqs2 with the following parameters: --cov-mode 0 c 0.8 --min-seq-id 0.3 (ref. 62). FeatureCounts was then used to estimate gene abundance, with KEGG annotations from gene annotations generated using Mantis[63], mapping to 17,536 KOs. We filtered KEGG data so that only those 8,518 KOs associated with prokaryotes were retained, as determined by KofamScan profiles[64]. For comparison with other cryospheric systems, we obtained 92 metagenomes from ref. 21 and processed these with the IMP workflow using the same parameters.

Data analysis

Gamma and alpha diversity. Rank abundance curves were computed for each mountain range. We explored patterns of gamma diversity by computing both observed and estimated (that is, asymptotic estimator) ASV richness using iNEXT[65]. For alpha diversity, we used the Hill numbers approach because it provides an effective comparison of diversity indices based on the 'equivalent number' of features[66]. We therefore present ASV observed richness (q = 0) and effective number of ASVs based on the exponential of Shannon entropy, calculated using the function hill_taxa within the package hillR[67] (v.0.5.2). For comparison of alpha diversity from GFSs with other cryospheric ecosystems, we used a published 16S rRNA gene amplicon sequence dataset of 178 samples spanning various cryospheric ecosystems (for example, snow, cryoconite, permafrost soil)[21]. Preprocessing of cryospheric-related data is further detailed in Supplementary Methods 3. To assess the effects of elevation and glacier coverage of the catchment on ASV richness, we computed generalized additive models and linear models using the package mgcv[68] (v.1.8-42). Assumptions of normality and homoscedasticity of variance were tested on data residuals using Shapiro–Wilk and Levene tests, with the function check_model in the package performance[69] (v.0.9.2.2).

Beta diversity and indicator taxa. For assessment of compositional differences between (1) the GFS microbiome and other cryospheric ecosystems (16S rRNA gene dataset of 178 cryospheric samples from ref. 21) and (2) mountain ranges and hemispheres for the GFS microbiome dataset, we computed non-metric multidimensional scaling (NMDS) analyses based on the Bray–Curtis index using metaMDS in the package vegan[70] (v.2.6-2). We used generalized additive models to visualize latitudinal variation in NMDS space using the function ordisurf in vegan. We tested for the effects of ecosystem type for (1) GFS mountain ranges and GFS hemispheres and (2) community composition using permutational analysis of variance (PERMANOVA) based on Bray–Curtis dissimilarity, with the function adonis2 in vegan. Pairwise differences were assessed using the package pairwiseAdonis[71] (v.0.4), and P values were adjusted for multiple comparisons following the Holm method. To account for issues pertaining to PERMANOVA—in which differences in microbiome dispersion could be confounded by composition effects[72]—we computed multivariate generalized linear models using the function manyglm in the package mvabund[73] (v.4.2.1), and further computed ANOVA with the function anova.manyglm for significance testing. In addition, we partitioned the contribution of ASVs to beta diversity[74] measured as total community variance, along with the contribution of GFSs in different mountain ranges to global beta diversity, using the function beta.div in the package adespatial[75] (v.0.3-21). Indicator ASVs significantly contributing to compositional differences among mountain ranges were calculated based on Bray–Curtis dissimilarity, 999 permutations and Benjamini–Hochberg correction[76,77], using the function dbMANOVAspecies implemented in the package adiv[78] (v.2.2).

Specific, unique and core ASVs. Amplicon sequence variants exclusively detected in a single mountain range (that is, accounting for multiple GFSs) were categorized as specific, and those identified in only one GFS as unique. The occurrences of unique ASVs among replicated sediment patches were assessed from the dataset before averaging and rarefaction. We tabulated the number of specific and unique ASVs in each mountain range, their contribution to total relative abundance, the number of GFSs per mountain range where specific ASVs occur and finally investigated the taxonomic composition of these specific ASVs. The taxonomic core microbiome was computed at the ASV level. ASVs present in at least one GFS across more than six out of ten mountain ranges (that is, omitting the sole GFS from the Rwenzori Mountains), and detected at a relative abundance equal to or greater than 0.1%, were categorized as core ASVs. To illustrate the relative abundance of the most abundant phyla and families across core, indicator and specific ASVs, we computed ridgeline plots using ggridges[79] (v.0.5.4) and agglomerated ASVs at the genus level using tax_glom in the phyloseq[80] package (v.1.41.1).

Distance-decay patterns. Distance-decay patterns were generated on both taxonomic and functional tables based on either geographic distances and dissimilarity indices (Bray–Curtis, Sørensen) or distance metrics (that is, weighted and unweighted UniFrac distances). Distances between GFSs were calculated using distm in the package geosphere[81] (v.1.5-18), including the parameter distGeo, which estimates the shortest distance between points on an ellipsoid. We computed Mantel tests to examine the statistical significance of distance-decay patterns. Moreover, we compared regression slopes using linear models or permutational analysis of covariance when assumptions were not met.

Variation partitioning. We partitioned the variance of the GFS microbiome (excluding the Rwenzori Mountains GFS) beta diversity[82–85] with db-RDA using the Bray–Curtis dissimilarity index. This analysis quantifies the relative importance of the spatial structure (that is, within and among mountain ranges), the GFS environment (that is, physicochemical parameters and glaciology) and climatic variables in explaining the variance in beta diversity. db-RDA analyses were executed to investigate the linear effects of latitude, longitude and elevation on microbiome composition.

We computed variation partitioning by combining a three-level, spatial-scale model with an environmental model. The former was built as follows: (1) a component accounting for the linear effects of latitude, longitude and elevation of the sample; (2) a component accounting for mountain range identity using dummy variables; and (3) a spatial component encompassing Moran's eigenvector maps (MEMs) to elucidate the spatial relationships among GFSs within a

particular mountain range. The type of MEM variables used are principal coordinates of neighbour matrices (PCNM)[85,86]. MEM analysis generates a set of orthogonal variables based on the geographic coordinates of GFSs[85]; these variables serving as explanatory variables in the db-RDA to model spatial structure in community data. Specifically, MEM variables from the spatial matrix were arranged in blocks, with each block allocated to a specific mountain range. Pools from the other mountain ranges were assigned a value of 0. The environmental model was computed using a forward selection procedure applied to a predetermined set of environmental variables[86].

Before variation partitioning, we applied natural log transformation to the community data matrix[87]. We included streamwater temperature, pH, specific conductivity, turbidity, DOC, chlorophyll *a*, DIN, SRP, benthic sediment mineral composition (that is, calcite, quartz, feldspar and clays) and glacier metrics (glacier surface area and percentage glacier cover of the catchment). In addition, we included monthly precipitation and snow cover days, metrics retrieved from the Chelsa climate database (https://chelsa-climate.org). We used the functions dbrda and varpart of the vegan package and forward.sel of the package packfor[88]. Variance inflation factors were used to detect multicollinearity amongst predictors, which were log transformed before analyses. Lastly, PCNM from the PCNM package was used to compute MEM spatial eigenfunctions[89].

Assembly processes and phylogenetic profiling. Phylogenetic turnover analyses can unravel community assembly processes (that is, ecological drift, selection and dispersal). To this end, we used the statistical framework iCAMP[47] to identify dominant assembly processes within each mountain region. iCAMP leverages null-model analysis of the beta-net-relatedness index, a metric that quantifies bin-level phylogenetic distances, which emphasizes deep-level phylogenetic turnover. Taxonomic turnover, assessed using a modified version of the Raup–Crick metric, is further used to distinguish between dispersal limitation, homogenizing dispersal and ecological drift. iCAMP uses phylogenetic tree-based binning, thereby facilitating the resolution of different assembly processes at a finer taxonomic resolution. Bin-based dominant community assembly processes—namely, dispersal limitation, HoS and ecological drift (Drift), along with the taxonomies of major clades— were visually represented on the phylogenetic tree using ggtreeExtra[90].

The assessment of autocorrelation of phylogenetic distances among ASVs with a similar contribution to beta diversity (that is, phylogenetic signal in species contribution to beta diversity) was carried out using the packages adespatial and phylosignal[91]. To unravel the phylogenetic depth at which different facets of beta diversity manifest, we consecutively agglomerated phylogenetic tips using the tree_glom function in the package speedyseq (v.0.5.3.9018). As a reference metric we used mean phylogenetic distance between ASVs, specifically mean nearest taxon distance (0.09), and evaluated the rate of change in beta diversity. This was quantified as Bray–Curtis dissimilarity (that is, the quantitative form of Sørensen dissimilarity, using the beta.div.comp function in the package adespatial. Beta diversity exhibited a change of less 1% at phylogenetic depths greater than 0.2. In light of these findings, and corroborated by previous work (ref. 46 and references therein), we screened the phylogeny from 0 to 0.2 in steps of 0.005 (in arbitrary units). For each incremental step, we calculated both global beta diversity and the number of unique taxa—here referring to unique phylogenetic tips present in a single GFS.

KEGG functional analysis. The KEGG counts per sample from GFSs were merged in R (v.4.1.0) and log transformed using the log1p function. The abundance-coverage estimator, and Shannon and Chao1 indices (Supplementary Table 5), were calculated using the trans_alpha function encoded in the package microeco (v.1.0.0). We used the MaAsLin2 (ref. 92) (v.1.7.3) R package to determine KOs significantly different between GFS samples and other cryospheric samples. The data were normalized with the total sum scaling method to account for differences in sequencing depth between samples. To further control for potential confounding variables, we added covariate data to the model to identify differences in sample type (GFS versus non-GFS). Beta diversity was calculated with trans_beta, generating a principal coordinate analysis (PCoA) figure using the built-in plot function. Subsequently, KEGG API (https://www.kegg.jp/kegg/rest/keggapi.html) from KEGGREST[93] (v.3.17) was used to retrieve pathway annotations for each KO number. The most abundant pathways were included in PCoA using geom_text from the package ggplot2 (ref. 94) (v.3.4.4). KO abundances were summarized based on KEGG pathways for both GFS and other cryospheric ecosystems, and a heatmap was generated using the package pheatmap[95] (v.1.0.12). Classification of KOs into generalist and specialist functional potential, using Levin's niche breadth index (derived from KO prevalence and abundance)[96], was performed with spec.gen within the package EcolUtils (v.0.1)[97]. Finally, investigation of the contribution of generalists and specialists to functional dissimilarity was conducted using the function simper within the package vegan. All statistical analyses were computed in R (v.4.1.0)[98]. The maps were generated through RStudio using the package rnaturalearth (v.1.0.1)[99,100].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequence data have been deposited in the Sequence Read Archives with NCBI BioProject accession no. PRJNA781406. Processed data, along with data tables associated with Figs. 1–5 and Extended Data Figs. 1–11, are available at Zenodo (https://doi.org/10.5281/zenodo.13897903)[101].

Code availability

R analysis scripts are provided at GitHub (github.com/laylaeb/GFS_biogeo2023).

51. Kohler, T. J. et al. Patterns and drivers of extracellular enzyme activity in New Zealand glacier-fed streams. *Front. Microbiol.* **11**, 591465 (2020).
52. Busi, S. B. et al. Optimised biomolecular extraction for metagenomic analysis of microbial biofilms from high-mountain streams. *PeerJ* **8**, e9973 (2020).
53. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
54. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
55. Bolyen, E. et al. QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *Nat. Biotechnol.* **37**, 852–857 (2019).
56. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
57. Piñeiro, C., Abuín, J. M. & Pichel, J. C. Very Fast Tree: speeding up the estimation of phylogenies for large alignments through parallelization and vectorization strategies. *Bioinformatics* **36**, 4658–4659 (2020).
58. Busi, S. B. et al. Genomic and metabolic adaptations of biofilms to ecological windows of opportunities in glacier-fed streams. *Nat. Commun.* **13**, 2168 (2022).
59. Narayanasamy, S. et al. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol.* **17**, 260 (2016).
60. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
61. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
62. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
63. Queirós, P., Delogu, F., Hickl, O., May, P. & Wilmes, P. Mantis: flexible and consensus-driven genome annotation. *Gigascience* **10**, giab042 (2021).
64. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
65. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).
66. Jost, L. Entropy and diversity. *Oikos* **113**, 363–375 (2006).
67. Li, D. hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *J. Open Source Softw.* **3**, 1041 (2018).

68. Wood, S. mgcv: mixed GAM computation vehicle with automatic smoothness estimation. R package v.1.8-42 (CRAN, 2023).
69. Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P. & Makowski, D. performance: An R package for assessment, comparison and testing of statistical models. *J. Open Source Softw.* **6**, 3139 (2021).
70. Oksanen, J., Kindt, R. & O'Hara, B. Vegan: R functions for vegetation ecologists. *Date of access* **15**, 2014 (2005).
71. Martinez Arbizu, P. pairwiseAdonis: pairwise multilevel comparison using adonis. R package v.0.4 (GitHub, 2017).
72. Warton, D. I., Wright, S. T. & Wang, Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* **3**, 89–101 (2012).
73. Wang, Y. I., Naumann, U., Wright, S. T. & Warton, D. I. mvabund—An R package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.* **3**, 471–474 (2012).
74. Legendre, P. & De Cáceres, M. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecol. Lett.* **16**, 951–963 (2013).
75. Dray, S. et al. adespatial: multivariate multiscale spatial analysis. R package v.0.3-21 (CRAN, 2023).
76. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
77. Ricotta, C., Pavoine, S., Cerabolini, B. E. L. & Pillar, V. D. A new method for indicator species analysis in the framework of multivariate analysis of variance. *J. Veg. Sci.* **32**, e13013 (2021).
78. Pavoine, S. adiv: An R package to analyse biodiversity in ecology. *Methods Ecol. Evol.* **11**, 1106–1112 (2020).
79. Wilke, C. O. ggridges: ridgeline plots in 'ggplot2'. R package v.0.5.4 (CRAN, 2021).
80. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
81. Hijmans, R. J., Williams, E., Vennes, C. & Hijmans, M. R. J. Package 'geosphere'. *Spherical trigonometry* **1**, 1–45 (2017).
82. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).
83. Legendre, P., Borcard, D. & Peres-Neto, P. R. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecol. Monogr.* **75**, 435–450 (2005).
84. Peres-Neto, P. R., Legendre, P., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**, 2614–2625 (2006).
85. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Model.* **196**, 483–493 (2006).
86. Blanchet, F. G., Legendre, P. & Borcard, D. Forward selection of explanatory variables. *Ecology* **89**, 2623–2632 (2008).
87. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280 (2001).
88. Dray, S., Legendre, P. & Blanchet, F. G. packfor: forward selection with permutation (Canoco p. 46). R package v.0.0-8 (R-Forge, 2007).
89. Legendre, P. et al. Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology* **90**, 663–674 (2009).
90. Xu, S. et al. ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol. Biol. Evol.* **38**, 4039–4042 (2021).
91. Keck, F., Rimet, F., Bouchez, A. & Franc, A. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol. Evol.* **6**, 2774–2780 (2016).
92. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* **17**, e1009442 (2021).
93. Tenenbaum, D. et al. KEGGREST: client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). R package v.1.32.0 (Bioconductor, 2021).
94. Wickham, H. et al. ggplot2: create elegant data visualisations using the grammar of graphics. R package v.3.5.0 (CRAN, 2024).
95. Kolde, R. pheatmap: pretty heatmaps. R package v.1.0.12 (CRAN, 2019).
96. Yan, Q. et al. Distinct strategies of the habitat generalists and specialists in sediment of Tibetan lakes. *Environ. Microbiol.* **24**, 4153–4166 (2022).
97. Salazar, G. EcolUtils: utilities for community ecology analysis. R package v.3 (2020).
98. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021).
99. RStudio Team. *RStudio: integrated development environment for R* (RStudio, 2021).
100. Massicotte, P. & South, A. rnaturalearth: world map data from Natural Earth. R package v.1.0.1 (CRAN, 2017).
101. Ezzat, L. et al. Diversity and biogeography of the glacier-fed stream bacterial microbiome. *Zenodo* https://doi.org/10.5281/zenodo.13897903 (2024).

Acknowledgements The Vanishing Glaciers project is supported by The NOMIS Foundation, to T.J.B. We thank A. McIntosh and L. Morris in New Zealand, J. Abermann and T. Juul-Pedersen in Greenland, O. Solomina and T. Kuderina Maratovna in Russia, V. Crespo-Pérez and P. Andino Guarderas in Ecuador, J. Yde and S. Leth Jørgensen in Norway, S. Sharma and P. Joshi in Nepal, N. Shaidyldaeva-Myktybekovna and R. Kenzhebaev in Kyrgyzstan, J. Nattabi Kigongo, R. Nalwanga and C. Masembe in Uganda, M. Gonzaléz and J. Luis Rodriguez in Chile and C. Kuhle and P. Tomco in Alaska for various logistical support; see https://www.glacierstreams.ch for all institutions involved in the logistics of the expeditions. We particularly acknowledge help from the porters and guides in Nepal, Uganda and Kyrgyzstan. We also acknowledge E. Oppliger for laboratory support and the Bioscience Core Laboratory at King Abdullah University and Technology (KAUST) for DNA sequencing. T.J.K. was supported by Charles University project no. PRIMUS/22/SCI/001. D.D. acknowledges financial support from KAUST through the baseline research fund. S.B.B. was supported by Swiss National Science Foundation grant no. CRSII5_180241 to T.J.B., and by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 863664) to P.W.

Author contributions L.E. carried out conceptualization, methodology, data curation, formal analysis, visualization and writing of the original draft. H.P. performed conceptualization, methodology, data curation, investigation, formal analysis, visualization and writing of the original draft. M.B. was responsible for conceptualization, methodology, data curation, formal analysis, visualization and writing review and editing. S.B.B. carried out conceptualization, methodology, data curation, formal analysis, visualization and writing review and editing. G.M. performed conceptualization, methodology, formal analysis, visualization and writing review and editing. S.F. was responsible for conceptualization, methodology, formal analysis and investigation. T.J.K. carried out conceptualization, investigation, data curation, methodology and writing review and editing. T.L. undertook methodology, formal analysis and writing review and editing. A.G. carried out formal analysis. P.P. carried out investigation and data curation. F.B. was responsible for investigation and data curation. D.D. was responsible for resources and writing review and editing. N.D. performed formal analysis and data curation. P.W. carried out methodology. V.D.S. was responsible for investigation and data curation. R.M. performed methodology, investigation and writing review and editing. M. Schön carried out investigation and data curation. M. Styllas undertook investigation, formal analysis and data curation. M.T. performed investigation and data curation. T.J.B. carried out conceptualization, methodology, investigation, writing of the original draft, supervision, project administration and funding acquisition.

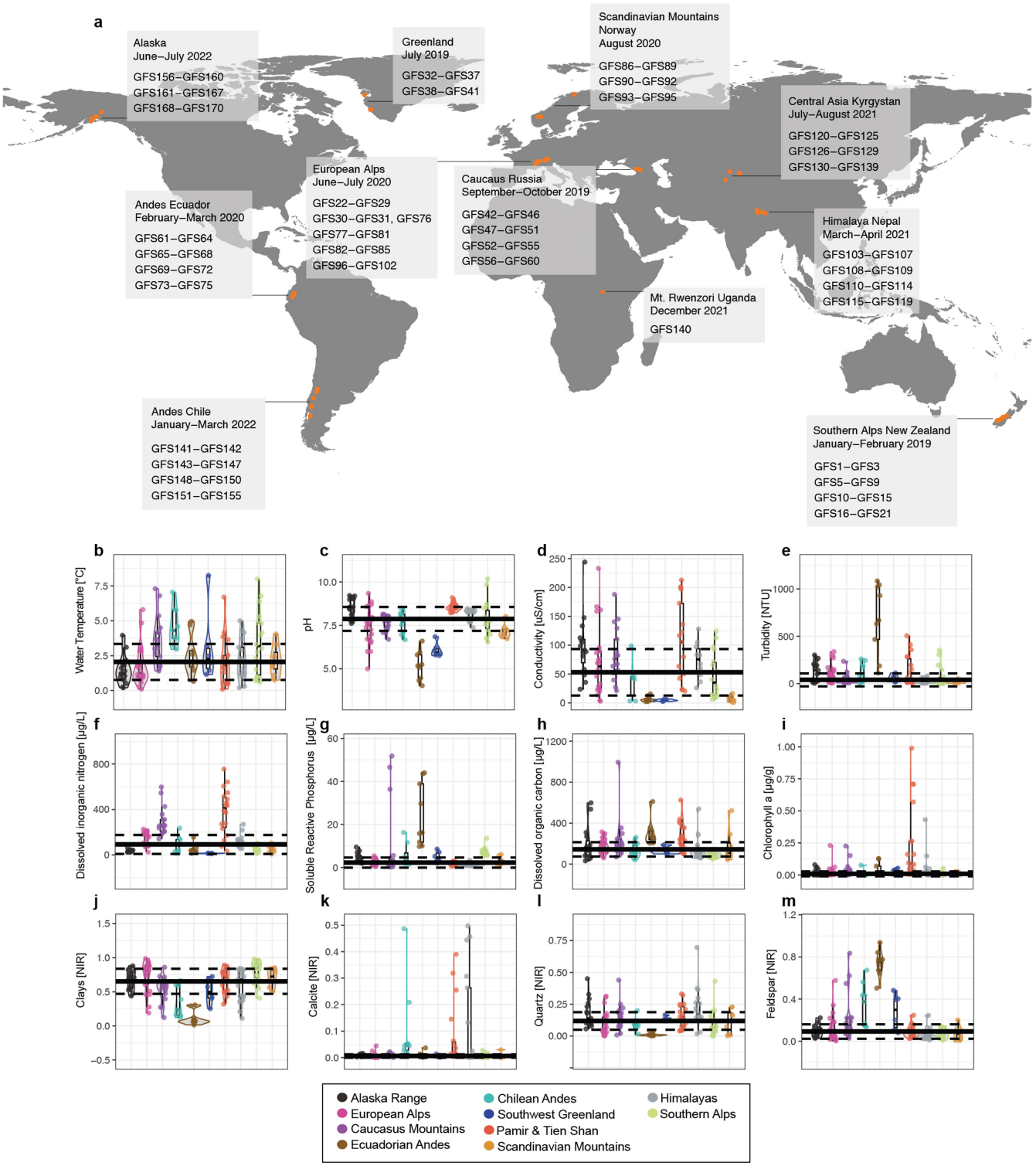
Competing interests The authors declare no competing interests.

Additional information

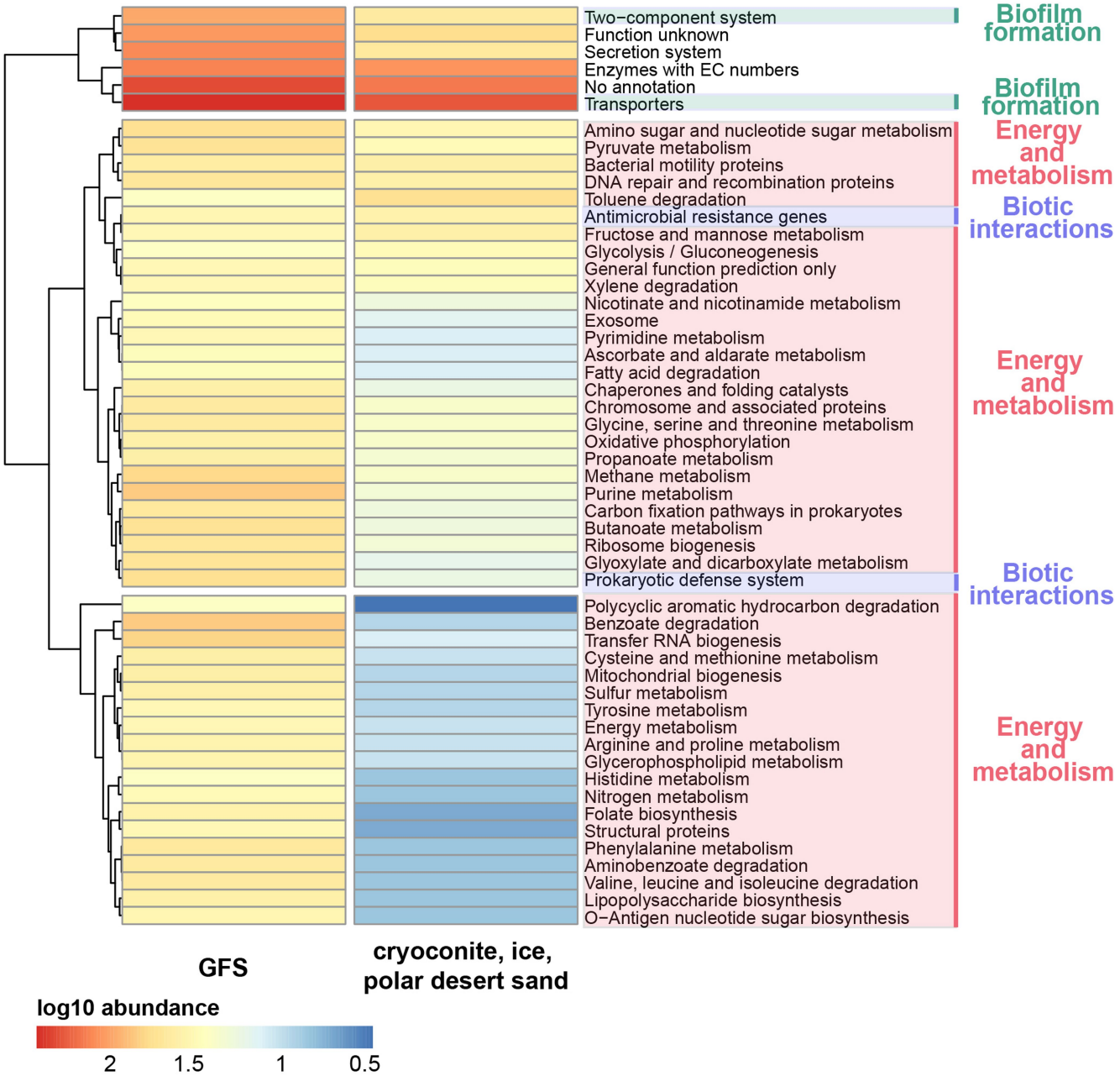
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08313-z.

Correspondence and requests for materials should be addressed to Leïla Ezzat or Tom J. Battin.

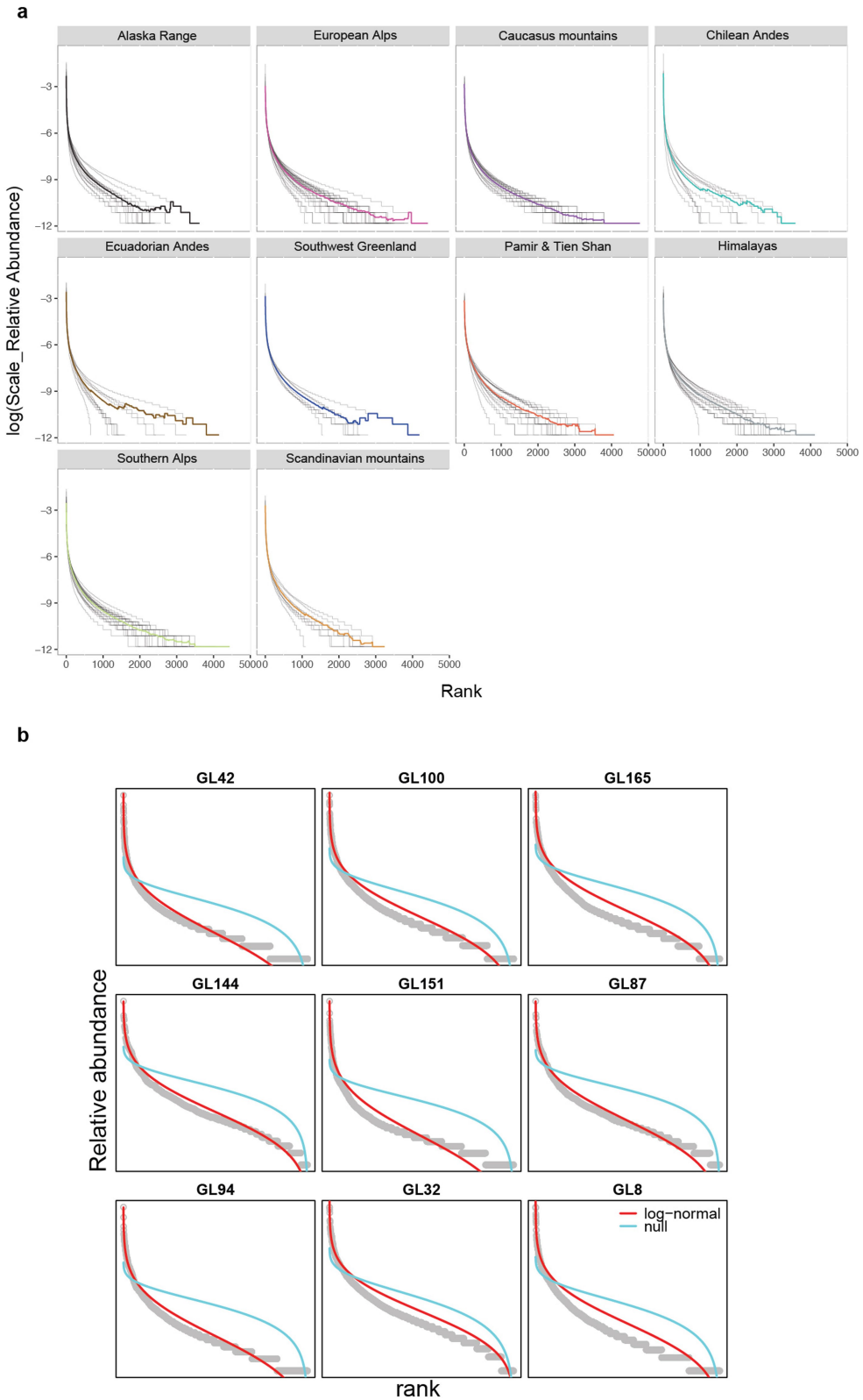
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.



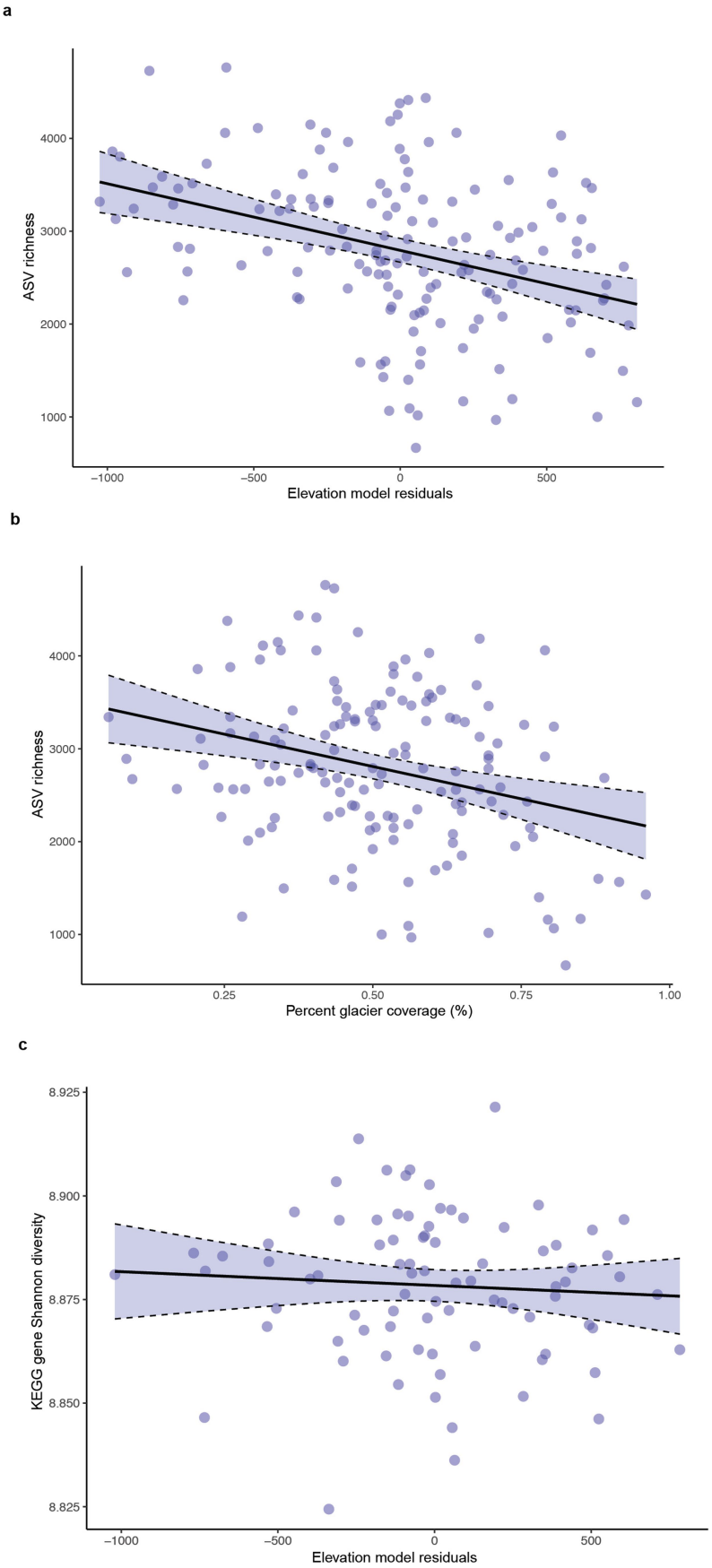
Extended Data Fig. 1 | World map of the glacier-fed streams (GFSs) sampled by the *Vanishing Glaciers* Project and environmental data associated with the study. (**a**) In total, the *Vanishing Glaciers* Project studied 170 GFSs from 11 of the largest mountain ranges worldwide between 2019 and 2022. Per mountain range, benthic sediment samples were collected from GFSs in up to three different regions (orange dots) to account for the variability within a given mountain range. For the present study, we retained 152 GFSs for microbiome analyses. Violin plots of environmental data. (**b**) Water temperature [°C] with n = 139 biologically independent samples, (**c**) pH (n = 139), (**d**) Conductivity [uS/cm] (n = 128), (**e**) Turbidity [NTU] (n = 138), (**f**) Dissolved inorganic nitrogen - DIN [ug/L] (n = 139), (**g**) Soluble reactive phosphorus - SRP [ug/L](n = 137), (**h**) Dissolved organic carbon - DOC [ug/L] (n = 139), (**i**) Chlorophyll a [ug/g] (n = 139), (**j**) Clays [NIR] (n = 139), (**k**) Calcite [NIR] (n = 139), (**l**) Quartz [NIR] (n = 139), (**m**) Feldspar [NIR] (n = 139) across the different mountain ranges. Each dot represents a GFS. The horizontal lines within the box plots represent the median, box height is the interquartile range, and whiskers represent the data range. The width of the violin plot represents the frequency of values in the data. Solid and dashed lines across each graph are median and quartile values computed across the different mountain ranges.



Extended Data Fig. 2 | Functional pathways across cryospheric ecosystems. Heatmap of illustrating functional pathways differentially associated (as log10 abundance) with metagenomes collected from GFSs and other cryospheric ecosystems. Shown are abundances of 53 pathways generated by collapsing the 8,518 KOs into their respective functional ortholog categories.

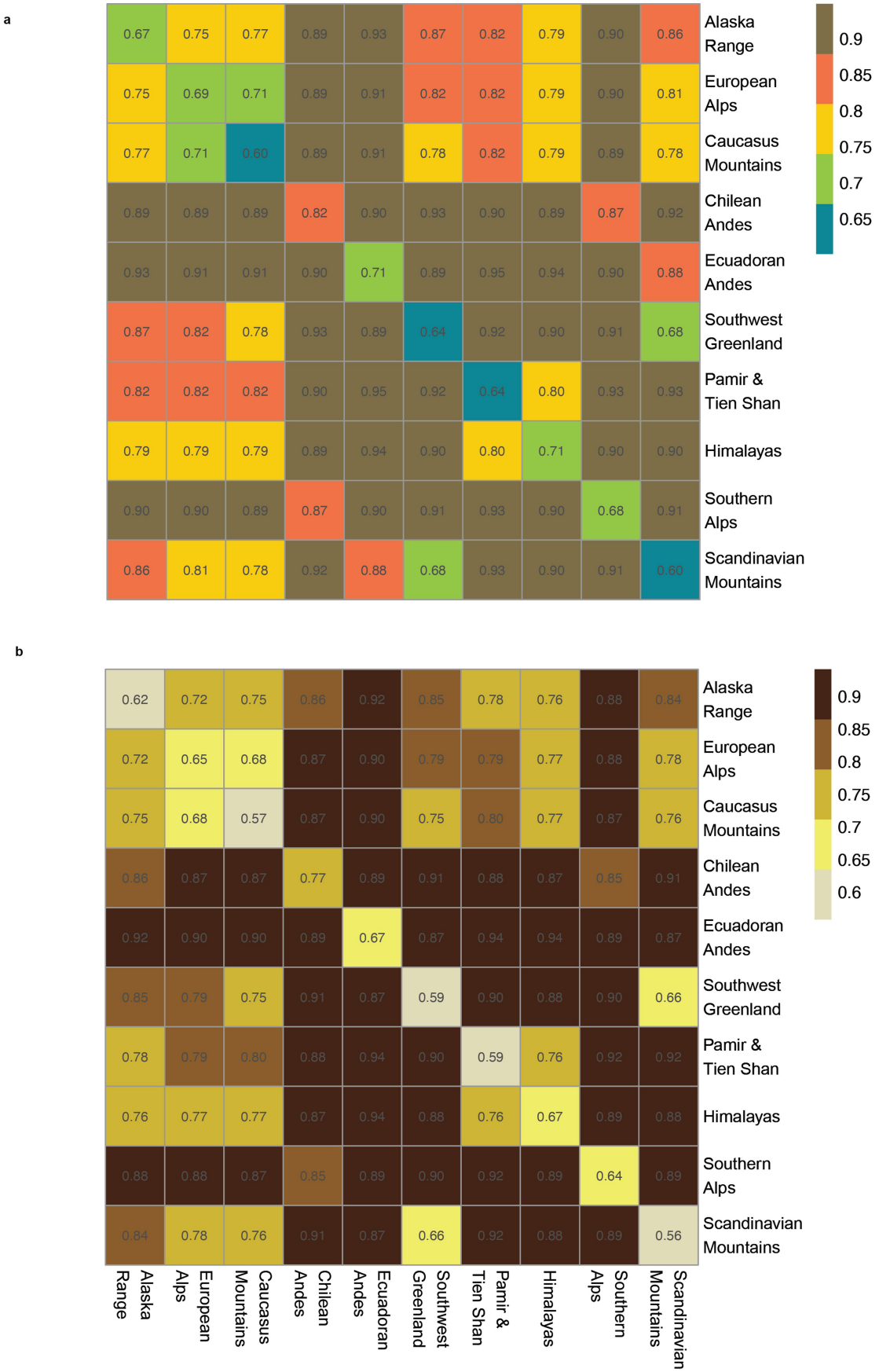


Extended Data Fig. 3 | Rank-abundance model fits. (**a**) Rank abundance curves (RAC) for the different mountain ranges. Bold lines represent the mean RAC for a given mountain range (n = 151 GFSs). (**b**) Rank-abundance distributions with log-normal model (red) and Brokenstick null model (cyan) fits for a subset of nine GFS for the filtered and rarefied dataset. The entire set of distributions for both filtered-rarefied and unfiltered datasets is found on Zenodo repository (see Data Availability section). For most GFS communities, the log-normal model fits the empirical abundance distributions (grey) well. The log-normal model (i.e., frequency distributions which approximate normality upon log-transformation) is expected to describe large communities because of the central limit theorem and is thought to emerge from interactions among stochastic processes, such as growth and dispersal. The log-normal model fit further suggests that communities have relatively few rare ASVs (compared for instance to the log-series model). While our stringent filtering criteria led to the preferential removal of rare taxa, they did not qualitatively influence rank-abundance model fits.

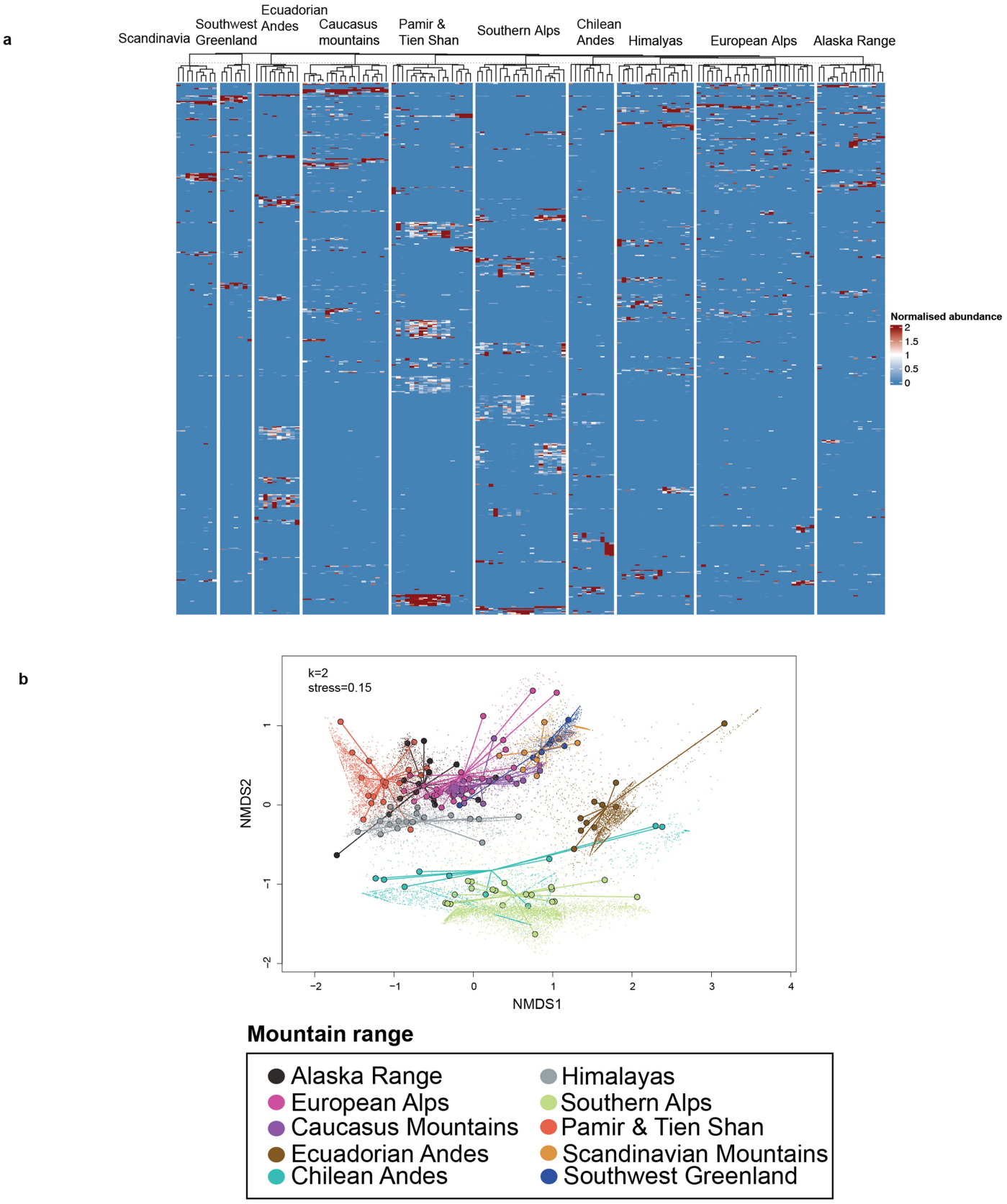


Extended Data Fig. 4 | See next page for caption.

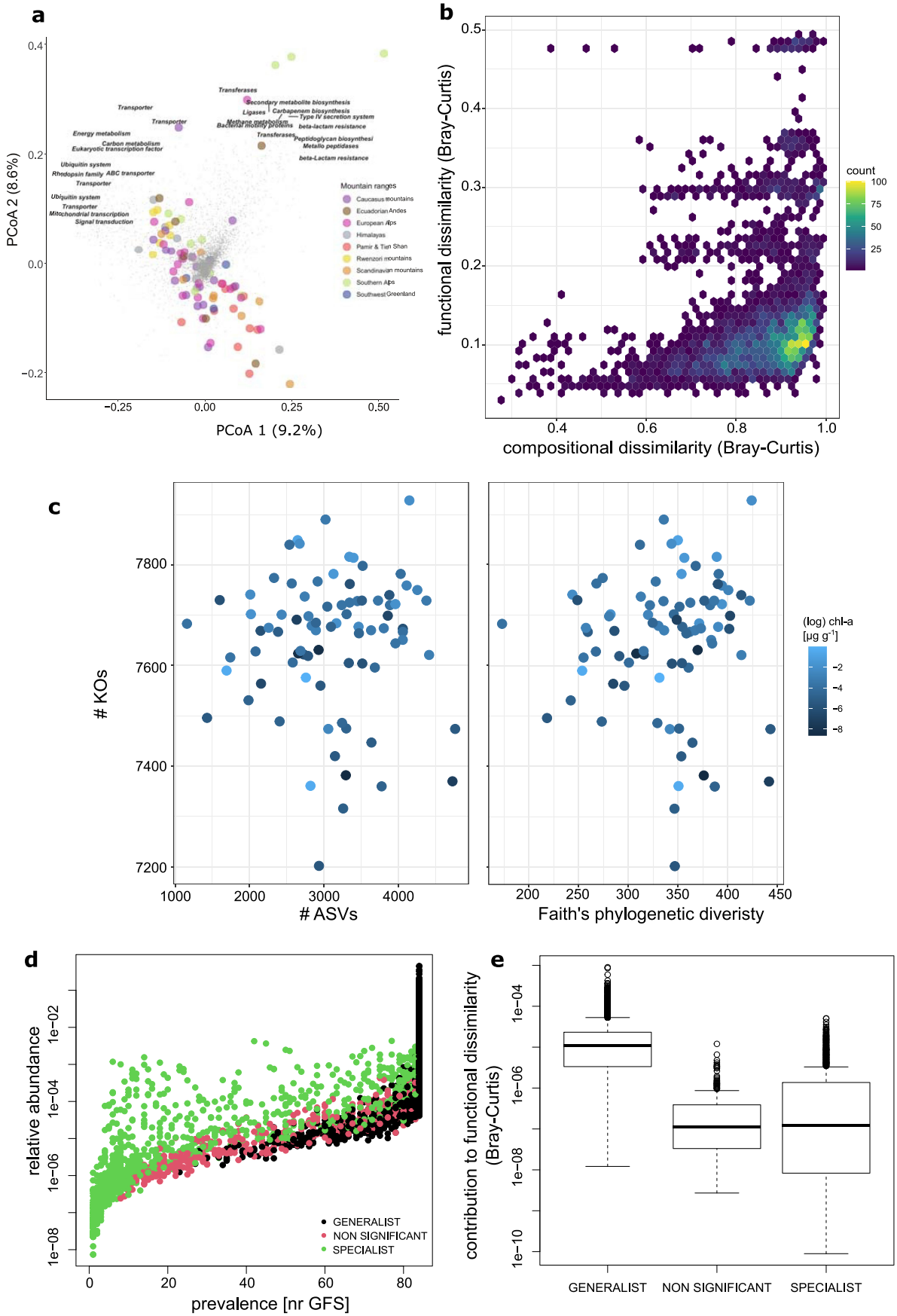
Extended Data Fig. 4 | Microbiome structure and function in relation to altitude and glacier coverage. Linear relationships between ASV richness and (**a**) elevation (taking into account the latitudinal effect), and (**b**) percent glacier coverage across all GFSs (n = 151; excluding the Rwenzori Mountains). Generalised additive models (GAMs) were built regressing elevation against a spline of latitude (bs=tp). The residuals of this regression were then fitted against ASV richness in a second model. Results from the latter showed that species richness decreased significantly with increasing elevation, when accounting for latitude (GAM elev_lat, spline F = 179.6, p < 2e-16, Deviance explained=91.3%, adj R^2 = 0.908; LM richness_elevation, F = 19.96, p = 0.000016, adj R^2 = 0.118), and glacier coverage (LM richness_glaciercov, F = 13.66, p = 0.0003, adj R^2 = 0.078). (**c**) The elevation model residuals were also fitted against the KEGG gene Shannon richness, displaying the more uniform functional diversity. (LM KEGG Shannon vs elevation residuals, F = 0.3898, p = 0.53341, adj R^2 = −0.0074). Confidence intervals (95% as dashed lines) are shown around the regression line, indicating the standard errors of the predicted value.



Extended Data Fig. 5 | Dissimilarities of the glacier-fed stream bacterial communities within and among mountain ranges. Average values of (**a**) Bray-Curtis and (**b**) Sørensen dissimilarity indices computed within and among mountain ranges (n = 151 GFSs).

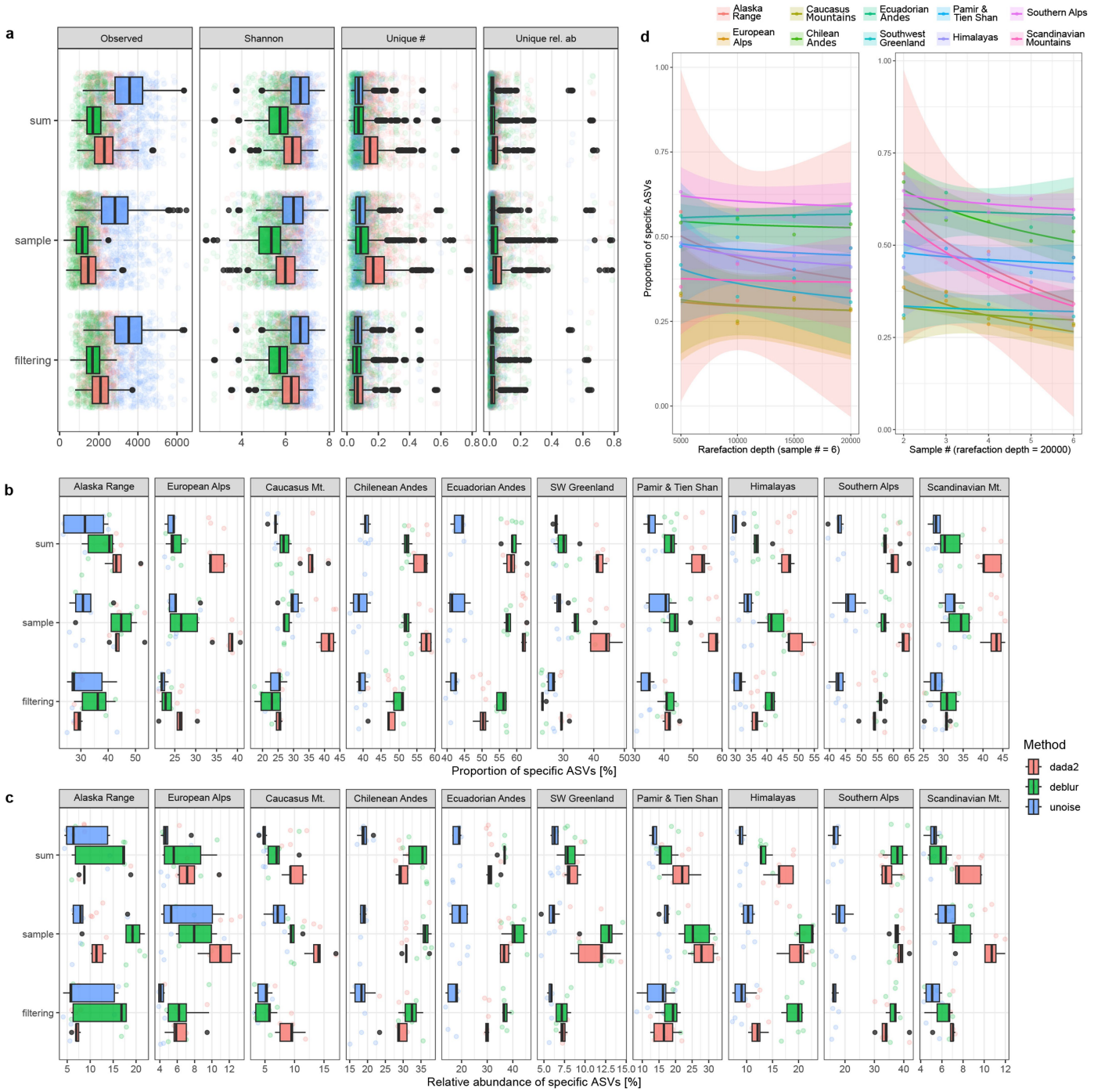


Extended Data Fig. 6 | Indicator amplicon sequence variants (ASVs) that significantly contribute to beta-diversity of the GFS benthic microbiome. (**a**) Heatmap illustrating the relative abundance of the 500 most abundant indicator ASVs across mountain ranges (n = 151 GFSs). (**b**) NMDS based on Bray-Curtis dissimilarity illustrating biogeographic patterns (large circles reflect individual GFS, lines connecting GFSs to the group centroid highlight the different regions). Species scores of the 500 indicator ASVs, which contribute to dissimilarities among mountain ranges, are shown (dots, coloured by region).

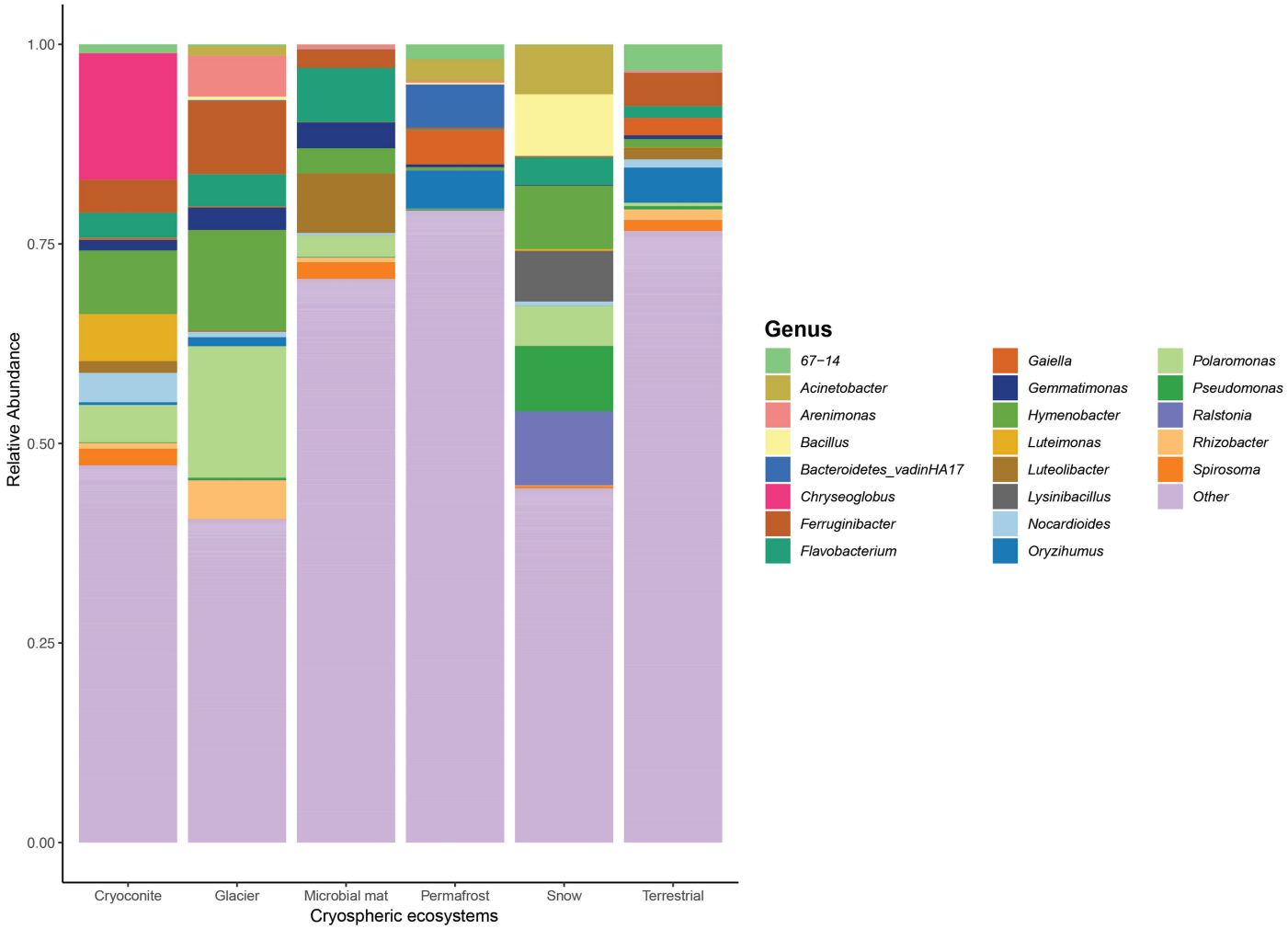


Extended Data Fig. 7 | See next page for caption.

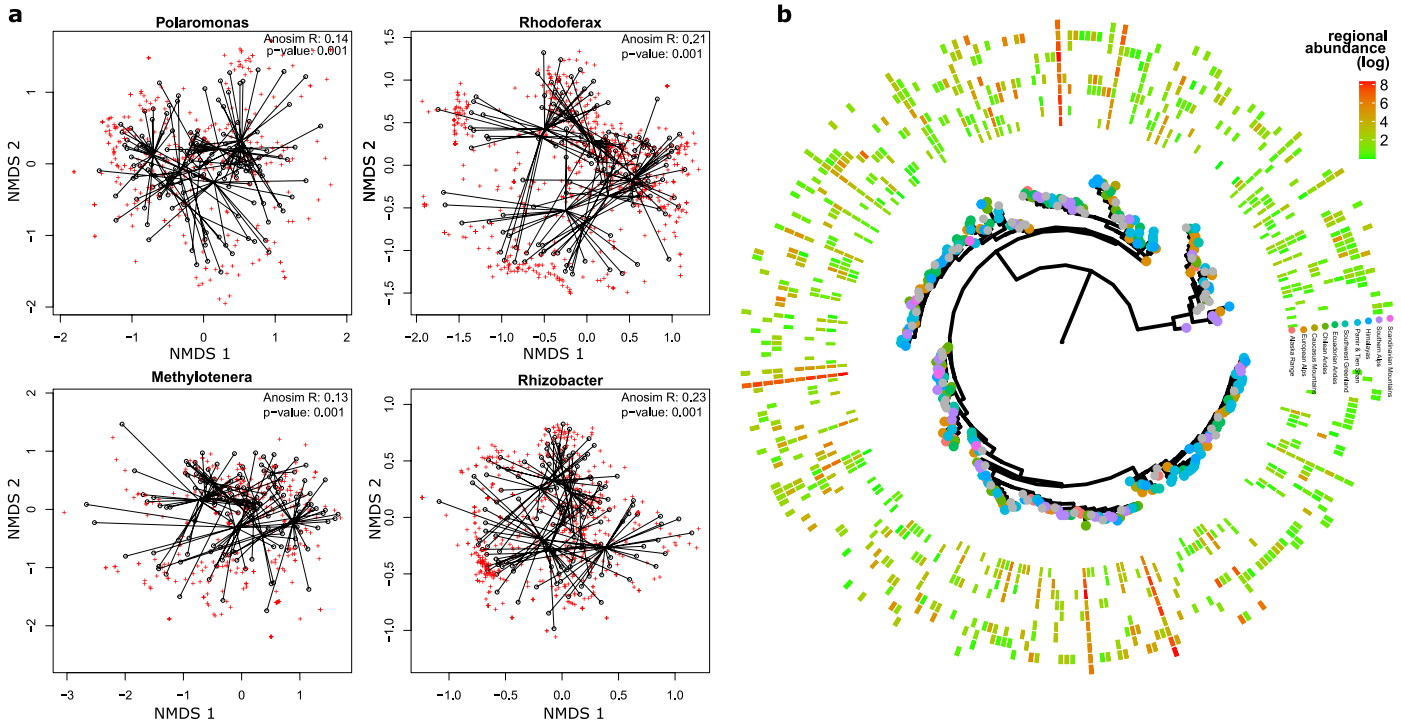
Extended Data Fig. 7 | Functional diversity observed in the GFSs. (**a**) Principal coordinates analysis (PCoA) based on KEGG orthology (KO) abundances representing GFSs' functional diversity. Filled circles represent individual samples across the mountain ranges. Grey dots indicate the underlying KOs contributing to the dissimilarity, calculated using Bray-Curtis distance. Pathways based on grouped KOs, contributing to the variation within the samples are indicated as text. (**b**) Functional dissimilarity across GFSs (as Bray-Curtis) was consistently low (average $BC_{dis} = 0.14$), contrasting the large taxonomic dissimilarity (average $BC_{dis} = 0.86$) and pointing to the replacement of functionally redundant taxa across samples. Shown is the density distribution of taxonomic and functional dissimilarity across all sample pairs. (**c**) Despite pronounced differences in GFS microbiome alpha diversity (in terms of number of ASVs and Faith's phylogenetic diversity), these communities encoded similar numbers of KOs (between 7202 and 7928 KOs). (**d**) Classification of KOs into generalist and specialist functional potential using Levin's niche breadth index revealed a large number of widely-distributed generalist KOs (n = 7133) and comparably few (n = 991) specialist KOs. Another 373 KOs could not be classified unambiguously using Levin's index and 999 permutations (i.e. non-significant). Notably, (**e**) similarity percentage (SIMPER) analysis, revealed that generalist KOscontribute most to the observed functional dissimilarity. The boxplot shows median (horizontal line), interquartile ranges (IQR, boxes) and outliers (>1.5 x IQR, circles) for generalist (n = 7133), specialist (n = 991) and non-significant (n = 377) KOs. Taken together, the presence of most KOs across GFS globally and that wide-spread KOs contribute to the (low) functional dissimilarity points towards substantial functional redundancy in GFS microbiomes.



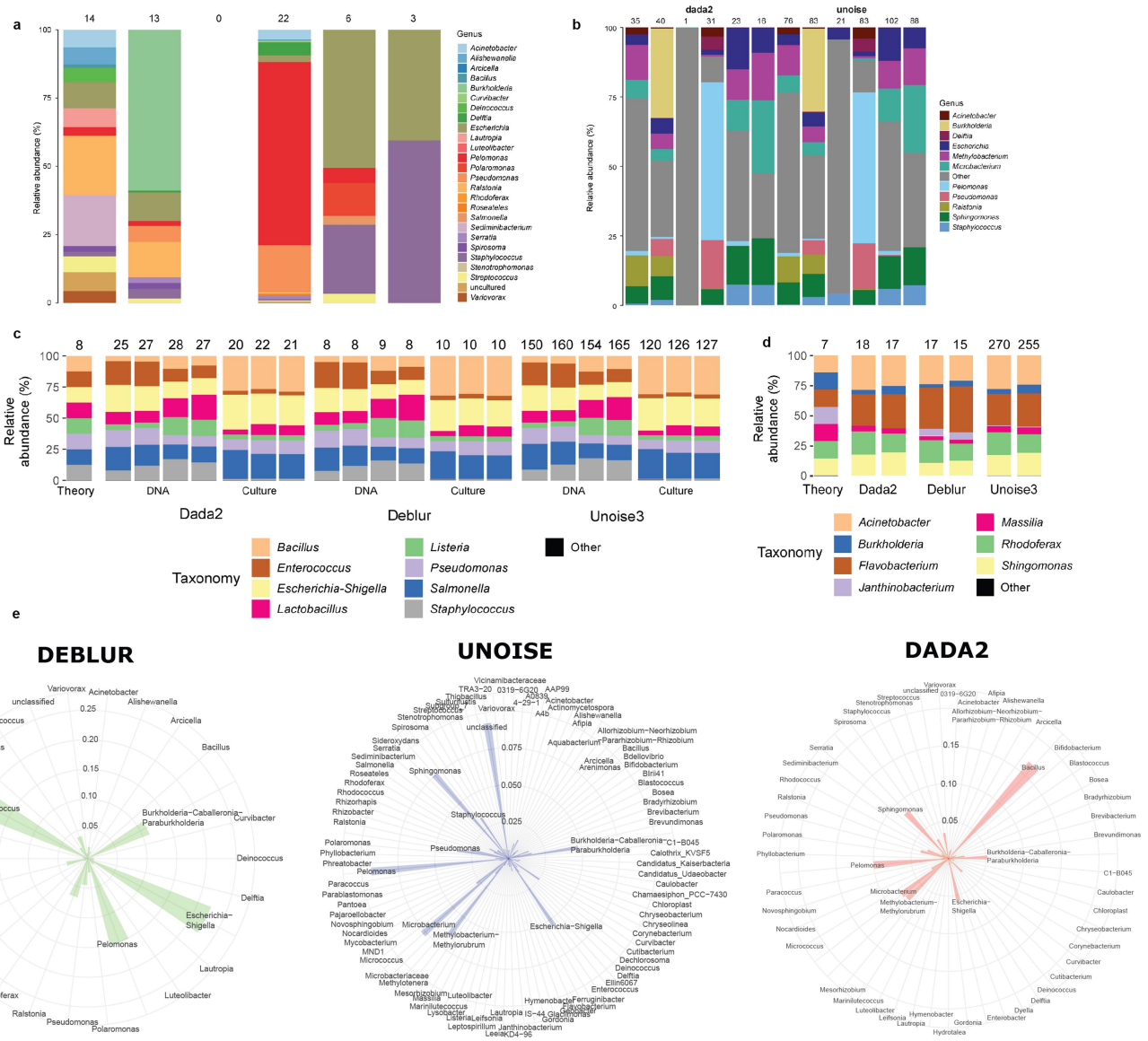
Extended Data Fig. 8 | Benchmarking denoising procedures and sensitivity analyses. (**a**) Comparison of diversity metrics (ASV richness, Shannon index, proportion of unique ASVs in a sample, and relative abundance of unique ASVs) across the whole GFS microbiome dataset for the three different denoising approaches (DADA2, Deblur, UNOISE3). Values were computed with a rarefaction threshold set at n = 135,665 (22,611 reads/sample x 6) after filtering. Comparisons of the effect of different sequence denoising approaches and replicate merging strategies (sum, sample, filtering) on the identification of specific ASVs in the GFS dataset at a rarefaction depth of 22,611 reads per sample. (**b**) Proportion and (**c**) relative abundance of specific ASVs across mountain ranges as computed with UNOISE3, DADA2 and Deblur, taking into account a filtering approach (see Methods) without further alteration of the data. For the method 'sum', counts from replicate samples were summed; for the method 'sample', one replicate was randomly sampled each time; and for the method 'filtering', we applied the filtering as described in the methods, and averaged the replicates. For the boxplots, given are median, interquartile ranges and whiskers extending 1.5× beyond the inter-quartile range. (**d**) Sensitivity analysis of the proportion of specific ASVs (within each mountain range) performed using Deblur, and for varying levels of rarefaction depth (right panel) and number of GFSs sampled per mountain range (left panel). Generally, and as expected, estimates of endemicity are sensitive to both low-rarefaction depth and a low number of samples per region. Lower bound estimates of the proportion of specific ASVs range between approximately 25-65% (using Deblur). Geographically isolated GFS (e.g., New Zealand's Southern Alps, Ecuadorian Andes) harbor consistently (and with high confidence) higher levels of specificity than other regions. Please note that for comparability, a maximum number of six samples was chosen. For most regions, we included many more samples and final rarefaction was adjusted to 22,611 ASVs, leading to overall conservative estimates of specificity.



Extended Data Fig. 9 | Taxonomy of cryospheric samples. Relative abundance of the most abundant genera detected across cryospheric ecosystems (n = 268 samples). Data were retrieved from a previous meta-analysis (Bourquin et al.[21]), and were agglomerated at the genus level.



Extended Data Fig. 10 | Microdiversity of globally present and abundant genera. (**a**) Globally dominant GFS microbiome members (i.e., *Polaromonas*, *Rhodoferax*, *Methylotenera* and *Rhizobacter*) resolve biogeographic differences. Shown are non-metric multidimensional scaling ordinations; lines connect the regional centroids with samples of the respective region (circles). Red symbols reflect the "species scores" of ASVs. Analysis of similarity (ANOSIM) testing for differences between regions are displayed for each genus. (**b**) The global distribution of *Polaromonas* ASVs shows phylogenetic signatures. Shown is the phylogenetic tree of 321 *Polaromonas* ASVs. Circles at the tips highlight regions in which the respective ASV dominates (in terms of relative abundance). Widespread ASVs (i.e. without dominance in a single region) are shown in grey. The outer ring displays mean regional abundances (log-transformed). Note that abundant *Polaromonas* ASVs tend to be abundant across many regions. Phylogenetically closely related ASVs tend to co-occur within regions.



Extended Data Fig. 11 | Taxonomic composition of blank and mock samples. Taxonomic composition of (**a**) blanks analysed with deblur with all the genera assigned and (**b**) with dada2 and unoise with an "other" category. (**c**) Taxonomic composition of commercial mock communities obtained both as DNA (DNA, Zymo, Cat. Nr. 6305) and cell cultures (Culture, Zymo Cat. Nr. 6300). (**d**) Mock community created based on isolates obtained from Swiss GFS. Taxonomic compositions were computed using the three denoising approaches (DADA2, Deblur and UNOISE3). Numbers above the bar plots correspond to the number of ASVs found in each community. (**e**) Visual representation of ASVs detected in blanks and retained by the three different denoising algorithms (Deblur, UNOISE3 and DADA2). These circularized barcharts show the mean relative abundance of all genera detected in the blanks. Abundant genera are highlighted (moved inside). Note that for UNOISE only the 100 most abundant genera could be visualized and that GFS taxonomies, including *Polaromonas*, *Rhodoferax* and *Methylotenera*, were either absent (*Methylotenera*) or of very low relative abundance in these blank samples.

nature portfolio

Corresponding author(s): Tom Battin

Last updated by author(s): Aug 24, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	N.A.
Data analysis	All data were analyses using R (version 4.4.1). All codes are available on github.com/laylaeb/GFS_biogeo2023

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequence data have been deposited in the Sequence Read Archives with NCBI BioProject accession no. PRJNA781406. Processed data along with data tables associated with main and extended figures can be retrieved from Zenodo at doi:10.5281/zenodo.13897903. Data retrieval URL: https://zenodo.org/records/13897903.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	non applicable
Population characteristics	non applicable
Recruitment	non applicable
Ethics oversight	non applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Field work at ca 150 glacier-fed streams around the world strictly following established standard operational procedures and trained personnel. Study systems were selected to cover the major mountain ranges, comply with geopolitical conditions, Covid-19 related context and field safety, covering altitudes from 2000 to 4800 m above sea level.
Research sample	We sampled the benthic sandy sediments of glacier fed streams as this fraction contributes most surface area for microbial colonisation per unit streambed surface area. DNA was extracted from these sediments. We also collected water samples for physico-chemical characterisation. We inferred glaciological metrics from satellite imagery and GIS. Sediments were also analysed for mineral composition.
Sampling strategy	As aforementioned, we covered the major glacierized mountain ranges across the continents. Per mountain range, we selected up to three regions to have representative sample at the scale of the mountain range. Per region, we identified up to 4 glacier-fed streams with different glacier size etc. Per stream, we sampled three sediment patches in a reach as close to the glacier snout as possible (depending on safety and accessibility). All samples for microbiological analyses flash-frozen (liquid nitrogen) immediately upon collection in the field. Temperature chain always (!!) kept until handling in the lab at EPFL, Switzerland.
Data collection	The PI's (Battin) lab collected all samples to ensure the highest level of reproducibility and comparability possible. All DNA was extracted and libraries constructed in the PI's lab by the same two well-trained technicians (again to reduce bias etc). Sequencing was done at KAUST, data analyses at EPFL (Scitas computer) and University of Luxembourg (HPC).
Timing and spatial scale	Expeditions (up to 2.5 months each) were conducted over four entire years, also during the Covid-19 pandemics. The design was such that we sampled glacier-fed streams once either during the vernal or autumnal 'window of opportunity' when environmental conditions were less extreme. For this, we basically traveled with the seasons depending on the hemisphere. This is critical allow comparability across systems.
Data exclusions	A small percentage of samples was excluded because not enough DNA was extracted, which simply reduced the overall number of cases. Few ASVs excluded because they were identified in mock communities
Reproducibility	This is a large-scale spatial survey of complex systems. Hence it is clear that reproducibility at the level of streams is simply not feasible. However (!), we did reproduce at the level of sediment patches (triplicates) sampled per stream reach.
Randomization	none
Blinding	Mock communities (DNA, Zymo, Cat Nr. 6305 and cell cultures Culture, Zymo Cat Nr. 6300) as well as sequencing blanks were used.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Challenging field conditions, well met by highly trained personnel. Safety regulations, regular medical checks and access respected.
Location	Detailed map shown in Extended Data Fig 1 with the Southern Alps of New Zealand, Himalayas, Caucasus Mountains, Pamir and Tian Shan, European Alps, Scandinavian Mountains, Southwest Greenland, Alaska Range, African Rwenzori and both Ecuadorian and Chilean Andes
Access & import/export	We had all required permits for sampling and sample exports. New Zealand: Ministry of Conservation (#72437 – RES); Greenland: Ministry of Industry and Energy (# G19-016); Russia: Ministry of Natural Resources and Environment (Permit granted to IGRAS base on the signed MoU); Ecuador: Ministerio del Ambiente, Direccion National de Biodiversidad; we worked under the global research permit of Escuela Politécnica Nacional); Norway: Norwegian Environment Agency; France (Permissions authorized through the collaboration between EPFL – RIVER Lab and academic and governmental and non-governmental entities in France); Nepal (Research permit from Department of National Parks and Wildlife Conservation; Trekking permits (TIMS) for Sagamartha, Langtang National Parks & Annapurna Conservation Area); Kyrgyzstan (Government of Kyrgyz Republic, Ministry of Environment; Border Zone Permits and Research permits obtained under MoU with CAIAG); Uganda (Uganda National Council for Science and Technology UNCST, # NS291ES, Uganda Wildlife Authority UWA, # COD/96/05); Chile (Ministère de l'Agriculture, La Corporación Nacional Forestal, CONAF # 9.494.705-7); Alaska (US Forest Service USFS Authorization permit GLA1207 FS, Department of Natural Resources DNR Division of Parks & Outdoor Recreation Special Park Use Permit 11 AAC 18.010)
Disturbance	none

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	not applicable
Wild animals	Benthic biofilm microorganisms
Reporting on sex	not applicable
Field-collected samples	As described above, all sediment sampled for microbiological analyses were deep-frozen at the spot in liquid nitrogen (dewars carried up to 5,200 m a.s.l.!), the temperature chain unbroken until handling in the lab. Liquid samples frozen or stored cool in the dark - prefiltered in precombusted vials. Sediments sampled with flame sterilized stainless equipment.
Ethics oversight	None required. However, we had permissions from the various authorities to enter national reserves or Maori territory in NZ, for instance.

Note that full information on the approval of the study protocol must also be provided in the manuscript.