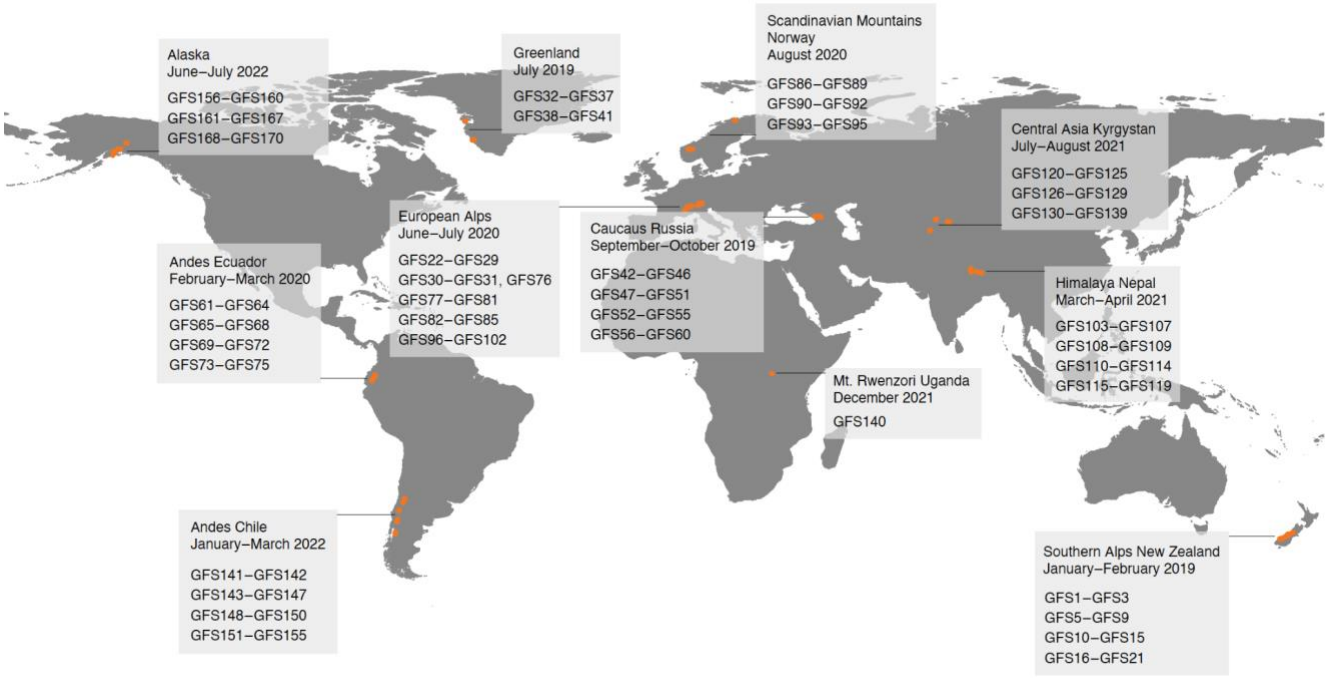
Extended Data

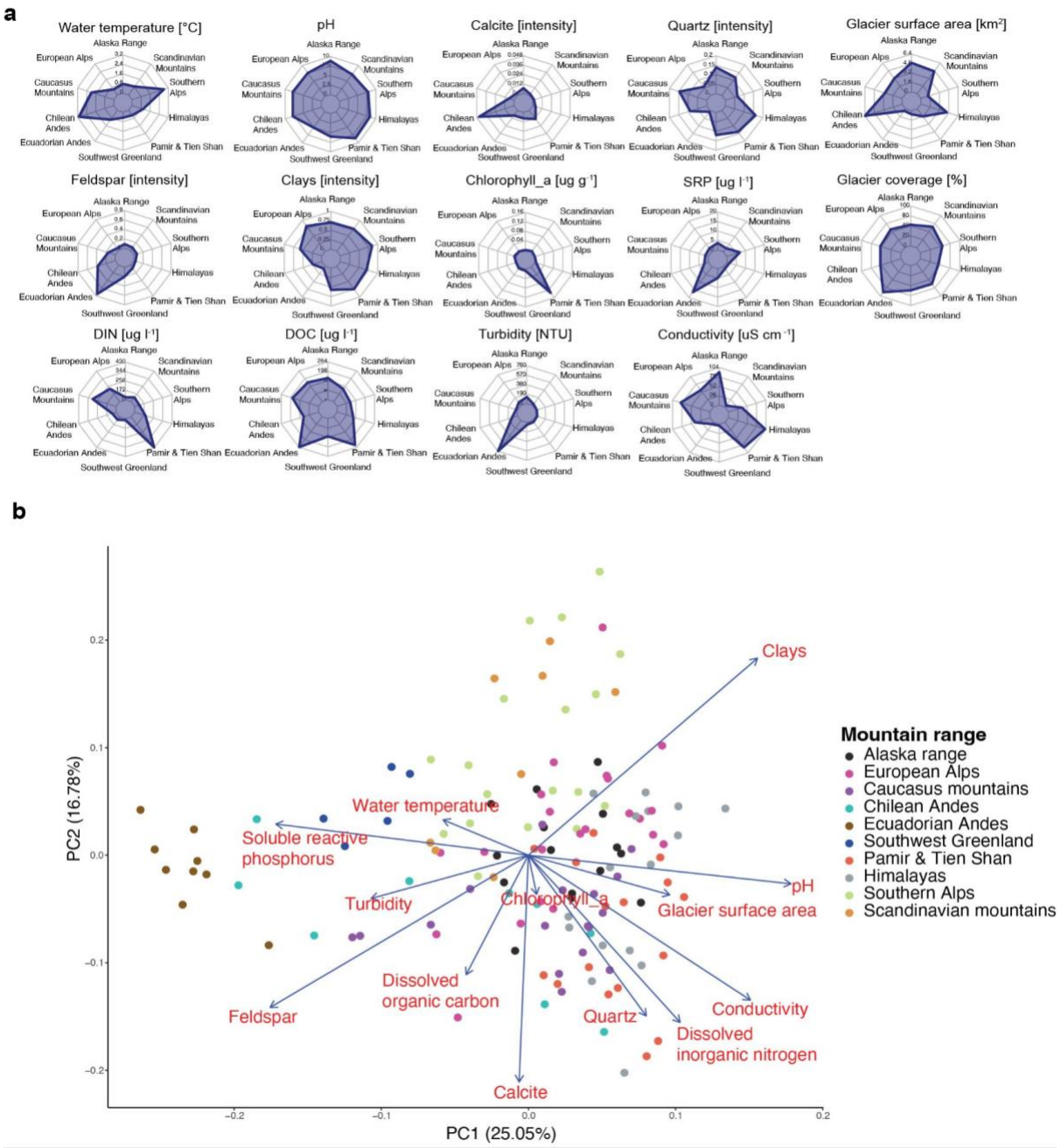
Towards a global biogeography of the glacier-fed stream benthic microbiome

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

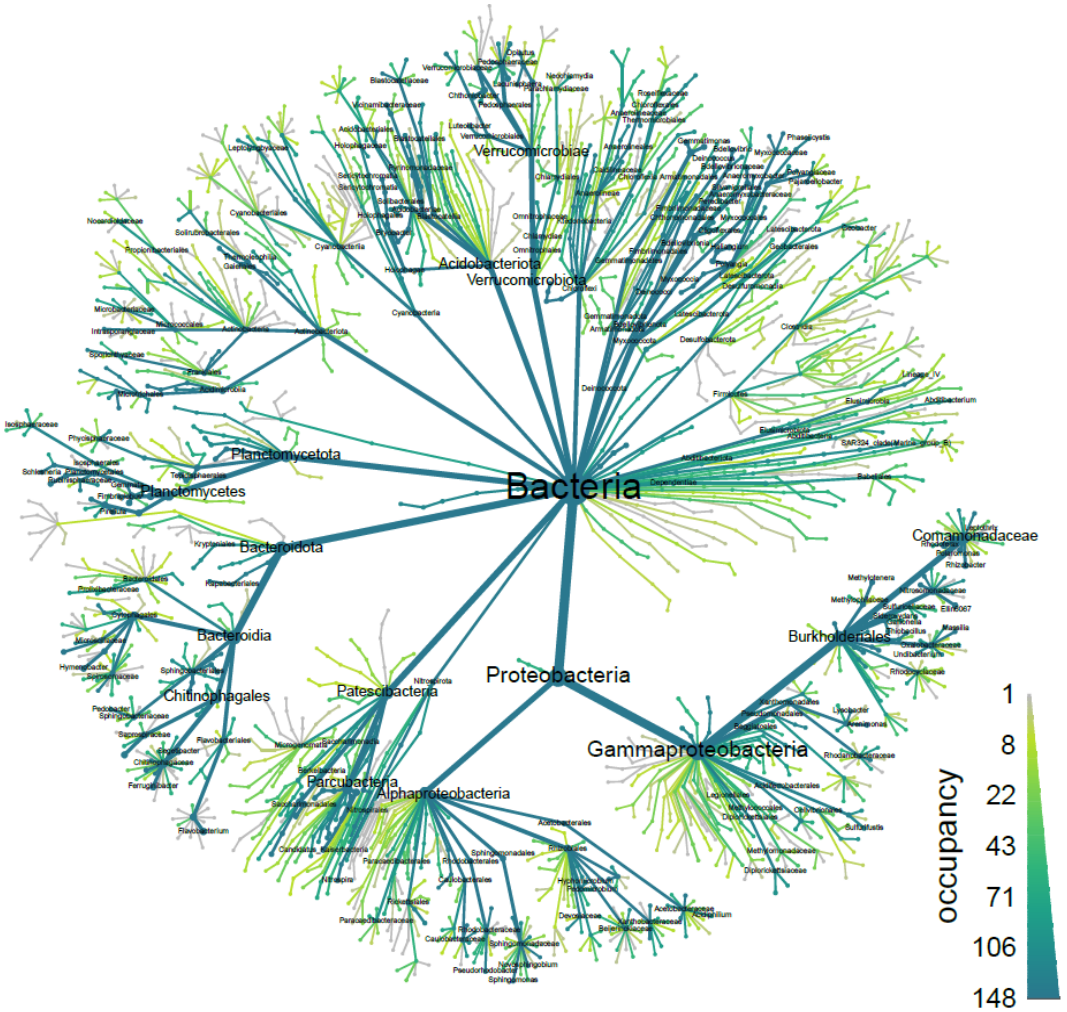
(1) River Ecosystems Laboratory, Alpine and Polar Environmental Research Center, Ecole Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland
(2) MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France
(3) Department of Ecology, Faculty of Science, Charles University, Prague, Czechia
(4) Systems Ecology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
(5) Marine Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia
(6) Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia



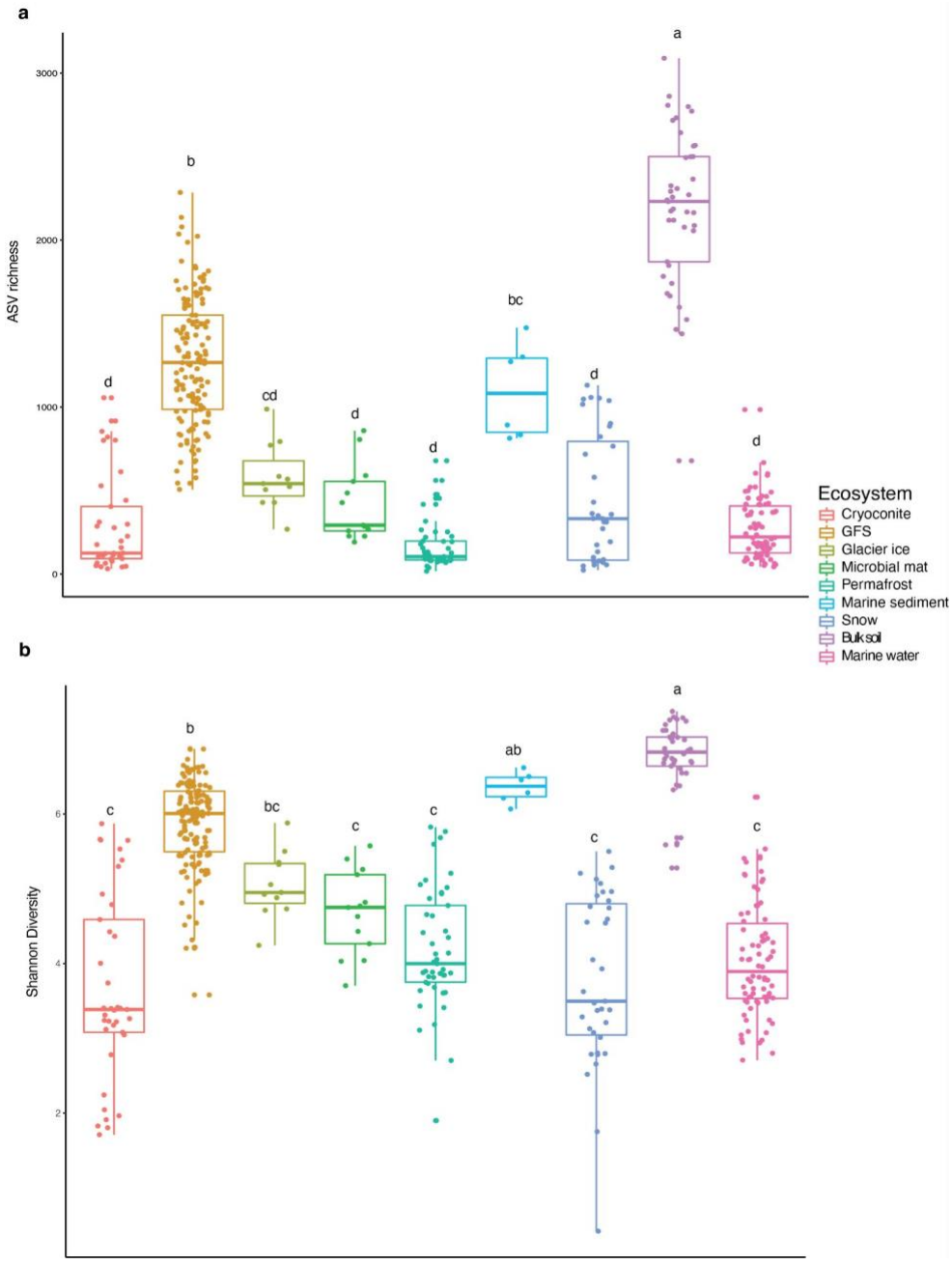
Extended Data Figure 1. World map of the glacier-fed streams (GFSs) sampled by the Vanishing Glaciers Project between 2019 and 2022. Per mountain range, we sampled GFSs from up to three different regions (orange dots). Of the 170 GFSs, we retained 148 for the present study; sediment samples from 22 GFSs could not be processed for microbiological analyses.



Extended Data Figure 2. Environmental parameters of the glacier-fed streams (GFS). (a) Spider plots depicting the relative variation of the GFS (n=147; 1 GFS from the Rwenzori Mountain not included) environmental parameters for each mountain range. **(b)** Principal component analysis (PCA) indicates the segregation of the mountain ranges, largely driven by benthic sediment mineralogy (i.e., calcite, feldspar and clays) and related streamwater geochemistry (e.g., pH).

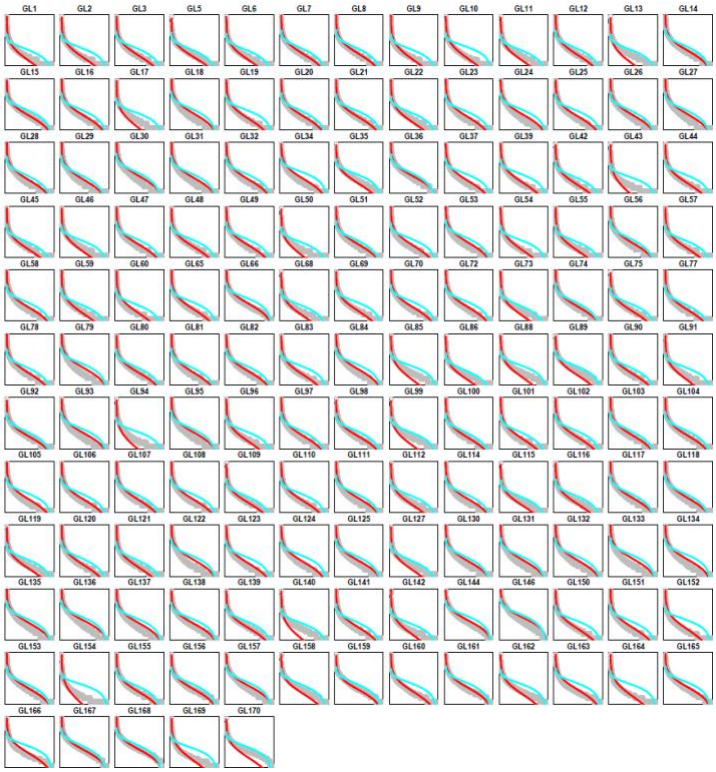


Extended Data Figure 3. Taxonomic diversity of the bacteria from the global GFS microbiome. The dendrogram shows the taxonomic structure from domain- to genus-level and highlights the most prevalent bacterial lineages. Edge color and node size reflect occupancy (number of GFS a taxonomic unit occurs).

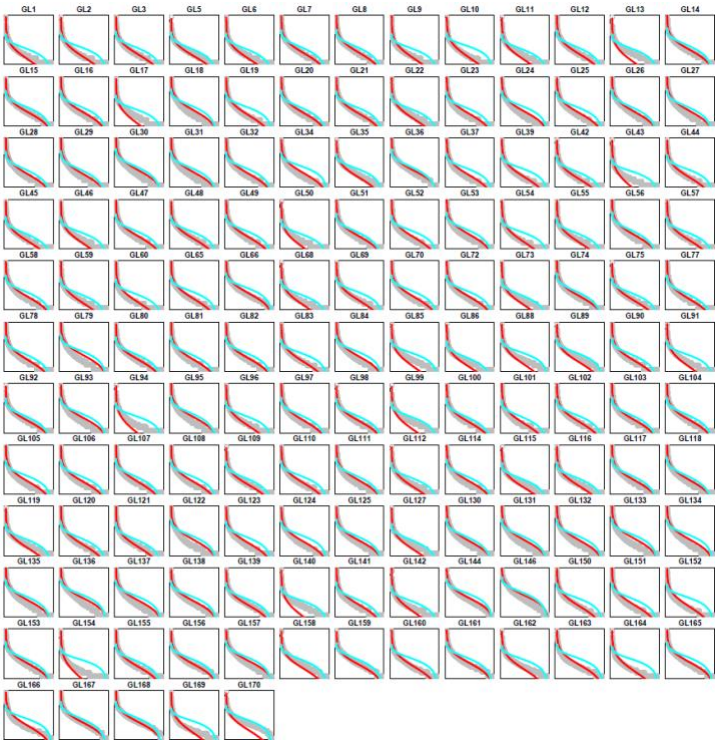


Extended Data Figure 4. Alpha diversity metrics of bacterial communities across cryospheric ecosystems. Patterns of alpha diversity for the different cryospheric ecosystems illustrated as box plots. Using the global inventory assessed by Bourquin and colleagues (ref.[18]), we compared patterns of alpha diversity between benthic biofilms in glacier-fed streams (n=148 GFSs) and other cryospheric ecosystems. We noted that values of (a) ASV richness and (b) Shannon diversity in glacier-fed streams were bracketed between the cryoconite ice environment (Tukey HSD, $p < 0.001$; Dunn Test, $z=-8.8$, $p < 0.001$) and other terrestrial ecosystems such as bulk soils (Tukey HSD, $p < 0.001$; Dunn Test, $z=-4.1$, $p < 0.01$) and further demonstrated similar values as marine arctic sediments (TukeyHSD, $p=0.9$; Dunn Test, $z=-1.1$, $p = 1$). Letters indicate significant groupings through posthoc permutational tests to account for multiple comparisons.

a)

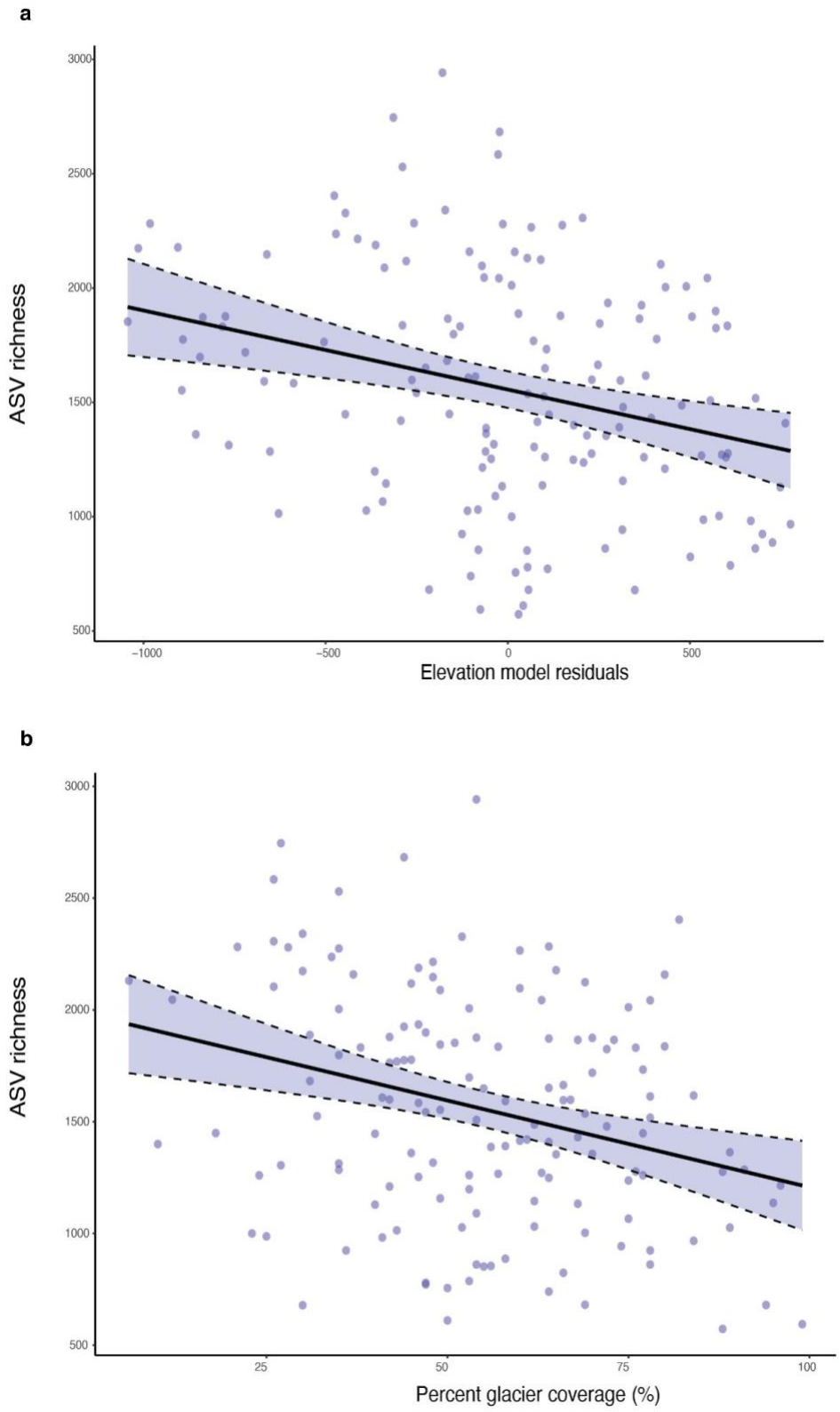


b)



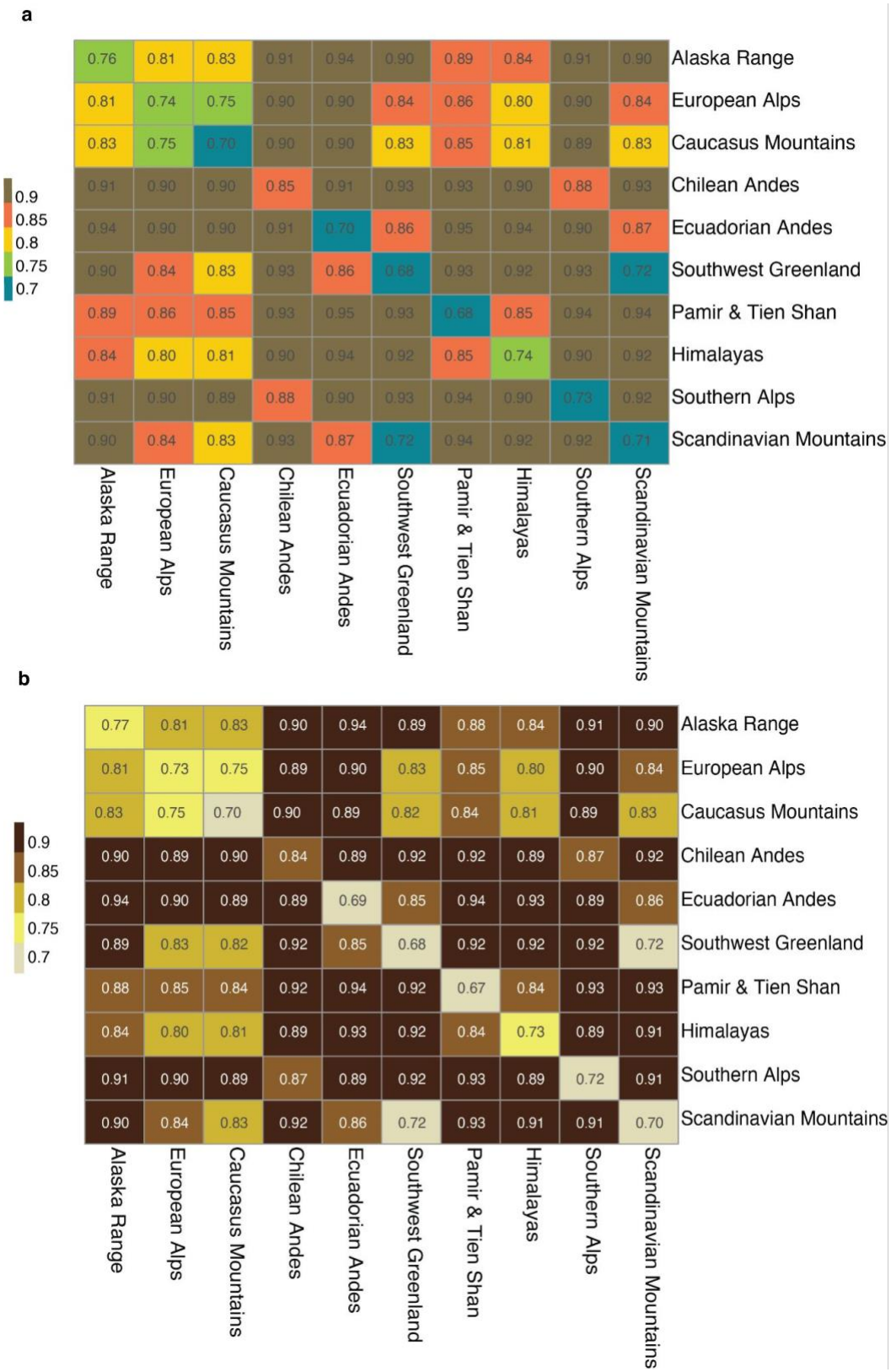
Extended Data Figure 5. Rank-abundance model fits. Shown are rank-abundance distributions with log-normal model (red) and Brokenstick null model (cyan) fits for each GFS for filtered and rarefied **(a)** and unfiltered datasets **(b)**. For most GFS

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. While our stringent filtering criteria led to the preferential removal of rare taxa, they did not qualitatively influence rank-abundance model fits.

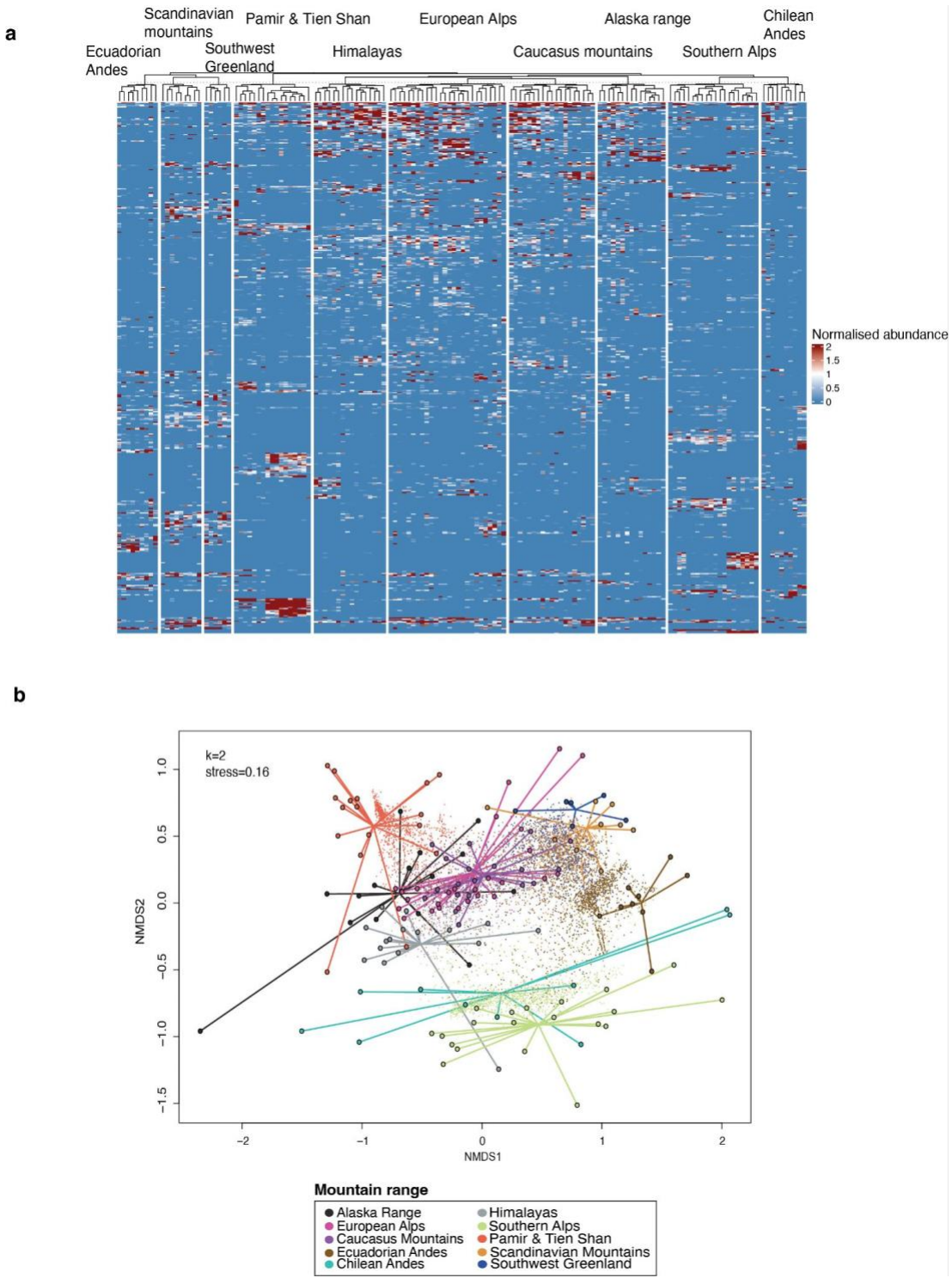


Extended Data Figure 6. Richness of amplicon sequence variants (ASV) in glacier-fed streams changes with elevation. Linear relationships between ASV richness and elevation **(a)** taking in account the latitudinal effect, and percent glacier coverage (b) across all GFSs (n=147; except the Rwenzori Mountains). Generalized 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_latitude, F=7.4, p<0.001; Deviance explained=90.3%; GAM richness_elevation, F=13.1, p<0.001, Deviance explained = 8.3%, adj R^2=0.077), and glacier coverage (GAM richness_glaciercov, spline F=12.4, Deviance explained=7.9%, adj R^2=0.072).

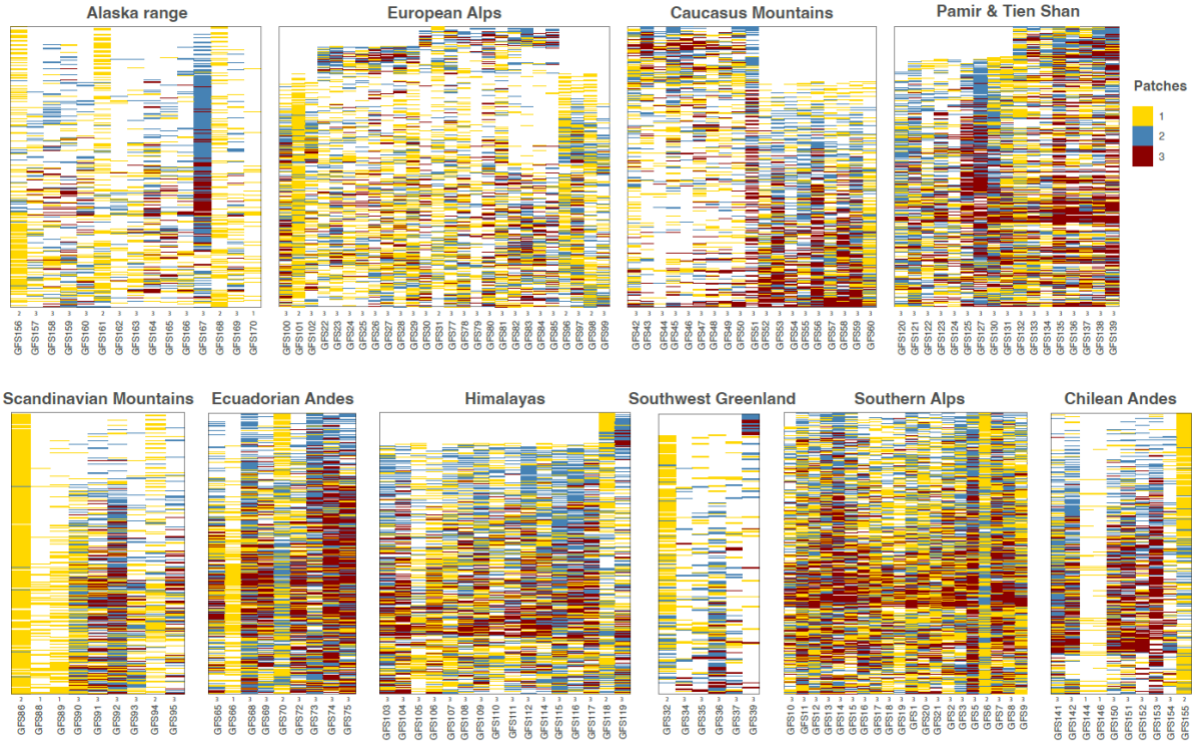


Extended Data Figure 7. Dissimilarities of the glacier-fed stream bacterial communities within and among mountain ranges. Average values of **(a)** Bray-Curtis and **(b)** Sorensen dissimilarity indices computed within and among mountain ranges (n=147 GFSs).

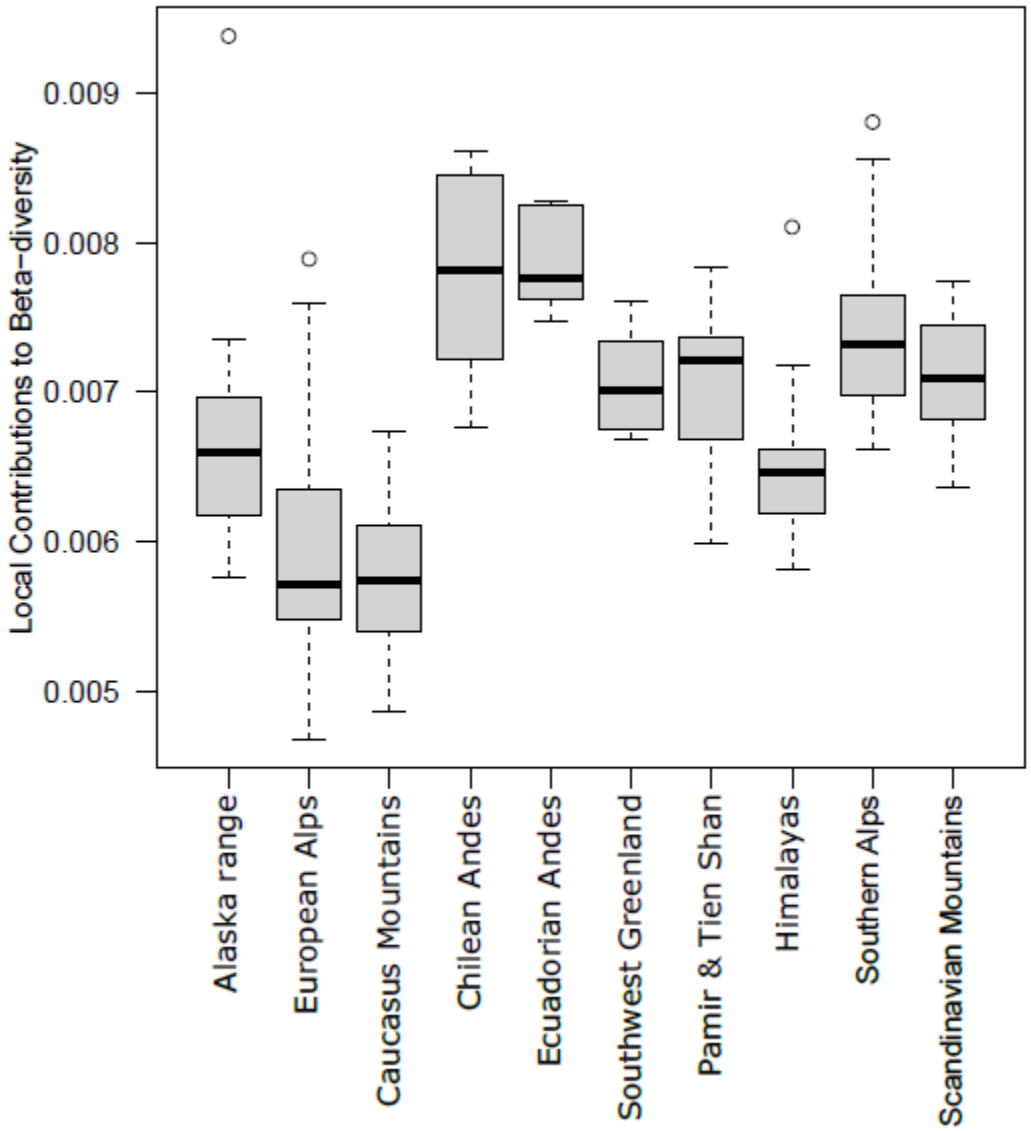


Extended Data Figure 8. Indicator amplicon sequences variants (ASVs) that significantly contribute to the beta-diversity of the benthic microbiome of the glacier-fed streams. (a) Heatmap illustrating the relative abundance of the 500 most

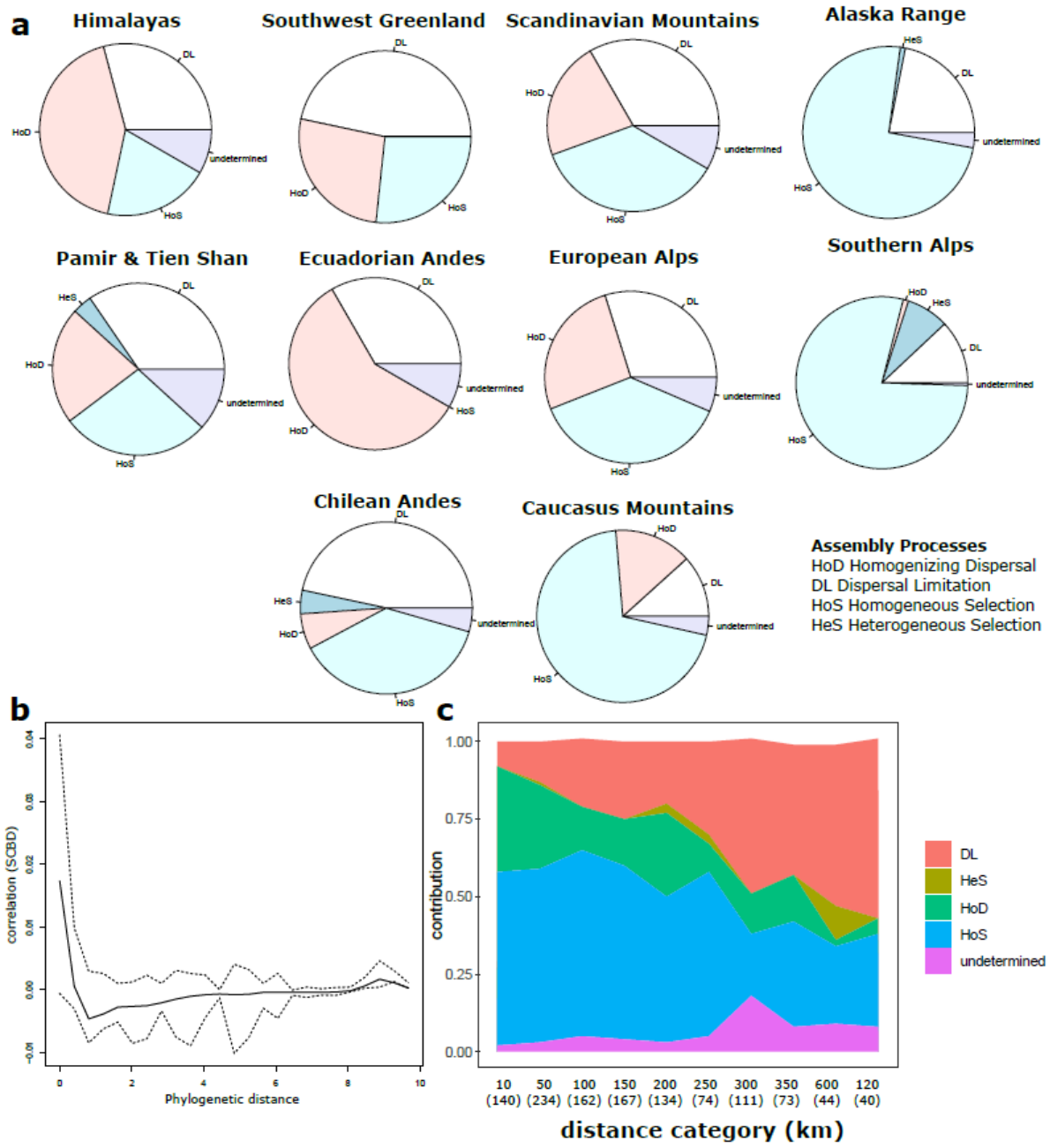
abundant indicator ASVs across mountain ranges (n=147 GFSs). **(b)** NMDS based on Bray-Curtis dissimilarity illustrating biogeographic patterns (large circles reflect individual GFS, lines connecting GFS to the group centroid highlight the different regions). Species scores of the 7800 indicator ASVs, which contribute to dissimilarities among mountain ranges are shown (dots, colored by region).



Extended Data Figure 9. Heatmaps illustrating the occurrence of endemic ASVs in replicate samples (i.e. patches) across regions. Rows reflect ASVs while columns depict different GFS (IDs given below). While the majority of endemic ASVs occurred only in a single GFS (see main text), endemic ASVs also occurred across several samples within the different mountain ranges and were detected in multiple patches of various GFS. Endemic ASVs that were detected in a single GFS (i.e. “unique” ASVs) were found, across all mountain ranges, in 17.7% (median, 12.3%-23.5%) in all three patches of a single GFS and in 32.9% (28.1%-36.1%) in 2 out of 3 patches. Endemic ASVs were clustered based on occupancy patterns. The number of patches (1-3) retained in the dataset (i.e. prior to averaging and rarefaction) is given beneath each GFS. Note that ASVs found in a single patch were retained for GFS with 1 or 2 sampled patches.

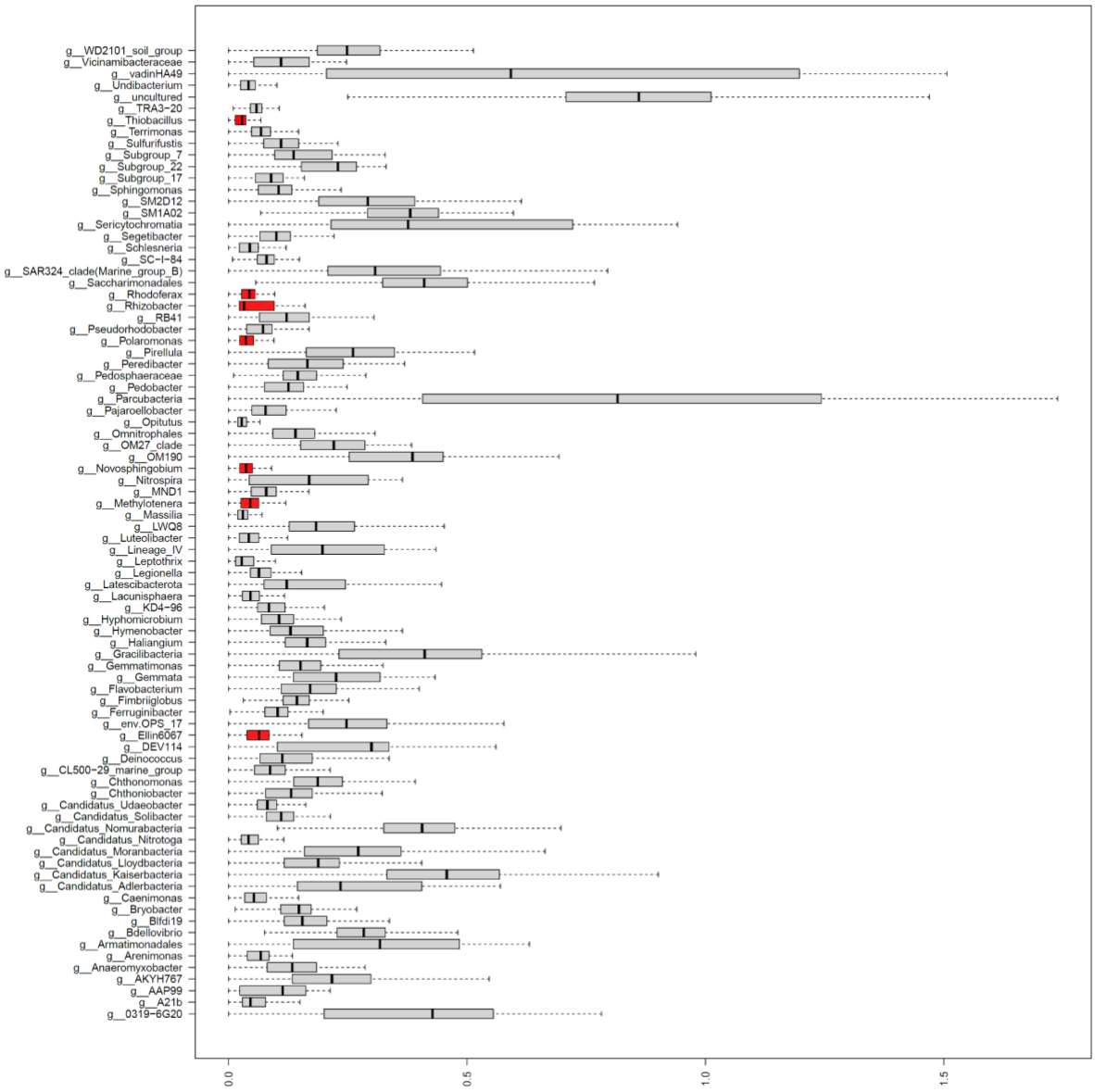


Extended Data Figure 10. Contribution of the mountain ranges to the global microbiome beta-diversity as total variance. Partitioning local contributions to beta-diversity based on total variance of the community matrix. Note that mountain ranges with similar GFS microbial communities (e.g., European Alps and Caucasus Mountains) contribute relatively less to global beta-diversity, whereas geographically isolated localities (such as GFS draining volcanoes in the Chilean and Ecuadorian Andes or GFS in New Zealand Southern Alps) contribute more to global GFS beta-diversity.

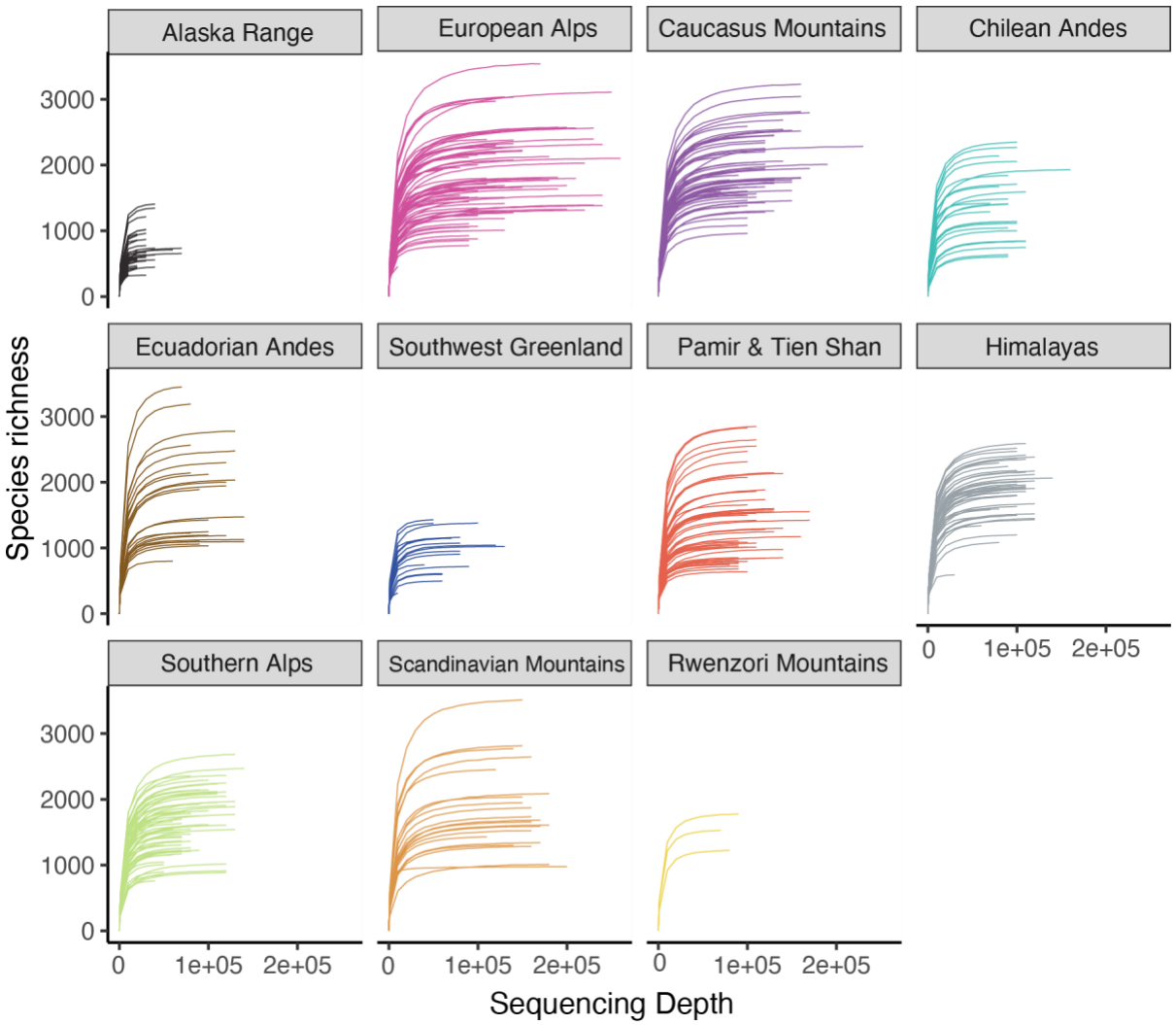


Extended Data Figure 11. Assembly processes. Shown are the relative contributions of assembly processes as derived from null-model comparisons of β-NTI and Raup-Crick$_{\text{Bray Curtis}}$ for pairs of GFS communities within each mountain range **(a)**. Overall, homogeneous environmental selection (HoS) dominated, accounting for 40.9% of all pairwise comparisons, followed by dispersal limitation (DL, 29.9%) and homogenizing dispersal (HoD, 22.0%). Phylogenetic signal in the contribution of ASVs to beta-diversity (as total variance; SCBD) is exemplified as a phylogenetic correlogram between phylogenetic distance and SCBD for a random subsample of

750 ASVs from the entire phylogeny **(b)**. Moran's I statistic for phylogenetic signal was obtained from 50 iterations of the random subsampling process. Median Moran's I was 0.008 (25^{th} quartile: 0.003, 75^{th} quartile: 0.014), p=0.01). Homogeneous environmental selection and homogenizing dispersal were particularly important at short geographic distances (>250 km), whereas dispersal limitation dominated at large, inter-mountain range scales **(c)**. Numbers in brackets indicate the number of sample pairs in each distance category.



Extended Data Figure 12. Bacterial genera under homogeneous environmental selection. Particularly *Polaromonas, Methylotenera, Rhodoferax, Thiobacillus, Rhizobacter, Novosphingobium* and Ellin6067 (highlighted in red) contain particularly many phylogenetically closely related ASVs and are thus considered microdiverse. Note that other clades under homogeneous environmental selection, such as the genera *Ferruginibacter, Flavobacterium* and *Sphingomonas* deviate from this definition of microdiversity. Shown are genera (n=84) with more than 100 classified ASVs, respectively. Box-whisker plots show pairwise phylogenetic distances among ASVs.



Extended Data Figure 13. Rarefaction curves. Rarefaction curves computed prior to filtering and averaging sample patches, illustrating the impact of sequencing depth on richness for all GFSs across the different mountain ranges (n=418).