Supplementary Information

Microbial biodiversity in glacier-fed streams

Linda Wilhelm[1], Gabriel A. Singer[1,2], Christina Fasching[1], Tom J. Battin[1,2] & Katharina Besemer[1,2]

[1]Department of Limnology, University of Vienna, Althanstr. 14, A-1090 Vienna, Austria

[2]WasserCluster Lunz, Dr Carl Kupelwieser Promenade 5, A-3293 Lunz am See, Austria

Supplementary methods

Data analysis

For biofilms we tested for a potential effect of an environmental variable (e.g. streamwater temperature) on beta diversity by two different methods. First, a formal test was done by multiple linear regression on distance matrices (Legendre *et al*., 1994; Lichstein, 2006; Ptacnik *et al*., 2010). For this, the BC dissimilarity matrix was unfolded to a vector of pairwise dissimilarities between any two communities and analysed as dependent on pairwise average temperature and pairwise temperature difference (both computed for the respective pairwise community combinations) as two predictors in a multiple linear regression. In this analysis, the hypothesis of interest concerns the effect of average temperature on dissimilarity. Importantly, a specific pairwise average temperature may result from a combination of communities with similar as well as a combination of communities with dissimilar temperatures, and the environmental differentiation given by the latter combination may also affect the dissimilarity (i.e., beta diversity) between the two communities. It is therefore necessary to include pairwise temperature difference as a covariate in a multiple linear regression approach alongside pairwise average temperature. Given the strong dependencies caused by pairwise habitat (community) combinations in the three involved matrices, significances were computed by permutation (Legendre *et al*., 1994).

Second, to graphically analyse the behaviour of beta diversity as a function of an environmental variable, for example streamwater temperature, we adapted a local polynomial regression fitting procedure known as LOESS (Cleveland *et al.*, 1992) to estimate 'local' beta diversity at a given temperature of interest (for instance, the streamwater temperature of a specific site). For a given temperature, we identified a subset of sites with similar temperatures (the "neighbourhood" in terms of temperature), whose communities were then included in the computation of the centroid and beta diversity with weights determined by their difference to the given temperature of interest. Thus, for each temperature there is a single 'local' value for beta diversity, which has to be understood as the average dissimilarity among communities with similar temperatures.

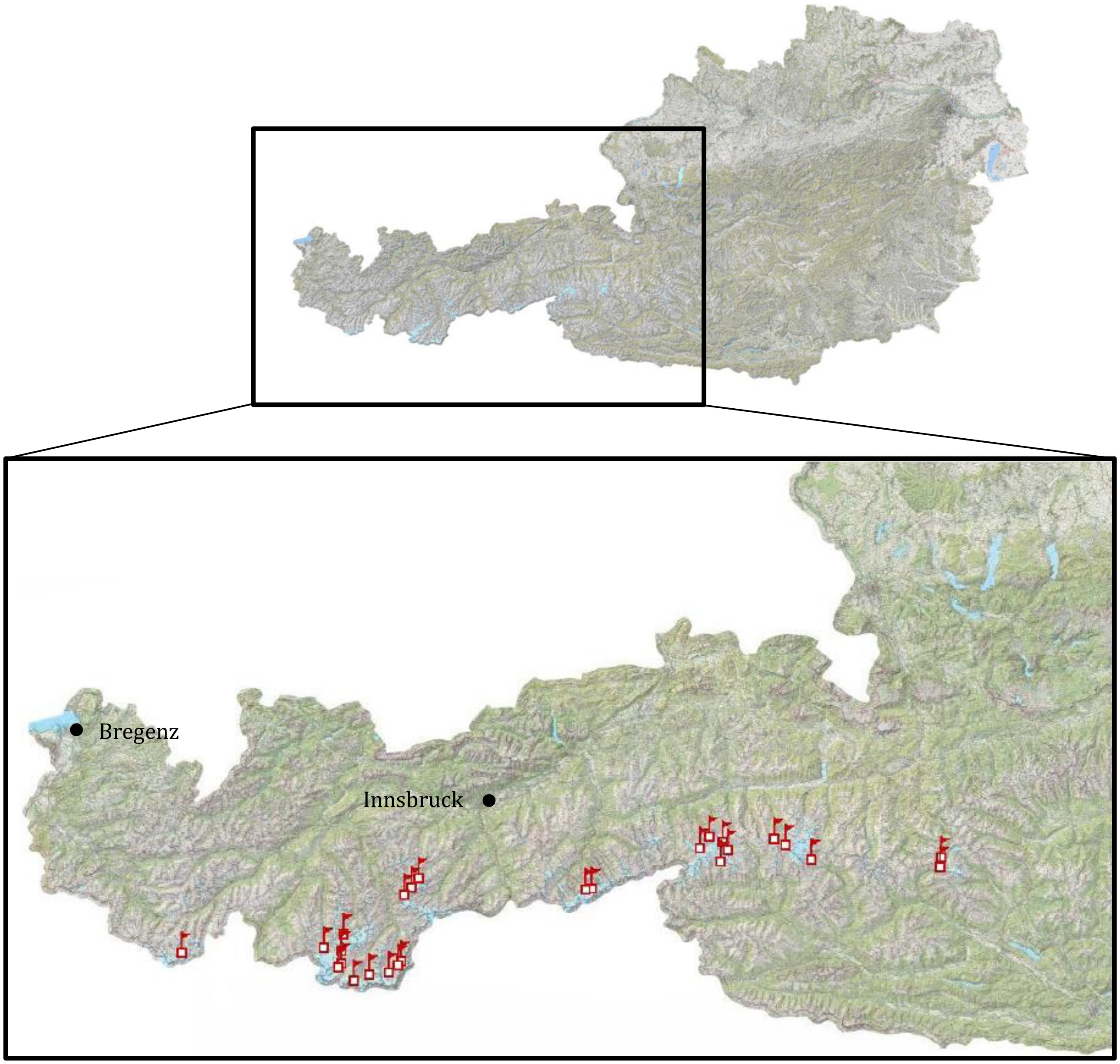
Generally formulated, for each temperature, beta diversity was computed from all communities with tricubic weighting given by

$$w_i = N\left(1 - \left(\frac{d_i}{\alpha\, d_{max}}\right)^3\right)^3$$

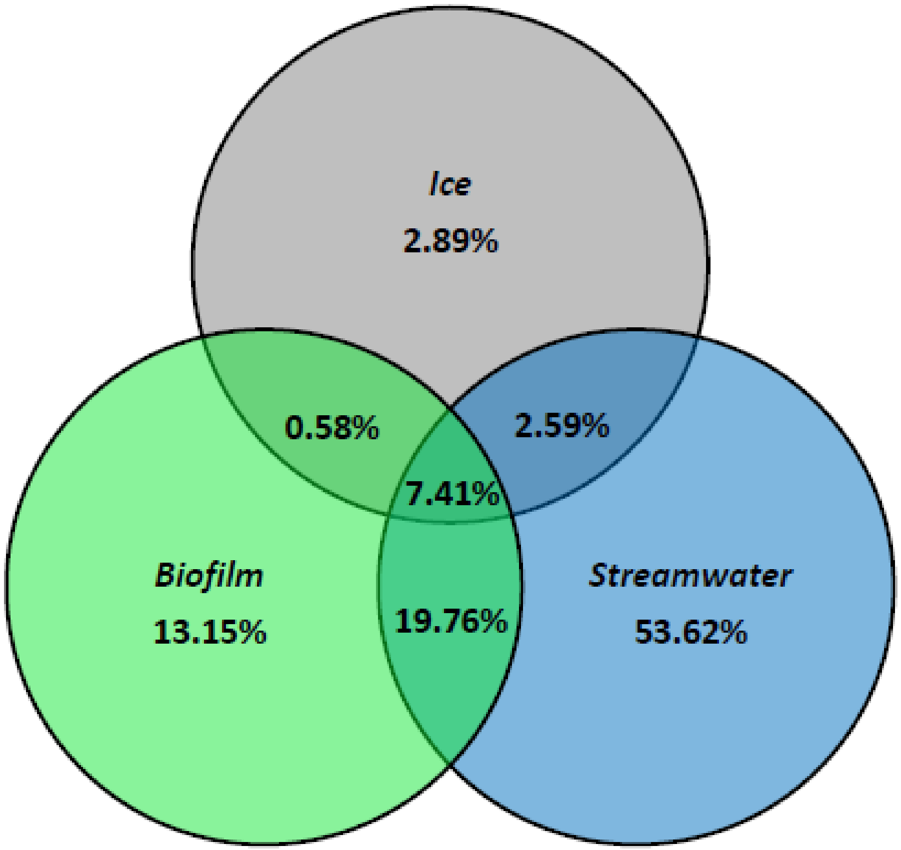
where w_i is the weight for community *i*, d_i is the temperature difference between community *i* and the temperature of interest, d_{max} is the temperature range and α (=0.75) defines the fraction of the temperature range to be considered for the neighbourhood definition. This definition of weights is adapted from the definition of weights in LOESS local polynomial regression fitting (Cleveland *et al.*, 1992). To allow usage of weights in the computation of beta-diversity, the weights must be transformed to counts by *N* (=20) and rounding to integers. As such, the weights become effective by defining how often a community is repeatedly used in the computation of (i) the centroid following principal coordinate analysis and of (ii) beta diversity as the average distance to the centroid (Anderson *et al.*, 2006a; Anderson, 2006b). The weights are thus effective twice: once for centroid computation and once for computation of the average distance. Communities with temperatures more similar to the temperature of interest are used more often and thus have more influence on the beta diversity. The distribution of weights is controlled by the parameter α and can be plotted for any temperature of interest (lower panel in Fig 5). For any temperature of interest, this distribution of weights defines which communities are included in the neighbourhood for a specific 'local' beta diversity (or in other words: how similar communities must be in terms of temperature to be considered). Similar ("neighbouring") temperatures of interest make use of similar, but not identical subsets of communities, which results in a smoothed description of the functional relationship between temperature and beta diversity. The degree of smoothing is controlled by the size of the neighbourhood (α) and by *N*. The latter may not be needlessly increased to avoid inflation of computing times due to large dissimilarity matrices. Neither α nor *N* affects the shape of the relationship between temperature and beta diversity.

Computation of local beta diversity was done for the observed streamwater temperature of each community or any temperature within the observed temperature range to give a smoothed trendline of temperature vs. beta diversity. Confidence intervals for the trendline over the whole observed temperature were estimated by bootstrapping and null model trendlines were constructed by recomputation after randomising the temperature values (Manly, 2006). The same analysis of beta diversity was carried out separately for other environmental variables, i.e., electrical conductivity, pH and elevation. Streamwater temperature and electrical conductivity were square root transformed to decrease skewness of the respective environmental gradient.

Supplementary Figures



SI Figure 1. Map with sampling sites. Upper image: Map of Austria; Lower image: Magnification of the western part of Austria including the 26 sampling sites. Each study site is indicated by a red flag. The map was drawn using digital maps of 'Kompass-Karten GmbH'.



SI Figure 2. OTU clustering in all three glacial habitats. Of all 11,032 OTUs, more than 50% were restricted to the streamwater, while less than 3% were exclusively found in the ice. Streamwater and biofilm communities shared a high number of OTUs, whereas the ice and biofilms had little in common. This suggests that the ice community was not the main source of microbial biodiversity in biofilms.

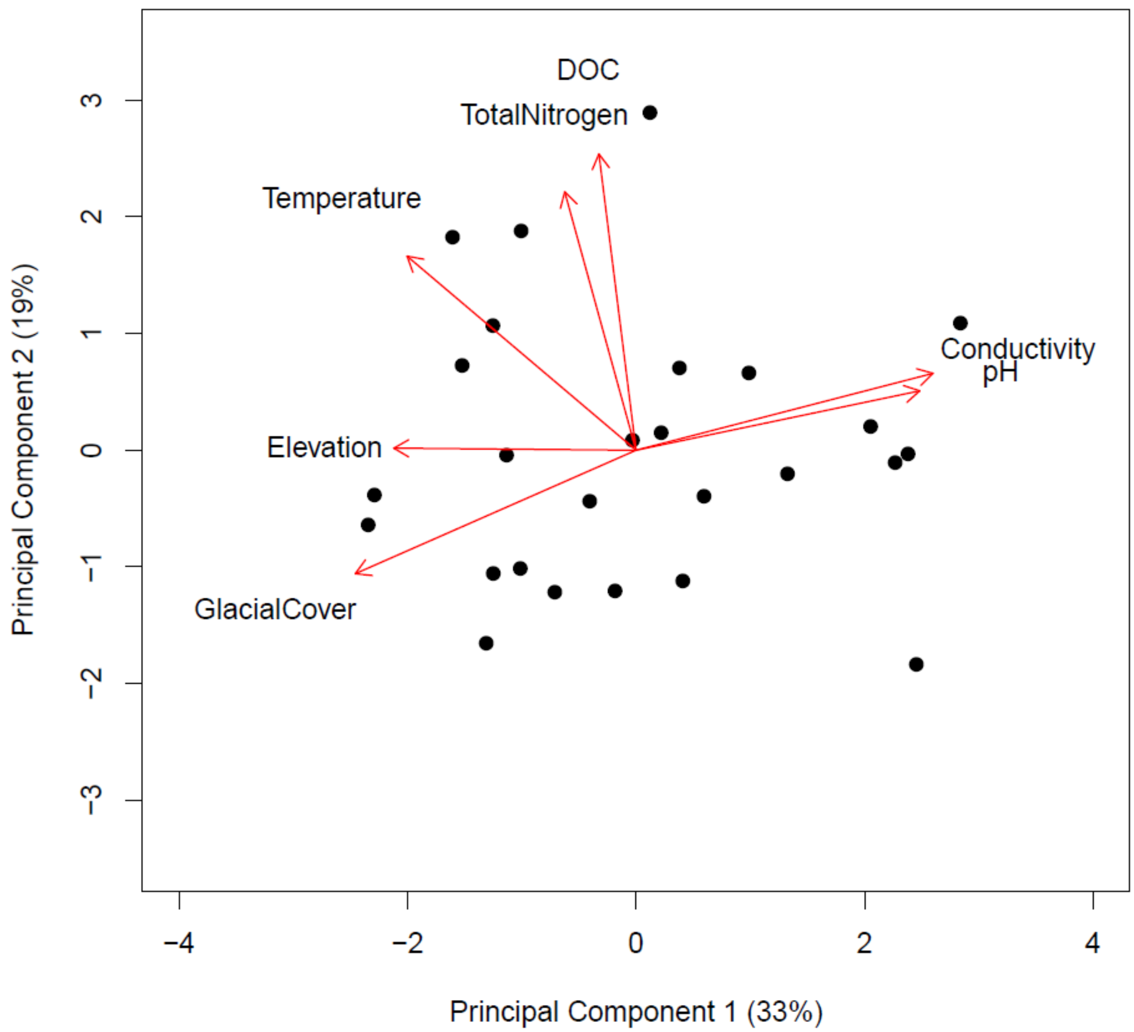
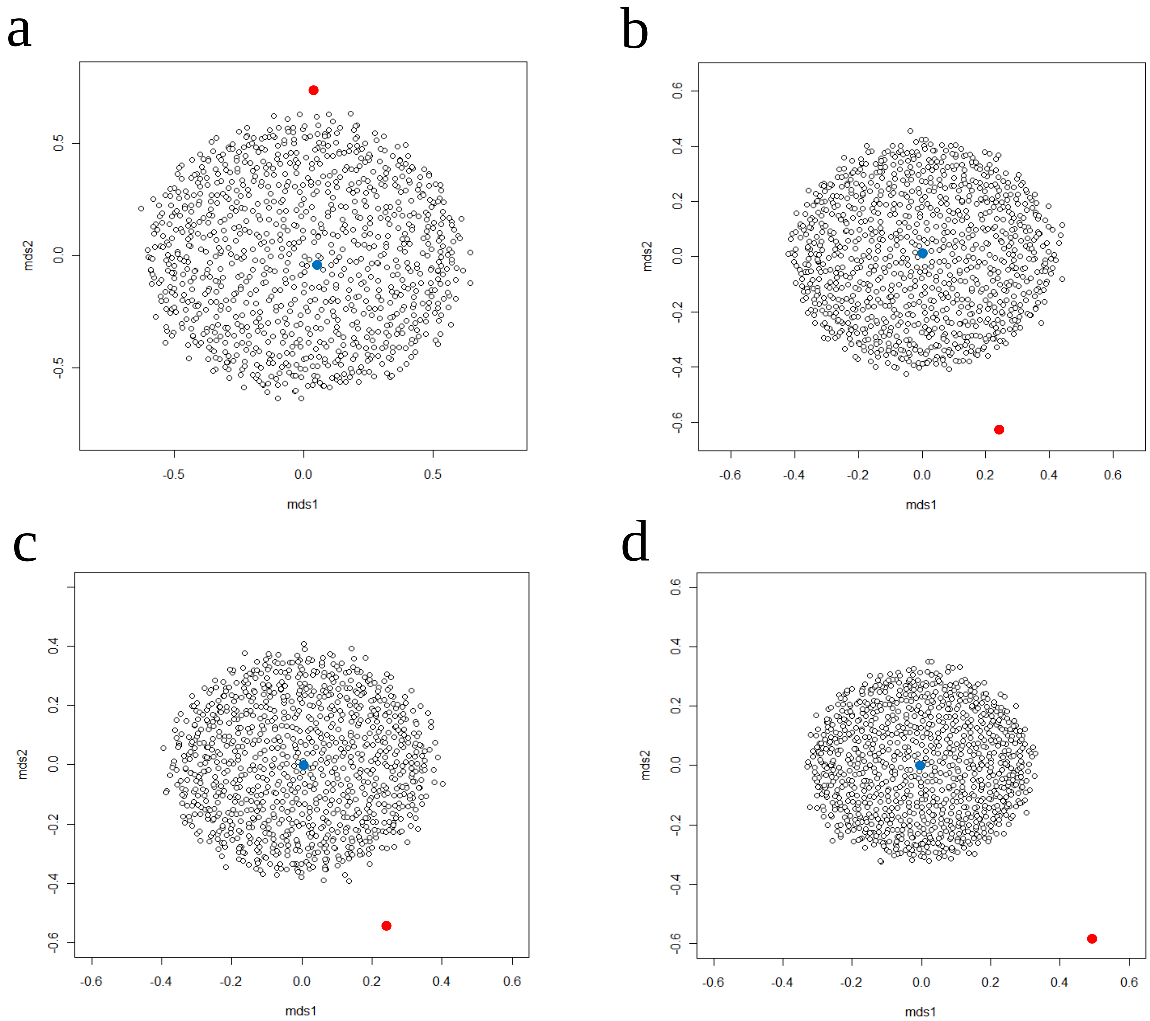
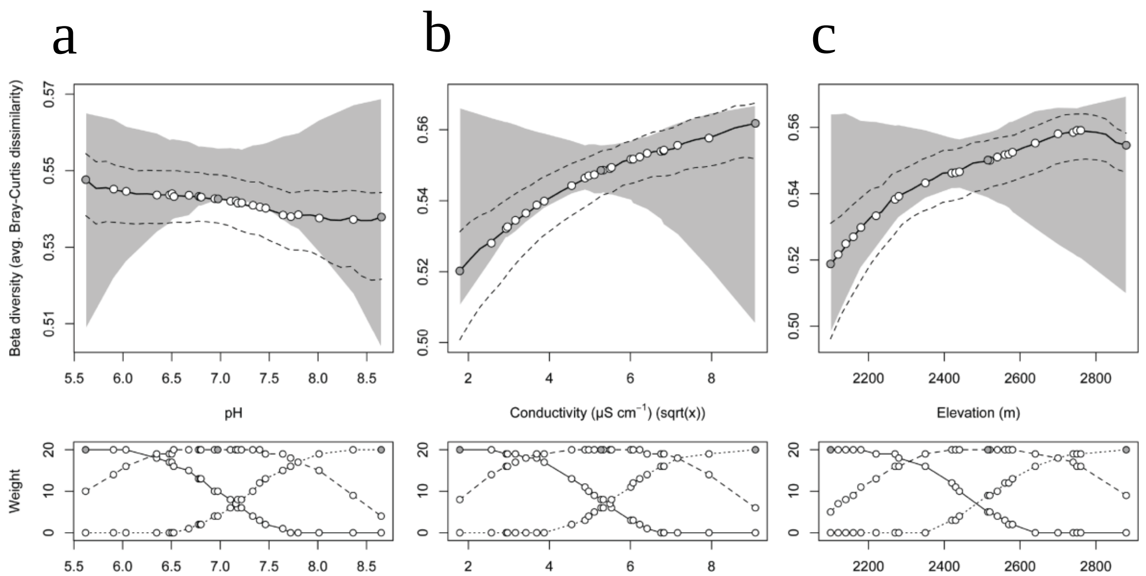


Figure 3 Principal component analysis of environmental variables of 26 glacial ecosystems. Data were z-transformed prior to analysis. Electrical conductivity, pH, elevation and glacial coverage contributed mainly to the first axis.



SI Figure 4. Biofilm communities assembled non-randomly from the streamwater communities. Non-metric multidimensional scaling (nMDS) analysis, visualising the results of a random sampling procedure, to estimate the probability that the biofilm communities represented random samples of their respective suspended source communities. A total of 1000 random subsamples of the streamwater communities were assembled for each sample pair. In all 26 glaciers the biofilm differed significantly from the random assemblages ($P < 0.001$ for each glacier), indicating that the biofilm communities are unlikely to represent a random sample of the streamwater community. Four examples **a**) Jamtalferner **b**) Langtaler Ferner **c**) Niederjochferner and **d**) Schalfferner are shown, to illustrate the distribution of the randomly produced communities in relation to the biofilm community. ○ represents the random subsamples of the streamwater communities, ● the biofilm community and ● the streamwater community.



SI Figure 5. Beta diversity of biofilms was insignificantly related to (**a**) pH (pseudo-t = -1.49, P = 0.45), (**b**) electrical conductivity (square root transformed, pseudo-t = 2.92, P = 0.18) and (**c**) elevation (pseudo-t = 1.93, P = 0.93). Each dot represents an at least once observed environmental situation for which local beta diversity was estimated. The shaded area gives a bootstrap confidence interval for the generated trendline; the trendline itself consists of beta diversities computed for any value within the range of the observed environmental gradient. Grey colored dots correspond to observed minimum, median and maximum of the respective environmental variable, and the lower panel gives the distributions of weights across all communities used for the computation of beta diversity at these three environmental conditions. For a specific environmental condition, communities with higher weights have similar environmental conditions and are disproportionately stronger used in the computation of the environment-dependent beta diversity. All test results refer to multiple linear regressions on distance matrices. See SI methods for computational details.

SI Figure 6. Glacier terminus with various flow paths. As glaciers shrink, glacial runoff flows over bare rock, which is exposed to sunlight and thus increases local streamwater temperature. Exposure to different sources (e.g., bedrock, groundwater and snowfields) of microbial diversity and ion loading (e.g. lateral moraines) are consequences of increased glacier retreat. (Sulztalferner, Stubaier Alps, Photo by Linda Wilhelm).

SI Table 1. Basic characteristics of the study glaciers in the Austrian Alps. Longitude and latitude of each glacier were derived from the World Glacier Inventory (http://nsidc.org/data/glacier_inventory/browse.html). Glacier cover (%) was calculated from total drainage area (Austrian Map 2001) and glacier area (World Glacier Inventory). Elevation (meters above sea level) refers to the glacier terminus.

Glacier	**Longitude (°E)**	**Latitude (°N)**	**Drainage area (km^2)**	**Glacier cover (%)**	**Elevation (m)**
Alpeiner Ferner	11.642	47.032	7.31	70.4	2560
Gaisbergferner	11.065	46.830	2.54	53.2	2560
Gepatschferner	10.758	46.855	30.80	57.4	2140
Hintereisferner	10.770	46.797	16.88	56.1	2430
Hornkees	11.823	47.000	5.96	50.9	2120
Jamtalferner	10.162	46.862	5.84	70.7	2420
Kälberspitzkees	13.278	47.030	1.27	64.8	2520
Kesselwandferner	10.798	46.838	2.69	100.0	2750
Kleinelendkees	13.257	47.063	2.37	100.0	2640
Langtaler Ferner	11.017	46.793	8.15	43.2	2540
Niederjochferner	10.860	46.778	1.04	79.8	2880
Obersulzbachkees	12.302	47.113	13.46	86.0	2270
Ödenwinkelkees	12.652	47.108	6.08	36.5	2160
Pasterze	12.695	47.102	30.09	65.7	2100
Rotmoosferner	11.048	46.817	3.85	82.4	2515
Schalfferner	10.935	46.785	12.35	69.8	2580
Schlatenkees	12.385	47.112	11.97	78.2	2220
Schwarzenbergferner	11.115	47.043	1.99	79.7	2740
Schwarzensteinkees	11.855	47.013	4.33	75.5	2350
Stubacher Sonnblickkees	12.600	47.132	1.56	76.3	2570
Sulztalferner	11.080	47.002	4.77	86.9	2560
Taschachferner	10.858	46.900	11.03	60.5	2440
Untersulzbachkees	12.350	47.135	8.66	49.3	2180
Vernagtferner	10.817	46.877	3.09	78.3	2760
Viltragenkees	12.372	47.128	3.47	87.9	2280
Zettalunitzachkees	12.511	47.106	3.52	92.4	2700

SI Table 2. Environmental parameters of the glacier-fed streams. pH, electrical conductivity and streamwater temperature were measured in the field using WTW-sondes (pH320, Cond340i). DOC concentrations were measured with a Sievers 900 TOC Analyser. Total nitrogen ($N\text{-}NO_2$ + $N\text{-}NO_3$ + $N\text{-}NH_4$) was measured using a Continuous Flow Analysis (Alliance instruments).

Glacier	pH	Conductivity [µS cm^{-1}]	Streamwater temperature [°C]	DOC [µg L^{-1}]	Total nitrogen [µg L^{-1}]
Alpeiner Ferner	7.641	30.2	1.9	50	210.9
Gaisbergferner	8.650	82.5	0.7	610	216.4
Gepatschferner	7.159	63.0	0.6	50	241.2
Hintereisferner	6.033	46.5	0.3	400	180.1
Hornkees	6.478	8.6	1.0	140	263.5
Jamtalferner	7.721	46.6	0.6	220	307.4
Kälberspitzkees	6.778	11.7	0.4	110	147.7
Kesselwandferner	7.178	28.4	0.5	160	201.8
Kleinelendkees	6.802	10.0	0.0	110	182.1
Langtaler Ferner	6.795	24.7	1.0	3160	205.2
Niederjochferner	6.503	38.8	1.1	470	218.1
Obersulzbachkees	6.974	13.6	3.7	380	223.6
Ödenwinkelkees	7.463	41.1	0.4	140	187.8
Pasterze	7.798	45.6	-0.1	70	45.8
Rotmoosferner	7.333	36.8	3.1	630	324.7
Schalfferner	6.348	36.1	1.1	270	180.7
Schlatenkees	7.104	26.1	0.9	160	131.2
Schwarzenbergferner	6.677	6.6	0.3	160	200.6
Schwarzensteinkees	6.523	8.8	1.5	180	391.8
Stubacher Sonnblickkees	5.905	3.2	3.1	140	128.5
Sulztalferner	5.619	15.0	2.1	160	221.2
Taschachferner	7.404	51.3	0.7	110	192.3
Untersulzbachkees	8.362	27.8	0.2	160	206.6
Vernagtferner	6.941	23.8	2.4	1800	229.8
Viltragenkees	8.013	30.6	1.0	600	291.9
Zettalunitzachkees	7.217	20.7	0.1	210	183.4

SI Table 3. Correlations between pH and electrical conductivity and phyla in biofilms. Representation of the Pearson correlation coefficient, calculated between the relative abundance of phyla and the environmental variables pH and electrical conductivity.

n.s. – not significant

Variables	Phyla	P	r
pH	Acidobacteria	< 0.05	-0.415
	Gemmatimonadetes	< 0.05	-0.425
	Proteobacteria	n. s.	-0.354
Electrical Conductivity	Actinobacteria	< 0.05	0.429
	Nitrospira	< 0.05	-0.493
	Verrucomicrobia	< 0.05	-0.463

References

Anderson MJ, Ellingsen KE, Mcardle BH. (2006a). Multivariate dispersion as a measure of beta diversity. *Ecol Lett* **9:** 683-693.

Anderson MJ. (2006b). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62:** 245-253.

Cleveland WS, Grosse E, Shyu WM. (1992). Local regression models. Chapter 8 of *Statistical Models in S*eds J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.

Legendre P, Lapointe FJ, Casgrain P. (1994). Modeling brain evolution from behavior - a permutational regression approach. *Evolution* **48:** 1487-1499.

Lichstein JW. (2006). Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol* **188:** 117-131.

Manly BFJ. (2006). *Randomization, bootstrap and Monte Carlo methods in biology,* London, UK, Chapmann & Hall.

Ptacnik R, Andersen T, Brettum P, Lepistö L, Willén E. (2010). Regional species pools control community saturation in lake phytoplankton. *Proc R Soc B* **277:** 3755-3764.