2010

Biogeography and habitat modelling of high-alpine bacteria

Andrew J. King
University of Colorado Boulder

Kristen R. Freeman
University of Colorado Boulder

Katherine F. McCormick
University of Colorado Boulder

Ryan C. Lynch
University of Colorado Boulder

Catherine Lozupone
Washington University School of Medicine in St. Louis

See next page for additional authors

Recommended Citation

King, Andrew J.; Freeman, Kristen R.; McCormick, Katherine F.; Lynch, Ryan C.; Lozupone, Catherine; Knight, Rob; and Schmidt, Steven K. ,"Biogeography and habitat modelling of high-alpine bacteria." Nature Communications. 1,. 53. (2010).
https://digitalcommons.wustl.edu/open_access_pubs/4329

Authors

Andrew J. King, Kristen R. Freeman, Katherine F. McCormick, Ryan C. Lynch, Catherine Lozupone, Rob Knight, and Steven K. Schmidt

Biogeography and habitat modelling of high-alpine bacteria

Andrew J. King[1], Kristen R. Freeman[1], Katherine F. McCormick[1], Ryan C. Lynch[1], Catherine Lozupone[2,3], Rob Knight[2] & Steven K. Schmidt[1]

Soil microorganisms dominate terrestrial biogeochemical cycles; however, we know very little about their spatial distribution and how changes in the distributions of specific groups of microbes translate into landscape and global patterns of biogeochemical processes. In this paper, we use a nested sampling scheme at scales ranging from 2 to 2,000 m to show that bacteria have significant spatial autocorrelation in community composition up to a distance of 240 m, and that this pattern is driven by changes in the relative abundance of specific bacterial clades across the landscape. Analysis of clade habitat distribution models and spatial co-correlation maps identified soil pH, plant abundance and snow depth as major variables structuring bacterial communities across this landscape, and revealed an unexpected and important oligotrophic niche for the Rhodospirillales in soil. Furthermore, our global analysis of high-elevation soils from the Andes, Rockies, Himalayas and Alaskan range shows that habitat distribution models for bacteria have a strong predictive power across the entire globe.

[1] Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA. [2] Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA. [3] Washington University School of Medicine, St Louis, Missouri 63108, USA. Correspondence and requests for materials should be addressed to S.K.S. (email: steve.schmidt@colorado.edu).

Despite the known importance of microorganisms to the maintenance of the Earth's biogeochemical cycles[1,2], the relationship between the ecological niche of microbial groups and the culture-independent identification of their presence in the environment is poorly understood[3–5]. This is because of lack of resolution in data collected using traditional methods, which obscures the identification of potentially important groups across a landscape[3]. However, Ettema and Wardle[3] point out that, given enough data, 'spatial variability is the key, rather than the obstacle, to understanding the structure and function of soil biodiversity'. Previous studies have shown that there are spatial patterns to microorganisms[6–10] and that some coarse-scale taxonomic groups (at the phylum or division level) show strong correlation with environmental parameters; for example, Acidobacteria are negatively correlated with soil pH, whereas Proteobacteria are positively correlated with pH[11,12]. However, groups such as the Acidobacteria and Proteobacteria are extremely large and functionally diverse; for example, Proteobacteria encompass almost all known microbial physiologies ranging from phototrophs to heterotrophs to chemoautotrophs, and recent studies indicate that the Acidobacteria may be equally metabolically diverse[13]. Thus, we gain very little information about the biogeochemistry of a specific system or of the global biosphere by knowing the spatial distribution of such large taxonomic groups. Similarly, we still have only a rudimentary understanding of the local spatial scale at which soil microbes and soil biogeochemical parameters are distributed. Furthermore, matching the relative abundance patterns of specific microbial taxa and biogeochemistry at both local and global scales has remained an elusive task.

Our study makes use of recent advances in high-throughput sequencing, bioinformatics and biogeochemical methods[14–19] to map the co-occurrence of microbial groups with biogeochemical soil properties across a highly heterogeneous, high-elevation landscape near the continental divide in the Rocky Mountains of Colorado, USA[20–22]. On the basis of our previous study of the spatial autocorrelation of soil biogeochemical properties[22], we collected 160 soil samples in a nested sampling scheme. This sampling scheme allowed us to determine spatial variation in microbial diversity (a random subset of 85 samples was pyrosequenced for the 16S gene) and its relationship to 21 soil biogeochemical properties at scales from 2 to 2,000 m in Colorado. These analyses were essential for obtaining spatially explicit, landscape habitat distributions (models based on co-variation of relative abundance with biogeochemistry) for bacterial community members, which were tested at a global scale by sampling similar soils in the Colorado Rockies, Himalayas, Andes and Alaskan range (Sanger clone libraries of the 16S gene).

In this study, we show that bacterial communities have significant spatial autocorrelation at distances up to 240 m; however, beyond that distance, community composition does not display significant spatial autocorrelation. In addition, the dominant bacterial clades from the landscape-scale survey display strong co-variation with biogeochemical parameters, such that their relative abundances across the globe are predictable using habitat distribution models.

Results

Landscape patterns in bacterial community relatedness. The first step in assessing habitat distributions for bacteria was to determine whether there was a significant spatial pattern to their distribution across the landscape (see Fig. 1 for sampling design). We used Unifrac phylogenetic analysis[18,19] to show that there was a significant change in community relatedness with increasing distance between any two samples across the landscape ($n = 85$, $P = 0.001$, Mantel test) up to a maximum autocorrelation distance of 240 m (Fig. 2). However, the change in community relatedness up to this scale was somewhat small (change in UniFrac distance over 240 m = 0.03), perhaps indicating that only a subset of the total

bacterial community is changing across the landscape. In contrast, beyond 240 m, there is a random scatter of data around the plateau value that is equal to the average community relatedness among all
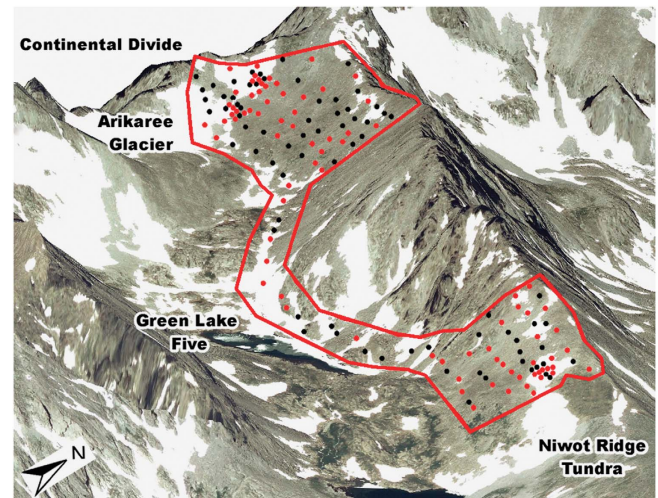

Figure 1 | Landscape sampling scheme. Satellite image of the sampling area in Green Lakes Valley, Colorado (40°3′24″N 105°37′30″W). Black dots indicate sampling locations and red dots indicate the samples that were sequenced for bacterial community composition. The distance between the farthest two samples was 2 km.
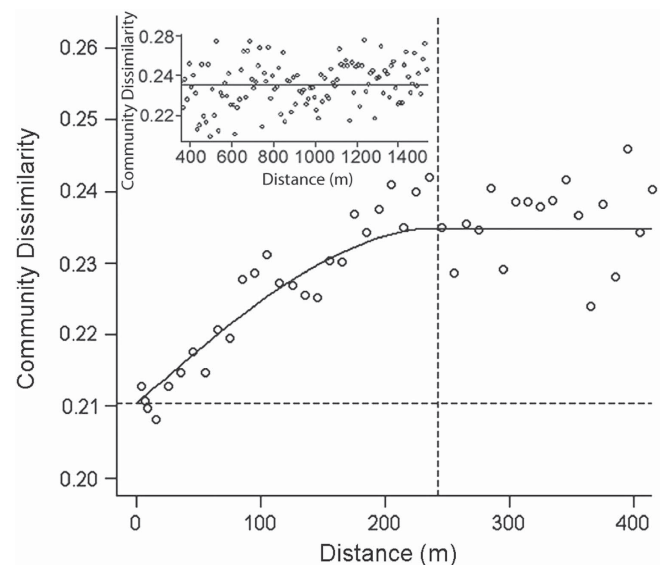

Figure 2 | Community-level spatial autocorrelation semivariogram. A semivariogram plot of the decay in community similarity (as measured by the UniFrac community dissimilarity metric on the y axis) with increasing distance between samples. A UniFrac value of 1 indicates no shared community members between two samples and a value of 0 indicates 100% shared community members. The solid line is the variogram model fit, which tracks the predictable change in shared community membership with distance. The vertical dashed line is the distance of spatial autocorrelation (range), which is the maximum distance, according to the model, at which similarity in community composition between samples is correlated (in this case, 240 m). The horizontal dashed line represents the 'nugget,' which is the proportion of the change in community composition not explained by the spatial model. The inset demonstrates that, past the autocorrelation distance, there is no predictable change in community composition with distance.

Table 1 | Habitat-model spatial and biogeochemical components.

AIC variables	Explained by spatial model (r^2)	Explained by predictor variables (r^2)	Explained by OLS (r^2)	Total explained by (predictor + space, r^2)	Residual autocorrelation (m)
Acidobacteria G4 β-Glucosidase, soil water, snow depth, MBN, pH	0.365	0.409	0.44	0.573	350
Rhizobiales Soil water, forb abundance, snow depth, pH, WHC	0.352	0.339	0.365	0.533	250
Rhodospirillales Soil water, snow depth, DOC, sand, forb abundance, pH	0.148	0.557	0.56	0.591	110
Saprospirales β-Glucosidase, soil water, DOC, pH, forb abundance, snow depth	0.29	0.587	0.593	0.621	250

AIC, Akaike Information Criteria; DOC, dissolved organic carbon; MBN, microbial biomass nitrogen; OLS, ordinary least squares; WHC, water holding capacity.

samples (see inset to Fig. 2); thus, at distances greater than 240 m, it is just as likely to find a closely related community as it is to find a distantly related community.

To determine which, if any, subset of the bacterial community was changing across the landscape, we examined the spatial autocorrelation in genetic relatedness and relative abundance for major bacterial clades containing more than 100 sequences across all sampling sites (30 clades in all). Genetic relatedness and/or relative abundance of clades may account for community biogeographical patterns[5]; however, across the high-alpine landscape, only the relative abundance of specific clades contributed to the community spatial pattern across the landscape (six clades, $P \leq 0.002$, Moran's I for relative abundance, Supplementary Data). These analyses show that high-alpine bacterial clades have distinct landscape-scale patterns in distribution, suggesting that bacterial clade relative abundance may be structured by patterns in biogeochemical parameters at the landscape scale.

Habitat distribution modelling for bacterial clades. Given the spatial patterning of microbial clades, we next determined which, if any, biogeochemical parameters underlie these patterns[4]. We used habitat distribution models[23,24] to analyse the habitat distributions of the 30 clades, in order to identify the major landscape-scale structuring factors. The models characterize the relationship between the relative abundance of bacterial clades and an extensive set of biogeochemical parameters across the landscape (21 different factors, including soil pH, plant cover, average annual snow depth, soil texture and extracellular enzyme activities, see Supplementary Data). These independent analyses identified some of the same clades that spatial autocorrelation analyses did, including the Rhodospirillales, Rhizobiales, Acidobacteria G4 and Saprospirales, which were identified as having the highest levels of correlation with soil biogeochemical parameters (Supplementary Data). This approach yielded strong model fits with r^2 values between 0.53 and 0.62 (Table 1), equivalent to the best r^2 values for models of plant and animal abundances at the landscape scale[23,24].

To visualize how these clades are related to soil biogeochemical parameters, we mapped the habitat distributions for each of the three most abundant clades across the landscape (Fig. 3); each mapped model describes the relative abundance of a clade on the basis of its response to the biogeochemical parameters at any one location in combination with a Kriged[25] spatial component that is a proxy for the influence of unmeasured biogeochemical parameters. In addition, out of the 21 biogeochemical parameters measured, we identified three parameters as the primary factors shaping microbial distribution in this environment (Table 2). These parameters were soil pH, snow depth and forb abundance (forbs are broadleaved flowering plants, not grasses). Snow depth and plant abundance are known drivers of landscape structure in this
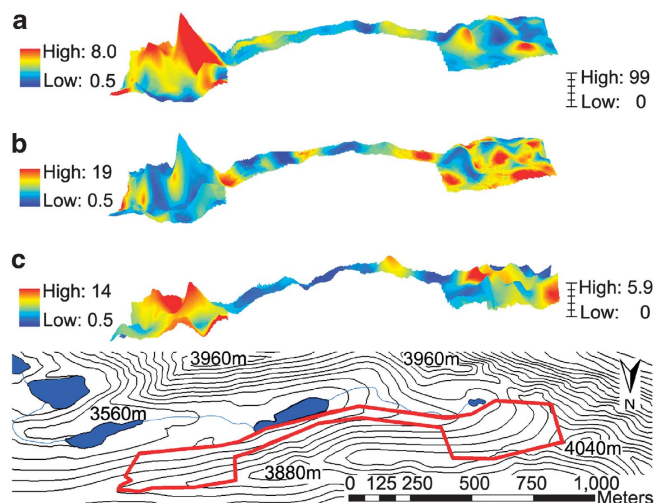

Figure 3 | Major clade habitat distribution maps. Maps of the sequence relative abundance as predicted by habitat distribution models for the three most abundant clades with a strong correlation with environmental variables. Each relative abundance map is depicted with four dimensions, the length and width representing geographical space, the vertical dimension representing forb abundance for the upper two maps and soil pH for the bottom map (hash mark scale on the right) and colour representing the relative abundance of each of the clades (scale on left, red = high; blue = low). The clades are (**a**) Rhizobiales, (**b**) Rhodospirillales and (**c**) Acidobacteria G4. The clade with the fourth highest relative abundance, the Saprospirales, had a distribution very similar to that of the Acidobacteria G4 and is not shown. The maps were created by cokriging[25], an interpolation method that uses the 85 relative abundance measurements in combination with the environmental predictors from our model at all 160 sample locations to create a continuous map of relative abundance in the sampling area. The bottom topographic map shows the two-dimensional extent of the landscape.

extreme alpine environment[20–22], and interact in that snow depth can control plant abundance in this system. In contrast, soil pH did not have a large interaction with other model variables, which suggests that pH measures a separate landscape process such as the composition of bedrock weathering products[26]. In addition, pH was the only model variable that showed a strong effect on the distribution of all four clades, which is in agreement with previous studies at continental scales[11,12,14], and suggests that soil pH is an important driver of microbial community composition at both small and large scales. Although spatial studies have been conducted for microorganisms at many scales[6–10], these models

Table 2 | Landscape and global habitat distribution model coefficients.

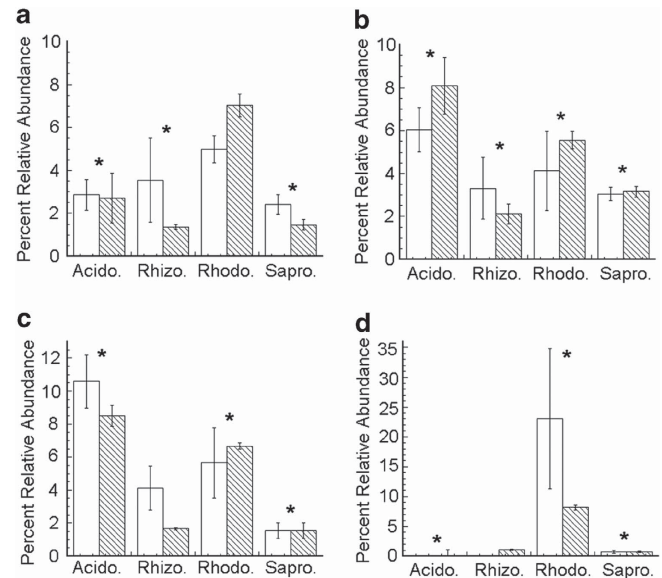|	Acidobacteria G4	Rhizobiales	Rhodospirillales	Saprospirales
β-Glucosidase	0.251 (0.231)	0 (0)	0 (−0.214)	0.326 (0.347)
Soil water	−0.129 (−0.014)	−0.264 (−1.145)	−0.23 (−0.251)	−0.074 (0.569)
Forb abundance	0 (−0.412)	0.372 (0.383)	0 (0)	0.173 (−0.16)
Snow depth	−0.544 (0)	−0.249 (0)	−0.295 (0)	−0.165 (0)
WHC	0 (−0.067)	0.32 (0)	0 (0)	0 (0.389)
MBN	−0.196 (0)	0 (0.305)	−0.077 (0)	0 (0)
DOC	0 (0)	0 (0)	−0.255 (−0.099)	−0.122 (−0.267)
MBC	0 (0)	0 (−0.139)	0 (0)	0 (0)
SAND	0 (0)	0 (0)	0.413 (0)	0 (0)
pH	0.315 (0.325)	0.059 (0.276)	−0.557 (−0.446)	0.274 (0.3)
Leucine peptidase	0 (0.944)	0 (0)	0 (0)	0 (0.307)
Global scale model predictive power				
Mean difference (predicted − actual)	−2.4	−1.15	−2.7	−1.13
Standard deviation of residuals (error-corrected model)	4.49 (3.06)	3.11 (NA)	14.47 (NA)	1.23 (0.98)

DOC, dissolved organic carbon; MBC, microbial biomass carbon; MBN, microbial biomass nitrogen; SAND, % sand; WHC, water holding capacity.
Landscape-scale habitat models are given as the first number and the revised global-scale models are given in parentheses.

represent the first successful description of predictive habitat distributions for bacteria.

The predicted ecological niches based on our habitat distribution models confirm what is already known about some groups and suggest an unexpected dominance and new niche for another large group of bacteria. For example, the Rhizobiales are known plant root symbionts[27]; hence it was not a surprise that their relative abundance was most highly correlated with plant abundance across the landscape. In contrast, our results pointed to the unexpected importance of the Rhodospirillales (our most abundant clade) across this alpine landscape. These organisms are mostly found in aquatic habitats where many of them fill a phototrophic niche, although they have extreme metabolic versatility encompassing photoheterotrophic, chemoorganotrophic and photoautotrophic lifestyles[28]. Although pH was the best predictor of the relative abundance of Rhodospirillales, the mechanism of this control for this group and for broader groups of microorganisms remains unknown[11,12]. However, although we do not yet know the Rhodospirillales' function in high-alpine soils, their negative correlation with plant abundance (the next strongest correlate) may indicate that they are outcompeted for light by plants in the alpine landscape and may be previously undocumented phototrophs in this environment. Rhodospirillales negatively correlated with soil nutrients and total microbial biomass as well (Supplementary Data), which, in combination with their metabolic versatility, suggests that these alpine microorganisms are well adapted to extremely oligotrophic areas. Regardless of their exact function, this is the first report of the widespread occurrence of Rhodospirillales in soil and we would not have predicted their importance across large expanses of the alpine landscape without detailed models of their habitat distribution.

Global-scale predictive power of habitat distribution models. To estimate global-scale applicability of bacterial habitat models to geographically separate high-elevation environments, we constructed Sanger clone libraries from six samples in each of four of the highest mountain ranges on Earth. These locations represent tests of our models against the extreme environmental limits of high-alpine systems (newly deglaciated soils near Mt Denali, late-melting snowbanks near the continental divide in the Colorado Rockies and alpine deserts in the high Andes and Himalayas; Supplementary Table S1). Taken together, the four clades identified in our pyrosequencing study made up a significant portion of the Sanger library microbial community in most of our sites, representing 26% of all bacteria in the Colorado Rockies, Himalayas and Alaskan range, but only 8% in the most extreme site, the high Andes.

Figure 4 | Major clade global-scale abundances and model predictions. The relative abundance of Colorado's four major clades across high-alpine soils at a global scale (Acido, Acidobacteria G4; Rhizo, Rhizobiales; Rhodo, Rhodospirillales; and Sapro, Saprospirales; **a**, Rocky Mountains; **b**, Alaska Range; **c**, Himalayas; and **d**, Andes). Actual relative abundance: open bars; predictive habitat model relative abundances: shaded bars; error bars represent standard error; *indicates nonsignificant difference between actual and predicted (*t*-test, $n = 6$, $P > 0.05$). Ordinary Least Squares (OLS) predictive habitat models using a restricted parameter set (see model parameters, Table 2). Acidobacteria G4 and Saprospirales had significant correlation between residuals (predicted relative abundance – actual relative abundance) and environmental variables, and are error corrected using OLS to predict residual error. Sanger relative abundances were rescaled because of the previously described biases in Sanger versus pyrosequencing[14,41].

Our habitat distribution models correctly predicted the relative abundance of the four major clades from our pyrosequencing study across the entire global data set (Fig. 4). The models in which the biogeochemical variables closely matched the Colorado Rockies environment had the highest predictive power; however, the models did not fit as well in areas with extreme differences in environment. In our most extreme global location, that is, the volcanic soils of the high Andes with almost no snowpack and no plant cover, the Rhodospirillales

were, predictably, the group with the highest relative abundance, whereas the other three groups were absent or had very low relative abundances. This suggests that, although extreme habitats result in lower accuracy of habitat modelling for alpine bacteria, these same major clades are predictable in their importance globally. This conclusion is supported by our recent findings that fungal communities are very similar in plant-free soils of the Rocky Mountains, Himalayas and Antarctica[29]. Thus, high elevation and high latitude environments seem to harbour globally distributed microbial clades and are proving to be ideal environments to test hypotheses about the biogeography of soil microbial community diversity and function.

Discussion

Although other studies have shown that (1) spatial patterns exist in soil microorganisms[6,7] and (2) steep gradients in soil chemistry are correlated with phylum-level changes in microbial community composition[20,21], our study is the first to successfully link spatial autocorrelation in microbial communities to the distribution of individual clades and to demonstrate that these distributions can be modelled with strong predictive power across the landscape and the globe. We did this by examining the relationship between narrowly defined bacterial clades and soil environmental and biogeochemical patterns, which affords greater power to identify ecological patterns than do previous operational taxonomic unit diversity[6,7] or phylum level[11,12] studies. By examining narrowly defined clades, we were able to provide the first environmental-sequencing-based description of ecological niches for bacteria and identify unexpectedly important bacterial clades such as the Rhodospirillales. In addition, the groups that showed the highest level of spatial structuring across the landscape have predictable distributions in high-elevation soils across the globe, suggesting that these groups are easily dispersed and are of significant importance to alpine biogeochemistry and bacterial community dynamics. These findings are evidence that soil microorganisms are not homogeneously distributed across landscapes but rather occur in patches the composition of which is related to the landscape distribution of biogeochemical properties. This approach is uniquely ecosystem focused and greatly expands our ability to link changes in community diversity with the relative abundance of individual bacterial clades and understand the ecology of soil organisms across the landscape and the Earth.

Methods

Sampling scheme. A total of 160 soil samples were collected from a continuous landscape on the south side of the Green Lakes Valley Watershed (GLV), CO, USA. We sampled a distinct and well-defined landscape unit within the GLV that is bound on the east by the tundra, on the south by alpine lakes, glaciers and meadows, on the west by the continental divide and on the north by steep cliffs. There is a large cliff face in the centre of our landscape, along the base of which exists a narrow 75 m wide corridor that connects the upper and lower parts of the landscape (Fig. 1). The sampled landscape is composed of a matrix of block slope, late-melting snow banks overlaying unvegetated gravel soils, fellfields and small patches of vegetation[20,22,30]. However, even in the most developed soils, the soil texture is high in sand content and the total soil depth is minimal. The valley receives the majority of its precipitation during winter months[14] and many snow banks do not completely melt until late July/early August. Our sampling was conducted from 4–8 September 2007 in order to minimize the effects of localized variation in soil water because of snowmelt subsidies.

The main goal of our sampling effort was to construct spatially explicit landscape models. Such models require a subset of samples to be collected at a small enough scale in order to establish a baseline for the spatial autocorrelation[31]. A preliminary study of GLV Watershed soils[22] spanning sampling distances from 10 cm to 1 km was used to determine an optimum sampling interval of 50 m. However, to generate accurate spatial models, we selected three locations for smaller spaced sampling, which was performed in a 5 m grid over 30 m×30 m plots. At each sampling location, a 10 cm diameter section of soil to 4 cm depth in the approximate centre of the soil patch closest to the predetermined grid point was mixed and ~75 g placed in a sterile conical tube. The location of each sampling point was recorded with a Garmine eTrex Vista gps unit (Garmin International). Soil samples were stored at 4 °C for a maximum of 1 week while soil-dissolved organic carbon and total dissolved nitrogen measurements were taken. Afterwards, soils were stored at − 20 °C until processing. Soils for soil texture analysis were collected from each location in September 2008.

Sequencing and biogeochemical measurements. Microbial diversity data for the GLV samples were obtained by pyrosequencing 85 randomly selected samples (out of 160 total) (Fig. 1) for the 16S gene using the method of Fierer et al.[32] and resulting in 16,894 sequences with an average length of 230 nucleotides. Dissolved organic C/N and microbial biomass C/N were determined, and analysis of extracellular enzymes N-aceytalglucosaminase, β-glucosidase, α-glucosidase, β-xylase, cellobiosidase, leucine amino-peptidase, organic phosphatase and lignin oxidase/peroxidase was performed using the methods of Weintraub et al.[33]. Soil pH was measured after the addition of 2 ml water to 2 g soil and shaking for 1 h. Soil water content and soil water holding capacity were measured gravimetrically. Soil texture (clay/sand/silt) was measured by South Dakota Soil Laboratory (South Dakota State University, Brookings, SD, USA). Plant diversity and abundance were measured by identifying and recording all vascular plants within a 1 m radius of a sampling location. Snow depth values at each point were obtained by averaging the snowpack depth from the kriging interpolations of snow surveys in the GLV from 1997 to 2003 (Niwot LTER database, http://culter.colorado.edu/exec/Database/gis_layer_query.cgi). Degree of slope was calculated on the basis of a 10 m digital elevation model also available from the Niwot LTER website.

Phylogenetics and habitat modelling. Clades were defined by selecting all nodes on the full community tree that aggregated at least 100 sequences (12,303 sequences were obtained, an average of 145 sequences per sample). Semivariograms, correlation matrices and correlation significance tests were performed in R[34] (version 2.8.1, 22 December 2008, The R Foundation for Statistical Computing http://www.r-project.org/index.html) with the aid of the spatial statistics add-on package geoR. Semivarogram models were fit in R for a spherical model[25] using a Nelder–Mead nonlinear algorithm[35]. Mantel tests for spatial autocorrelation models[36] were evaluated from 0 m to the modelled distance of autocorrelation for each semivariogram in R using the statistics add-on package ade4 using 1,000 iterations. Moran's I-tests for spatial autocorrelation[37] in clade relative abundance were conducted in R using in spatial statistics add-on package ape. Clade habitat distribution models were constructed in the Spatial Analyses in Macroecology (SAM[38]) programme using the Akaike information criterion[39] to select environmental variables and a generalized least squares spatial partial regression to add the spatial component. Maps were generated in ArcGIS 9.3 (ESRI) using cokriging of the relative abundance for each clade on 85 sequenced samples, in combination with the three most significant environmental variables in each habitat distribution model for all 160 samples[25]. Cokriging was chosen to generate the maps of the distribution of the clades because it creates a continuous map surface using a linear least squares model similar to our SAM models and has an estimation error that is dependent on the spatial autocorrelation distance for the variable of interest (relative abundance). Thus, the error is relatively low for estimations at distances less that the clade's autocorrelation distance from sample locations[25].

Global-scale sampling and analysis. For the global biogeographical analysis, we collected six samples from each of four sites during the regional dry season using the same methods as for the main Colorado Rocky Mountains data set (Table 2). The sites were GLV, Denali National Park & Preserve, AK, USA (DNP&P); Annapurna Conservation Area, Nepal; and Llullaillaco Volcano, Argentina. Samples from outside the United States were frozen in the field, kept frozen during transportation from the field, shipped frozen through express airmail and stored at − 20 °C until processed. For each sample, DNA was extracted and Sanger clone libraries for the 16S gene were constructed according to the methods of Freeman et al.[29], resulting in 3,429 sequences. A restricted set of biogeochemical properties was measured for each site using the same methods as for the primary Colorado data set (soil water content, soil water holding capacity, all eight extracellular enzymes, forb abundance, soil pH, microbial biomass C&N, total dissolved nitrogen and dissolved organic carbon).

After the initial (Niwot) models were used to predict the relative abundance, we looked for additional predictor variables that showed a broader range of variation and had significant correlation with the residuals (predicted relative abundance—actual relative abundance) on the global scale. Clades with significant correlation between residuals and environmental variables (Acidobacteria G4 and Saprospirales) were error corrected using ordinary least squares regression to predict residual error. The reason that these variables were not adequately weighted in the original model is because the DNP&P and Annapurna soils are formed from calcareous shale bedrock, whereas the GLV (and Llullaillaco) site is formed from igneous bedrock. The shales of the DNP&P and Annapurna sites create soils with a significantly more basic pH (7.5 versus 4.5, t-test, $P < 0.001$). This difference in soil pH is known to have a significant effect on extracellular enzyme activity, particularly for leucine peptidase, which, similar to most peptidases, has its activity optimum in basic pH solutions[40] (11 versus 0.01, t-test, $P < 0.001$). In addition, DNP&P and Annapurna soils had lower water holding capacities than found in the main GLV data set (0.27 versus 0.46, t-test, $P = 0.016$). As a result, the habitat models for Acidobacteria G4 and Saprospirales are error corrected by adding leucine amino-peptidase activity, which was only appreciably active at pH > 7, as a model parameter, and reweighting the contribution of water holding capacity and soil pH. In essence, we had to broaden the range of predictability of the models once we had data across a wider range of pH values.

Relative abundances were rescaled for Figure 4 because of a bias in Sanger versus pyrosequencing that we observed in the six samples from Colorado that were analysed using both sequencing approaches (Sanger relative abundance * factor: Acidobacteria G4, *0.667; Rhizobilaes, *2.85; Rhodospirillales, *6.54; Saprospirales, *0.204; raw relative abundances are given in Supplementary Table S2). Similar effects have been observed in previous comparisons of Sanger versus pyro-sequencing, although the cause of this bias is still the subject of debate[14,41]. However, the fact that these rough rescalings enabled the accurate prediction of Sanger rela-tive abundances based on models of pryosquencing data suggests that these biases are consistent across samples and, given absolute abundance estimates derived from a method such as fluorescence *in situ* hybridization, similar types of correc-tion factors could be used in future to estimate actual abundances of microbial clades in soil samples.

For additional details on DNA extraction, sequence processing and phylogenetic determination, see Supplementary Methods.

References

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320,** 1034–1039 (2008).
2. Houghton, R. A. Balancing the global carbon budget. *Annu. Rev. Earth Planet Sci.* **35,** 313–347 (2004).
3. Ettema, C. H. & Wardle, D. A. Spatial Soil Ecology. *Trends Ecol. Evol.* **17,** 177–183 (2002).
4. Green, J. L., Bohannan, B. J. M. & Whitaker, R. J. Microbial biogeography, from taxonomy to traits. *Science* **320,** 1039–1043 (2008).
5. Martiny, J. B. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4,** 102–111 (2006).
6. Green, J. L. *et al.* Spatial scaling of microbial eukaryote diversity. *Nature* **432,** 747–750 (2004).
7. Horner-Devine, M. C., Large, M., Hughes, J. B. & Bohannan, B. J. M. A taxa-area relationship for bacteria. *Nature* **432,** 750–753 (2004).
8. Ramette, A. & Tiedje, A. M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl Acad. Sci. USA* **104,** 2761–2766 (2007).
9. Robertson, G. P. & Freckman, D. W. The spatial distribution of nematode trophic groups across a culitivated ecosysytem. *Ecology* **76,** 1425–1432 (1995).
10. Rodrigues, D. F. *et al.* Biogeography of two cold-adapted genera: *Psychrobacter* and *Exiguobacterium*. *ISME J.* **3,** 658–665 (2009).
11. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103,** 626–631 (2006).
12. Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A. & Cleveland, C. C. Global patterns in belowground communities. *Ecol. Lett.* **12,** 1238–1249 (2009).
13. Bryant, D. A. *et al.* Candidatus *Chloracidobacterium thermophilum*: an aerobic phototrophic acidobacterium. *Science* **317,** 523–526 (2007).
14. Jones, R. T. *et al.* A comprehensive survey of soil Acidobacteria diversity using pyrosequencing and clone library analyses. *ISME J.* **3,** 442–453 (2009).
15. Monson, R. K. *et al.* Winter forest soil respiration controlled by climate and microbial community composition. *Nature* **439,** 711–714 (2006).
16. Sinsabaugh, R. L., Gallo, M. E., Lauber, C., Waldrop, M. & Zak, D. R. Extracellular enzyme activities and soil carbon dynamics for northern hardwood forests receiving simulated nitrogen deposition. *Biogeochemistry* **75,** 201–215 (2005).
17. Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J. & Knight, R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* **5,** 235–237 (2008).
18. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).
19. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the under explored 'rare biosphere'. *Proc. Natl Acad. Sci. USA* **103,** 12115–12120 (2006).
20. Ley, R. E., Williams, M. W. & Schmidt, S. K. Microbial population dynamics in an extreme environment: Controlling factors in talus soils at 3750 m in the Colorado Rocky Mountains. *Biogeochemistry* **68,** 313–335 (2004).
21. Nemergut, D. R. *et al.* The effects of chronic nitrogen fertilization on alpine tundra soil microbial communities: Implications for carbon and nitrogen cycling. *Environ. Microbiol.* **10,** 3093–3105 (2008).
22. King, A. J., Meyer, A. M. & Schmidt, S. K. High levels of microbial biomass and activity in unvegetated tropical and temperate alpine soils. *Soil Biol. Biochem.* **40,** 2605–2610 (2008).
23. van Buskirk, J. Local and landscape influence on amphibian occurrence and abundance. *Ecology* **86,** 1936–1947 (2005).
24. Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Model.* **135,** 147–186 (2000).
25. Bailey, T. C. & Gatrell, A. C. *Interactive Spatial Data Analysis* (Prentice Hall, 1995).
26. Jacobson, A. D., Blum, J. D. & Walter, L. M. Reconciling the elemental and Sr isotope composition of Himalayan weathering fluxes: insights from the carbonate geochemistry of stream waters. *Geochim. Cosmochim. Acta* **66,** 3417–3429 (2002).
27. Garg, N. & Geetanjali, N., Symbiotic nitrogen fixation in legume nodules: process and signaling. A review. *Agron. Sust. Devel.* **27,** 59–68 (2006).
28. Madigan, M. T. & Jung, D. O. Systematics, physiology, and habitats, in *The Purple Phototrophic Bacteria* (eds Hunter, C. N., Daldal, F., Thurnauer, M. C. & Beatty, J. T.) 1–15 (Springer Science, 2008).
29. Freeman, K. R. *et al.* Evidence that chytrids dominate fungal communities in high-elevation soils. *Proc. Natl Acad. Sci. USA* **106,** 18315–18320 (2009).
30. Hood, E. W., Williams, M. W. & Caine, N. Yields of dissolved C, N, and P, from three high-elevation catchments, Colorado Front Range, USA. *Water Air Soil Pollut.: Focus* **2,** 165–180 (2002).
31. Legendre, P. & Fortin, M. J. Spatial pattern and ecological analysis. *Plant Ecol.* **80,** 107–138 (1989).
32. Fierer, N., Hamady, M., Lauber, C. L. & Knight, R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc. Natl Acad. Sci. USA* **105,** 17994–17999 (2008).
33. Weintraub, M. N., Scott-Denton, L. E., Schmidt, S. K. & Monson, R. K. The effects of tree rhizodeposition on soil exoenzyme activity, dissolved organic carbon, and nutrient availability in a subalpine forest ecosystem. *Oecologia* **154,** 327–338 (2007).
34. R Development Core Team. *R: A Language and Environment for Statistical Computing*, (R Foundation for Statistical Computing, 2009) http://www.r-project.org/index.html.
35. Nelder, J. A. & Mead, R. A simplex algorithm for function minimization. *Comput. J.* **7,** 308–313 (1965).
36. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27,** 209–220 (1967).
37. Anselin, L. Local indicators of spatial association LISA. *Geogr. Anal.* **27,** 93–115 (1995).
38. Rangel, T. F. L. V. B., Diniz-Filho, J. A. F. & Bini, L. M. Towards an integrated computational tool for spatial analysis in marcoecology and biogeography. *Glob. Ecol. Biogeogr.* **15,** 321–327 (2006).
39. Buckland, S. T., Burnham, K. P. & Augustin, N. H. Model selection: and integral part of inference. *Biometrics* **53,** 603–618 (1997).
40. Sinsabaugh, R. L., Hill, B. H. & Shah, J. J. F. Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment. *Nature* **462,** 795–799.
41. Morgan, J. L., Darling, A. E. & Eisen, J. A. Metagenomic sequencing of an *in vitro*-simulated microbial community. *PLOS One* **5,** e10209.

Author contributions

A.J.K. collected samples, performed biogeochemical measurements, performed bioinformatic, geospatial and statistical analysis and wrote the paper. K.R.F. and K.F.M. extracted DNA, performed PCR and prepared DNA for Sanger- and pyrosequencing. R.C.L. extracted DNA, performed PCR, prepared DNA for Sanger sequencing and performed bioinformatic analysis. C.L. and R.K. prepared pyrosequencing data and performed bioinformatic analysis. S.K.S. collected samples and wrote the paper.

Additional information

Sequences are available at the Alpine Microbial Observatory Website: http://amo.colorado.edu/database.html

Sequences have been deposited in NCBI's Genbank database under accession numbers HM780503-HM797396.

Supplementary Information accompanies this paper on http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: King, A. J. *et al.* Biogeography and habitat modelling of high-alpine bacteria. *Nat. Commun.* 1:53 doi: 10.1038/ncomms1055 (2010).