

Supplementary Tables

Supplementary Table S1: Global Sampling Sites

Range	Site	UTM coordinates (zone)	Elevation (m.a.s.l.)	Number of Sequences
Rocky Mountains	Green Lakes Valley, Colorado USA	446772E, 4434307N (13)	3740	1052 (6 samples)
Alaska Range	Denali National Park & Preserve, Alaska, USA	354260E, 7032590N (6)	1300	949 (6 samples)
Himalayas	Annapurna Conservation Area, Nepal	787000E, 3188900N (44)	5450	1040 (6 samples)
Andes	Llullaillaco, Argentina	547606E, 7266160S (19)	6330	388 (5 samples)

Sampling sites used in this study. Sequence counts are from Sanger clone libraries. UTM: Universal Transverse Mercator coordinate system; m.a.s.l.: meters above sea level.

Table S2: Sanger Clone Library Relative Abundances

	Acidobacteria G4	Rhizobiales	Rhodospirillales	Saprospirales
Niwot	4.29 (1.87)	1.24 (1.19)	0.76 (0.17)	11.84 (3.95)
Nepal	15.86 (5.87)	1.45 (1.14)	0.86 (0.79)	7.60 (5.67)
Llullaillaco	0	0	3.53 (4.03)	3.53 (2.82)
Denali NP&P	9.06 (3.78)	1.16 (1.23)	0.63 (0.69)	14.97 (3.72)

Means for raw Sanger clone-library relative abundances uncorrected for biases between Sanger and pyro sequencing. Standard deviations are in parentheses.

Supplementary Methods

DNA Extraction and Amplification for Pyrosequencing: 85 samples were randomly selected from the 160 total samples for sequencing. DNA was extracted from 1 gram of each soil sample using MO BIO Power Soil™ DNA Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA). Partial 16S gene sequences were amplified using fusion primers (Integrated DNA Technologies, Inc., Coralville, IA, USA) similar to those described by Hamady *et al.*⁴², targeted toward the 27-338 region. Each reverse primer consisted of a 454® adapter sequence, a unique 12 nt error-correcting barcode, a two base pair linker sequence, and the 338R bacterial 16S primer as described elsewhere^{33,43}.

Forward primers consisted of a 454® adapter sequence, a two base pair linker sequence, and the 27F bacterial 16S primer. Polymerase chain reactions were performed using a reaction mix consisting of 7 µl nuclease free water, 13 µl HotStarTaq® Matster Mix (Qiagen, Valencia, CA, USA), 1.5 µl of each primer (5 µM), and 3 µl DNA extract. For samples that would not amplify using this formulation, 2 µl of DNA extract and 1 µl BSA were used in place of 3 µl extract. The samples were initially denatured at 94° C for 3 minutes, followed by 35 cycles of 94° C for 45 seconds, 50° for 30 seconds, and 72° for 90 seconds. Finally, samples were held at 72° for ten minutes to promote complete amplification of the target sequence. PCR was performed in triplicate for all samples, and the PCR products for each sample were pooled prior to clean-up and normalization.

PCR products were cleaned and concentrations standardized to an approximate concentration of 1 ng/µl using the SequalPrep™ Normalization Kit (Invitrogen Corp., Frederick, MD, USA). Equal volumes (5 µl) of all samples were pooled, and 100 µl of

the pooled product was sent to the Consortium for Comparative Genomics pyrosequencing facility at the University of Colorado in Denver.

Sequence Analysis and Tree Building: The raw pyrosequencing data were preprocessed as described in Hamady *et al.*⁴², assigning each sequence to the sample it came from using an error-correcting barcode. Sequences removed from the analysis include: sequences of length < 200 nt or >300 nt, with low quality scores (<25), containing ambiguous characters, or that were from samples containing fewer than 50 sequences. The remaining sequences were clustered using cd-hit⁴³, an alignment was built from the resulting representative set of sequences using NAST⁴⁴, a tree was then built from this alignment using FastTree⁴⁵, and Fast UniFrac⁴⁶ was used to generate the UniFrac distance matrix of all pairs of samples. Taxonomy assignments were made using the RDP classifier⁴⁷ with a minimum support of 60%. For clade specific UniFrac analyses, these taxonomy assignments were used to prune all but members of the clade of interest from the full tree before applying Fast UniFrac.

Sanger Clone Library Comparisons: DNA was extracted from samples from Green Lakes Valley of the Colorado Rocky Mountains (GLV), Denali National Park & Preserve, AK, USA (DNP&P), Annapurna Conservation Area, Nepal (ACA), and Llullaillaco Volcano, Argentina. The GLV sequences were obtained from six samples collected in late July, 2006 from unvegetated soils underlying late melting snowbanks; these samples were collected at the same locations as the depth separated samples from the current study⁴⁸. The DNP sequences were obtained from six samples collected in from a glacial recession chronosequence at the Middle Fork Toklat Glacier (63°23'50"N, 150°07'50"E) at 1350 m. a. s. l. in July of 2008 and represent an average of 25 yrs

exposure. The DNP&P site is located on shale bedrock. The ANC samples were collected in October, 2008; three from a lower elevation site and three from a higher elevation site within the region. The low-elevation samples in the ACA are from unvegetated, south-east facing slopes (28°43'N, 83°55'E) at 5122 m above sea level (m.a.s.l.) and approximately 8 km east of the northern edge of Tilicho Lake. The high-elevation samples in the ACA are from unvegetated south-east facing slopes from 5503 to 5516 m.a.s.l., 1 km north of Thorong Pass (28°48'N, 83°56'E). The Thorong Pass crosses the divide separating the Marsyangdi River to the east and the Kali Gandaki River to the west. All of the ACA sites are located on shale bedrock. The Llullaillaco sequences were collected from two elevations on the Llullaillaco Volcano in Argentina (24°42'S, 68°31'E). Three samples were collected at 6034 m.a.s.l. and two samples were collected at 6330 m.a.s.l. The Llullaillaco sites are located on basalt bedrock.

DNA was extracted from the soils using the MO BIO Ultra-Clean Mega Prep Soil DNA kit (MO BIO Laboratories, Carlsbad, CA, USA). Bacterial community SSUrDNAs were amplified using the *Bacteria*-specific primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1391R (5'-GACGGGCGGTGWGTRCA-3') and primers 4Fa (5'-TCCGGTTGATCCTGCCRG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). PCRs were performed with 2.5 mM MgCl₂, 0.2 mM each deoxynucleoside triphosphate (dNTP), 0.4 mM each primer, 1 U Taq polymerase (Promega) and buffer supplied with the enzyme using a range of template concentrations. Gradient thermal cycling was carried out for 25 cycles to minimize PCR bias. Amplicons from six different reactions, with different annealing temperatures ($\pm 3^{\circ}\text{C}$ of optimal), were pooled and gel purified using isolated bands from agarose gels and QIAquick gel

purification columns (Qiagen, Valencia, CA, USA). Purified products were ligated into the pCR 2.1 vector (Invitrogen, Carlsbad, CA, USA) and transformed into *Escherichia coli* following the manufacturer's instructions. Transformants were inoculated into a 96-well deep-dish plate containing 1.5 ml of TB Dry nutrient broth (MO BIO Laboratories, Carlsbad, CA, USA). Cultures were shaken at 200 r.p.m. for 16 h at 37°C and then centrifuged. Cell pellets were sent to Functional Biosciences (Madison, WI, USA) for plasmid extraction and were sequenced bidirectionally using sequencing primers T7 and M13R. For each sample, a clone library of 192 sequences was constructed.

Sequences were edited (vector-trimmed and assembled into contigs) using Sequencher 4.1 (Gene Codes, Ann Arbor, MI, USA). Bacterial sequences were aligned using the NAST alignment tool⁴³ and preliminary classifications were obtained using the Greengenes Database⁴⁹. This alignment was then imported into a 16S rRNAARB database⁵⁰. All sequences were chimera checked using Bellerophon⁵¹ and Mallard⁵².

Supplementary References

42. 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).
43. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659 (2006).
44. DeSantis, T. Z. *et al.*, NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* 34, W394-399 (2006).
45. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* doi:10.1093/molbev/msp077 (2009)

46. Hamady, M., Lozupone, C. & Knight, R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* 4, 17-27 (2010).
47. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261 (2007).
48. Freeman, K. R. *et al.*, Soil CO₂ flux and photoautotrophic community composition in high-elevation, 'barren soil.' *Environ. Microbiol.* 11, 674-686 (2009).
49. DeSantis, T. Z. *et al.* Greengenes, a chimerachecked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072 (2006).
50. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371 (2004).
51. Huber, T., Faulkner, G. & Hugenholtz, P. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20, 2317–2319 (2004).
52. Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., & Weightman, A. J. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ Microbiol.* 72: 5734–5741 (2006).