

Cross-biome metagenomic analyses of soil microbial communities and their functional attributes

Noah Fierer^{a,b,1}, Jonathan W. Leff^b, Byron J. Adams^c, Uffe N. Nielsen^d, Scott Thomas Bates^b, Christian L. Lauber^b, Sarah Owens^{e,f}, Jack A. Gilbert^{e,g}, Diana H. Wall^h, and J. Gregory Caporaso^{e,i}

^aDepartment of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309; ^bCooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309; ^cDepartment of Biology and Evolutionary Ecology Laboratories, Brigham Young University, Provo, UT 84602; ^dHawkesbury Institute for the Environment and School of Science and Health, University of Western Sydney, Penrith, NSW 2751, Australia; ^eInstitute of Genomic and Systems Biology, Argonne National Laboratory, Argonne, IL 60439; ^fComputation Institute, University of Chicago, Chicago, IL 60637; ^gDepartment of Ecology and Evolution, University of Chicago, Chicago, IL 60637; ^hDepartment of Biology and School of Global Environmental Sustainability, Colorado State University, Fort Collins, CO 80523; and ⁱDepartment of Computer Science, Northern Arizona University, Flagstaff, AZ 86011

Edited by Peter M. Vitousek, Stanford University, Stanford, CA, and approved November 12, 2012 (received for review September 5, 2012)

For centuries ecologists have studied how the diversity and functional traits of plant and animal communities vary across biomes. In contrast, we have only just begun exploring similar questions for soil microbial communities despite soil microbes being the dominant engines of biogeochemical cycles and a major pool of living biomass in terrestrial ecosystems. We used metagenomic sequencing to compare the composition and functional attributes of 16 soil microbial communities collected from cold deserts, hot deserts, forests, grasslands, and tundra. Those communities found in plant-free cold desert soils typically had the lowest levels of functional diversity (diversity of protein-coding gene categories) and the lowest levels of phylogenetic and taxonomic diversity. Across all soils, functional beta diversity was strongly correlated with taxonomic and phylogenetic beta diversity; the desert microbial communities were clearly distinct from the nondesert communities regardless of the metric used. The desert communities had higher relative abundances of genes associated with osmoregulation and dormancy, but lower relative abundances of genes associated with nutrient cycling and the catabolism of plant-derived organic compounds. Antibiotic resistance genes were consistently threefold less abundant in the desert soils than in the nondesert soils, suggesting that abiotic conditions, not competitive interactions, are more important in shaping the desert microbial communities. As the most comprehensive survey of soil taxonomic, phylogenetic, and functional diversity to date, this study demonstrates that metagenomic approaches can be used to build a predictive understanding of how microbial diversity and function vary across terrestrial biomes.

shotgun metagenomics | soil microbial ecology | 16S rRNA gene sequencing | biogeography

Soil microorganisms play critical roles in regulating soil fertility, plant health, and the cycling of carbon, nitrogen, and other nutrients. Every gram of soil harbors thousands of bacterial, archaeal, and eukaryotic taxa, and this taxonomic diversity is mirrored by the diversity of their protein-encoded functions, encompassing a seemingly limitless array of physiologies and life history strategies. Although these characteristics of soil microbial communities have been known for decades, the ongoing development of high-throughput molecular tools (and the tools necessary to analyze the associated flood of data) allow microbial ecologists to characterize the taxonomic, phylogenetic, and functional diversity of soil microbial communities to an extent that was unimaginable only a few years ago. We can now move beyond detailed studies of individual soils to conduct detailed comparative studies of soils across broad spatial gradients.

Perhaps the most dramatic and well-studied spatial gradients in biological diversity are those that exist across the major global terrestrial biomes. Different biomes typically harbor distinct assemblages of macrobial (plant and animal) taxa and ecologists have spent many decades describing the apparent differences in biological diversity. Although comparable research on the

biogeographical patterns exhibited by microbial taxa has lagged far behind research on plant and animal communities (1), we are beginning to understand how soil microbial diversity varies across the globe and how this diversity is related to the physical, chemical, and biological characteristics of ecosystems. In particular, we now know that soil bacterial communities are strongly influenced by pH, which explains a large proportion of the variance in soil bacterial diversity and community composition at local (2, 3), regional (4–6), and continental scales (7). Soils with near-neutral pH typically have higher bacterial diversity than more acidic or more basic soils and the relative abundances of many bacterial phyla have been shown to be strongly correlated with soil pH (7). Of course, soil pH is not the only factor that can influence bacterial communities and there is evidence that other microbial taxa that are abundant in soil (including Archaea, fungi, and protists) do not necessarily exhibit the same biogeographical patterns observed for bacteria (2, 8). Changes in the types and quantities of organic carbon added to soil can have considerable influences on soil microbial communities (9, 10) and, depending on the gradients being studied or the experimental treatments imposed, other factors such as soil temperature, moisture, and nutrient availability have also been shown to influence microbial structure in soil.

Although our understanding of the phylogenetic and taxonomic biogeography of soil microbial communities continues to expand, there has been limited progress in understanding how the functional capabilities of soil microbial communities change across biomes. For individual well-studied soil microbial processes (e.g., N₂ fixation) (11) or specific extracellular enzymes (12), researchers have been able to document their interbiome characteristics. Likewise, previous work has demonstrated how specific functional groups or gene categories can vary across space (e.g., ref. 13). However, we lack an integrated understanding of how the functional genes encoded in their collective genomes act to structure communities across environmental gradients. Although we might expect an overall correlation between taxonomic composition and the functional attributes of soil microbial communities, this may not always be the case as distinct taxa can share specific functional attributes and closely related taxa may have very different physiologies and environmental tolerances (14). As

Author contributions: N.F., B.J.A., U.N.N., and D.H.W. designed research; N.F. and J.G.C. performed research; U.N.N., S.O., and J.A.G. contributed new reagents/analytic tools; N.F., J.W.L., S.T.B., C.L.L., and J.G.C. analyzed data; and N.F., J.W.L., B.J.A., U.N.N., S.T.B., C.L.L., S.O., J.A.G., D.H.W., and J.G.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Rapid Annotation using Subsystems Technology for Metagenomes database (MG-RAST). Accession numbers are listed in Table S2.

¹To whom correspondence should be addressed. E-mail: noah.fierer@colorado.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1215210110/-DCSupplemental.

a result, we cannot rely entirely on our understanding of the biogeographical patterns in the taxonomic or phylogenetic structure of soil microbial communities to predict the functional attributes or the functional diversity of these communities (15). Using shotgun metagenomic sequencing, the direct sequencing of the collective genomes found in a given environmental sample, researchers have gained important insight into the potential functions of microbial communities from individual soil types (16, 17) and soils that have been experimentally manipulated in the laboratory (18). The value of shotgun metagenomic analyses is that a far more comprehensive understanding of the traits microbes may use to survive in an individual soil can be identified, traits which are often very difficult to measure using biogeochemical or culture-based approaches (19). To our knowledge, such tools have not yet been used to directly compare microbial metagenomes across soils representing a range of different biomes.

The current study was designed to test the hypothesis that the microbial communities found in desert soils are taxonomically and functionally distinct compared with those found in other biomes and that the variability across different desert sites is less than the variability between desert and nondesert biomes. This was predicated on the fact that at least three of the main factors known to shape the composition of soil microbial communities: pH, moisture availability, and inputs of plant-derived organic carbon, are often very different between desert and nondesert soils. Desert soils are drier, typically have higher pH soils than other biomes, and the paucity (or complete absence) of plant biomass reduces the inputs of organic carbon. We used shotgun metagenomic sequencing of soils from 16 sites representing a wide range of ecosystems (forests, grassland, tundra, and deserts) to determine how the functional capabilities of soil microbial communities vary across the major global terrestrial biomes and the extent to which these capabilities are predictable. We addressed three basic questions: Do deserts (both cold and hot) harbor microbial communities that are taxonomically, phylogenetically, and functionally distinct from those found in forests, grasslands, and tundra? What functional attributes distinguish desert and nondesert soil microbial communities? Can we use information on the taxonomic diversity and composition of soil microbial communities to predict their functional attributes?

Results and Discussion

General Characteristics of the Soil Microbial Communities. Soils were collected from 16 sites: 3 from hot deserts, 6 from Antarctic cold deserts, and 7 from temperate and tropical forests, a prairie grassland, a tundra, and a boreal forest (Table S1). The sites were selected to span a wide range of ecologically distinct biomes to examine how cold desert soils compare with hot deserts, and to forests, prairie, and tundra. Using a shotgun metagenomic approach, we obtained a total of 3.9–11 million 100-bp sequences per sample (390–1,100 Mbp per sample). Only 13–23% of the sequences (688,000–1,900,000 reads per sample) could be annotated using the technique applied (Table S2), a percentage similar to that reported in previous studies that used shotgun metagenomic sequencing to characterize soil microbial communities (17, 20) and communities in other highly diverse microbial habitats (21, 22). As survey depth can affect estimation of the relative abundances of gene categories, all of the shotgun metagenomic datasets were rarefied by randomly subsampling 688,000 annotated reads per sample before downstream analyses.

The majority of the shotgun metagenomic reads were derived from bacteria, as shown by the analysis of both the large-subunit (LSU) and small-subunit (SSU) rRNA reads recovered from the metagenomic data. Between 74% and 96% of either the SSU or the LSU reads were assigned to bacterial or archaeal taxa (Table S3). Although fungi and other eukaryotes can represent a large portion of the microbial biomass contained within soils, their representation in the metagenomic data was low. A similar pattern has been observed in comparable shotgun metagenomic datasets obtained from other soils (17, 20) and is most likely a product of many eukaryotic taxa (including fungi), having a far

lower ratio of rRNA gene copies per unit biomass than bacterial cells. However, the ratio of fungal:bacterial rRNA reads did vary across soils in this study, with temperate and boreal forests having the highest fungal:bacterial ratios (Table S3), a pattern similar to that noted previously (23).

An amplicon survey of a portion of the 16S rRNA gene was performed to provide a higher resolution and more in-depth analysis of the composition and diversity of the soil bacterial communities. We used barcoded primers that target the V4 region of the 16S rRNA gene from both bacteria and Archaea with the resulting amplicons sequenced on the Illumina HiSeq platform (24). All samples were compared at an equivalent sequencing depth of 118,000 randomly selected 16S rRNA gene amplicons per sample. These results show that all of the communities were dominated by *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, and *Verrucomicrobia* (Fig. S1), bacterial phyla that are known to be relatively abundant and ubiquitous in soil (25). Additional phyla including *Chloroflexi*, *Cyanobacteria*, *Firmicutes*, and *Gemmatimonadetes* were also found in nearly all soils, but their relative abundances were highly variable and typically represented less than 5% of the 16S rRNA reads in any individual soil (although *Cyanobacteria* were more abundant in some of the desert soils, Fig. S1). Archaea were relatively rare in all soils (0.01–6.7% of reads) but were most abundant in the three hot desert sites and one of the tropical rainforest sites (Fig. S1). The observed range in archaeal abundances and the observation that *Thaumarchaeota* were the dominant archaeal group in nearly all of the soils validates results reported previously (8).

Comparison of Community Structure Determined via the Amplicon and Shotgun-Metagenomic Approaches. The 16S rRNA gene data obtained from both the amplicon and shotgun sequencing were used to directly compare the taxonomic results obtained using these two very different methods. We conducted this comparison to determine whether biases introduced by both approaches may influence the determination of bacterial community structure, as suggested previously (26). Such biases may be derived from the PCR process itself or because the 16S rRNA gene regions recovered from the metagenomic data can span the entire length of the gene, whereas the PCR-based amplicon approach only targets the V4 region. Because different regions of the 16S rRNA gene vary in the accuracy of their taxonomic assignments (27), the two approaches may not necessarily give identical results. However, this was not the case; the two methods generated nearly identical estimates of bacterial community composition. This is evident from the strong correlation between the Bray–Curtis distance matrices (Spearman $r = 0.91$, $P < 0.001$), and by directly comparing the relative abundances of the dominant taxa (Fig. S2). In addition, although the 16S rRNA amplicon dataset contained orders of magnitude more of the 16S rRNA reads than the shotgun-derived 16S rRNA dataset (118,000 and 1,884 reads per sample, respectively), the estimates of taxonomic richness for each dataset were significantly correlated ($r^2 = 0.81$, $P < 0.001$). The strong concordance between these two very different approaches suggests that, at least across the wide range of soils examined here, the two methods yield nearly identical estimates of the overall differences in soil bacterial community diversity and composition.

Alpha Diversity Patterns. Alpha diversity, the richness and/or evenness of taxa or lineages contained within an individual community, was highly variable across the 16 soils (Fig. 1). The richness of the bacterial and archaeal communities ranged from <4,000 to >12,000 phylotypes per sample (Table S2) with all samples compared at an identical sequencing depth. The cold desert soils harbored far lower diversity than the other soils regardless of the taxonomic or phylogenetic metric used (Fig. 1 and Table S2). This trend is similar to that observed for invertebrates, with soils from the McMurdo Dry Valleys in Antarctica having very low levels of invertebrate diversity and extremely simple food webs (28). Although it has been reported that these cold desert soils harbor surprisingly high levels of bacterial diversity (29), we find

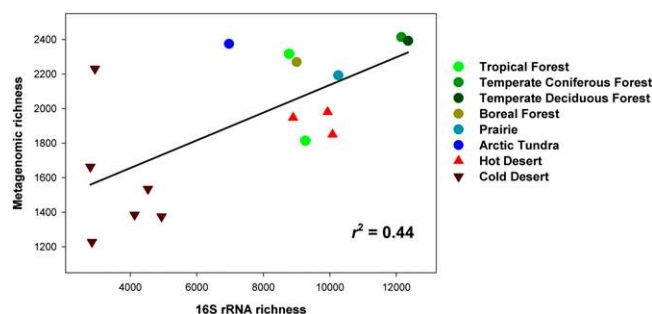


Fig. 1. Differences in alpha diversity levels across 16 soils. X axis shows taxonomic richness of the bacterial communities (number of phylotypes out of 118,000 amplicon reads per sample). Y axis shows the functional gene richness (number of functional gene categories identified from 688,000 annotated shotgun metagenomic reads per sample). See Table S2 for additional information on diversity levels across the 16 soils.

that their diversity is actually far lower, on average, than that found in other biome types.

As has been demonstrated previously (7), soil pH is a reasonably good predictor of prokaryotic diversity across the 16 soils ($y = -533x^2 + 6,371x - 8,823$; $r^2 = 0.6$, where x = soil pH and y = phylotype richness). Soils close to neutral had the highest diversity levels, whereas soils that were either very basic (the desert soils) or acidic (the Peruvian tropical rainforest soil and the Arctic tundra soil) had lower levels of diversity. As this study only included 16 soils that differ in a wide variety of ways, we cannot use this sample set to definitively identify the edaphic or site factors responsible for the diversity patterns observed here—indeed, there are many possible reasons why these soils harbor such different levels of bacterial diversity. For example, it is possible that the low diversity of the cold desert soils is not directly related to their very high pH levels, but rather due to their high salinities, negligible plant-carbon inputs, or the extreme moisture and temperature conditions encountered at those sites (29–31).

Although functional alpha diversity is less frequently measured, it is increasingly common for both macrobial ecologists (32, 33) and microbial ecologists (15, 34) to consider the diversity and distributions of functional traits (or functional genes) across communities. Functional diversity (the richness of protein-coding gene categories identified out of 688,000 reads per metagenome, Fig. 1) was typically lowest in the cold desert soils, intermediate in the hot desert soils, and highest in the nondesert soils (a pattern unrelated to the percentage of reads that could be annotated from each soil, Table S2). However, there was notable variation within these broadly defined categories. For example, one of the cold desert soils (EB026) had far higher functional diversity than the other cold desert soils. This is likely a result of that soil having a broader array of genes associated with photosynthesis and carbon-fixation pathways than the other cold desert soils, as evidenced from both the metagenomic data (Fig. S3) and from the higher abundances of *Cyanobacteria* in that soil compared with the other soils (Fig. S1).

There were significant correlations between functional diversity and both the taxonomic (Fig. 1) and phylogenetic diversity of the bacterial communities ($P < 0.001$ in both cases), with the cold desert soils consistently harboring the lowest levels of diversity. This finding highlights that the overall diversity of functional gene categories found in a given sample is, to some degree, predictable from the taxonomic or phylogenetic diversity of the microbial communities. A similar pattern has been observed in other studies of microbial communities (16, 35, 36), demonstrating that functional redundancy at the genomic level is not so pervasive as to obscure any relationship between these very different metrics of diversity. However, the correlations between functional diversity and taxonomic or phylogenetic diversity were largely driven by the cold desert soils and were not significant

when the cold desert soils were omitted from the analyses ($r^2 < 0.2$, $P > 0.1$ in both cases). This suggests that functional diversity is not necessarily predictable from the taxonomic or phylogenetic diversity of communities when comparing vegetated soils. Likewise, it is worth noting that one of the samples with the highest levels of metagenomic richness (the cold desert soil EB026) had nearly the lowest level of taxonomic richness (Table S2), suggesting that the types of taxa found in a community are also important to consider when trying to predict functional diversity. Although it is often observed that macrobial communities with lower taxonomic or phylogenetic diversity have reduced functional diversity (32, 33), this paradigm does not necessarily hold true for microbial communities.

Beta Diversity Patterns—Bacterial Community Composition. Biome-specific differences between the 16 soil communities were evident from the 16S rRNA amplicon data (Fig. S1 and Fig. 2). The desert soils harbored communities that clustered apart from the nondesert communities when community differences were measured using either a taxonomic metric (Bray–Curtis distance, Fig. 2) or a phylogenetic metric (unweighted Unifrac) with both metrics yielding nearly identical patterns. The hot desert and cold desert soils were taxonomically (Bray–Curtis analysis of similarity, ANOSIM $R = 0.91$ and 0.89 , respectively, $P < 0.002$ in both cases) and phylogenetically (unweighted Unifrac ANOSIM $R = 0.98$ and 0.89 , respectively, $P < 0.005$ in both cases) distinct from those found in the nondesert soils. Also, whereas the cold and hot desert soil communities were distinct (unweighted Unifrac ANOSIM $R = 0.65$, $P = 0.01$), these differences were less than the differences between the desert and nondesert soils. Although different biomes clearly harbor distinct bacterial communities (Fig. 2), the largest distinction was between the desert and nondesert biomes with the cold and hot desert soils harboring relatively similar bacterial communities.

The general taxonomic patterns evident in Fig. 2 are largely driven by differences in the abundances of major taxonomic

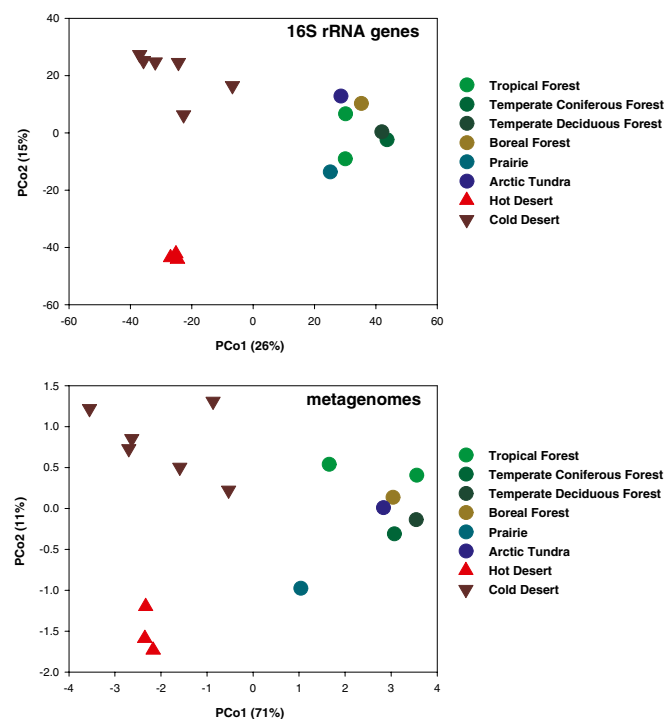


Fig. 2. Ordination plots derived from principal coordinates analyses of Bray–Curtis distances between bacterial community composition (Upper, based on amplicon 16S rRNA gene data) and metagenome composition (Lower, based on annotated shotgun metagenomic data).

(averaging 4.8% of the annotated reads, Fig. S4). Likewise, murein hydrolases, which cleave bacterial peptidoglycan and are frequently associated with bacterial cell lysis (43), were consistently more abundant in nondesert soils than in the desert soils (Fig. S4). We hypothesize that these patterns reflect reduced microbial competition in the desert soils. The production of antibiotics and resistance to antibiotics are traits that are widespread among soil bacteria and fungi (44, 45), with both models and experimental data suggesting that elevated microbe–microbe competition should select for increased antibiotic production and resistance (46–48). Likewise, murein hydrolase production has been linked to antagonistic interactions between microbes and a range of antimicrobial defenses (43). In the desert soils, where conditions are less conducive to microbial growth, adaptations that enhance microbial competition may be less important than adaptations that allow for persistence of cells under adverse environmental conditions or the ability to respond rapidly to pulses in moisture availability. Although additional work is required to verify this hypothesis, our results do suggest that the intensity of competitive interactions within microbial communities varies as a function of environmental conditions, a phenomenon that has frequently been observed in plant and animal communities (49, 50).

Only three major gene categories were significantly different in abundance between the cold and hot desert communities (Fig. S5): genes associated with the metabolism of carbohydrates and aromatic compounds being relatively more abundant in the hot desert soils. This pattern is further supported by the determination of functional gene abundances at a higher level of resolution (Fig. S6) as we found genes associated with monosaccharide utilization, carbohydrate transporters, and aromatic compound catabolism to be relatively more abundant in the hot desert soils. As described above, these functional differences are likely linked to differences in the quantity or quality of plant-carbon inputs, because the hot desert soil microbial communities likely receive far more plant-derived carbon than the cold deserts where plants are absent.

Caveats. The shotgun metagenomic results presented above should be considered carefully given that the technique has clear limitations. First, with only 688,000 annotated metagenomic reads per sample, we have not captured the full extent of the genomic diversity contained within individual samples and deeper sequencing would have allowed us to describe changes in the relative abundances of rarer (yet potentially important) genes or gene categories. Nevertheless, we were still able to detect clear differences across biomes suggesting that, for certain questions, shallower sequencing of many samples may be more useful than deeper sequencing of fewer samples (51). Second, only 13–23% of the sequence reads in this study could be annotated; the genomes of many important soil taxa have not been sequenced and even fewer have been appropriately annotated. We are invariably misannotating genes or ignoring genes that may have important functions or may account for key differences across biomes, a problem that plagues every study that uses shotgun metagenomic analyses (52). Third, even though this study represents one of the largest cross-site terrestrial metagenomic surveys conducted to date, we recognize that the sites sampled here do not necessarily represent each of the biomes in question and that even more samples are required to adequately assess intrabiome variability. However, given the strength of the patterns observed, particularly the clear separation between desert and nondesert soils, we suspect that more comprehensive analyses will further confirm the general patterns observed here. Finally, we only examined soils at a single time point per site, as it was not our goal to also quantify the temporal variability in the soil metagenomes. Although the metagenomes are unlikely to be static over time, previous work has demonstrated that the temporal variability in the composition of soil bacterial communities is typically far lower than the spatial variability (53–55) so we would expect the general patterns observed here to persist across seasons.

Conclusions. This study represents one of the most comprehensive analyses of soil metagenomes conducted to date with >1.2 Mbp of 16S rRNA gene data and >390 Mbp shotgun metagenomic data obtained from each of 16 soils (~20 Mbp of 16S rRNA data, and 6.2 Gbp of shotgun metagenomic data). However, even at these sequencing depths, we have not surveyed the full extent of microbial taxonomic, phylogenetic, or functional diversity found within individual soil samples and, with only 16 soils, we have not described the full range of soil microbial community types found across the globe. Nevertheless, we were still able to detect strong patterns in the datasets that highlight the predictability of soil microbial community attributes across biomes. Like plant and animal communities, the diversity and relative abundances of major soil microbial taxa and functional gene categories can be related to broad-scale gradients in biotic and abiotic characteristics. Functional diversity was significantly correlated with phylogenetic and taxonomic diversity across the 16 soils, but these patterns were driven by the very low levels of diversity observed in the soils from the cold desert sites. The microbial metagenomes obtained from the cold and hot desert soils were relatively similar to one another, suggesting that the composition and functional attributes of the microbial communities in these two desert types may be more comparable than often assumed (56). The metagenomes recovered from the nondesert soils clustered together apart from the desert soils even though they represented a wide range of biomes that included tropical forests, tundra, and a prairie.

Microbial ecology continues to lag far behind plant and animal ecology in our ability to resolve large-scale biogeographical patterns in diversity, community composition, and functional attributes. However, this work highlights how coupling metagenomic analyses with extensive cross-site sampling efforts can reduce this disparity. As sequencing capacities continue to increase and tools for analyzing the resulting data become more effective, we will soon be able to expand upon the work presented here and gain a more comprehensive understanding of how soil microbial communities vary across time and space.

Materials and Methods

Additional information on sample collection and analytical methods is provided in *SI Materials and Methods*.

The cold desert soils were collected from various sites with the McMurdo Dry Valleys region of Antarctica with the hot desert soils collected from sites in the southwestern United States. The seven “nondesert” soils were collected from tropical forests in Peru and Argentina, an arctic tundra in Alaska, a native tallgrass prairie in Kansas, a temperate deciduous forest in South Carolina, a temperate coniferous forest in North Carolina, and a boreal forest in Alaska (Table S1). For both the 16S rRNA gene analyses and the shotgun metagenomic analyses DNA was extracted from each soil sample using the approach described in Fierer et al. (20). To determine the diversity and composition of the bacterial communities in each of these soils, we used the PCR-based protocol described in Caporaso et al. (24) that targets the V4–V5 region of the 16S rRNA gene. Amplicon sequencing was conducted on an Illumina HiSeq2000 with processing of the reads conducted as described in Caporaso et al. (57). For all downstream analyses, we rarefied to 118,000 randomly selected reads per sample to correct for differences in sequencing depth. Reads were assigned to phylotypes at the $\geq 97\%$ sequence similarity level using the open-reference phylotype picking protocol in QIIME (58). Shotgun metagenomic analyses were conducted on the soil DNA extracts following the Illumina Paired-End Prep kit protocol with sequencing performed using a 2×100 bp sequencing run on the Illumina GAIIx. Sequences were uploaded to MG-RAST (59) for downstream analyses and data accession numbers are provided in Table S2. Sequences were annotated to functional categories against the MSNR database using BLASTX at an e-value cutoff of 1×10^{-2} and the SEED subsystems hierarchy. Downstream analyses were performed on the metagenomes evenly sampled at random to 688,000 annotated reads per sample.

ACKNOWLEDGMENTS. We thank Jessica Henley and Donna Berg-Lyons for their assistance with the molecular analyses. This work was funded by a Department of Agriculture Grant 2008-34158-04713 and National Science Foundation (NSF) Grant DEB-0953331 (to N.F.). Funding for the cold desert research was provided in part by the NSF McMurdo Dry Valleys Long-Term Ecological Research Program Award OPP-0423595 (to D.H.W. and B.J.A.). The Department of Energy supported J.A.G. and S.O. under Contract DE-AC02-06CH11357.

1. Fierer N, Ladau J (2012) Predicting microbial distributions in space and time. *Nat Methods* 9(6):549–551.
2. Rousk J, et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* 4(10):1340–1351.
3. Osborne CA, Zwart AB, Broadhurst LM, Young AG, Richardson AE (2011) The influence of sampling strategies and spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiol Ecol* 78(1):70–79.
4. Chu H, et al. (2010) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ Microbiol* 12(11):2998–3006.
5. Kuramae EE, et al. (2012) Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol Ecol* 79(1):12–24.
6. Griffiths RI, et al. (2011) The bacterial biogeography of British soils. *Environ Microbiol* 13(6):1642–1654.
7. Lauber C, Knight R, Hamady M, Fierer N (2009) Soil pH as a predictor of soil bacterial community structure at the continental scale: A pyrosequencing-based assessment. *Appl Environ Microbiol* 75:5111–5120.
8. Bates S, Caporaso JG, Walters WA, Knight R, Fierer N (2011) A global-scale survey of archaeal abundance and diversity in soils. *ISME J* 5:908–917.
9. Goldfarb KC, et al. (2011) Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Front Microbiol* 2:94.
10. Fierer N, Bradford MA, Jackson RB (2007) Toward an ecological classification of soil bacteria. *Ecology* 88(6):1354–1364.
11. Cleveland CC, et al. (1999) Global patterns of terrestrial biological nitrogen (N_2) fixation in natural ecosystems. *Global Biogeochem Cycles* 13:623–645.
12. Sinsabaugh RL, et al. (2008) Stoichiometry of soil enzyme activity at global scale. *Ecol Lett* 11(11):1252–1264.
13. Bru D, et al. (2011) Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J* 5(3):532–542.
14. Philippot L, et al. (2010) The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* 8(7):523–529.
15. Green JL, Bohannan BJM, Whitaker RJ (2008) Microbial biogeography: From taxonomy to traits. *Science* 320(5879):1039–1043.
16. Tringe SG, et al. (2005) Comparative metagenomics of microbial communities. *Science* 308(5721):554–557.
17. Delmont TO, et al. (2012) Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* 6(9):1677–1687.
18. Mackelprang R, et al. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480(7377):368–371.
19. Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 6(11):805–814.
20. Fierer N, et al. (2012) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* 6(5):1007–1017.
21. Quaiser A, Zivanovic Y, Moreira D, López-García P (2011) Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J* 5(2):285–304.
22. Dinsdale EA, et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* 452(7187):629–632.
23. Fierer N, Strickland MS, Liptzin D, Bradford MA, Cleveland CC (2009) Global patterns in belowground communities. *Ecol Lett* 12(11):1238–1249.
24. Caporaso JG, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8):1621–1624.
25. Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* 72(3):1719–1728.
26. Steven B, Gallegos-Graves LV, Starkenburg SR, Chain PS, Kuske CR (2012) Targeted and shotgun metagenomic approaches provide different descriptions of dryland soil microbial communities in a manipulated field study. *Environ Microbiol Rep* 4:248–256.
27. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* 35(18):e120.
28. Wall DH, Virginia RA (1999) Controls on soil biodiversity: Insights from extreme environments. *Appl Soil Ecol* 13:137–150.
29. Cary SC, McDonald IR, Barrett JE, Cowan DA (2010) On the rocks: The microbiology of Antarctic Dry Valley soils. *Nat Rev Microbiol* 8(2):129–138.
30. Lee CK, Barbier BA, Bottos EM, McDonald IR, Cary SC (2012) The Inter-Valley Soil Comparative Survey: The ecology of Dry Valley edaphic microbial communities. *ISME J* 6(5):1046–1057.
31. Wall D (2005) Biodiversity and ecosystem functioning in terrestrial habitats of Antarctica. *Antarct Sci* 17:523–531.
32. Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. *Ecol Lett* 5:402–411.
33. Diaz S, Cabido M (2001) Vive la difference: Plant functional diversity matters to ecosystem processes. *Trends Ecol Evol* 16:646–655.
34. Raes J, Letunic I, Yamada T, Jensen LJ, Bork P (2011) Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol Syst Biol* 7:473.
35. Gilbert JA, et al. (2010) The taxonomic and functional diversity of microbes at a temperate coastal site: A ‘multi-omic’ study of seasonal and diel temporal variation. *PLoS ONE* 5(11):e15545.
36. Bryant JA, Stewart FJ, Eppley JM, DeLong EF (2012) Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone. *Ecology* 93(7):1659–1673.
37. Pointing SB, et al. (2009) Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci USA* 106(47):19964–19969.
38. Jones RT, et al. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* 3(4):442–453.
39. Harris R (1981) Effect of water potential on microbial growth and activity. *Water Potential Relations in Soil Microbiology*, eds Parr J, Gardner W, Elliott L (Soil Science Society of America, Madison, WI), pp 23–95.
40. Muegge BD, et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332(6032):970–974.
41. Manzoni S, Schimel JP, Porporato A (2012) Responses of soil microbial communities to water stress: Results from a meta-analysis. *Ecology* 93(4):930–938.
42. Grandy A, Neff J (2008) Molecular C dynamics downstream: The biochemical decomposition sequence and its impact on soil organic matter structure and function. *Sci Total Environ* 404:221–446.
43. Vollmer W, Joris B, Charlier P, Foster S (2008) Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev* 32(2):259–286.
44. Allen HK, et al. (2010) Call of the wild: Antibiotic resistance genes in natural environments. *Nat Rev Microbiol* 8(4):251–259.
45. Dantas G, Sommer MOA, Oluwasegun RD, Church GM (2008) Bacteria subsisting on antibiotics. *Science* 320(5872):100–103.
46. Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* 74(3):417–433.
47. Williams ST, Vickers JC (1986) The ecology of antibiotic production. *Microb Ecol* 12:43–52.
48. Czárán TL, Hoekstra RF, Pagie L (2002) Chemical warfare between microbes promotes biodiversity. *Proc Natl Acad Sci USA* 99(2):786–790.
49. Wiens JA (1977) Competition and variable environments. *Am Sci* 65:590–597.
50. Chesson P, Huntly N (1997) The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *Am Nat* 150(5):519–553.
51. Knight R, et al. (2012) Unlocking the potential of metagenomics through replicated experimental design. *Nat Biotechnol* 30(6):513–520.
52. Gilbert JA, Meyer F, Bailey MJ (2011) The future of microbial metagenomics (or is ignorance bliss?). *ISME J* 5(5):777–779.
53. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103(3):626–631.
54. Wallenstein MD, McMahon S, Schimel J (2007) Bacterial and fungal community structure in Arctic tundra tussock and shrub soils. *FEMS Microbiol Ecol* 59(2):428–435.
55. Krave AS, et al. (2002) Stratification and seasonal stability of diverse bacterial communities in a *Pinus merkusii* (pine) forest soil in central Java, Indonesia. *Environ Microbiol* 4(6):361–373.
56. Pointing SB, Belnap J (2012) Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* 10(8):551–562.
57. Caporaso JG, et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 108(Suppl 1):4516–4522.
58. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
59. Meyer F, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386.

Supporting Information

Fierer et al. 10.1073/pnas.1215210110

SI Materials and Methods

Sample Collection. The cold desert soils were collected from various sites with the McMurdo Dry Valleys region of Antarctica with the hot desert soils collected from sites in the southwestern United States. The seven “nondesert” soils were collected from tropical forests in Peru and Argentina, an arctic tundra in Alaska, a native tallgrass prairie in Kansas, a temperate deciduous forest in South Carolina, a temperate coniferous forest in North Carolina, and a boreal forest in Alaska (see Table S1 for additional information on each of the collected soils). We did not analyze replicate samples from individual sites as it was not our objective to assess intrasite variability, but rather to assess the general patterns structuring soil microbial communities across major biome types. At each site, the upper 5 cm of mineral soil was collected from 5 to 10 locations within a given plot of $\sim 100\text{ m}^2$ and composited into a single bulk sample. For those sites with plants, soils were collected during the peak of the plant growing season. All soils were stored at -20°C immediately after collection.

DNA Extractions. For both the 16S rRNA gene analyses and the shotgun metagenomic analyses DNA was extracted from each soil sample using the approach described in Fierer et al. (1). Briefly, 10 g of soil from each sample was homogenized in a mortar and pestle with liquid N_2 , and DNA was extracted from 0.25-g subsamples of each soil using the MoBio PowerSoil DNA extraction kit modified with an additional incubation step at 65°C for 10 min followed by 2 min of bead beating (2). To obtain sufficient DNA for the shotgun metagenomic analyses, we conducted 4–12 replicate extractions per soil, with the replicates pooled together using the approach described previously (1).

16S rRNA Gene Analyses via Amplicon Sequencing. To determine the diversity and composition of the bacterial communities in each of these soils, we used the protocol described in Caporaso et al. (3). PCR amplifications were conducted in triplicate reactions for each of the 16 soil samples with the 515f/806r primer set that amplifies the V4–V5 region of the 16S rRNA gene. The primer set was selected as it exhibits few biases against individual bacterial taxa and even the 100-bp Illumina reads should yield accurate phylogenetic and taxonomic information (4). The primers contain the appropriate Illumina adapters and the reverse primer contains a 12-bp error-correcting barcode unique to each sample. DNA was amplified in triplicate PCR reactions following the protocol described previously (3), the triplicate reactions were composited, and the amplicons from all samples were pooled together in equimolar concentrations. Sequencing was conducted on an Illumina HiSeq2000 at the University of Colorado Genomics Core Facility following the 2×100 bp paired-end protocol (3). Quality filtering of reads and processing of the reads was conducted as described in Caporaso et al. (5). After quality filtering, we obtained 118,000–750,000 forward reads per sample; only the forward reads were used for downstream analyses as it has been shown that including the reverse read adds little additional information (6). For all downstream analyses, we rarefied to 118,000 randomly selected reads per sample to correct for differences in sequencing depth. Reads were assigned to phylotypes at the $\geq 97\%$ sequence similarity level using the open-reference phylotype picking protocol in QIIME (7). QIIME was used to estimate both taxonomic and phylogenetic metrics of the pairwise distances between communities (Bray–Curtis and unweighted Unifrac distances, respectively). Alpha diversity was determined using both taxonomic metrics (numbers of phylotypes and

Shannon index, H') and a phylogenetic metric (Faith's phylogenetic diversity).

Shotgun Metagenomic Analyses. Shotgun metagenomic analyses were conducted on the same 16 DNA extracts used for the PCR-based 16S rRNA gene analyses described. The laboratory protocol followed that described in the Illumina Paired-End Prep kit protocol. Aliquots of each DNA sample were mechanically sheared and products were size-selected to 170–180 bp and gel purified. Sequencing was performed at Argonne National Laboratory in the Institute for Genomics and Systems Biology Next Generation Sequencing (IGSB-NGS) Core using a 2×100 bp sequencing run on the Illumina GAIIx. Sequences were uploaded to MG-RAST (Rapid Annotation using Subsystems Technology for Metagenomes) (8) for downstream analyses. Data accession numbers are provided in Table S2. Sequences were annotated to functional categories against the M5NR database using BLASTX at an e-value cutoff of 1×10^{-2} and the SEED subsystems hierarchy. Reads that could not be annotated by MG-RAST were discarded and subsequent analyses were performed on the metagenomes evenly sampled at random to 688,000 annotated reads per sample. Pairwise distances between each of the 16 metagenomes were determined by calculating Bray–Curtis distances from the relative abundances of reads in each of the level 2 gene categories (of which there were a total of 417 across the 16 rarefied datasets). We used the level 2 abundances instead of the abundances of individual annotated genes (of which there were $\sim 5,000$) as this should yield a more conservative estimate of the distances between metagenomes; individual genes should exhibit less overlap between samples than analyses based on more broadly defined functional categories. For taxonomic analyses of the small-subunit (SSU) and large-subunit (LSU) rRNA reads, the shotgun metagenomic data were compared with the Silva SSU and LSU databases available through MG-RAST using a maximum e-value of $1\text{e-}5$, a minimum identity of 60%, and a minimum alignment length of 15. For the more detailed analyses of the metagenomic SSU (16S) rRNA reads obtained from bacteria and archaea, we extracted 16S reads from these samples using QIIME by applying a closed-reference phylotype picking process where we search reads against the Greengenes database and discard reads that fail to hit any sequences at greater than or equal to 90% identity. This resulted in 83,486 16S reads across the 16 samples. The taxonomy of each phylotype was assigned as the taxonomy of the best Greengenes hit during the closed-reference phylotype picking process.

Statistical Analyses. We conducted principal coordinates analyses in PRIMER (9), using as input the pairwise distances between metagenomes (Bray–Curtis distances) or bacterial communities (Bray–Curtis and UniFrac distances calculated from the 16S rRNA gene amplicon data). To test whether sample categories harbored significantly different metagenomes or microbial communities, we used analysis of similarities (ANOSIM) tests as implemented in PRIMER. Mantel tests were run to assess correlations between metagenomic and bacterial community distance matrices. Mantel tests were also used to compare the Bray–Curtis distances between taxonomic distributions determined via 16S rRNA amplicon sequencing versus shotgun metagenomic sequencing. To determine whether the relative abundances of individual taxa or functional gene categories were significantly different between sample categories (or between the shotgun metagenomic and PCR-based analyses), we conducted pairwise t tests with P values calculated using a Bonferroni correction for multiple comparisons.

- Fierer N, et al. (2012) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* 6(5): 1007–1017.
- Lauber C, Knight R, Hamady M, Fierer N (2009) Soil pH as a predictor of soil bacterial community structure at the continental scale: A pyrosequencing-based assessment. *Appl Environ Microbiol* 75:5111–5120.
- Caporaso JG, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8):1621–1624.
- Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* 35(18):e120.
- Caporaso JG, et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 108(Suppl 1):4516–4522.
- Werner JJ, Zhou D, Caporaso JG, Knight R, Angenent LT (2012) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *ISME J* 6(7):1273–1276.
- Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
- Meyer F, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386.
- Clarke K, Gorley R (2006) PRIMER (PRIMER-E, Plymouth, UK), version 6.

Phylum	Class	Cold Desert					Hot Desert			Tropical Forest		Temperate Forest		Prairie	Boreal Forest	Tundra	
		EB017	EB019	EB020	EB021	EB024	EB026	MD3	SF2	SV1	AR3	PE6	CL1	DF1	KP1	BZ1	TL1
Crenarchaeota (Archaea)		0.22	0.01	0.33	0.11	0.86	0.01	6.03	6.62	6.59	4.54	0.75	0.04	0.00	0.78	0.01	0.07
	Thaumarchaeota	0.22	0.01	0.33	0.11	0.86	0.01	6.03	6.62	6.59	4.54	0.45	0.02	0.00	0.78	0.00	0.01
Euryarchaeota (Archaea)		0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.06	0.05	0.04	0.12	0.02	0.01	0.00	0.00	0.01
Acidobacteria		45.88	0.14	17.39	6.99	19.65	4.50	8.18	17.47	13.49	13.08	43.14	31.62	16.82	17.31	30.67	37.76
	Acidobacteria	1.88	0.03	2.38	0.81	1.29	0.54	0.96	4.74	2.48	6.67	3.67	7.06	6.36	10.42	5.14	3.26
	Chloracidobacteria	42.36	0.08	13.35	5.83	16.90	3.60	4.44	7.73	5.12	2.51	0.08	1.04	1.29	2.88	0.28	0.29
	Solibacteres	0.77	0.01	0.32	0.04	1.04	0.06	2.16	3.78	4.86	1.16	22.78	16.26	5.80	1.82	16.44	15.17
Actinobacteria		8.57	40.63	31.12	39.81	26.07	2.33	19.30	14.55	16.12	3.61	4.71	3.98	3.54	11.46	3.69	2.08
	Actinobacteria	8.57	40.63	31.12	39.81	26.07	2.33	19.30	14.55	16.11	3.61	4.71	3.98	3.54	11.46	3.69	2.08
Bacteroidetes		12.69	40.49	13.04	9.40	15.23	24.79	13.90	11.81	15.42	2.06	1.61	2.71	3.50	6.87	5.56	4.68
	Flavobacteria	0.09	27.96	0.16	0.36	0.17	6.91	0.06	0.06	0.08	0.16	0.01	0.05	0.21	0.23	0.02	0.03
	Sphingobacteria	11.19	12.24	12.42	8.20	14.35	17.48	13.50	11.57	15.09	1.76	1.44	2.51	3.15	6.44	5.03	3.76
Chloroflexi		1.29	1.59	1.97	3.22	1.42	0.60	1.70	2.16	1.46	1.01	1.87	0.78	0.69	1.53	0.63	2.20
	Anaerolineae	0.18	0.04	0.16	0.29	0.18	0.30	0.09	0.53	0.18	0.25	0.11	0.40	0.25	0.24	0.18	0.51
	Bifidi12	0.05	0.00	0.05	0.01	0.02	0.00	0.10	0.14	0.12	0.22	0.25	0.07	0.10	0.39	0.13	0.10
	SOGA31	0.66	0.04	1.26	1.62	0.76	0.05	0.25	0.36	0.31	0.29	0.01	0.10	0.17	0.63	0.16	0.35
	Thermomicrobia	0.14	1.47	0.23	1.19	0.22	0.11	0.11	0.21	0.18	0.01	0.00	0.00	0.00	0.04	0.00	0.00
Cyanobacteria		0.25	0.49	0.06	0.03	1.02	25.25	14.48	9.64	2.09	0.09	0.40	0.30	0.24	0.03	0.25	0.13
	Chloroplast	0.01	0.19	0.02	0.00	0.23	0.59	1.32	0.17	0.06	0.03	0.04	0.01	0.03	0.00	0.03	0.00
	Oscillatoriothrycidea	0.14	0.17	0.01	0.01	0.48	23.18	12.89	8.99	1.91	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Firmicutes		0.02	0.57	0.06	0.02	0.03	0.00	1.26	0.54	0.17	3.74	0.42	0.49	0.34	0.74	0.27	0.07
	Bacilli	0.02	0.41	0.05	0.02	0.03	0.00	1.24	0.44	0.16	3.62	0.39	0.48	0.32	0.73	0.24	0.01
Gemmatimonadetes		3.32	0.14	3.96	1.62	2.35	1.06	1.31	2.12	2.37	0.55	0.84	0.77	0.45	1.40	1.81	3.24
Nitrospirae		0.14	0.00	0.12	0.01	0.13	0.03	0.29	0.51	0.49	1.03	0.61	0.05	0.02	0.07	0.04	0.10
Planctomycetes		1.61	0.08	2.88	1.57	2.28	1.53	1.12	2.68	2.44	2.02	2.36	3.56	3.39	3.27	3.63	3.98
	Phycisphaerae	0.72	0.01	1.71	0.92	1.38	0.46	0.51	0.98	1.16	0.56	0.13	0.57	0.61	1.52	0.98	0.80
	Planctomyceta	0.82	0.06	1.14	0.65	0.89	1.07	0.60	1.64	1.27	1.13	2.12	2.76	2.43	1.53	2.42	2.94
Proteobacteria		11.87	5.72	9.22	5.58	10.50	23.26	15.96	19.32	24.02	19.06	30.04	30.56	28.77	23.51	31.08	28.31
	Alphaproteobacteria	3.44	3.92	5.52	1.03	6.36	8.63	9.57	9.29	14.46	4.53	10.82	13.24	12.63	11.09	14.83	7.49
	Betaproteobacteria	3.27	0.88	1.86	1.32	1.80	6.33	2.35	2.11	3.45	1.68	3.52	5.27	5.83	5.63	4.26	10.61
	Deltaproteobacteria	1.46	0.11	0.64	0.19	1.02	2.16	2.65	5.18	4.20	10.71	4.25	4.97	5.88	3.38	2.72	4.79
	Gammaproteobacteria	3.18	0.73	0.94	2.27	0.99	5.65	0.91	2.21	1.43	1.14	10.42	6.02	3.34	2.24	8.11	4.76
Thermi		0.03	0.36	0.02	0.65	0.02	0.02	0.02	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Verrucomicrobia		4.92	0.19	12.17	4.05	13.60	10.39	2.63	4.04	3.74	40.29	5.22	17.61	34.60	28.12	14.14	9.49
	Opitutae	0.05	0.04	0.08	0.16	0.27	0.29	0.03	0.12	0.09	0.04	0.05	0.14	0.14	0.10	0.99	0.36
	Spartobacteria	2.94	0.05	11.26	3.56	12.05	1.07	2.06	2.31	2.47	37.46	2.08	14.34	30.73	26.43	8.61	4.79
	Verrucomicrobiae	1.78	0.07	0.61	0.29	1.11	8.92	0.35	1.23	0.82	2.28	2.97	2.54	2.72	1.12	2.87	3.73

Fig. S1. Relative abundances of the major bacterial and archaeal taxa across the 16 soils examined here. Except for the crenarchaeal and euryarchaeal groups, all phyla are bacterial. Numbers in cells indicate percentages, cells colored based on relative abundances (highest relative abundances in red).

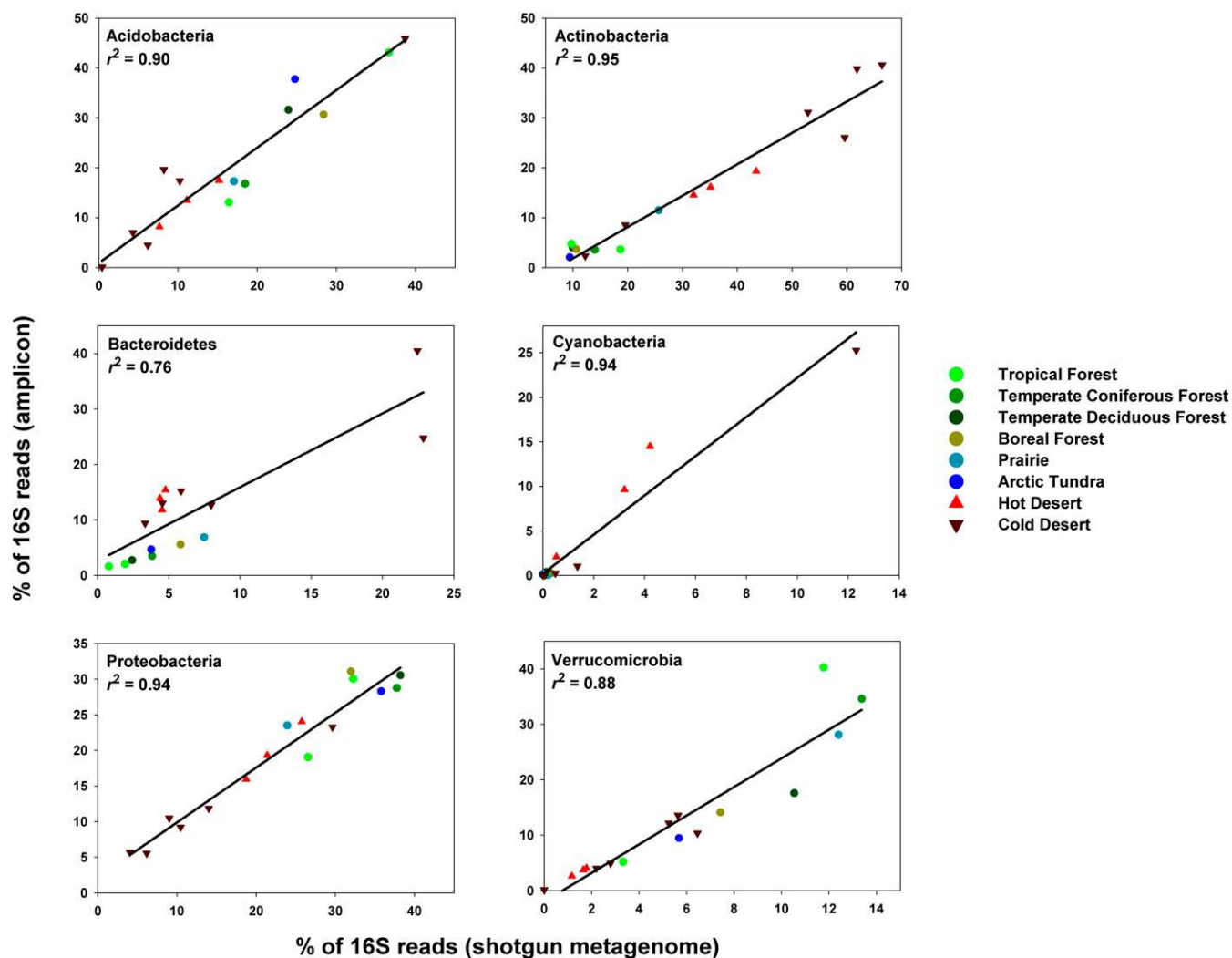


Fig. S2. Comparisons of the relative abundances (percentage of reads) of the dominant soil bacterial phyla determined using 16S rRNA gene data obtained using the shotgun metagenomic approach (x axes) versus the 16S rRNA gene data obtained using the amplicon sequencing approach (y axes).

	Cold Desert Soils						Hot Desert Soils			Non-Desert Soils						
	EB017	EB019	EB020	EB021	EB024	EB026	MD3	SF2	SV1	AR3	BZ1	CL1	DF1	KP1	PE6	TL1
Amino Acids and Derivatives	15.68	15.52	15.44	15.65	15.24	16.48	15.47	15.43	15.70	15.30	15.58	15.54	15.87	15.88	16.13	15.52
Carbohydrates	15.90	15.96	15.66	15.75	15.26	14.63	17.78	17.72	16.89	16.08	15.76	16.48	16.07	17.89	17.12	16.16
Cell Division and Cell Cycle	2.68	2.58	2.95	3.07	2.78	2.62	2.47	2.50	2.30	2.24	1.78	1.71	1.80	2.02	1.83	1.85
Cell Wall and Capsule	3.37	2.73	2.72	2.56	2.60	3.30	2.78	2.98	2.92	3.08	3.19	3.58	3.41	2.80	3.11	3.91
Clustering-based subsystems	17.53	19.09	18.88	19.88	19.81	18.41	18.48	17.92	17.97	16.92	16.24	15.62	16.11	16.57	15.65	16.41
Cofactors, Vitamins, Prosthetic Groups, Pigments	4.73	6.47	5.57	5.23	5.36	5.64	5.57	5.89	5.63	4.84	5.41	5.80	5.96	5.76	5.42	5.39
DNA Metabolism	4.42	5.18	4.64	4.73	4.57	4.78	4.30	3.62	4.45	2.96	3.88	3.59	3.68	4.02	3.27	3.71
Dormancy and Sporulation	0.03	0.05	0.06	0.08	0.06	0.10	0.07	0.03	0.06	0.01	0.01	0.01	0.03	0.02	0.01	0.02
Fatty Acids, Lipids, and Isoprenoids	1.19	1.48	1.29	1.32	1.32	2.01	1.44	1.35	1.38	1.24	1.57	1.53	1.39	1.32	1.22	1.54
Iron acquisition and metabolism	0.11	0.71	0.18	0.15	0.13	0.17	0.20	0.26	0.29	0.28	0.18	0.22	0.19	0.22	0.14	0.14
Membrane Transport	1.95	1.46	1.65	1.45	1.86	1.63	1.72	1.69	1.88	2.49	3.02	2.65	3.04	2.15	1.92	2.64
Metabolism of Aromatic Compounds	0.63	0.48	0.47	0.39	0.42	0.66	0.82	0.73	0.92	1.37	1.22	1.44	1.39	1.20	1.21	1.18
Miscellaneous	3.41	3.38	3.15	2.96	3.36	3.93	4.14	3.89	4.28	3.94	3.89	4.14	3.95	4.16	4.39	3.76
Motility and Chemotaxis	0.03	0.09	0.08	0.10	0.09	0.12	0.19	0.06	0.26	0.06	0.38	0.34	0.37	0.19	0.17	0.29
Nitrogen Metabolism	0.43	0.40	0.45	0.31	0.77	0.62	0.71	0.63	0.69	0.98	0.78	0.94	0.98	0.94	0.96	0.99
Nucleosides and Nucleotides	3.10	2.85	3.32	3.15	3.11	2.91	2.88	3.17	2.92	2.90	2.59	2.42	2.50	2.64	2.43	2.35
Phages, Prophages, Transposable elements, Plasmids	2.10	2.33	2.27	1.88	1.92	1.87	2.17	2.01	1.87	1.53	1.82	1.64	1.62	1.52	1.20	1.98
Phosphorus Metabolism	1.15	0.54	1.01	1.01	0.94	0.63	0.90	1.10	0.97	1.08	0.88	0.82	0.88	0.83	0.89	0.78
Photosynthesis	0.01	0.01	0.00	0.00	0.02	0.18	0.05	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Potassium metabolism	0.15	0.06	0.14	0.01	0.15	0.21	0.10	0.17	0.23	0.62	0.56	0.62	0.66	0.41	0.59	0.90
Protein Metabolism	10.55	9.60	10.87	11.54	10.76	9.36	9.64	9.89	9.88	8.86	7.73	7.02	7.02	8.15	7.90	7.22
Regulation and Cell signaling	0.08	0.19	0.10	0.10	0.12	0.19	0.14	0.15	0.15	0.13	0.20	0.17	0.20	0.14	0.19	0.23
Respiration	3.75	3.74	3.78	4.06	3.82	2.91	3.58	3.80	3.51	4.20	3.61	3.80	3.89	3.69	3.99	3.88
RNA Metabolism	2.82	2.25	2.77	2.70	2.88	2.86	2.47	2.60	2.66	2.18	2.27	2.10	1.91	2.19	2.41	2.28
Secondary Metabolism	0.20	0.14	0.18	0.15	0.28	0.29	0.19	0.23	0.20	0.25	0.22	0.28	0.24	0.24	0.22	0.19
Stress Response	1.16	0.85	0.96	1.11	0.99	0.86	0.93	1.14	1.00	1.39	1.36	1.42	1.36	1.28	0.93	1.35
Sulfur Metabolism	0.24	0.25	0.30	0.28	0.32	0.26	0.44	0.38	0.47	0.56	0.65	0.62	0.71	0.67	0.58	0.55
Virulence, Disease and Defense	2.61	1.62	1.11	0.36	1.05	2.39	0.39	0.60	0.50	4.51	5.20	5.49	4.77	3.10	6.09	4.80

Fig. S3. The relative abundances of functional gene categories as determined from the shotgun metagenomic data with gene categories defined at the lowest level of resolution. Values in each cell indicate percent abundance with the color of each cell indicating the z score for that particular gene category (blue colors, negative z scores; red colors, positive z scores).



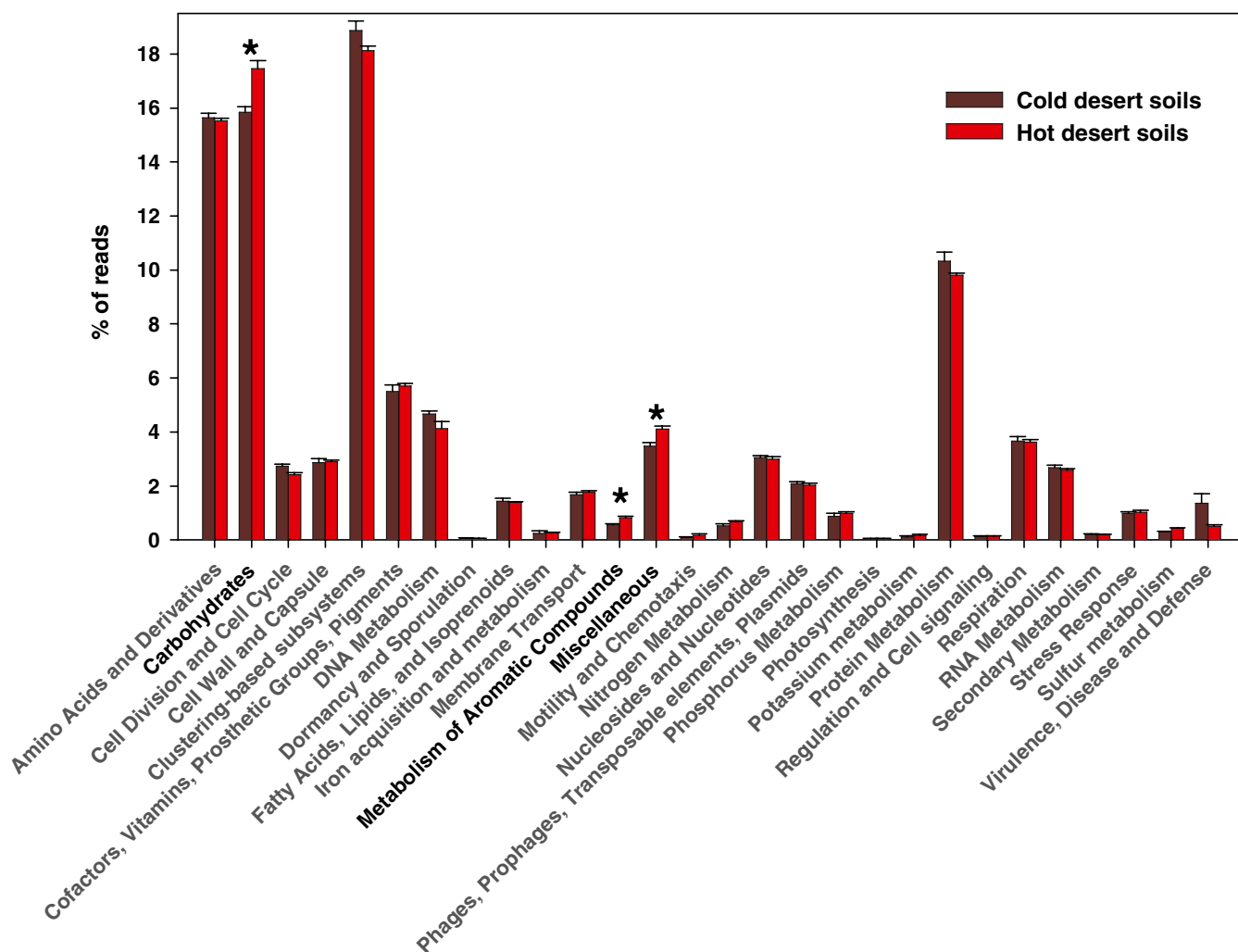


Fig. S5. The relative abundances of major categories of functional genes in metagenomes from cold desert soils from the McMurdo Dry Valley sites versus those soils collected from hot deserts in North America. Asterisks and bold type indicate those categories with significantly different relative abundances between the two desert types (Bonferroni corrected P values < 0.05 , uncorrected P values < 0.002).

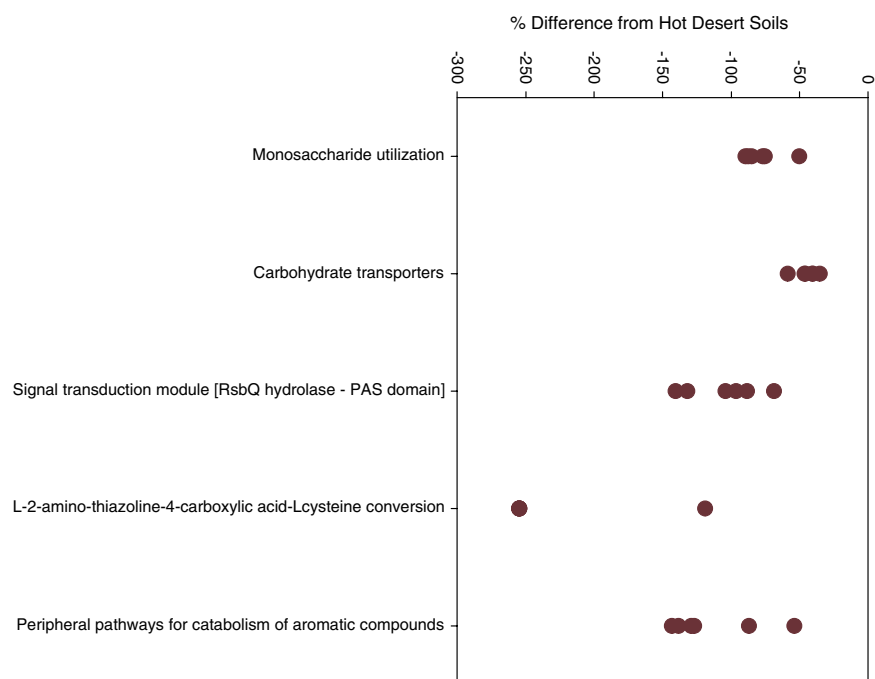


Fig. S6. Specific gene categories that are relatively less abundant in cold desert soils than in soils from the hot deserts. Only those gene categories that are significantly different between the two categories of samples (Bonferroni corrected P values <0.05) are shown here. At this level of resolution, no individual gene categories were significantly more abundant in the cold desert soils than in the hot desert soils.

Table S2. Diversity metrics, MG-RAST IDs, and percent of metagenomic reads annotated

Biome type	Sample ID	MG-RAST ID	% of quality reads annotated	Metagenomic richness (S)	Metagenomic diversity (H')	Bacterial 16S richness (S)	Bacterial 16S diversity (H')	Bacterial 16S phylogenetic diversity (PD)
Polar desert	EB017	4477900.3	14.5	1,535	6.39	4,527	5.31	300.0
Polar desert	EB019	4477901.3	23.6	1,663	6.52	2,796	3.60	261.1
Polar desert	EB020	4477902.3	17.3	1,376	6.33	4,936	5.79	305.6
Polar desert	EB021	4477903.3	15.9	1,228	6.17	2,845	4.57	195.3
Polar desert	EB024	4477904.3	17.2	1,386	6.34	4,124	5.56	270.0
Polar desert	EB026	4477803.3	20.5	2,231	6.78	2,935	4.92	232.9
Hot desert	MD3	4477805.3	16.4	1,948	6.60	8,895	6.72	485.8
Hot desert	SF2	4477872.3	14.4	1,850	6.56	10,078	6.93	554.4
Hot desert	SV1	4477873.3	17.3	1,981	6.68	9,929	7.14	527.4
Tropical forest	AR3	4477875.3	13.3	1,814	6.51	9,264	5.72	537.1
Boreal forest	BZ1	4477876.3	17.5	2,270	6.79	9,002	6.54	512.9
Temperate deciduous forest	CL1	4477877.3	18.2	2,393	6.81	12,352	7.06	675.0
Temperate coniferous forest	DF1	4477899.3	18.3	2,414	6.81	12,150	6.68	664.6
Temperate grassland	KP1	4477804.3	17.2	2,193	6.72	10,253	6.60	557.4
Tropical forest	PE6	4477807.3	15.6	2,317	6.70	8,772	6.66	476.8
Arctic tundra	TL1	4477874.3	18.8	2,375	6.84	6,965	6.27	437.6

Percentages of quality-filtered shotgun metagenomic reads that could be annotated to functional gene categories and diversity indices calculated from both the shotgun metagenomic data and the 16S rRNA gene amplicon data.

Table S3. Overall structure of the soil microbial communities

		Cold desert						Hot desert			Other nondesert biomes						
		EB017	EB019	EB020	EB021	EB024	EB026	MD3	SF2	SV1	AR3	BZ1	CL1	DF1	KP1	PE6	TL1
SSU	Archaea	0.30	0.04	0.20	0.16	0.31	0.09	1.85	3.06	3.08	3.37	0.14	0.00	0.00	1.28	0.53	0.00
	Bacteria	80.24	94.34	94.80	96.60	95.24	95.30	90.62	85.47	87.97	85.56	73.64	76.63	76.99	86.84	87.76	89.81
	Fungi (all)	7.54	0.15	0.85	0.21	1.09	1.21	2.36	3.93	3.26	2.56	11.21	10.33	9.32	3.54	3.63	3.53
	Fungi (Basidiomycota)	0.30	0.00	0.12	0.05	0.07	0.09	0.32	1.02	0.28	0.54	6.45	5.48	3.01	0.10	0.30	0.71
	Fungi (Ascomycota)	7.24	0.15	0.73	0.16	0.99	1.12	1.99	2.75	2.90	2.02	4.69	4.53	5.89	3.24	3.10	2.52
	Streptophyta	1.48	4.46	1.30	0.63	1.06	1.52	1.15	1.78	0.89	2.83	1.83	1.27	2.05	3.63	1.89	1.51
	Other eukaryotes	2.91	0.86	1.99	2.20	1.23	0.67	1.71	1.99	1.61	3.10	2.04	1.75	2.74	1.38	2.79	1.92
LSU	Archaea	0.06	0.00	0.04	0.00	0.03	0.03	0.11	0.35	0.42	1.24	0.00	0.00	0.11	0.24	0.18	0.07
	Bacteria	91.82	98.84	96.47	97.33	95.91	94.03	92.18	89.49	91.19	89.74	74.78	75.39	81.72	92.82	90.21	89.75
	Fungi (all)	2.38	0.12	0.63	0.30	0.95	0.80	2.75	3.42	2.87	2.22	11.01	10.50	7.88	2.02	3.02	4.07
	Fungi (Basidiomycota)	0.31	0.00	0.00	0.00	0.00	0.03	0.07	0.39	0.42	0.62	7.29	6.66	3.47	0.32	0.67	1.38
	Fungi (Ascomycota)	1.53	0.12	0.60	0.30	0.66	0.62	2.28	2.86	2.38	1.48	3.41	3.42	3.68	1.37	1.51	2.69
	Streptophyta	2.44	0.45	1.26	1.39	1.84	3.45	1.99	2.77	2.19	3.83	2.25	3.06	1.79	2.02	3.08	1.45
	Other eukaryotes	1.47	0.47	1.00	0.69	0.60	1.05	0.62	0.74	0.53	0.87	1.26	0.96	1.37	1.21	1.33	0.58

Percentages of small-subunit (SSU) and large-subunit (LSU) rRNA gene reads from the shotgun metagenomic data that could be assigned to major taxonomic groups. Category "Other eukaryotes" includes Mollusca, Nematoda, Arthropoda, Rotifera, and Annelida.