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

Noah Fierera,b,1, Jonathan W. Leftaf, Byron J. Adamsc, Uffe N. Nielsen d, Christian L. Lauberb, Scott Thomas Batesb, Sarah Owenses,f,1, Jack A. Gilberta,b,9, Diana H. Wallf, and J. Gregory Caporasoa,b,1

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; dHavensbury 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 Colorado, 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 86001

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 biochemical 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 non-desert communities regardless of the metric used. The desert communities had higher relative abundances of genes associated with ammophilaion and were lower relative abundances of genes predicted to be involved 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 non-desert 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.

S

oil 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 in turn has a large influence on the diversity 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 prostos) 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


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.orglookup/suppl/doi:10.1073/pnas.1215210110/-/DCSupplemental.
and their relative abundances were less than 3% in the 16S rRNA reads in any individual sample and less than 10% in any individual phylum. Additionally, the relatively abundant phyla in the cold desert soils included Bacteroidetes (20–40%) and Acidobacteria (15%), whereas in the tropical soils the relative abundance of Bacteroidetes was less than 20% and Actinobacteria (10–30%) was relatively well represented. Overall, these results showed that of the 16S rRNA reads 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 composition. This is evident from the strong correlation between the Bray-Curtis distance of the resulting amplicon 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, and Proteobacteria, and had very similar microbial community compositions. In addition, although the 16S rRNA reads contained a much higher proportion of reads from the taxonomically rich Bacteroidetes and Actinobacteria phyla (0.1–3.5% of reads) but were most abundant in the three hot desert soils (Fig. S1). The results showed that the diversity of the microbial communities found in the cold desert soils was substantially lower than that in the tropical soils, with the cold desert soils containing only a few dominant bacterial phyla, whereas in the tropical soils the bacterial community composition was more diverse and included a wider range of phyla. However, the only cold desert soils were not the case; the two methods generated nearly identical estimates of the overall differences in soil bacterial community diversity and composition. Concordance between these two very different approaches suggests that, at least across the wide range of soils examined here, the microbial communities found in the cold desert soils are taxonomically rich. The functional capabilities of the soil microbial community were determined using the functional metagenomic approach, which provides a higher resolution and more in-depth view of the microbial communities. The results showed that the microbial communities in the cold desert soils were composed of a wide range of functional gene categories, including genes for nitrogen fixation, carbon fixation, and alternative carbon sources. The results suggest that the microbial communities in the cold desert soils are well adapted to the harsh conditions of the desert environment, with the ability to survive and thrive in the low temperatures and limited resources.
As has been demonstrated previously (7), soil pH is a reasonably good predictor of prokaryotic diversity across the 16 soils (y = -2.875 + 0.37x, r² = 0.62, where x = soil pH and y = phylogeny 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 macrobiol communities (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 sample) 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 nitrogen 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 (r² < 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 microbial 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 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 microbial 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 non-desert 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 microbial communities with lower taxonomic or phylogenetic diversity have reduced functional diversity (32, 33), this paradigm does not necessarily hold true for microbial communities.
groups. The *Actinobacteria*, *Bacteroidetes*, and *Cyanobacteria* phyla were generally more abundant in the desert soils than in the non-desert soils, whereas *Verrucomicrobia* and *Acidobacteria* showed the opposite pattern (Fig. S1). Overall, the composition of the desert soil communities surveyed here was similar to those reported in other studies of cold and hot desert microbial communities (30, 37). More generally, the results shown here confirm the broad-scale patterns we would expect based on pH differences; high pH soils (such as those found in the desert soils included in this study) typically have higher relative abundances of *Actinobacteria* and *Bacteroidetes* with lower abundances of *Acidobacteria* compared with more acidic soils (6, 7). We note that the cold desert soil EB017 has a high abundance of *Acidobacteria* but these *Acidobacteria* belong to the class Chloracidobacteria that is distinct from the acidobacterial group (Solibacteres), which dominates in low pH soils; this is a pattern we would expect based on the results reported in Jones et al. (38). Factors other than pH may also be driving the bacterial community patterns evident in Fig. S1 and Fig. 2. For example, taxa known to be tolerant of low moisture conditions, including *Actinobacteria* (39), were more abundant in the desert soils surveyed here, whereas those taxa commonly associated with soils receiving higher rates of organic carbon input (e.g., beta- and gamma-proteobacteria) (10), were relatively less abundant in the desert soils.

Although *Cyanobacteria* and *Proteobacteria* were typically more abundant in the hot desert soils than in the cold desert soils (Fig. S1), the hot and cold deserts harbored relatively similar bacterial communities (as noted above). Despite large differences in site and edaphic characteristics, including the complete absence of plants in the cold desert sites and very low mean annual temperatures, cold and hot desert soil communities were relatively similar. This suggests that other factors common across these desert types (such as high soil pHs and low moisture levels) are most important in structuring these communities.

**Beta Diversity Patterns—Functional Genes.** The beta diversity patterns determined from the 16S rRNA gene analyses were nearly identical to the patterns determined from a comparison of functional gene abundances across the 16 soil metagenomes (Fig. 2). The Bray-Curtis distances calculated from taxon abundances and functional gene abundances were significantly correlated (Mantel $r = 0.76$, $P < 0.001$). Likewise, there was a strong correlation between unweighted Unfrcat distances, a phylogenetic metric of community similarity, and the Bray-Curtis distances in functional gene abundances (Mantel $r = 0.82$, $P < 0.001$). Therefore, as with the alpha diversity patterns, the concordance in beta diversity patterns highlights that the overall functional differences between the soil microbial communities were significantly correlated with the differences in the composition of these communities. Our findings are in line with comparable studies conducted in soil (20) and other habitats that also found strong correlations between metagenome composition and taxonomic composition (34, 35, 40). Although individual functional genes may not necessarily be correlated with community structure, the overall functional attributes of soil microbial communities appear to be predictable across broad gradients in soil and biome types if one has information on the taxonomic or phylogenetic structure of the communities.

Both the cold desert soils and hot desert soils had metagenomes distinct in composition from those found in the non-desert soils (ANOSIM $R = 0.97$ and 0.98 respectively, $P < 0.005$ in both cases), a pattern clearly evident from the ordination plot (Fig. 2) and the corresponding heatmap (Fig. S3). The large differences between desert and non-desert soils were also evident from a comparison of the relative abundances of functional genes classified at the lowest level of resolution (Fig. 3). After correction for multiple comparisons, 13 of 28 major gene categories were significantly different in abundance between desert and non-desert soils (Fig. 3), patterns that were examined in more detail by identifying the 35 specific gene categories (out of 417 in total) that strongly differentiated the desert soil metagenomes from the non-desert metagenomes (Fig. S4). The cold and hot desert microbial communities also had metagenomes that were distinct in composition from one another (Fig. 2; ANOSIM $R = 0.41$, $P = 0.03$), but the differences between these desert soils were less than the differences between the desert and non-desert soils.

Many of the gene categories that were more abundant in the desert soils than in the non-desert soils were those related to core metabolic functions (Fig. 3 and Figs. S3 and S4). Given that we were determining relative abundances, the overrepresentation of these gene categories in the desert soils may simply be a product of the desert soils having reduced diversity; lower phylogenetic or metagenomic diversity would presumably lead to an increase in the relative abundances of those core genes that are shared by nearly all cells and are required for cell survival and replication. However, some of the observed differences in functional gene abundances between the desert and non-desert soils may be more directly related to the unique conditions found in deserts, including lower moisture availability and reduced plant biomass. For example, we would expect nutrient cycling rates to be lower in desert systems than in more mesic systems due to moisture constraints (41), a pattern that was confirmed by the higher relative abundances of genes associated with nitrogen, potassium, and sulfur metabolism in the non-desert soils (Fig. 3 and Fig. S4). Likewise, exposure to frequent moisture stress may explain why the desert soils have higher relative abundances of genes associated with dormancy/sporulation, stress proteins, and amino acid metabolism (amino-acid-based solutes are commonly used by bacteria for osmoregulation) (39). The desert soils had lower relative abundances of genes associated with the degradation of complex organic compounds, including aromatics (Fig. S4), a pattern likely related to the lower levels of plant biomass found in the desert soils. Plants typically represent major sources of organic carbon to soil and these pools of organic carbon are often distinct (and more enriched in aromatics) (42) in soils supporting more plant biomass than in soils where plants are less abundant or nonexistent where we would expect microbe-derived organic carbon pools to dominate.

One of the most striking differences between desert and non-desert soil microbial communities was the differential abundance of antibiotic resistance genes and other genes likely associated with microbe-microbe competition. Genes associated with antibiotic resistance were far less abundant in the desert soils (averaging 1.5% of the annotated reads) than in the non-desert soils.
We aimed to elucidate the mechanisms underlying the high productivity of cold desert soils. To this end, we collected soil samples from various sites in the cold deserts across the globe and conducted comprehensive metagenomic and metatranscriptomic analyses on these samples. The cold desert soils were characterized by a high proportion of actinomycete and bacterial communities compared to other desert soils (Fig. S4). In addition, the cold desert soils exhibited a lower diversity of microbial communities compared to other desert soils (Table S1). These findings suggest that actinomycete and bacterial communities play a crucial role in the high productivity of cold desert soils.

The cold desert soils were also characterized by a high abundance of functional genes encoding enzymes involved in the degradation of plant material (Fig. S5). This finding is consistent with the observation that cold desert soils are rich in plant material due to the presence of >>100 species of plant communities (47). The cold desert soils also exhibited a high abundance of genes encoding enzymes involved in the synthesis of antibiotics and the degradation of antibiotics (Fig. S6). These findings suggest that the cold desert soils may be a reservoir of novel antibiotics that could be used in the treatment of antibiotic-resistant diseases.

In conclusion, the cold desert soils are characterized by a high abundance of actinomycete and bacterial communities, a lower diversity of microbial communities, and a high abundance of functional genes encoding enzymes involved in the degradation of plant material. These findings suggest that the cold desert soils play a crucial role in the high productivity of cold desert soils and that they may be a reservoir of novel antibiotics.