Ecological and Genomic Attributes of Novel Bacterial Taxa That Thrive in Subsurface Soil Horizons

Tess E. Brewer,[a] Emma L. Aronson,[b] Keshav Arogyaswamy,[b] Sharon A. Billings,[c] Jon K. Botthoff,[d] Ashley N. Campbell,[e] Nicholas C. Dove,[f] Dawson Fairbanks,[g] Rachel E. Gallery,[h] Stephen C. Hart,[i] Jason Kaye,[j] Gary King,[k] Geoffrey Logan,[b] Kathleen A. Lohse,[l] (ID) Mia R. Maltz,[d] Emilio Mayorga,[m] Caitlin O'Neill,[n] Sarah M. Owens,[o] Aaron Packman,[p] Jennifer Pett-Ridge,[e] Alain F. Plante,[q] Daniel D. Richter,[r] Whendee L. Silver,[s] Wendy H. Yang,[t] Noah Fierer[u]

[a]Cooperative Institute for Research in Environmental Sciences and Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA

[b]Department of Microbiology and Plant Pathology, University of California, Riverside, California, USA

[c]Department of Ecology and Evolutionary Biology, Kansas Biological Survey, University of Kansas, Lawrence, Kansas, USA

[d]Center for Conservation Biology, University of California, Riverside, California, USA

[e]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA

[f]Environmental Systems Graduate Group, University of California, Merced, California, USA

[g]Department of Soil, Water and Environmental Science, University of Arizona, Tucson, Arizona, USA

[h]School of Natural Resources and the Environment, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA

[i]Department of Life & Environmental Sciences, Sierra Nevada Research Institute, University of California, Merced, California, USA

[j]Department of Ecosystem Science and Management, Pennsylvania State University, University Park, Pennsylvania, USA

[k]Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA

[l]Department of Biological Sciences, Idaho State University, Pocatello, Idaho, USA

[m]Applied Physics Laboratory, University of Washington, Seattle, Washington, USA

[n]Department of Plant Biology, University of Illinois at Urbana Champaign, Urbana, Illinois, USA

[o]Biosciences Division, Argonne National Laboratory, Argonne, Illinois, USA

[p]Civil and Environmental Engineering, Northwestern University, Evanston, Illinois, USA

[q]Department of Earth & Environmental Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[r]Nicholas School of the Environment, Duke University, Durham, North Carolina, USA

[s]Department of Environmental Science, Policy, and Management, University of California, Berkeley, California, USA

[t]Departments of Plant Biology and Geology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[u]Department of Ecology and Evolutionary Biology, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA

ABSTRACT While most bacterial and archaeal taxa living in surface soils remain undescribed, this problem is exacerbated in deeper soils, owing to the unique oligotrophic conditions found in the subsurface. Additionally, previous studies of soil microbiomes have focused almost exclusively on surface soils, even though the microbes living in deeper soils also play critical roles in a wide range of biogeochemical processes. We examined soils collected from 20 distinct profiles across the United States to characterize the bacterial and archaeal communities that live in subsurface soils and to determine whether there are consistent changes in soil microbial communities with depth across a wide range of soil and environmental conditions. We found that bacterial and archaeal diversity generally decreased with depth, as did the degree of similarity of microbial communities to those found in surface horizons. We observed five phyla that consistently increased in relative abundance with depth across our soil profiles: *Chloroflexi*, *Nitrospirae*, *Euryarchaeota*, and candidate phyla GAL15 and *Dormibacteraeota* (formerly AD3). Leveraging the unusually high abundance of *Dormibacteraeota* at depth, we assembled genomes representative of this candidate phylum and identified traits that are likely to be beneficial in low-nutrient environments, including the synthesis and storage of carbohydrates, the potential to

use carbon monoxide (CO) as a supplemental energy source, and the ability to form spores. Together these attributes likely allow members of the candidate phylum *Dormibacteraeota* to flourish in deeper soils and provide insight into the survival and growth strategies employed by the microbes that thrive in oligotrophic soil environments.

IMPORTANCE Soil profiles are rarely homogeneous. Resource availability and microbial abundances typically decrease with soil depth, but microbes found in deeper horizons are still important components of terrestrial ecosystems. By studying 20 soil profiles across the United States, we documented consistent changes in soil bacterial and archaeal communities with depth. Deeper soils harbored communities distinct from those of the more commonly studied surface horizons. Most notably, we found that the candidate phylum *Dormibacteraeota* (formerly AD3) was often dominant in subsurface soils, and we used genomes from uncultivated members of this group to identify why these taxa are able to thrive in such resource-limited environments. Simply digging deeper into soil can reveal a surprising number of novel microbes with unique adaptations to oligotrophic subsurface conditions.

KEYWORDS soil microbiology, metagenomics, microbial traits, critical zone, microbial ecology

Subsurface soils often differ from surface horizons with respect to their pH, texture, moisture levels, nutrient concentrations, clay mineralogy, pore networks, redox state, and bulk densities. Globally, the top 20 cm of soil contains nearly five times more organic carbon (C) than soil in the bottom 20 cm of meter-deep profiles (1). In addition, residence times of organic C pools are typically far longer in deeper soil horizons (2), suggesting that much of the soil organic matter found in the subsurface is not readily utilized by microbes. Unsurprisingly, the strong resource gradient observed through most soil profiles is generally associated with large declines in microbial biomass (3–8); per gram soil, microbial biomass is typically 1 to 2 orders of magnitude lower in the subsurface than surface horizons (4, 6, 7). Although microbial abundances in deeper soils are relatively low on a per gram soil basis, the cumulative biomass of microbes inhabiting deeper soil horizons can be on par with that living in surface soils, owing to the large mass and volume of subsurface horizons (3, 5). Moreover, those microbes living in deeper horizons can play important roles in mediating a myriad of biogeochemical processes, including processes associated with soil C and nitrogen (N) dynamics (9, 10), soil formation (11), iron redox reactions (12, 13), and pollutant degradation (14).

Given that soil properties typically change dramatically with depth, it is not surprising that the composition of soil microbial communities also generally changes with depth through a given profile (4–6, 8, 15, 16). In some cases, the differences observed in microbial communities with depth through a single soil profile can be large enough to be evident even at the phylum level of resolution. For example, both *Chloroflexi* (15, 17) and *Nitrospirae* (15) may increase in relative abundance with depth. However, while previous work suggests that particular taxa can be relatively more abundant in deeper soils, it is unclear if such patterns are consistent across distinct soil and ecosystem types. We hypothesized that there are specific groups of soil bacteria and archaea that are typically rare in surface horizons but more abundant in deeper soils. Taxa that are proportionally more abundant in deeper soil horizons likely have slow-growing, oligotrophic life history strategies due to the lack of disturbance at depth and the low-resource conditions typical of most deeper soil horizons (18). Likewise, we expect deeper soils to harbor higher proportions of novel and undescribed microbial lineages, given that oligotrophic taxa are typically less amenable to *in vitro*, cultivation-based investigations (19).

We designed a comprehensive study to investigate how soil bacterial and archaeal communities change with soil profile depth, to identify taxa that are consistently more abundant in deeper horizons, and to determine what life history strategies enable these
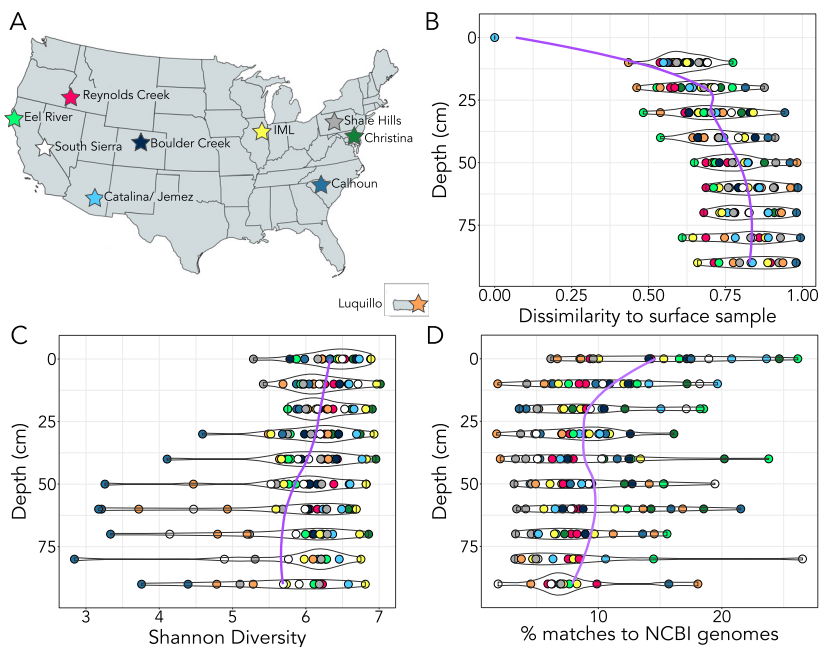
FIG 1 (A) Site map of sampling locations. We analyzed bacterial and archaeal communities from two soil profiles located at each of 10 different CZOs across the United States. Each profile was sampled in 10-cm intervals from surface soils to 1 m in depth (where possible). (B) Bray-Curtis dissimilarity to surface samples increases with depth. As depth increases, soil bacterial and archaeal communities become less similar to those communities at the surface. (C) Bacterial and archaeal diversity generally decreases with depth. Colors of points match the colors of the CZO sites indicated in panel A with two profiles sampled per site ($n = 20$). (D) The proportion of 16S rRNA gene sequences from the sampled soils for which representative genome data are available decreases with depth. We matched our 16S rRNA gene amplicon sequences to 16S rRNA genes from finished bacterial and archaeal genomes in the NCBI database. At deeper soil depths, we found that fewer taxa in our data set had matches to publicly available genomes, indicating that the bacterial and archaeal taxa found in deeper soil horizons are less represented in genomic databases than those found in surface soils. More details on these analyses are presented in Materials and Methods. The purple trend lines represent smoothed conditional means, generated using the loess modeling method.

taxa to thrive in the resource-limited conditions typical of most subsurface horizons. We collected soil samples at 10-cm increments from 20 soil profiles representing a wide range of ecosystem types throughout the United States, with most of the profiles sampled to 1 m in depth. We examined the bacterial and archaeal communities of these soil profiles by pairing amplicon 16S rRNA gene sequencing with shotgun metagenomic sequencing on a subset of samples. We found that deeper soil horizons typically harbored more undescribed bacterial and archaeal lineages, and we identified specific phyla (including *Dormibacteraeota*, GAL15, *Chloroflexi*, *Euryarchaeota*, and *Nitrospirae*) that consistently increased in relative abundance with depth across multiple profiles. Moreover, we found one candidate phylum (*Dormibacteraeota*, formerly AD3) to be particularly abundant in deeper soil horizons with low organic C concentrations. From our metagenomic data, we were able to assemble genomes from representative members of this candidate phylum and document the life history strategies, including low maximum growth rates and spore-forming potential, that are likely advantageous under low-resource conditions.

RESULTS AND DISCUSSION

Sample descriptions and soil properties linked to soil depth. We collected soils from a network of 10 current and former Critical Zone Observatories (CZOs) located across the United States (Fig. 1A) that span a broad range of hydrogeological provinces, soil orders, and ecosystem types, including tropical forest, temperate forest, grassland, and cropland sites. Soils were sampled from two distinct profiles per CZO for a total of 20 different soil profiles. Details of the site characteristics and edaphic properties for
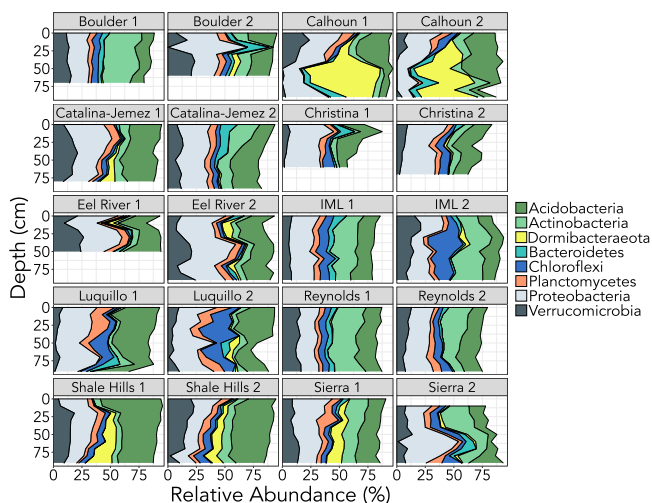
FIG 2 Different soil profiles have distinct microbial communities. Here, we show the relative abundances of the eight most abundant phyla identified from our 16S rRNA gene amplicon data. Not all profiles were sampled to 1 m due to variable bedrock depth. Note that the two profiles sampled from each CZO site were selected to represent distinct soil types (details on soil characteristics are available in Data Set S1 in the supplemental material).

each of the 20 soil profiles are provided in Data Set S1 in the supplemental material. Soils were collected from the first meter (where possible) of freshly excavated profiles, sampling at 10-cm increments and focusing on mineral soil horizons only (O horizons, if present, were not sampled). Together, this collection yielded 179 individual soil samples collected across sites with a wide range of different climatic conditions (e.g., mean annual temperatures ranging between 5 and 23°C and mean annual precipitation ranging from 26 to 402 cm yr^{-1}) (Data Set S1). The sampled profiles ranged from poorly developed Entisols and Inceptisols to highly developed Oxisols and Ultisols (per the USDA soil taxonomy system) and reflected an extremely broad range of soil properties. For example, in the 0- to 10-cm depth increment, soil pH ranged from 3.3 to 9.8, organic carbon concentrations ranged from 1.3% to 21.6%, and texture ranged from 0% to 45% silt plus clay across the profiles.

Some soil properties changed consistently with depth across all 20 profiles. Total N and organic C concentrations were both negatively correlated with soil depth, in agreement with previous observations (1, 20) (depth versus %C rho = −0.61, $P <$ 0.001; depth versus %N rho = −0.56, $P <$ 0.001; Spearman). On average, soil total organic C concentrations below 50 cm were 4.4 times lower than in surface soils, while total N concentrations were 6.3 times lower. While we measured a suite of additional chemical and soil properties (Data Set S1), only clay concentrations exhibited consistent changes with depth (with percent clay generally increasing with depth; rho = 0.29, $P <$ 0.001; Spearman). Given that our sampling effort included a wide range of different soil types and the expectedly high degree of variability in inter- and intraprofile edaphic characteristics, our goal was not to determine if distinct soil samples harbored distinct microbial communities or to characterize the factors related to shifts in overall community composition. Rather, our goal was to determine if there were any consistent changes in soil microbial communities with depth across the 20 sampled profiles.

Community characteristics linked to soil depth. Unsurprisingly, we found that the location of each soil profile had a strong influence on the composition of soil bacterial and archaeal communities, as determined by 16S rRNA gene amplicon sequencing ($r =$ 0.47, $P <$ 0.001, permutational multivariate analysis of variance [PERMANOVA]). Individual soil profiles generally harbored distinct microbial communities (Fig. 2; Fig. S1). In addition to this variation across the profiles, soil depth also had a significant effect on the composition of the bacterial and archaeal communities within individual profiles ($P <$ 0.01 for 16 of 20 profiles, rho values ranging from 0.24 to 0.45). In general,
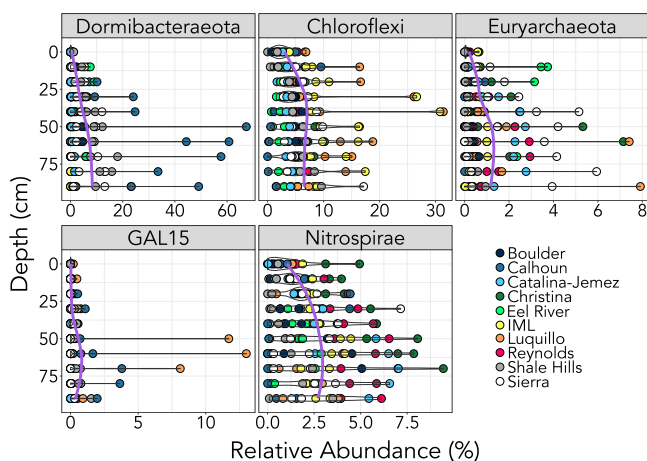
FIG 3 Five bacterial and archaeal phyla that consistently increased in relative abundance with soil depth. These phyla were identified via Spearman rank correlations against depth (FDR-corrected P values < 0.02, rho > 0.22). The purple trend lines represent smoothed conditional means, generated using the loess modeling method. For details on all phylum level abundances in each individual soil profile, see Table S5.

the variation in community composition with depth within a given profile, while significant, was less than the differences in soil communities observed across different profiles when all profiles and soil depths were examined together (depth, $r = 0.02$, $P < 0.001$; location, $r = 0.47$, $P < 0.001$, PERMANOVA).

Several characteristics of the bacterial and archaeal communities changed consistently with depth despite the high degree of heterogeneity observed across the different soil profiles. As soil depth increased, microbial communities found at depth became increasingly dissimilar to those found in surface horizons (Fig. 1B). When we analyzed the entire sample set together, dissimilarity to surface soils (0- to 10-cm depth) was positively correlated with depth ($P < 0.001$, rho = 0.73, Spearman). This trend also held for 17 out of 20 individual soil profiles (depth was not significant in both Eel River sites and IML site 1). We also found that, in general, the diversity of microbial communities decreased with depth, with several CZOs exhibiting stronger declines with depth than others (Calhoun, Luquillo, and South Sierra) (Fig. 1C). Lastly, when we compared the 16S rRNA gene sequences from this study to those 16S rRNA gene sequences from finished bacterial and archaeal genomes in the NCBI database, we found that the proportion of taxa for which genomic data are available declined with depth (from 6.2 to 26.1% in surface soils to 1.9 to 18.0% in the deepest horizons sampled) (Fig. 1D). Although representative genomes are currently unavailable for the majority of soil bacterial and archaeal taxa (21), we found that this problem is exacerbated for taxa living in deep soils.

Taxonomic shifts with soil depth. Although each soil profile harbored distinct microbial communities (Fig. 2), we identified five phyla that consistently increased in abundance with soil depth, as measured by Spearman correlations across the entire data set: *Chloroflexi*, *Euryarchaeota*, *Nitrospirae*, and the candidate phyla *Dormibacteraeota* and GAL15 (Fig. 3) (false discovery rate [FDR]-corrected P values < 0.02, rho > 0.22). For example, *Dormibacteraeota* were, on average, 27 times more abundant in soils at 90 cm than in surface horizons. The candidate phylum *Dormibacteraeota*, *Chloroflexi*, and *Nitrospirae* have previously been found to increase in abundance with increasing soil depth in individual profiles (15, 17), while candidate phylum GAL15 has been shown to be abundant in oxic subsurface sediments (22). Members of these phyla are likely oligotrophic taxa adapted to survive under the resource-limited conditions found in deeper horizons. Indeed, soil *Euryarchaeota* (23), *Chloroflexi*, and *Nitrospirae* (24) have been shown to decrease in relative abundance upon soil fertilization. These five phyla are also underrepresented in public genome databases; together, they

account for only 2.8% of bacterial and archaeal genomes deposited in the IMG database (as of December 2018), reinforcing our observation (highlighted in Fig. 1D) that bacteria and archaea living in deeper soils are underrepresented in genome databases.

Community-level shotgun metagenomic analyses. We selected one soil profile from each of the CZOs for metagenomic sequencing, selecting the profile that displayed the most community dissimilarity through depth (Eel River samples were not analyzed for logistical reasons). In total, we obtained shotgun metagenomic data from 67 soil samples with an average of 7.84 million quality-filtered reads per sample (see Table S1 for details). We first used these metagenomic data to quantify changes in the relative abundances of the bacterial, archaeal, and eukaryotic domains with depth. The overwhelming majority of rRNA gene sequences that we detected were from bacteria (89.2% to 98.7% of reads), followed by archaea (0.03% to 7.70%), and then eukaryotes (0.04% to 4.27%). Interestingly, we found that the proportion of eukaryotic sequences in our samples decreased with depth (rho = −0.32, $P = 0.05$). Most of these eukaryotic rRNA gene reads were classified as Fungi (58%), followed by Charophyta (16%), Metazoa (9.3%), and Cercozoa (7.0%). These results are in line with previous work showing that the contributions of eukaryotes, most notably fungi, to microbial biomass pools typically decrease with soil depth (25).

We also directly compared the results obtained from our 16S rRNA amplicon and shotgun metagenomic sequencing across the same set of samples. We did this to check whether our PCR primers introduced significant biases in the estimation of taxon relative abundances. We found that the shotgun and amplicon-based estimations of abundance for the most ubiquitous and abundant phyla we observed across the sampled profiles (Fig. 2) were well correlated (Fig. S2, mean rho values = 0.70). Next, we checked whether our primers missed any major groups of bacteria or archaea, as it has been noted that many taxa from the Candidate Phyla Radiation (CPR; recently assigned to the superphylum *Patescibacteria*) (26) are not detectable with the primer set used here (27). While we found that our primer pair did fail to recover sequences from the superphylum *Patescibacteria*, these taxa were rare in our data—the entire superphylum accounted for only 0.5% of 16S rRNA gene reads across the whole metagenomic data set.

Candidate phylum *Dormibacteraeota* is relatively more abundant in soils with low organic carbon. We found that members of candidate phylum *Dormibacteraeota* were consistently more abundant in deeper soil horizons and particularly abundant in subsurface horizons from the Calhoun and Shale Hills CZOs (Fig. 4). In these soils, *Dormibacteraeota* dominated the microbial communities—in some samples, over 60% of 16S rRNA sequences were classified as belonging to members of the *Dormibacteraeota* candidate phylum. The high abundances of *Dormibacteraeota* were confirmed with shotgun metagenomic analyses (Fig. S2), indicating that the abundances of this phylum were not inflated by PCR primer biases. The candidate phylum *Dormibacteraeota* was first observed in a sandy, highly weathered soil from Virginia, United States (28), and does not yet have a representative cultured isolate. The phylum was previously known as "AD3" but was renamed *Dormibacteraeota* after three genomes from the phylum were assembled from Antarctic soils (29). Other representative genomes from this phylum have also become available with the recent addition of 47 genomes assembled from thawing permafrost (30). The phylum *Dormibacteraeota* has been observed in subsurface soil horizons previously (31, 32), and its relative abundance has been found to be negatively correlated with water content, C, N, and total potential enzyme activities (17).

While the abundance of members of the phylum *Dormibacteraeota* was generally positively correlated with depth across all samples included in this study (rho = 0.22, $P = 0.02$, Spearman), this pattern did not hold for all profiles (Fig. 4). Instead, we found organic C concentrations to be the best predictor of the abundance of *Dormibacteraeota* in these soil communities (Fig. S3); *Dormibacteraeota* were typically eight times more abundant in soils with less than 1% organic C than in soils where organic C
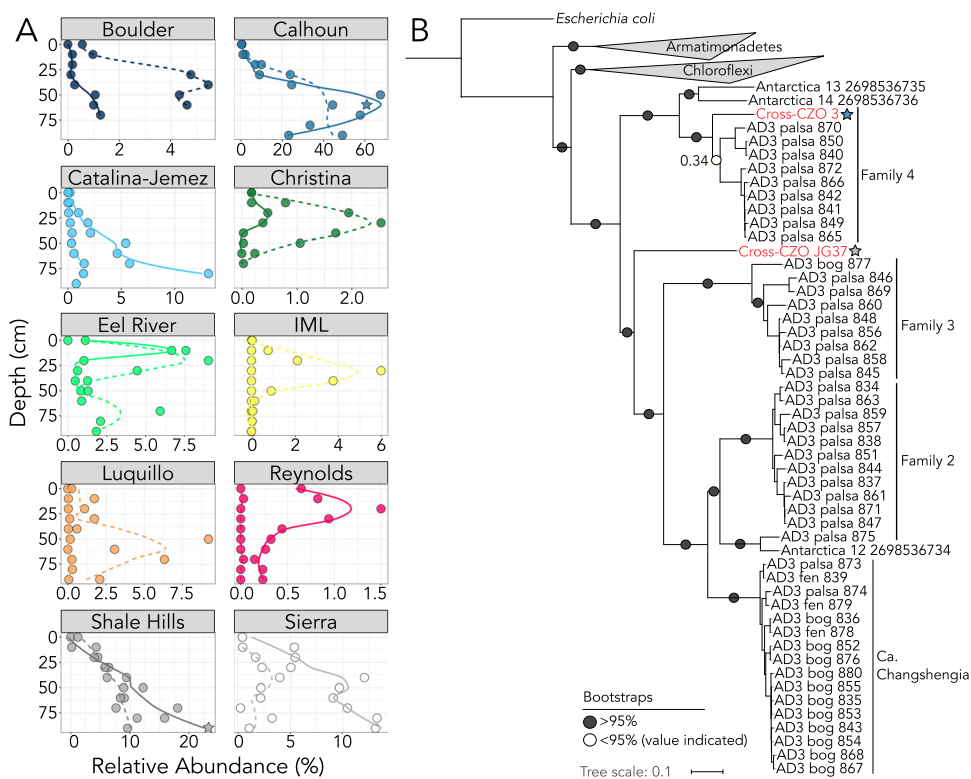
FIG 4 (A) The 16S rRNA gene relative abundance of phylum *Dormibacteraeota* is variable across different soil profiles but generally increases with depth. The samples used for the *Dormibacteraeota* genome assemblies are noted with stars. The trend lines represent smoothed conditional means, generated using the loess modeling method. (B) The two *Dormibacteraeota* genomes we assembled from soil profile metagenomic data cluster phylogenetically with previously published *Dormibacteraeota* genomes. Our deep soil genomes also fall near the known sister phyla *Chloroflexi* and *Armatimonadetes*, validating their identity as members of candidate phylum *Dormibacteraeota*. This tree was created using the concatenated marker gene phylogeny generated from GTDBTk (26) and was plotted using iTOL (70). Only closely related phyla are included in the tree. Genomes assembled in this study are indicated in red, and all other AD3/*Dormibacteraeota* genomes originated from either reference 29 or 30. The family groupings for the *Dormibacteraeota* tree were first presented in reference 30.

concentrations were greater than 2%. Because soil depth and organic C concentrations were correlated across the profiles studied here, we used an independent data set of surface soils (0- to 10-cm depth) collected from 1,006 sites across Australia to determine if the abundances of *Dormibacteraeota* were also correlated with organic C concentrations when analyses were restricted to a broad range of distinct surface soils (33). Indeed, we found that the relative abundances of *Dormibacteraeota* in Australian surface soils (which ranged from 0.0 to 7.0% of 16S rRNA gene sequences) were also negatively correlated with soil organic carbon concentrations (Fig. S3). Together, these results indicate that members of the *Dormibacteraeota* phylum are typically most abundant in surface or subsurface soils where organic C concentrations are relatively low.

Dormibacteraeota draft genomes recovered from metagenomic data. To gain more insight into the potential traits and genomic attributes of soil *Dormibacteraeota*, we conducted deeper shotgun metagenomic sequencing on several soils where *Dormibacteraeota* were found to be particularly abundant (Fig. 4A), with the goal of assembling draft genomes from members of this group. We assembled two *Dormibacteraeota* genomes, both from deep soils (Fig. 4). These genomes are considered medium-quality drafts, according to published genome reporting standards for metagenome-assembled genomes (MAGs) (34); bin 3 is estimated to be 72.6% complete at 3.43 Mb, while bin JG-37 is 69.9% complete at 2.48 Mb (see further genome details in Table S1). These genomes are similar in size to those previously assembled from the phylum (range of 3.0 to 5.3 Mb, all >90% complete [29]; range of 1.6 to

4.3 Mb, all >70% complete [30]). These genomes share only 45.1% average amino acid identity (AAI) (35) and cluster phylogenetically with the *Dormibacteraeota* genomes assembled from Antarctic soil metagenomes (29) and those from permafrost metagenomes (30) (Fig. 4B).

Analyses of the *Dormibacteraeota* genomes that we recovered indicate that members of this phylum are aerobic heterotrophs adapted to nutrient-poor conditions. Both *Dormibacteraeota* genomes encode high-affinity terminal oxidases, indicative of an aerobic metabolism (cbb$_3$ oxidase, bin JG-37; bd oxidase, bin 3). These genomes contain no markers of an autotrophic metabolism, with no RuBisCO or hydrogenase genes detected in either of the assembled genomes. Both genomes encode glycosyl hydrolases (with bin 3 containing 14 of these genes in total), indicating an ability to use polysaccharides for growth. Specifically, both genomes contain glycogen catalysis (alpha-amylase, glucoamylases) and synthesis (glycogen synthase) genes. The ability to synthesize, store, and break down glycogen has been shown to promote the survival of bacteria during periods of starvation (36, 37). Additionally, both *Dormibacteraeota* genomes contain the trehalose 6-phosphate synthase gene, a key gene in the pathway for the synthesis of trehalose, a C storage compound that also confers resistance to osmotic stress and heat shock (37) and can protect cells from oxidative damage, freezing, thermal injury, or desiccation (38). These attributes likely confer an advantage in resource-limited soils, as the ability to store C for later use may be advantageous in environments where organic C is infrequently available or of low quality.

Based on several lines of evidence, soil-dwelling *Dormibacteraeota* appear to be oligotrophic taxa with low maximum growth rates. First, as mentioned above, these taxa have the highest relative abundances in soils with low organic C concentrations, where we would expect oligotrophic lifestyles to be advantageous. Second, both *Dormibacteraeota* genomes appear to contain a single rRNA operon, a feature often linked to low maximum potential growth rates (39). Third, although we cannot directly measure the maximum growth rate of uncultivated bacterial cells, we can estimate maximum growth rate from genomes by measuring codon usage bias with the ΔENC' metric (40). ΔENC' is a measure of codon bias in highly expressed genes and has been shown to correlate strongly with growth rate for both bacteria and archaea (41). We calculated ΔENC' for our *Dormibacteraeota* genomes, the Antarctic *Dormibacteraeota* genomes (29), the thawing permafrost *Dormibacteraeota* genomes (30), and a set of bacterial and archaeal genomes which matched the 16S rRNA gene amplicon sequences recovered from our soil profile samples at ≥99% sequence similarity. The ΔENC' values for all the *Dormibacteraeota* genomes clustered together toward the lower end of the spectrum for our set of soil bacteria and archaea, indicating that members of the phylum *Dormibacteraeota* are likely to exhibit low potential growth rates (Fig. S4).

To our knowledge, all previous *Dormibacteraeota* genomes were recovered from either Antarctic desert (29) or permafrost soils (30), while our genomes hail from subsurface soils collected from temperate regions. Despite these disparate origins, some central characteristics of the phylum *Dormibacteraeota* appear to be consistent. Similar to the Antarctic and permafrost *Dormibacteraeota* genomes, our *Dormibacteraeota* genomes also contained carbon monoxide (CO) dehydrogenase genes. There are two forms of CO dehydrogenases, which differ in their ability to oxidize CO and the rate at which they do so (42). While the active site of form I is specific to CO dehydrogenases, form II active sites also occur in many molybdenum hydroxylases that do not accept CO as a substrate (42). Using sequence data from our assembled *Dormibacteraeota* genomes, the Antarctic and permafrost *Dormibacteraeota* genomes, and selected CO dehydrogenase large-subunit sequences (CoxL), we generated a phylogenetic tree based on the amino acid sequence of CoxL (Fig. S5). With these analyses, we found that both of the *Dormibacteraeota* genomes recovered here possess form II CO dehydrogenase genes, as do many of the Antarctic and permafrost *Dormibacteraeota* genomes. Although it has been shown that form II CO dehydrogenases can permit growth with CO as a sole C and energy source in some cases (43),

further work is needed to determine whether the form II CO dehydrogenase genes allow *Dormibacteraeota* to actively oxidize CO or if these genes code for molybdenum-containing hydroxylases responsible for other metabolic processes (44). Interestingly, one Antarctic and many permafrost *Dormibacteraeota* genomes also encode form I CoxL, indicating that some members of this phylum are capable of CO oxidation (Fig. S5). CO oxidation with form I CO dehydrogenases has been shown to improve the survival of bacterial cells under nutrient-limited conditions (45).

Analyses of our assembled *Dormibacteraeota* genomes also reveal that these soil bacteria may be capable of spore formation. All together, our *Dormibacteraeota* genomes contain 34 spore-related genes scattered across a variety of spore generation phases (Table S2). We also found spore-forming genes among the Antarctic and permafrost *Dormibacteraeota* genomes, most notably those encoding SpoIIE, SpoIIM, SpoIIIE, and SpoVS. Nutrient-limiting conditions are known to trigger spore formation (46), and sporulation can allow bacterial cells to persist until environmental conditions become more favorable. Additionally, members of the *Chloroflexi*, a sister phylum to *Dormibacteraeota*, are capable of spore formation (47). Because there are no *Dormibacteraeota* isolates available to test for sporulation, we adapted a method previously used in stool samples (48) to identify potential spore-forming taxa by using a culture-independent approach. We incubated three soil samples from our study in 70% ethanol to kill vegetative cells and then used propidium monoazide (PMA) to block the amplification of DNA from these dead cells (49). We then sequenced these soils using our standard 16S rRNA gene amplicon method both with and without the ethanol and PMA treatment. We found that the abundances of the two dominant *Dormibacteraeota* phylotypes were significantly higher in the spore-selected treatment than the untreated controls (Table S3). Other known sporeformers were enriched in the spore selection treatment as well, including taxa from the orders *Actinomycetales*, *Bacillales* (48), *Myxococcales* (50), and *Thermogemmatisporales* (51). While the enrichment of *Dormibacteraeota* in ethanol-treated samples shows that these cells are hardy, it is not conclusive proof of spore formation and further testing is needed to verify our findings (there are other methods of ethanol resistance in bacteria, such as biofilm formation [52] and residence inside other cells [53]).

Conclusions. Our results indicate that as soil depth increases, not only do bacterial and archaeal communities become less diverse and change in composition, but novel, understudied taxa become proportionally more abundant in deeper soil horizons. We identified five poorly studied bacterial and archaeal phyla that become more abundant in deeper soils across a broad range of locations and investigated one of these further (the candidate phylum *Dormibacteraeota*, formerly AD3) to determine what characteristics may allow *Dormibacteraeota* to survive in resource-limited soil environments. We found that members of *Dormibacteraeota* are likely slow-growing aerobic heterotrophs capable of persisting under low-resource conditions by putatively storing and processing glycogen and trehalose. Members of this candidate phylum also contain type I and II carbon monoxide dehydrogenases, which can potentially enable the use of trace amounts of CO as a supplemental energy source. We also found that soil-dwelling *Dormibacteraeota* are likely capable of sporulation, another trait that may allow cells to persist during periods of limited resource availability. More generally, analyses of these novel members of understudied phyla suggest life history strategies and traits that may be employed by oligotrophic microbes to thrive under resource-limited soil conditions.

MATERIALS AND METHODS

Sample collection and processing. Samples were collected from the network of 10 Critical Zone Observatories (CZOs; http://criticalzone.org) across the United States: Southern Sierra (CA), Boulder Creek (CO), Reynolds Creek (ID), Shale Hills (PA), Calhoun (SC), Luquillo (PR), Intensively Managed Landscapes (IL/IA/MN), Catalina/Jemez (AZ/NM), Eel River (CA), and Christina River (DE/PA). Volunteers from each CZO excavated two separate soil profiles ("sites") selected to represent distinct soil types and landscape positions. Soils were collected at peak greenness (as estimated from NASA's MODIS [moderate-resolution imaging spectroradiometer]) between April 2016 and November 2016, with the exception of the Eel River CZO samples, which were collected in May 2017. Volunteers were asked to sample in 10-cm increments

to a depth of at least 100 cm or to refusal. Site details are available in Data Set S1 in the supplemental material.

All soil samples were sent to the University of California, Riverside, for processing. A portion of each field sample was sieved (<2 mm, ASTM no. 10), homogenized, and divided into subsamples for further analyses, with subsamples stored at either 4°C, −20°C, or −80°C. For some soils (particularly some wet, finely textured depth intervals), sieving was not practical. These samples were homogenized by mixing, with larger root and rock fragments removed by hand. In addition, as samples from Shale Hills site 2 (70- to 100-cm depth) consisted almost entirely of medium-sized rocks, soil was collected by manually crushing rocks with a ceramic mortar and pestle; this material was then passed through a 2-mm sieve.

DNA was extracted from subsamples frozen at −20°C using the DNeasy PowerLyzer PowerSoil kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions, with minor modifications to increase yield and final DNA concentration based on the assumption that some sites and depths would have a relatively low microbial biomass. Specifically, 0.25 g of soil was weighed in triplicate (i.e., three 0.25-g aliquots = 0.75 g total soil per sample) from one frozen aliquot of sieved soil. Extractions on each 0.25-g replicate aliquot proceeded in parallel, until the stage when DNA was eluted onto the spin filter; replicates were pooled at this point onto a single filter, and extractions proceeded from this point as a single sample. In addition, the final step of elution of the DNA from the filter was conducted with 50 μl of elution buffer instead of 100 μl; the initial flowthrough was reapplied to the filter once to increase yield.

Soil characteristics. Frozen subsamples (stored at −20°C) were shipped to the University of Illinois at Urbana-Champaign for characterization of soil physicochemical properties. Soil C and N concentrations were measured on freeze-dried, sieved, and ground subsamples using a Vario Micro Cube elemental analyzer (Elementar, Hanau, Germany). Approximately 1 g of each subsample was also extracted in 30 ml of 0.5 N HCl for determination of Fe(III) and Fe(II) concentrations by using a modified ferrozine assay (54). Soil texture was measured for oven-dried and sieved soil in accordance with the method of Gee and Bauder (55).

Soil pH and gravimetric water content were measured using modified Long Term Ecological Research (LTER) protocols, as described by Robertson et al. (56). Soil pH was determined using 15 g of field-wet soil and 15 ml of Milli-Q water (Millipore Sigma, Burlington, MA) and was measured on a Hannah Instruments (Woonsocket, RI) HI 3220 pH meter with an HI 1053B pH electrode, designed for use with semisolids. For determining gravimetric water content, we oven-dried 7 g of soil at 105°C for a minimum of 24 h.

Amplicon-based 16S rRNA gene analyses. To characterize the bacterial and archaeal communities in each sample, we used the barcoded primer pair 515f/806r for sequencing the V4-V5 region of the 16S rRNA gene. We amplified this gene region three times per sample, combined these products, and normalized the concentration of each sample to 25 ng using SequalPrep normalization plate kits (Thermo Fisher Scientific, Waltham, MA). All samples were then pooled and sequenced on the Illumina MiSeq (2 × 150 paired-end chemistry) at the University of Colorado next-generation sequencing facility. The sample pool included several kit controls and no template controls to check for possible contamination.

Sequences were processed using a combination of QIIME (57) and USEARCH (58) commands to demultiplex, quality-filter, remove singletons, and merge paired-end reads. Sequences were classified into exact sequence variants (ESVs) using UNOISE2 (59) with default settings, and taxonomy was assigned against the Greengenes 13_8 database (60) using the RDP classifier (61). ESVs with greater than 1% average abundance across all sequenced controls were classified as contaminants and removed from further analyses, along with ESVs identified as mitochondria and chloroplasts. The entire data set was then rarefied to 3,400 sequences per sample. All statistical analyses were done in R version 3.5.1 (62), and all figures were created with ggplot2 (63) unless otherwise noted. We used the R package Vegan (64) to calculate Bray-Curtis dissimilarity (vegdist, method="bray") on Hellinger transformed ESV tables (decostand, method="hellinger") and to calculate Shannon diversity (diversity, index="shannon"). We calculated Spearman and Pearson correlations (cor.test) and corrected P values using base R functions (p.adjust, method="fdr").

We checked if our amplicon sequences had "representative genomes" in public databases by matching the 16S rRNA gene amplicon sequences generated in this study to 16S rRNA genes from finished bacterial and archaeal genomes in the NCBI database using the USEARCH10 command "usearch_global." We considered 16S rRNA gene amplicons to have a "representative genome" if they matched a genome sequence with ≥97% identity.

Shotgun metagenomic analyses. One soil profile from each CZO was selected for shotgun sequencing; we chose the profile that exhibited the most dissimilarity in microbial community composition through depth to sequence. The Eel River CZO samples were not collected in time to be included in these analyses. Using the same DNA as used for the amplicon sequencing, we generated metagenomic libraries with the TruSeq DNA LT library preparation kit (Illumina, San Diego, CA). All samples were pooled and sequenced on an Illumina NextSeq run using 2 × 150-bp paired-end chemistry at the University of Colorado next-generation sequencing facility. Prior to downstream analysis, we merged and quality filtered the paired-end metagenomic reads with USEARCH. After quality filtering, we had an average of 8.8 million quality-filtered reads per sample (range, 1.9 to 15.4 million reads; we included only samples with at last 1 million reads). These sequences were uploaded to MG-RAST (65) for public access. We used Metaxa2 (66) with default settings to extract small-subunit (SSU) rRNA gene sequences (bacterial, archaeal, and eukaryotic) in each sample and assigned taxonomy as described above using the Greengenes 13_8 database (60) and the RDP classifier (61).

Assembly, annotation, and characterization of *Dormibacteraeota* genomes. We assembled two genomes belonging to the candidate phylum *Dormibacteraeota* (29) from individual metagenomes obtained from Calhoun site 1 (60 to 70 cm) and Shale Hills site 1 (90 to 100 cm). These two soil samples were selected for deeper sequencing based on the high abundance of the phylum *Dormibacteraeota* (~60% of amplicon 16S rRNA gene reads at Calhoun, ~23% at Shale Hills). This sequencing effort yielded 57.7 million paired-end reads for Calhoun site 1 (60 to 70 cm) and 65.6 million paired-end reads for Shale Hills site 1 (90 to 100 cm).

Genomes were assembled using unmerged, paired-end reads that had been filtered using Sickle version 1.33 (-q 20 -l 50). We used Megahit version 1.1.4-2 (67) with the "bulk" preset to build the assembly and MaxBin 2.2.1 (68) for binning. We used a script to cycle through several MaxBin conditions (-min_contig_length 1100–1500, -prob_threshold 0.95–0.99) and used checkM version 1.0.7 (69) to pick the most complete bins. Selected bins were then manually curated by removing contigs that fell below the 2.5 percentile or above the 97.5 percentile in either scaffold abundance, tetranucleotide frequency, or GC content. After selecting the highest-quality bins from each sample, we ran Metaxa2 on the bins themselves to detect SSU or large-subunit (LSU) rRNA genes that could be used to determine taxonomic affiliations. Bin 3 contained one 16S rRNA sequence, which matched with 97.2% sequence identity a *Dormibacteraeota* sequence within the Greengenes database (sequence ID 151897). The bin 3 16S rRNA also matched the amplicon sequence for ESV1 at 98% identity. This ESV was the most abundant *Dormibacteraeota* sequence in our amplicon data set; its maximum relative abundance was 57% at Calhoun site 1 (50 cm), and it reached ≥1% relative abundance in 43% of our sites. Bin JG-37 contained only small fragments of the 5S and 23S rRNA genes, which were insufficient for taxonomic classification.

To verify that these two bins were affiliated with the *Dormibacteraeota* candidate phylum, we used the concatenated marker gene phylogeny generated from GTDBTk (26) to compare the placement of our genomes to previously published *Dormibacteraeota* genomes (29, 30). While GTDBTk could taxonomically classify our genomes only to the bacterial domain (a problem replicated in 45/53 currently available *Dormibacteraeota* genomes), both genomes clearly fall within the *Dormibacteraeota* phylum in the tree generated from concatenated marker genes (Fig. 4B). This tree was plotted in iTOL (70). Both *Dormibacteraeota* genomes were submitted to IMG for annotation under the taxon IDs 2824080494 (bin JG-37) and 2767802471 (bin 3). Based on CheckM (69) estimates, both genomes are substantially complete, with medium to high contamination (bin JG-37, 69.9% complete, 7.7% contamination; bin 3, 72.6% complete, 10.7% contamination). See Table S1 for additional genome details.

Phylogenetic tree of CoxL genes. The evolutionary history of the *Dormibacteraeota* CoxL genes was inferred by using the maximum likelihood method based on the JTT matrix-based model (71). The tree with the highest log likelihood (−24,038) is shown (Fig. 4). The percentage of trees in which the associated taxa clustered together is shown below the branches. An initial tree(s) for the heuristic search was obtained automatically by applying neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with the superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories [+G, parameter = 1.10]). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 6.75% sites). The resulting tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 71 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 658 residues in the final data set. Evolutionary analyses were conducted in MEGA7 (72), and the tree was plotted in iTOL (70).

Calculation of maximum growth rate proxy ΔENC. Because there are no cultivated members of the phylum *Dormibacteraeota*, we calculated ΔENC′ to estimate potential growth rate, as described previously (40, 41). We also calculated ΔENC′ on complete genomes in NCBI that matched amplicon sequences in our data set with ≥99% sequence similarity using the USEARCH10 command "usearch_global." We used this set of genomes to represent the bacteria found in the same soil profiles studied here to establish a range for potential microbial growth rates in soil. We ran ENCprime (40) with default options on both concatenated ribosomal protein sequences and concatenated genome sequences, and calculated ΔENC′ (73) as described by Vieira-Silva and Rocha (41).

Spore selection treatment. We adapted a method previously used for human stool samples (48) to select for spores in a culture-independent manner in three soil samples from our study (Calhoun site 1, soils of 50- to 60-cm and 60- to 70-cm depths, and Calhoun site 2, soil of 50- to 60-cm depth). To select for spores, we incubated 0.04 g of each sieved soil (as described above) in 70% ethanol for 4 h under aerobic conditions and constant agitation, with the goal of killing vegetative cells. After the incubations, we washed both sets of samples with phosphate-buffered saline (PBS) three times and then applied propidium monoazide (PMA) to the ethanol-treated samples, as described previously (49). We used PMA to block the amplification of DNA from cells with compromised membranes, ensuring that only those cells capable of surviving the harsh ethanol treatment would be amplified in subsequent PCRs. We PCR amplified, sequenced, and processed these samples as previously described. We restricted our analysis to the top 1,000 most abundant phylotypes to remove rare taxa and used the Wilcoxon test to identify enriched taxa, scoring taxa as "possible sporeformers" if they had false discovery rate (FDR)-corrected P values of >0.05. These taxa are presented in Table S3.

Data availability. Both *Dormibacteraeota* genomes and the metagenomes they were assembled from are publicly available in the IMG database under taxon IDs 2767802471 (bin 3), 2824080494 (bin JG-37), 3300022691 (Calhoun 60-cm metagenome), and 3300021046 (Shale Hills 90-cm metagenome). The merged, quality-filtered, and unassembled shotgun sequences are available under MG-RAST project ID mgp80869. The raw, unmerged 16S amplicon sequences are available on figshare at https://doi.org/10.6084/m9.figshare.4702711.

SUPPLEMENTAL MATERIAL

ACKNOWLEDGMENTS

REFERENCES

1. Jobbágy EG, Jackson RB. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol Appl 10:423–436. https://doi.org/10.2307/2641104.
2. Balesdent J, Basile-Doelsch I, Chadoeuf J, Cornu S, Derrien D, Fekiacova Z, Hatté C. 2018. Atmosphere-soil carbon transfer as a function of soil depth. Nature 559:599–602. https://doi.org/10.1038/s41586-018-0328-3.
3. Schütz K, Kandeler E, Nagel P, Scheu S, Ruess L. 2010. Functional microbial community response to nutrient pulses by artificial groundwater recharge practice in surface soils and subsoils. FEMS Microbiol Ecol 72:445–455. https://doi.org/10.1111/j.1574-6941.2010.00855.x.
4. Eilers KG, Debenport S, Anderson S, Fierer N. 2012. Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. Soil Biol Biochem 50:58–65. https://doi.org/10.1016/j.soilbio.2012.03.011.
5. Fierer N, Schimel JP, Holden PA. 2003. Variations in microbial community composition through two soil depth profiles. Soil Biol Biochem 35:167–176. https://doi.org/10.1016/S0038-0717(02)00251-1.
6. Blume E, Bischoff M, Reichert JM, Moorman T, Konopka A, Turco RF. 2002. Surface and subsurface microbial biomass, community structure and metabolic activity as a function of soil depth and season. Appl Soil Ecol 20:171–181. https://doi.org/10.1016/S0929-1393(02)00025-2.
7. Spohn M, Klaus K, Wanek W, Richter A. 2016. Microbial carbon use efficiency and biomass turnover times depending on soil depth—implications for carbon cycling. Soil Biol Biochem 96:74–81. https://doi.org/10.1016/j.soilbio.2016.01.016.
8. Stone MM, DeForest JL, Plante AF. 2014. Changes in extracellular enzyme activity and microbial community structure with soil depth at the Luquillo Critical Zone Observatory. Soil Biol Biochem 75:237–247. https://doi.org/10.1016/j.soilbio.2014.04.017.
9. Kramer C, Gleixner G. 2008. Soil organic matter in soil depth profiles: distinct carbon preferences of microbial groups during carbon transformation. Soil Biol Biochem 40:425–433. https://doi.org/10.1016/j.soilbio.2007.09.016.
10. Banning NC, Maccarone LD, Fisk LM, Murphy DV. 2015. Ammonia-oxidising bacteria not archaea dominate nitrification activity in semi-arid agricultural soil. Sci Rep 5:11146. https://doi.org/10.1038/srep11146.
11. Oh N-H, Richter DD. 2005. Elemental translocation and loss from three highly weathered soil–bedrock profiles in the southeastern United States. Geoderma 126:5–25. https://doi.org/10.1016/j.geoderma.2004.11.005.
12. Fimmen RL, Richter DD, Jr, Vasudevan D, Williams MA, West LT. 2008. Rhizogenic Fe–C redox cycling: a hypothetical biogeochemical mechanism that drives crustal weathering in upland soils. Biogeochemistry 87:127–141. https://doi.org/10.1007/s10533-007-9172-5.
13. Hall SJ, Liptzin D, Buss HL, DeAngelis K, Silver WL. 2016. Drivers and patterns of iron redox cycling from surface to bedrock in a deep tropical forest soil: a new conceptual model. Biogeochemistry 130:177–190. https://doi.org/10.1007/s10533-016-0251-3.
14. Schwarz A, Adetutu EM, Juhasz AL, Aburto-Medina A, Ball AS, Shahsavari E. 2018. Microbial degradation of phenanthrene in pristine and contaminated sandy soils. Microb Ecol 75:888–902. https://doi.org/10.1007/s00248-017-1094-8.
15. Will C, Thürmer A, Wollherr A, Nacke H, Herold N, Schrumpf M, Gutknecht J, Wubet T, Buscot F, Daniel R. 2010. Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol 76:6751–6759. https://doi.org/10.1128/AEM.01063-10.
16. Kramer S, Marhan S, Haslwimmer H, Ruess L, Kandeler E. 2013. Temporal variation in surface and subsoil abundance and function of the soil microbial community in an arable soil. Soil Biol Biochem 61:76–85. https://doi.org/10.1016/j.soilbio.2013.02.006.
17. Tas N, Prestat E, McFarland JW, Wickland KP, Knight R, Berhe AA, Jorgenson T, Waldrop MP, Jansson JK. 2014. Impact of fire on active layer

and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. ISME J 8:1904–1919. https://doi.org/10.1038/ismej.2014.36.

18. Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol 15:579–590. https://doi.org/10.1038/nrmicro.2017.87.

19. Vartoukian SR, Palmer RM, Wade WG. 2010. Strategies for culture of "unculturable" bacteria. FEMS Microbiol Lett 309:1–7. https://doi.org/10.1111/j.1574-6968.2010.02000.x.

20. Marty C, Houle D, Gagnon C, Courchesne F. 2017. The relationships of soil total nitrogen concentrations, pools and C:N ratios with climate, vegetation types and nitrate deposition in temperate and boreal forests of eastern Canada. Catena 152:163–172. https://doi.org/10.1016/j.catena.2017.01.014.

21. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems 3:e00055-18. https://doi.org/10.1128/mSystems.00055-18.

22. Lin X, Kennedy D, Fredrickson J, Bjornstad B, Konopka A. 2012. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. Environ Microbiol 14:414–425. https://doi.org/10.1111/j.1462-2920.2011.02659.x.

23. Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, Harpole WS, Hobbie SE, Hofmockel KS, Knops JMH, McCulley RL, La Pierre K, Risch AC, Seabloom EW, Schütz M, Steenbock C, Stevens CJ, Fierer N. 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proc Natl Acad Sci U S A 112:10967–10972. https://doi.org/10.1073/pnas.1508382112.

24. Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R. 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J 6:1007–1017. https://doi.org/10.1038/ismej.2011.159.

25. Turner S, Mikutta R, Meyer-Stüve S, Guggenberger G, Schaarschmidt F, Lazar CS, Dohrmann R, Schippers A. 2017. Microbial community dynamics in soil depth profiles over 120,000 years of ecosystem development. Front Microbiol 8:874. https://doi.org/10.3389/fmicb.2017.00874.

26. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.

27. Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC. 2016. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol 1:15032. https://doi.org/10.1038/nmicrobiol.2015.32.

28. Zhou J, Xia B, Huang H, Treves DS, Hauser LJ, Mural RJ, Palumbo AV, Tiedje JM. 2003. Bacterial phylogenetic diversity and a novel candidate division of two humid region, sandy surface soils. Soil Biol Biochem 35:915–924. https://doi.org/10.1016/S0038-0717(03)00124-X.

29. Ji M, Greening C, Vanwonterghem I, Carere CR, Bay SK, Steen JA, Montgomery K, Lines T, Beardall J, van Dorst J, Snape I, Stott MB, Hugenholtz P, Ferrari BC. 2017. Atmospheric trace gases support primary production in Antarctic desert surface soil. Nature 552:400–403. https://doi.org/10.1038/nature25014.

30. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, Hoelzle RD, Lamberton TO, McCalley CK, Hodgkins SB, Wilson RM, Purvine SO, Nicora CD, Li C, Frolking S, Chanton JP, Crill PM, Saleska SR, Rich VI, Tyson GW. 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560:49–54. https://doi.org/10.1038/s41586-018-0338-1.

31. Kim HM, Jung JY, Yergeau E, Hwang CY, Hinzman L, Nam S, Hong SG, Kim O-S, Chun J, Lee YK. 2014. Bacterial community structure and soil properties of a subarctic tundra soil in Council, Alaska. FEMS Microbiol Ecol 89:465–475. https://doi.org/10.1111/1574-6941.12362.

32. Billings SA, Hirmas D, Sullivan PL, Lehmeier CA, Bagchi S, Min K, Brecheisen Z, Hauser E, Stair R, Flournoy R, Richter DD. 2018. Loss of deep roots limits biogenic agents of soil development that are only partially restored by decades of forest regeneration. Elem Sci Anth 6:34. https://doi.org/10.1525/elementa.287.

33. Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, Breed MF, Brown B, Brown MV, Brugger J, Byrne M, Caddy-Retalic S, Carmody B, Coates DJ, Correa C, Ferrari BC, Gupta VVSR, Hamonts K, Haslem A, Hugenholtz P, Karan M, Koval J, Lowe AJ, Macdonald S, McGrath L, Martin D, Morgan M, North KI, Paungfoo-Lonhienne C, Pendall E, Phillips L, Pirzl R, Powell JR, Ragan MA, Schmidt S, Seymour N, Snape I, Stephen JR, Stevens M, Tinning M, Williams K, Yeoh YK, Zammit CM, Young A. 2016. Introducing BASE: the Biomes of Australian Soil Environments

34. soil microbial diversity database. GigaSci 5:21. https://doi.org/10.1186/s13742-016-0126-5.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.

35. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264. https://doi.org/10.1128/JB.187.18.6258-6264.2005.

36. Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM, Pozueta-Romero J. 2010. Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev 34:952–985. https://doi.org/10.1111/j.1574-6976.2010.00220.x.

37. Fung T, Kwong N, van der Zwan T, Wu M. 2013. Residual glycogen metabolism in Escherichia coli is specific to the limiting macronutrient and varies during stationary phase. J Exp Microbiol Immunol 17:83–87.

38. Kandror O, DeLeon A, Goldberg AL. 2002. Trehalose synthesis is induced upon exposure of Escherichia coli to cold and is essential for viability at low temperatures. Proc Natl Acad Sci U S A 99:9727–9732. https://doi.org/10.1073/pnas.142314099.

39. Roller BRK, Stoddard SF, Schmidt TM. 2016. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. Nat Microbiol 1:16160. https://doi.org/10.1038/nmicrobiol.2016.160.

40. Novembre JA. 2002. Accounting for background nucleotide composition when measuring codon usage bias. Mol Biol Evol 19:1390–1394. https://doi.org/10.1093/oxfordjournals.molbev.a004201.

41. Vieira-Silva S, Rocha E. 2010. The systemic imprint of growth and its uses in ecological (meta)genomics. PLoS Genet 6:e1000808. https://doi.org/10.1371/journal.pgen.1000808.

42. King GM, Weber CF. 2007. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nat Rev Microbiol 5:107–118. https://doi.org/10.1038/nrmicro1595.

43. Lorite MJ, Tachil J, Sanjuán J, Meyer O, Bedmar EJ. 2000. Carbon monoxide dehydrogenase activity in Bradyrhizobium japonicum. Appl Environ Microbiol 66:1871–1876. https://doi.org/10.1128/aem.66.5.1871-1876.2000.

44. Hille R. 2005. Molybdenum-containing hydroxylases. Arch Biochem Biophys 433:107–116. https://doi.org/10.1016/j.abb.2004.08.012.

45. Cordero PRF, Bayly K, Leung PM, Huang C, Islam ZF, Schittenhelm RB, King GM, Greening C. 2019. Carbon monoxide dehydrogenases enhance bacterial survival by oxidising atmospheric CO. bioRxiv https://doi.org/10.1101/628081.

46. Fujita M, Losick R. 2005. Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. Genes Dev 19:2236–2244. https://doi.org/10.1101/gad.1335705.

47. Cavaletti L, Monciardini P, Bamonte R, Schumann P, Rohde M, Sosio M, Donadio S. 2006. New lineage of filamentous, spore-forming, gram-positive bacteria from soil. Appl Environ Microbiol 72:4360–4369. https://doi.org/10.1128/AEM.00132-06.

48. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of "unculturable" human microbiota reveals novel taxa and extensive sporulation. Nature 533:543–546. https://doi.org/10.1038/nature17645.

49. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat Microbiol 2:16242. https://doi.org/10.1038/nmicrobiol.2016.242.

50. Shimkets LJ. 1999. Intercellular signaling during fruiting-body development of Myxococcus xanthus. Annu Rev Microbiol 53:525–549. https://doi.org/10.1146/annurev.micro.53.1.525.

51. Yabe S, Aiba Y, Sakai Y, Hazaka M, Yokota A. 2011. Thermogemmatispora onikobensis gen. nov., sp. nov. and Thermogemmatispora foliorum sp. nov., isolated from fallen leaves on geothermal soils, and description of Thermogemmatisporaceae fam. nov. and Thermogemmatisporales ord.

nov. within the class *Ktedonobacteria*. Int J Syst Evol Microbiol 61: 903–910. https://doi.org/10.1099/ijs.0.024877-0.

52. Luther MK, Bilida S, Mermel LA, LaPlante KL. 2015. Ethanol and isopropyl alcohol exposure increases biofilm formation in *Staphylococcus aureus* and *Staphylococcus epidermidis*. Infect Dis Ther 4:219–226. https://doi.org/10.1007/s40121-015-0065-y.

53. Zinniel DK, Lambrecht P, Harris NB, Feng Z, Kuczmarski D, Higley P, Ishimaru CA, Arunakumari A, Barletta RG, Vidaver AK. 2002. Isolation and characterization of endophytic colonizing bacteria from agronomic crops and prairie plants. Appl Environ Microbiol 68:2198–2208. https://doi.org/10.1128/aem.68.5.2198-2208.2002.

54. Liptzin D, Silver WL. 2009. Effects of carbon additions on iron reduction and phosphorus availability in a humid tropical forest soil. Soil Biol Biochem 41:1696–1702. https://doi.org/10.1016/j.soilbio.2009.05.013.

55. Gee GW, Bauder JW. 1986. Particle-size analysis, p 383–411. *In* Page AL (ed), Methods of soil analysis. Part 1, Physical and mineralogical methods. SSSA Book Series 5.1. SSSA, ASA, Madison, WI.

56. Robertson GP, Coleman DC, Bledsoe CS, Sollins P. 1999. Standard soil methods for long-term ecological research. Oxford University Press, New York, NY.

57. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.

58. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

59. Edgar RC. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv https://doi.org/10.1101/081257.

60. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.

61. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42:D633–D642. https://doi.org/10.1093/nar/gkt1244.

62. R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

63. Wickham H. 2009. ggplot2: elegant graphics for data analysis (Use R). Springer Nature, New York, NY.

64. Dixon P 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

65. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386–388. https://doi.org/10.1186/1471-2105-9-386.

66. Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ, Nilsson RH. 2015. metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. Mol Ecol Resour 15:1403–1414. https://doi.org/10.1111/1755-0998.12399.

67. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

68. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.

69. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

70. Letunic I, Bork P 2016. Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. https://doi.org/10.1093/nar/gkw290.

71. Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

72. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.

73. Rocha E. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res 14:2279–2286. https://doi.org/10.1101/gr.2896904.