AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

GEOMICROBIOLOGY

Distinct Depth-Discrete Profiles of Microbial Communities and Geochemical Insights in the Subsurface Critical Zone

Xiaoqin Wu,[a] Sara Gushgari-Doyle,[a] Lauren M. Lui,[b] Andrew J. Hendrickson,[b] Yina Liu,[c,d] Sindhu Jagadamma,[e] Torben N. Nielsen,[b] Nicholas B. Justice,[b] Tuesday Simmons,[f] Nancy J. Hess,[d] Dominique C. Joyner,[e] Terry C. Hazen,[e,g] Adam P. Arkin,[b,f] Romy Chakraborty[a]

[a]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[b]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[c]Department of Oceanography, Texas A&M University, College Station, Texas, USA
[d]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, USA
[e]University of Tennessee, Knoxville, Tennessee, USA
[f]University of California, Berkeley, Berkeley, California, USA
[g]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

ABSTRACT Microbial assembly and metabolic potential in the subsurface critical zone (SCZ) are substantially impacted by subsurface geochemistry and hydrogeology, selecting for microbes distinct from those in surficial soils. In this study, we integrated metagenomics and geochemistry to elucidate how microbial composition and metabolic potential are shaped and impacted by vertical variations in geochemistry and hydrogeology in terrestrial subsurface sediment. A sediment core from an uncontaminated, pristine well at Oak Ridge Field Research Center in Oak Ridge, Tennessee, including the shallow subsurface, vadose zone, capillary fringe, and saturated zone, was used in this study. Our results showed that subsurface microbes were highly localized and that communities were rarely interconnected. Microbial community composition as well as metabolic potential in carbon and nitrogen cycling varied even over short vertical distances. Further analyses indicated a strong depth-related covariation of community composition with a subset of 12 environmental variables. An analysis of dissolved organic carbon (DOC) quality via ultrahigh resolution mass spectrometry suggested that the SCZ was generally a low-carbon environment, with the relative portion of labile DOC decreasing and that of recalcitrant DOC increasing along the depth, selecting microbes from copiotrophs to oligotrophs and also impacting the microbial metabolic potential in the carbon cycle. Our study demonstrates that sediment geochemistry and hydrogeology are vital in the selection of distinct microbial populations and metabolism in the SCZ.

IMPORTANCE In this study, we explored the links between geochemical parameters, microbial community structure and metabolic potential across the depth of sediment, including the shallow subsurface, vadose zone, capillary fringe, and saturated zone. Our results revealed that microbes in the terrestrial subsurface can be highly localized, with communities rarely being interconnected along the depth. Overall, our research demonstrates that sediment geochemistry and hydrogeology are vital in the selection of distinct microbial populations and metabolic potential in different depths of subsurface terrestrial sediment. Such studies correlating microbial community analyses and geochemistry analyses, including high resolution mass spectrometry analyses of natural organic carbon, will further the fundamental understanding of microbial ecology and biogeochemistry in subsurface terrestrial ecosystems and will benefit the future development of predictive models on nutrient turnover in these environments.

KEYWORDS carbon transformation, microbial ecology, subsurface

Editor Jennifer B. Glass, Georgia Institute of Technology

Address correspondence to Romy Chakraborty, rchakraborty@lbl.gov.

The authors declare no conflict of interest.

Received 29 March 2023
Accepted 16 April 2023
Published 5 June 2023

The Earth's critical zone has been evolving as an emerging and rapidly growing research area since the term "critical zone" arose about two decades ago (1). The critical zone, which is considered to be Earth's outer skin, refers to a permeable layer from the tops of the tree canopy to the bottom of the groundwater, an environment in which rock, soil, water, air, and living organisms interact and shape the Earth's surface (2). The subsurface critical zone (SCZ), which extends from the ground surface down to the fresh, unweathered bedrock (3), harbors more than half of all global microorganisms (4), and these associate to form complex microbial communities that colonize subsurface environments and control key ecological processes, such as the carbon (C) and nitrogen (N) cycles (5). However, much is yet to be explored, and, despite decades of study, subsurface key microorganisms and their ecological functions still remain largely unknown. A large number of microbial species in the terrestrial subsurface remain uncultivated, and their physiologies and ecological impacts continue to remain an enigma (6–9).

The microbial assembly and metabolic potential in the SCZ are substantially impacted by sediment geochemical gradients and subsurface hydrogeology. In surface soils, microbial abundance and activity are often associated with "hot spots" (10, 11), which are typically close to plant roots (i.e., the rhizosphere) or are associated with decaying plant material. It is traditionally assumed that the availability of labile organic matter as well as the microbial abundance and activity are lower in the SCZ than in surface soils and that they decline as a function of depth (12–14). A recent study (15) reported that the localized distribution of microbes in the subsurface of a desert, where oligotrophic microbes were ubiquitous and highly diverse in metabolic potential, was correlated with an increase in the concentrations of Fe, K, Mg, and Ti in the sediment as well as with a change in lithology and groundwater capillary action. This study, together with many others (16–18), provides evidence that microbial communities can adapt to oligotrophic subsurface environments and that vertical variations in microbial activities and C turnover are shaped by a variety of geochemical factors that select for unique microbes that are distinct in sequential depths. For example, it was reported that microbial communities that are spatially close (within the same soil profile) but are separated by only 10 to 20 cm in depth can be as distinct from one another as they are from communities that are thousands of kilometers away (18).

Hydraulic conditions are one of the key factors influencing microbial communities in the SCZ. Microbes in the capillary fringe and groundwater are usually quite distinct from those in surface soils and vadose sediments due to geochemical differences (e.g., temperature, redox conditions, nutrient availability, and hydrogeology) (19–21). As an example, it was reported that psychrophiles tended to inhabit groundwater, whereas thermophiles and mesophiles were likely to be present in sediments (15). Hydrogeological processes may transport chemical energy in the form of gases (such as methane) (22), reduced metals (23), and nutrients (24) to the subsurface saturated zone. The capillary action and fluctuation of groundwater would cause the mixing of chemicals and microbes (15), thereby resulting in increased microbial biomass, diversity, and activity in the capillary fringe and groundwater zones, compared to the above vadose zone (12).

The Oak Ridge Reservation Field Research Center (ORR-FRC) in Oak Ridge, Tennessee, was established as a part of the U.S. Department of Energy's Natural and Accelerated Bioremediation Research (NABIR) program (25) to evaluate *in situ* strategies for the long-term treatment of mixed radionuclide wastes. The ORR-FRC includes five contaminated sites and an uncontaminated background site. In this study, we investigated the depth-wise profiles of microbial features and geochemistry in the SCZ of the ORR-FRC uncontaminated, pristine background site. We hypothesized that (i) microbial communities are highly localized with low transport among different zones in the SCZ; and (ii) the compositional and functional changes in microbial communities in the SCZ are constrained by geochemical gradients, even across close spatial scales. To prove these hypotheses, we collected sediment samples spanning from the shallow subsurface to the groundwater table (i.e., the saturated zone) at the ORR-FRC background site, and we conducted both

metagenomic and geochemical investigations, including organic matter characterization via ultrahigh resolution mass spectrometry, to draw connections between microbial community characteristics and geochemical gradients and to demonstrate the connection between genotype and ecotype. The nature and form of the C compounds as a function of depth are not well-characterized in subsurface systems, nor are the microbial mechanisms for its utilization (26). The results from this study provide a fundamental understanding of microbial ecology and biogeochemistry, which will benefit the future development of predictive models on nutrient turnover in the SCZ, especially at sites such as the ORR-FRC.

RESULTS

Metagenome assembly. In this study, we applied metagenomics to investigate depth-related variations in the microbial composition and metabolic potential in the SCZ of an ORR-FRC background site. The binning of bacterial and archaeal genomes from metagenomic data resulted in a set of 28 high-quality metagenome assembled genomes (MAGs) and 80 medium-quality MAGs, according to MIMAG standards (27). Assembling the reads from all of the samples, amounting to 187.7 Gbp of sequence data, resulted in 1.748 million contigs longer than 1 kbp. After quality filtering, we obtained a total of 215 small-subunits (SSUs) from the contigs of all of the sediment samples.

Distinct depth-wise microbial community composition. The microbial diversity (Shannon and Simpson) indexes showed no significant differences across the six sediment segments (one-way ANOVA, P value > 0.1) (Fig. S1), suggesting that the microbial communities were comparably diverse across depths. However, a phylogenetic analysis revealed distinct patterns in microbial community composition in different depths. Overall, the communities were dominated by bacteria, with an increasing relative abundance of archaea from $<1\%$ in the shallow subsurface to $32.7 \pm 6.5\%$ in the saturated zone (Fig. S2). This is also reflected in the bins obtained: few archaeal bins were obtained from the shallow subsurface, whereas approximately 20 to 30% of the bins from the saturated zone were classified as archaeal.

Specifically, the microbial communities clearly differed among depths at the phylum level (Fig. 1A), with a strong divide among the shallow subsurface (0.1 to 0.9 m below ground surface [bgs]), vadose zone (0.9 to 1.5, 1.5 to 2.4, 2.4 to 3.0 m bgs), capillary fringe (3.0 to 4.0 m bgs), and saturated zone (4.0 to 4.6 m bgs) on the basis of nonmetric multidimensional scaling (NMDS) (Fig. 1B–E). Out of the 14 classified bacterial phyla, Latescibacteria tended to occur in the shallow subsurface, Firmicutes tended to be present in the capillary fringe, and Patescibacteria and Bacteroidetes were present in the saturated zone (Fig. 1B–E). All three archaeal phyla, namely, Euryarchaeota, Thaumarchaeota, and Crenarchaeota, were relatively abundant in the capillary fringe and saturated zone. We noted the variance among replicates in two segments (2.4 to 3.0 and 3.0 to 4.0 m bgs) (Fig. 1A), which is probably due to the structural heterogeneity of the sediment environments that leads to the physical niche space and spatial isolation of microbial communities.

The further investigation of microbial community composition at the order level showed a clear depth-wise pattern (Fig. 2). The orders "uncultured AD3", Myxococcales, and Rokubacteriales were widely present ($>1\%$) across depths. The subgroup 13 and "uncultured GAL15" were generally abundant (2.7 to 7.7% and 3.3 to 14.4%, respectively) in the SCZ, except the shallow subsurface (not detected). Some orders occurred only in the shallow subsurface, including Pyrinomonadales, Micromonosporales, Microtrichales, Solirubrobacterales, "uncultured Latescibacteria", Tepidisphaerales, Dongiales, NB1-j, and Sphingomonadales, whereas some orders were only found in the capillary fringe and saturated zone, with most of these being uncultured or poorly characterized members ("uncultured Bathyarchaeia", "uncultured Marine Benthic Group A", and "uncultured JG30-KF-CM66"). Two orders of Chloroflexi (Ktedonobacterales and "uncultured AD3") particularly dominated the vadose zone (0.9 to 3.0 m bgs), representing 8.9 to 17.9% and 3.4 to 25.4% of the communities, respectively.

Distinct depth-wise microbial cooccurrence patterns. We generated predictions for the SSUs from the metagenome contigs, calculated associations between the SSUs,

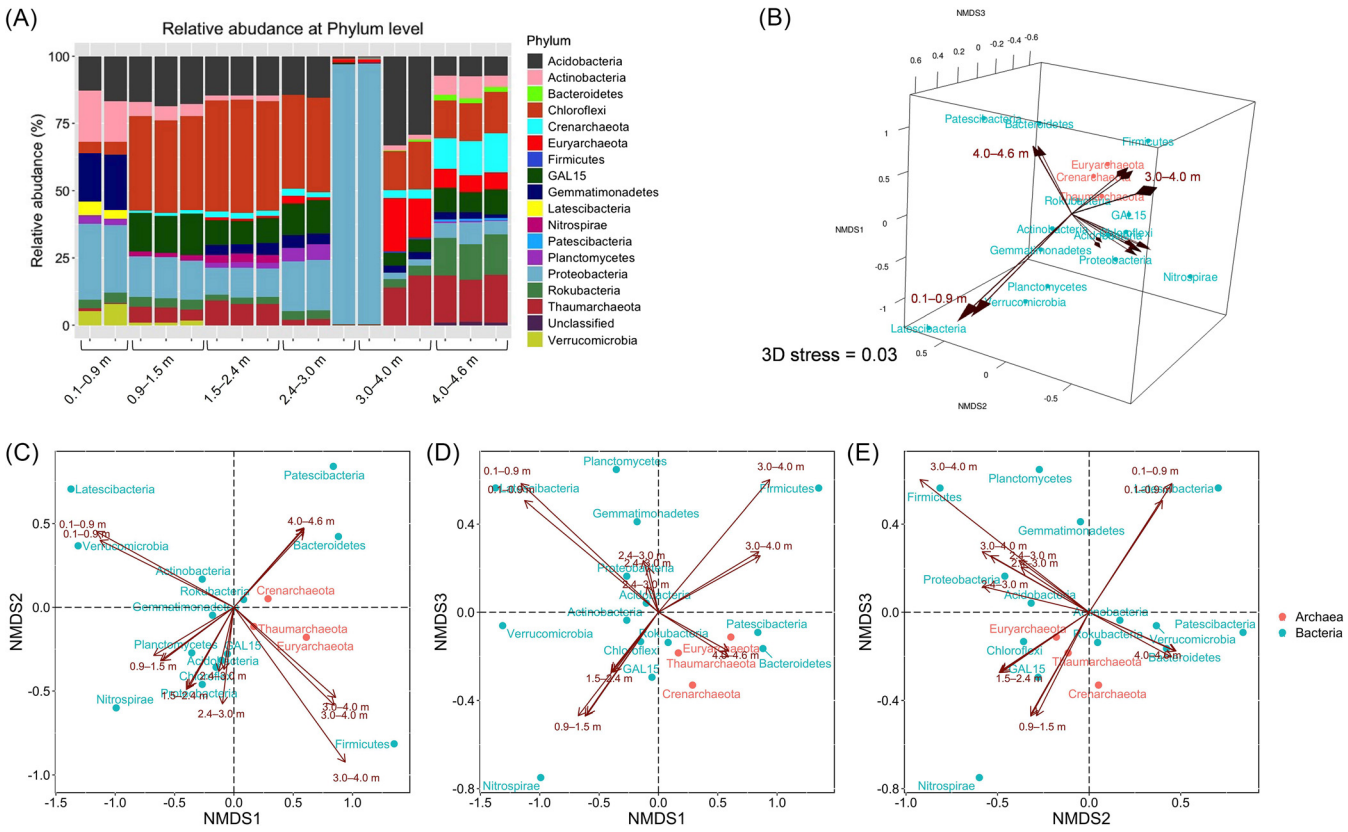

FIG 1 Microbial community compositions at the phylum level in the SCZ of the ORR-FRC background site. (A) Relative abundances of classified phyla in sediment samples. (B) 3D-nonmetric multidimensional scaling (NMDS) ordination with projections (C–E) of microbial community compositions. The arrows are replicates of each depth, and they indicate the directions in which the gradient changes of the classified phyla were the greatest.

and applied a network analysis based on strong (Spearman threshold > 0.6) and significant ($P < 0.05$) correlations to explore the cooccurrence patterns of the microbial communities along the depth of the SCZ. The resulting microbial cooccurrence networks (Fig. 3) consisted of 215 nodes (SSUs) and 3,821 edges, and they were clustered into 6 modules that represented 6 sediment segments, respectively (Table S1). The module representing the saturated zone possessed more nodes (n = 59) and edges (n = 1,711) than did other modules (nodes, 29 to 34; edges, 210 to 558), suggesting that the saturated zone had a more complex microbial cooccurrence network than did other sediment segments.

The cooccurrence networks (Fig. 3) demonstrate that the microbial communities at different depths are highly distinct in composition and are barely connected to each other, even between communities that are spatially adjacent. The shallow subsurface (0.1 to 0.9 m bgs) had no shared SSUs with other segments and exhibited a separated network. The up-vadose zone (0.9 to 1.5 m bgs) and mid-vadose zone (1.5 to 2.4 m) shared only one SSU, whereas communities at the variable groundwater table showed weak connections: the low-vadose zone (2.4 to 3.0 m bgs) shared one SSU with the capillary fringe (3.0 to 4.0 m bgs), which shared three SSUs with the adjacent saturated zone (4.0 to 4.6 m bgs). The existence of shared SSUs among these zones might be mediated by transport through the fluctuation of the groundwater table.

Depth-wise variations in microbial metabolic potential in the C and N cycles. Complex C oxidation (i.e., organic C $\rightarrow$ CO_2), was found to be most prevalent in the shallow subsurface (0.1 to 0.9 m bgs) and vadose zone (0.9 to 3.0 m bgs), whereas C fixation (i.e., CO_2 $\rightarrow$ organic C) was found to be most prevalent in the vadose zone (0.9 to 3.0 m bgs) (Fig. 4). Methane oxidation (i.e., CH_4 $\rightarrow$ CO_2) was most prevalent in the shallow subsurface and mid- to low-vadose zones (1.5 to 3.0 m bgs). The potential for methanogenesis from CO_2, acetate, and hydrogen to CH_4, while observed in all depths,

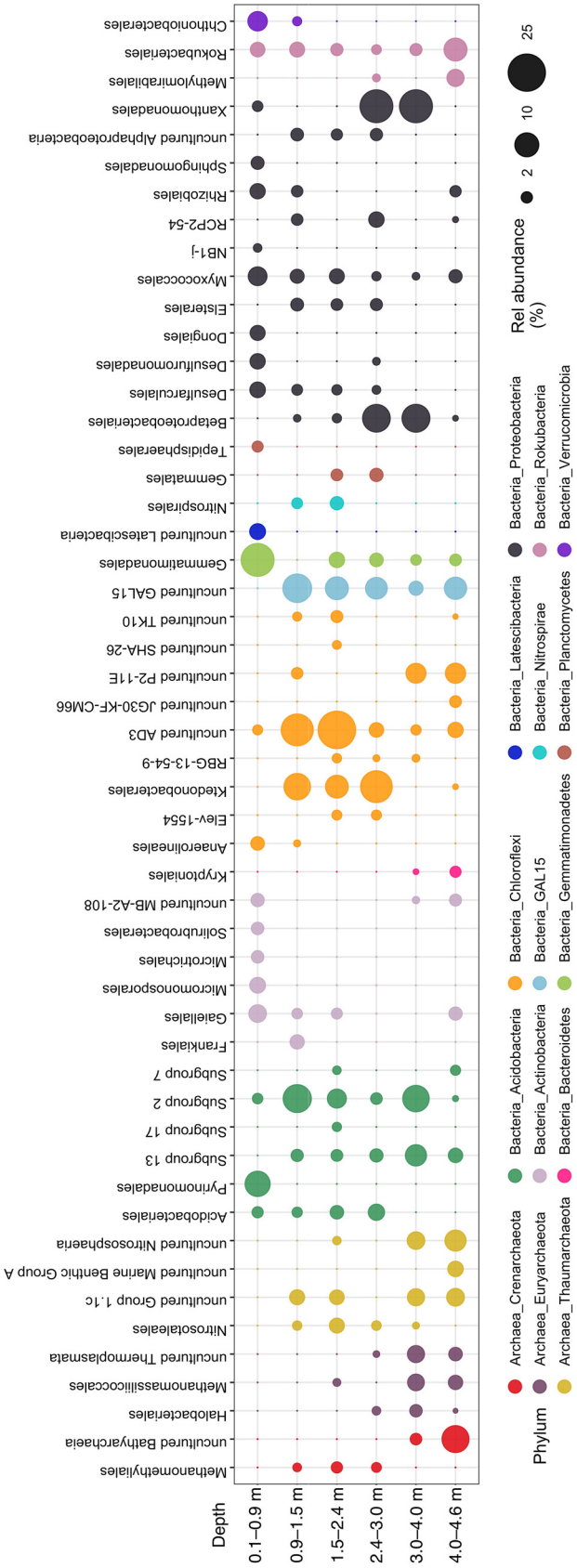

FIG 2 Bubble plot showing the mean values of the relative abundance (%) of the majority orders (>1% in any sediment segment) in the SCZ of the ORR-FRC background site.

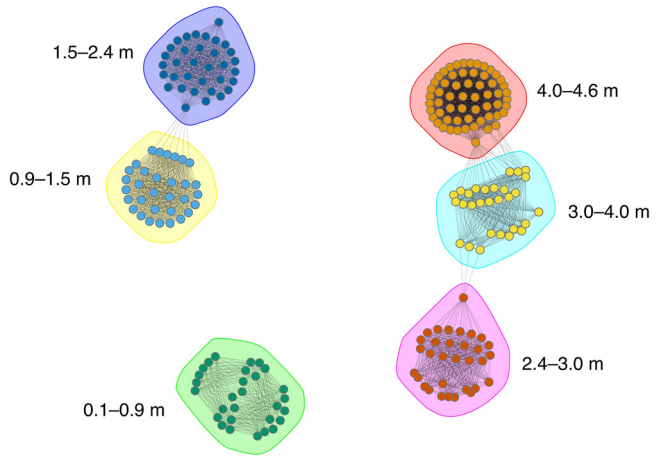

FIG 3 Layout of six clustered modules in the microbial cooccurrence networks of sediment microbes.

was found to be most prevalent in the low-vadose zone (2.4 to 3.0 m bgs) and in the saturated zone (4.0 to 4.6 m bgs). A notably lower abundance of C-cycling genes was observed in the capillary fringe (3.0 to 4.0 m bgs), which is possibly due to the fact that the dominant C metabolism in this zone was not captured by the marker genes that were used in this study. A deeper evaluation of the metagenomic data, including a genome-resolved analysis, will be performed in the future to investigate the low relative abundance of C-cycling genes in the capillary fringe.

Both the reduction of nitrate to nitrite and nitrite ammonification (i.e., $NO_2^- \rightarrow NH_4^+$) were most prevalent in the shallow subsurface (0.1 to 0.9 m bgs), whereas the reduction of nitrite to nitric oxide was least prevalent in that zone and most prevalent in the saturated zone (4.0 to 4.6 m bgs) (Fig. 4). N fixation (i.e., $N_2 \rightarrow NH_4^+$) was mostly observed in the shallow subsurface, but it was not detected or weak among other depths. AmoA, which is a marker gene for the oxidation of ammonia to nitrite, was more abundant in both the shallow subsurface and the vadose zone (0.9 to 3.0 m bgs) than in the deeper sediments.

Depth profiles of sediment geochemistry. We assessed a total of 27 geochemical factors (Table S2) to elucidate the factors that shape the microbial community composition in sediments. A statistical analysis shows that the microbial community composition strongly correlates (Spearman correlation, $r = 0.8461$) with a combination of 12 geochemical factors, including pH, cation exchange capacity (CEC), total organic carbon (TOC), dissolved organic carbon (DOC), total nitrogen (TN), nitrate, P, and five metals (Ca, Mg, Na, Ni, Zn). The results were validated via a Mantel test (Spearman's *r*, 0.6769; *P* value, 0.014; 719 permutations). The depth profiles of these selected geochemical factors are presented in Fig. 5.

The sediment pH decreased from neutral (pH = 6.48) in the shallow subsurface to weakly acidic (pH 3.93 to 4.41) in the layers beneath it. The CEC increased from 11 meq/100 g in the shallow subsurface and reached its peak (44.8 meq/100 g) in the mid-vadose zone, suggesting an increase of the clay content or a change in the clay type. As expected, the organic C content, including TOC and DOC, decreased sharply with depth, indicating an oligotrophic environment throughout the SCZ below the shallow subsurface.

The nitrate concentration in SCZ declined abruptly from the vadose zone to the capillary fringe and saturated zones (Fig. 5), indicating the potential presence of nitrate-reducing organisms and the enhanced denitrification potential of microbial communities in groundwater and aquifer sediments. Conversely, the concentrations of P and metals (Mg, Na, Ni, Zn, Ca) increased in the capillary fringe/saturated zone (Fig. 5), suggesting that groundwater may continuously introduce essential nutrients and metals that support the growth and activities of microbes in the oligotrophic subsurface environment.

Molecular composition of DOC. The water-extractable DOC represents a natural suite of organic substrates that microbes may easily access in sediment and is therefore relevant to microbial communities, despite being a small fraction of the bulk TOC (28, 29). Since the sediment DOC concentration shows a strong correlation with the microbial

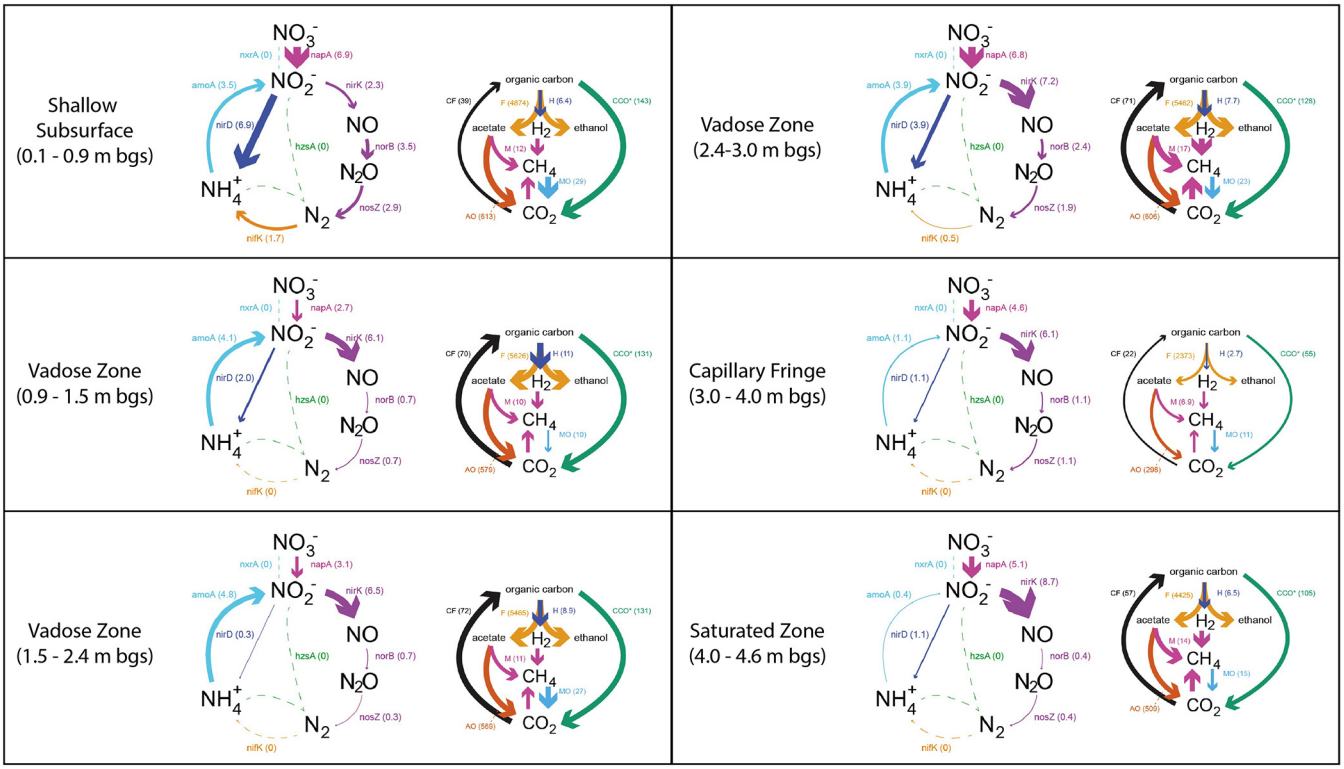

FIG 4 Nitrogen and carbon biogeochemical cycling diagrams profiling the community-level metabolic potential of the six sediment segments. The C cycle steps are categorized as follows: C fixation (CF), hydrogenesis (H), fermentation (F), methanogenesis (M), acetate oxidation (AO), methane oxidation (MO), and complex C oxidation (CCO). Listed next to each metabolic step are the normalized relative abundance of the genes capable of carrying out the step and the marker gene(s) that were used to quantify the step. The arrow sizes are drawn proportionally to the normalized relative abundance of the genes that are capable of carrying out the metabolic step. A dashed line indicates that no genes were identified for that step in the metagenome. The gene relative abundance was normalized to the number of reads binned to the metagenome. Lists of the marker genes for steps in which multiple genes were used as markers can be found in Table S4.

community composition, as discussed above, we further investigated the DOC composition at the molecular level using Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS). The chemical groups of DOC, including lignin, amino sugar, protein, lipid, carbohydrates, tannin, and condensed aromatics, were classified based on the assigned molecular formulae of masses from the FT-ICR MS analysis (28). As shown in Fig. 6, labile C, such as carbohydrate-like compounds, mainly existed in the shallow subsurface, with the average relative abundance decreasing from 21.2% in the shallow subsurface to $<$2% in the saturated zone, whereas the average relative abundance of recalcitrant C, such as condensed aromatic-like compounds, increased from 3.4% in the shallow subsurface to 12.3% in the saturated zone. Protein-like compounds existed consistently throughout the SCZ with a relative abundance of 2.1 to 5.4%, suggesting the origin of microbial cell death (30) throughout the depth.

DISCUSSION

Compared to surface soils, the SCZ is considered to be a unique ecosystem that usually has low concentrations of substrates and nutrients as well as low microbial biomass and activity (26, 31). Yet, it harbors a large number of distinct microbes that are, thus far, mostly uncultured or uncharacterized (6), and these assemble and function, depending to an extent on the stratigraphy, geochemistry, and hydrogeology of the site (31). In this study, we integrated metagenomics and geochemical analyses to elucidate how microbial community composition and metabolic potential are shaped and impacted by geochemical factors in the SCZ.

Our results demonstrate that the microbes in the uncontaminated ORR-FRC subsurface are highly localized and that communities are rarely interconnected. Spatially

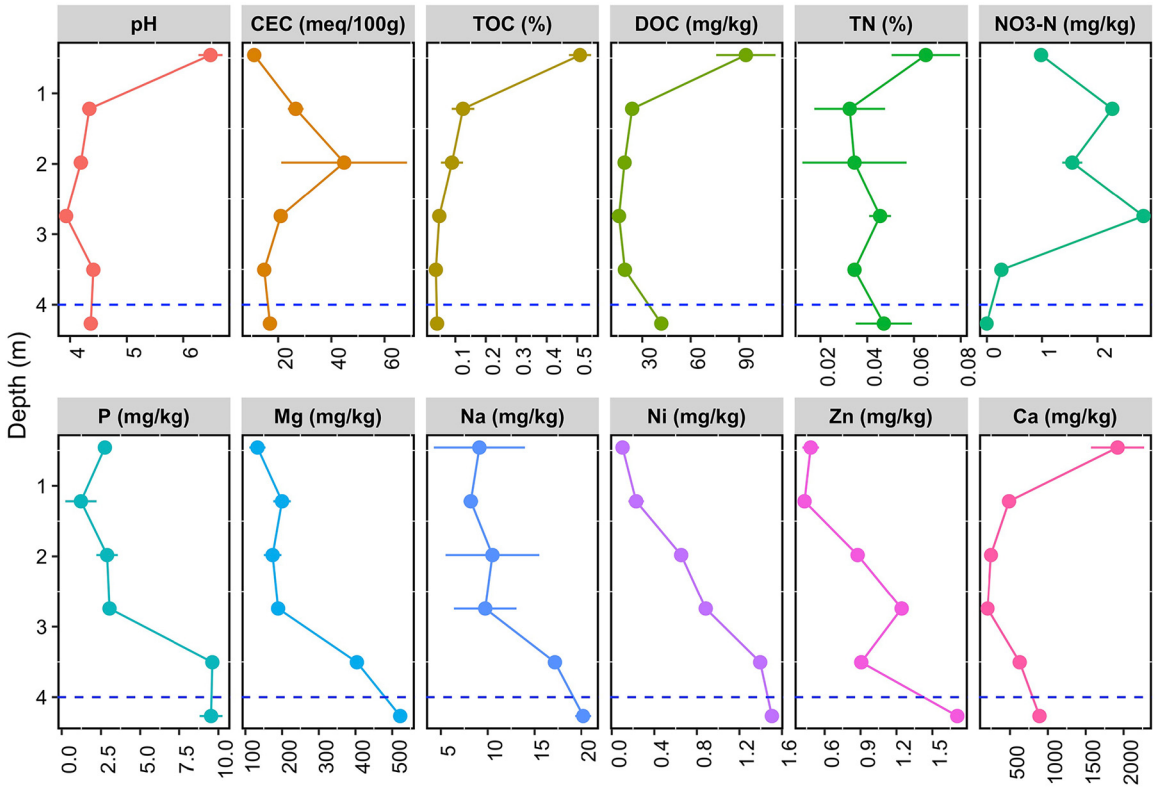

FIG 5 Depth profiles of selected geochemical factors in the SCZ of the ORR-FRC background site. The blue dashed line indicates the groundwater table.

localized subcommunities likely provide different "services" as resources change with depth and conditions become more selective. The community composition varies vertically from layer to layer, even over short distances. The significant Mantel test results indicate that the differences in microbial community composition among sediment segments are strongly correlated or, rather, "covary" with the differences in a subset of 12 environmental variables, including pH, CEC, TOC, DOC, TN, nitrate, P, and five metals (Ca, Mg, Na, Ni, Zn), suggesting that the sediment geochemistry is vital in the selection of the distinct microbial communities in the SCZ. Our observation with natural sediment significantly differs from those that were made with a model soil system, with those results indicating that the microbial community and its associated physiology were stronger drivers of DOC dynamics than was the associated mineralogy (30).

The unique microbial composition pattern along the depth of sediment shows a correlation with the quantity and quality of sediment organic matter. As the quantities of bulk TOC and DOC decline down the depth, the property of DOC also transitions toward recalcitrant C, and dominant microbes accordingly shift from copiotrophs to oligotrophs. Similar observations were made by Fierer et al. working with subsurface sediment in the Santa Ynez Valley (26) as well as by others with different sediment types, from permafrost to forest to coastal environmental sites (32, 33, 34) The phyla Actinobacteria, Latescibacteria, and Verrucomicrobia, which were found mostly in the shallow subsurface and upper vadose zone (Fig. 1), significantly correlated with the labile DOC components such as carbohydrates, amino sugar, and tannin (Spearman correlation, $r > 0.6$; $P < 0.05$) (Fig. 6B), suggesting that these phyla appear to be copiotrophic organisms and therefore tend to grow in the shallow subsurface, where labile C is relatively abundant. For example, although the order Chthoniobacterales (phylum: Verrucomicrobia) detected in this study is not well-described in the literature, other members of Verrucomicrobia have been reported to be "cosmopolitans" in the rhizosphere (35) and highly prevalent in soils (36), suggesting that Verrucomicrobia tend to live in environments that are rich in labile C,

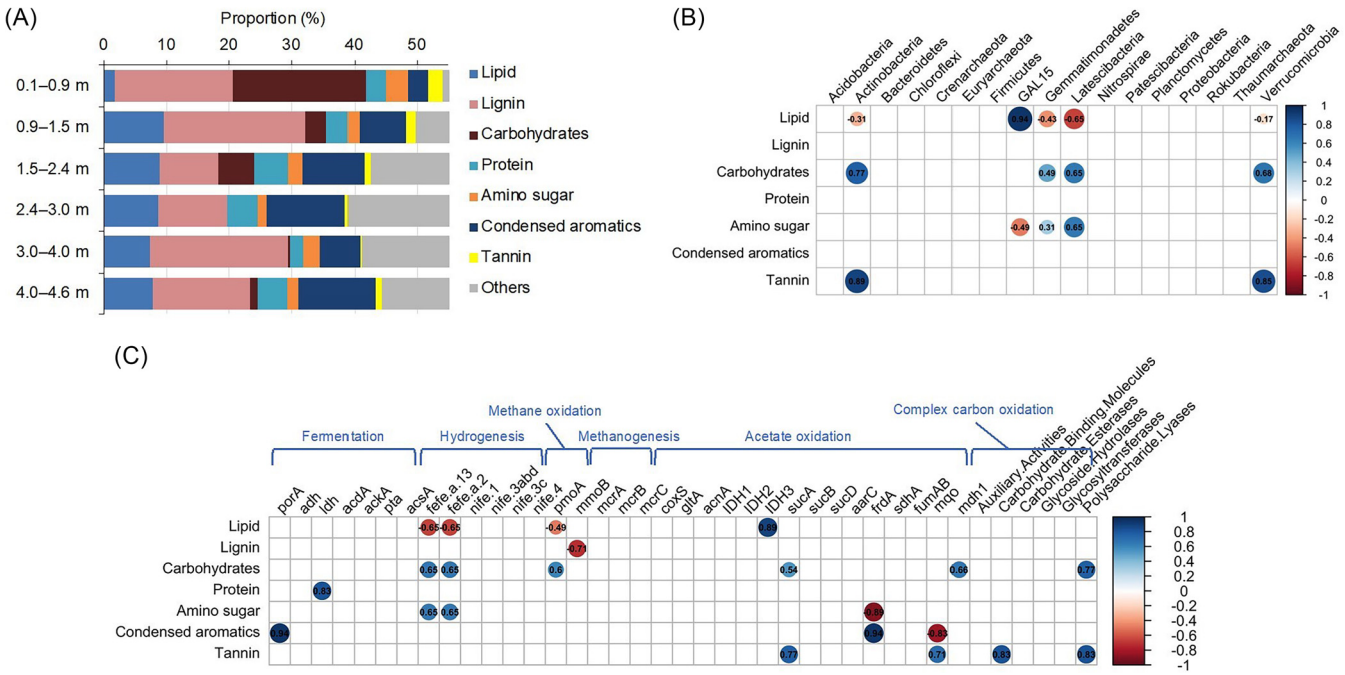

FIG 6 (A) Biochemical composition of the sediment DOC, as measured via FT-ICR MS. The relative proportion is the mean value of the replicates. "Others" refers to the rest of the detected masses that were assigned molecular formulas but could not be classified as any of the biochemical groups. (B) Correlation matrix of the relative abundances of the DOC biochemical groups and the relative abundances of the classified phyla. (C) Correlation matrix of the relative abundances of the DOC biochemical groups and the relative abundances of the marker genes that are related to C-cycling. In panels B and C, the numbers indicate Spearman correlation coefficients, and only statistically significant correlations with $P < 0.05$ are shown.

such as root exudates. This is in agreement with our recent published study, in which we found that Verrucomicrobia was highly enriched by DOC that was extracted from shallow subsurface sediment (1 m bgs) collected from the ORR-FRC background site (37). On the other hand, members of the phyla Euryarchaeota, Thaumarchaeota, Crenarchaeota, Acidobacteria, Chloroflexi, GAL15, and Rokubacteria, most of which belong to uncultured/uncharacterized clades, dominate in the layers below the shallow subsurface (Fig. 2), suggesting that these organisms might be capable of utilizing relatively recalcitrant substrates. As an example, the phylum GAL15 was found to significantly correlate with the relatively recalcitrate, lipid-like compounds in the subsurface (Spearman correlation, $r = 0.94$; $P < 0.05$) (Fig. 6B). Members of the order Ktedonobacterales (phylum: Chloroflexi) were reported to preferentially predominate in oligotrophic and extreme environments (38), which could be explained by their large genomes (7.7 to 13.7 Mb) and broad metabolic potential (39).

Archaea were found to be abundant in the capillary fringe and saturated zone, but they rarely exist in the shallow layer (Fig. 2), probably because of the low-carbon and low-oxygen environments in the subsurface. Many members of Crenarchaeota were identified in anaerobic environments (39), and the orders Methanomethyliales and Methanomassiliicoccales are known to be methanogenic anaerobic archaea (40, 41). Thus far, the characterized representatives of Thaumarchaeota are chemolithotrophs with an oligophilic lifestyle (42–44), which may explain why this phylum can thrive in the oligotrophic subsurface, as observed in this study.

In accordance with the changes in the microbial composition, the microbial C and N metabolic potential also changes with the depth. The shallow subsurface is most influenced by the spatially close rhizosphere, where N-fixing organisms, such as Sphingomonas harboring a nitrogenase gene (45), are known to thrive. We found that members of the order Sphingomonadales were exclusively present in the shallow layer (Fig. 2), which might explain the observed greater N fixation potential of the community in the shallow layer, compared to those in deeper sediments (Fig. 4). As discussed above, the C cycling genes related to complex C oxidation and methane oxidation were

found to be mostly prevalent in the shallow subsurface and vadose zone, where labile C is relatively abundant. We found that some of these genes and two labile DOC components, namely, carbohydrates and tannin, were significantly correlated (Spearman correlation, $r > 0.6$; $P < 0.05$) (Fig. 6C), suggesting that these two microbial metabolic processes are more likely to be influenced and regulated by these labile C in the subsurface. On the other hand, members of the genes related to acetate oxidation were found to be significantly correlated with diverse types of DOC, including labile carbohydrates and tannin as well as recalcitrant lipid and condensed aromatics (Spearman correlation, $r > 0.5$; $P < 0.05$) (Fig. 6C), suggesting that microbes may utilize a diverse group of OC to fuel this process in the subsurface. In the deep saturated zone, where the environment has less oxygen and more recalcitrant C, compared to the layers above, the acetate could be produced from the anaerobic fermentation of recalcitrant C, as suggested by the strong correlation between the fermentation related gene *porA* and recalcitrant condensed aromatics (Spearman correlation, $r = 0.94$; $P < 0.05$) (Fig. 6C). Most of the C-cycling pathways, both oxidizing and reducing, were present across all depths of the sediment core, highlighting the microheterogeneity in the sediment environments due to the differing redox potentials required to support these diverse metabolisms. For example, methanogenesis genes, though present across all depths, were most abundant in the deep vadose and saturated zones, whereas methane oxidation was most prevalent in the shallow subsurface and mid-vadose zones, suggesting the formulation of discrete zones with both adequate methane production and redox potential to fuel these metabolisms.

In the SCZ, the fluctuation of groundwater introduces nutrients and solutes that benefit the growth of microbes, as reflected by the increased concentrations of P and metals (Fig. 5) as well as the increased metabolic potential in the saturated zone, compared to those in the capillary fringe (Fig. 4). These changes in microbial communities may lead to the enhanced denitrification that was observed in the saturated zone. As shown in Fig. 4, the normalized abundances of nitrate and nitrite reduction genes (*napA* and *nirK*) in the saturated zone are higher than those in the capillary fringe, which may explain the decreased nitrate concentration in the saturated zone (Fig. 5). We also observed a notably lower ammonia oxidation potential in the capillary fringe and saturated zone, compared to the shallow subsurface and vadose zone (Fig. 4), which may be a result of complications in annotating *amoA* versus *pmoA* genes using HMM approaches, as has been previously reported (46, 47). Further development in gene annotation techniques is needed to resolve issues such as these.

The SCZ is a dynamic and critical zone that plays an important role in ecological C and N cycles. This study demonstrates that the highly localized and barely interconnected subsurface microbial communities, including many uncultured/uncharacterized clades, vary widely in composition and metabolic potential, regarding C and N cycling along the depth. Overall, our research demonstrates that sediment geochemistry and hydrogeology are vital in the selection of distinct microbial populations and metabolic potential in different depths of subsurface terrestrial sediment. It highlights the microbial members that are vital in these biogeochemical processes as well as certain geochemical factors, including specific classes of sediment DOC that regulate and select for these unique microbes in the SCZ. These results of environmental constraints on microbial assembly and metabolic potential are critical in the enhancement of our predictive understanding of subsurface ecosystem function and resilience in terrestrial subsurface environments.

MATERIALS AND METHODS

Study site and sample collection. This study was conducted on the subsurface zone at the ORR-FRC background area. The ORR-FRC has a complex geochemical and hydrological structure that has been probed with more than 800 groundwater wells that were drilled in both contaminated and uncontaminated background areas (48).

We collected sediment samples from a borehole that was drilled at the well FW306 at ORR-FRC in June of 2015. The borehole was advanced using a dual tube (DT22) direct-push Geoprobe drill rig. During dual tube sampling, one set of rods was driven into the ground as an outer casing, which

received the driving force from the hammer. It also provided a sealed casing through which undisturbed sediment samples were recovered using disposable polyvinyl chloride liners (152.4 cm length, 2.86 cm I.D.) that were attached to 3.18 cm (outside diameter) inner rods. Six sediment segments were taken at different depths, including the shallow subsurface (0.1 to 0.9 m below ground surface [bgs]), vadose zone (0.9 to 1.5 m, 1.5 to 2.4 m, 2.4 to 3.0 m bgs), capillary fringe (3.0 to 4.0 m bgs), and saturated zone (4.0 to 4.6 m bgs). The sediment segments were immediately cut, capped, and stored under a nitrogen atmosphere at 4°C and shipped immediately to the lab on dry ice. On receipt, the samples were homogenized in a mixing tray with a sterile plastic liner. The subsamples were then preserved at −80°C for further analyses. For each sediment segment, two or three replicates were taken for DNA extraction (Table S3) as well as geochemical analyses.

DNA extraction from sediment samples. Multiple DNA extraction kits and methods were tested to extract the largest amount of DNA for our type of sample without sacrificing DNA quality (checked on a Bioanalyzer). Our samples often had a high clay composition, which can interfere with DNA extraction. The best results came from a method that was similar to that which was used by Hug et al., 2015 (49), and Kantor et al., 2013 (50). Briefly, genomic DNA was extracted from homogenized sediment using a Qiagen PowerMax Soil Kit (Qiagen, USA), according to the manufacturer's instructions, with modifications. For each replicate, 10 g of homogenized sediment was used, but it was split into two 5 g portions as input to the PowerMax Soil Kit. Extraction with water was conducted as a control. As a modification, after the addition of Solution C1 of the Soil Kit, the sample was subjected to a freeze-thaw and was vortexed at max speed for 3.5 min after thawing. The samples were then incubated in a 65°C water bath for 30 min and mixed by inversion every 10 min. The rest of the DNA extraction was followed, according to the manufacturer's instructions, except during the elution step. For eluting DNA from the filter, 5 mL of Solution C6 that had been prewarmed to 55°C was added to the Spin Filter membrane. The filter was incubated for a minute, and then the DNA was eluted via centrifugation into a low retention tube. This elution was repeated to improve the yield.

To concentrate the DNA sample via ethanol precipitation, a 1/10 volume of sodium acetate solution (3 M, pH 5.2, with 0.4 mg/mL glycogen) was added to each sample. Then, 2.5 volumes of chilled 100% ethanol were added to the solution, mixed by inversion, and incubated overnight at 4°C. The sample was allowed to warm to room temperature and was centrifuged for 30 min at 15,000 $\times$ *g* to pellet the DNA. Overnight incubation as well as a longer centrifugation time at room temperature with a higher force would provide more time for most of the DNA fragments to pellet (51). The DNA pellet was washed with 10 mL of 70% ethanol and was resuspended in sterile nuclease-free water overnight. The DNA concentration was measured using a Qubit 3 Fluorometer (Thermo Fisher Scientific, USA). Detailed information about the amount of DNA that was extracted from each sample can be found in Table S1.

Library preparation and metagenomics sequencing. The sequencing libraries were prepared using an Illumina DNA Prep Kit (Illumina, USA), according to the manufacturer's instructions. The library concentration was measured using a Qubit 3 Fluorometer (Thermo Fisher Scientific, USA), and the library fragment size distribution was collected using a Bioanalyzer High Sensitivity Kit (Agilent, USA), according to the manufacturer's instructions. The samples were sequenced by Novogene Corporation, Inc. (CA, USA), using 2 $\times$ 150 bp on an Illumina NovaSeq 6000 (Illumina, USA).

Metagenomics read processing and genome assembly. To improve the assemblies, we performed coassemblies of the replicates for each segment. That is, for each sediment segment, the reads from replicates were combined and used as the input for assembly. The reads were processed and assembled as previously described (52). Briefly, the metagenomic reads were preprocessed using BBtools version 38.60 (https://jgi.doe.gov/data-and-tools/bbtools/) to remove Illumina adapters, perform quality filtering and trimming, and remove PhiX174 spike-ins. Assembly was performed using SPAdes (53, 54) version 3.13.0 with the following parameters: *-meta -k 21,33,55,77,99,127*. Following assembly, from BWA version 0.7.17-r1188 (55), we used the BWA-MEM algorithm with the default parameters to map the reads to the set of contigs that was produced by the assembly. We did this to obtain the BAM file that was required by MetaBAT 2 version 2.0 (56). We used MetaBAT 2 with the parameters *-unbinned -minContig 1500 -maxEdges 500* to bin the contigs. The parameters set were as follows: *-NW:cac = warn*, *-CO:fnic = yes -AS:nop = 6:sdlpo = no -KS:fenn = 0.3*. We used Pilon version 1.23 (57) with the default parameters to run final read coherence checks and clean up minor indels. CheckM was used to assess the bin completeness and contamination (58), and the output can be found in Table S2. The taxonomic classification of the metagenome bins was generated using GTDB-Tk version 1.5 (59), and it can be found in Table S4 in Supplemental file 2. Bins with less than 10% completeness were excluded from further analysis.

Gene annotation, taxonomic assignment, and metagenomics analysis. Small subunit rRNA (SSU rRNA) gene predictions were generated by Infernal 1.1.2 (60), using *cmsearch* with the parameters *-notextw -cut_tc* on the contigs from the metagenomics assembly. Only the 16S rRNA genes were used for further analysis. We used RFAM version 14.1 for the rRNA models (archaeal and bacterial) and kept the hits with bitscores that were greater than or equal to 1,000. We used VSEARCH to dereplicate the sequences (61). The taxonomic classification was carried out by loading the sequences into QIIME2 (62) v2019.7 and using the naive Bayes classifier that was built from the SILVA132 (63) database. SSUs that were classified as chloroplasts or mitochondria were removed from the downstream analyses. The coverage of the SSUs were calculated by finding the contigs on which they were found and calculating the average coverage of the contigs. The coverage and taxonomic information can be found in Table S5 in Supplemental file 3. See Supplemental file 5 for the fasta sequences of the SSUs.

The protein-coding genes were annotated via eggnog-mapper (v2) (64). The carbohydrate-active enzyme-coding genes were annotated via dbCAN2 (65). The genomic capacity of each community for N and C cycling was evaluated. Marker genes were identified to represent each step of the N and C cycles. In

each metagenome, the number of times that each marker gene appeared was tabulated and normalized via the metagenome sequence coverage (per 10^8 mapped reads) to represent the relative abundance of the marker gene in the microbial community. For N cycling, the marker genes napA, nirK, norB, nosZ, nifK, amoA, nxrA, nirD, and hzsA were used to represent their respective steps of the N cycle. The C cycle steps were delineated into C fixation, complex C oxidation, fermentation, hydrogenesis, acetate oxidation, methane oxidation, and methanogenesis. For these categories, multiple marker genes were tabulated, normalized, and summed to obtain the final relative abundance. The genes that were utilized as marker genes for each of these steps, as well as the normalized gene abundance, are summarized in Table S6 in Supplemental file 4.

Geochemical analyses of sediments. A total of 26 geochemical factors, including pH, cation exchange capacity (CEC), 14 metals (Ca, Cd, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Pb, Zn, U), bulk total organic carbon (TOC), bulk total inorganic carbon (TIC), bulk total nitrogen (TN), nitrate, ammonium, phosphorus, chloride, fluorine, phosphate, and sulfate, were measured by the Agricultural and Environmental Services Laboratories at the University of Georgia (http://aesl.ces.uga.edu).

Sediment water-extractable dissolved organic carbon (DOC) was extracted and measured in our lab, and it was then characterized via ultrahigh resolution mass spectrometry, namely, the Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) that was described previously (28). Briefly, the sediment samples were freeze-dried and extracted using Milli-Q water (18.2 MΩ·cm, 0.22 μm membrane filtered) via rotary shaking overnight at 35°C, and this was followed by sonication in a water bath for 2 h. The ratio of water to sediment was 4:1 (wt/wt). The sediment extracts were centrifuged at 6,000 $\times$ *g* for 20 min. Then, the supernatant was decanted and filtered through a polycarbonate filter (0.2 μm pore-sized, Whatman), and this was followed by a second filtration step using a polyethersulfone filtration system (0.22 μm pore size, Corning). For the FT-ICR MS analysis, an aliquot of the filtrate (15 mL) was freeze-dried, redissolved in 1 mL of methanol (HPLC grade, Fisher Scientific), and then filtered using a polytetrafluoroethylene filter (0.2 μm pore size, Pall Corporation) prior to its injection into an FT-ICR MS system. The details of the FT-ICR MS instrumental analysis and the data processing have been described previously (28, 66).

Data analysis, network generation, and visualization. The processing of the FT-ICR MS data set, microbial community analysis, and all statistical analyses were performed in R v3.6.1 and RStudio. The generation of the abundance plots, 3D ordination, calculation of the Shannon and Simpson indices, and Mantel tests were conducted using the "*vegan*" package in R. The selection of the best subset of environmental variables that was correlated with the microbial community dissimilarities was performed using BIOENV, as implemented in the bio.env function in the "*vegan*" package, using the Bray-Curtis index and a Euclidean distance matrix. The correlation matrix was computed in R, using Spearman as the method, and it was visualized using the "*corrplot*" package with a significance level of 0.05.

The abundance matrix of all 215 SSUs from all of the sediment samples was loaded into the CoNet ensemble app in Cytoscape v. 3.7.2 for the network generation (67). The Spearman correlation coefficient between two SSUs was considered to be statistically robust if the threshold value was $>$0.6 with a corresponding *P* value of $<$0.05. The generated network was then loaded into R for subsequent analysis and visualization. The cooccurrence networks, with each node representing one SSU, each edge denoting a strong and significant correlation, and clustering being used to generate modules that are densely connected within themselves but sparsely connected to others, were built and visualized using the "*igraph*" package in R (67, 68).

Data availability. The raw reads and assemblies have been submitted to NCBI under the BioProject accession number PRJNA962994. The coassemblies of the replicates for each sediment segment (FW306_01, FW306_02, etc.) have the BioSample IDs SAMN34430745, SAMN34590182, SAMN34590183, SAMN34590184, SAMN34590185, and SAMN34590186.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, DOCX file, 0.5 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.03 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 4, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 5, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

This research by ENIGMA – Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under contract number DE-AC02-05CH11231. A portion of the research was performed using EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

REFERENCES

1. Science BROiE. 2001. Basic research opportunities in earth science. Available online at: https://www.nap.edu/read/9981/chapter/1#x.
2. The U.S. NSF-supported national CZO Program. https://czo-archive.criticalzone.org/national/research/the-critical-zone-1national/.
3. Hahm WJ, Rempe DM, Dralle DN, Dawson TE, Lovill SM, Bryk AB, Bish DL, Schieber J, Dietrich WE. 2019. Lithologically controlled subsurface critical zone thickness and water storage capacity determine regional plant community composition. Water Resour Res 55:3028–3055. https://doi.org/10.1029/2018WR023760.
4. Flemming HC, Wuertz S. 2019. Bacteria and archaea on Earth and their abundance in biofilms. Nat Rev Microbiol 17:247–260. https://doi.org/10.1038/s41579-019-0158-9.
5. Lohmann P, Benk S, Gleixner G, Potthast K, Michalzik B, Jehmlich N, von Bergen M. 2020. Seasonal patterns of dominant microbes involved in central nutrient cycles in the subsurface. Microorganisms 8:1694. https://doi.org/10.3390/microorganisms8111694.
6. Lloyd KG, Steen AD, Ladau J, Yin JQ, Crosby L. 2018. Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. Msystems 3. https://doi.org/10.1128/mSystems.00055-18.
7. Steen AD, Crits-Christoph A, Carini P, DeAngelis KM, Fierer N, Lloyd KG, Thrash JC. 2019. High proportions of bacteria and archaea across most biomes remain uncultured. ISME J 13:3126–3130. https://doi.org/10.1038/s41396-019-0484-y.
8. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, Klingl A, Woyke T, Ryan MC, Banfield JF. 2018. Differential depth distribution of microbial function and putative symbionts through sediment- hosted aquifers in the deep terrestrial subsurface. Nat Microbiol 3:328–336. https://doi.org/10.1038/s41564-017-0098-y.
9. Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, Wanner G, Klingl A, Berg IA, Yoshinaga M, Viehweger B, Hinrichs K-U, Thomas BC, Meck S, Auerbach AK, Heise M, Schintlmeister A, Schmid M, Wagner M, Gribaldo S, Banfield JF, Moissl-Eichinger C. 2014. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. Nat Commun 5. https://doi.org/10.1038/ncomms6497.
10. Bundt M, Widmer F, Pesaro M, Zeyer J, Blaser P. 2001. Preferential flow paths: biological 'hot spots' in soils. Soil Biology & Biochemistry 33:729–738. https://doi.org/10.1016/S0038-0717(00)00218-2.
11. Kuzyakov Y, Blagodatskaya E. 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology & Biochemistry 83:184–199. https://doi.org/10.1016/j.soilbio.2015.01.025.
12. Holden PA, Fierer N. 2005. Microbial processes in the vadose zone. Vadose Zone j 4:1–21. https://doi.org/10.2136/vzj2005.0001.
13. Smith HJ, Zelaya AJ, De Leon KB, Chakraborty R, Elias DA, Hazen TC, Arkin AP, Cunningham AB, Fields MW. 2018. Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments. FEMS Microbiol Ecol 94. https://doi.org/10.1093/femsec/fiy191.
14. Tecon R, Or D. 2017. Biophysical processes supporting the diversity of microbial life in soil. FEMS Microbiol Rev 41:599–623. https://doi.org/10.1093/femsre/fux039.
15. Glamoclija M, Ramirez S, Sirisena K, Widanagamage I. 2019. Subsurface microbial ecology at sediment-groundwater interface in sulfate-rich playa; White Sands National Monument, New Mexico. Front Microbiol 10:2595. https://doi.org/10.3389/fmicb.2019.02595.
16. Colman DR, Poudel S, Stamps BW, Boyd ES, Spear JR. 2017. The deep, hot biosphere: twenty-five years of retrospection. Proc Natl Acad Sci U S A 114:6895–6903. https://doi.org/10.1073/pnas.1701266114.
17. Sirisena KA, Ramirez S, Steele A, Glamoclija M. 2018. Microbial diversity of hypersaline sediments from Lake Lucero playa in White Sands National Monument, New Mexico, USA. Microb Ecol 76:404–418. https://doi.org/10.1007/s00248-018-1142-z.
18. 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 Biology & Biochemistry 50:58–65. https://doi.org/10.1016/j.soilbio.2012.03.011.
19. Le Lay H, Thomas Z, Rouault F, Pichelin P, Moatar F. 2019. Characterization of diffuse groundwater inflows into streamwater (part I: spatial and temporal mapping framework based on fiber optic distributed temperature sensing). Water 11:2389. https://doi.org/10.3390/w11112389.
20. Hansen B, Voutchkova DD, Sandersen PBE, Kallesøe A, Thorling L, Møller I, Madsen RB, Jakobsen R, Aamand J, Maurya P, Kim H. 2021. Assessment of complex subsurface redox structures for sustainable development of agriculture and the environment. Environ Res Lett 16:025007. https://doi.org/10.1088/1748-9326/abda6d.
21. Gushgari-Doyle S, Lui LM, Nielsen TN, Wu X, Malana RG, Hendrickson AJ, Carion H, Poole FL, Adams MW, Arkin AP, Chakraborty R. 2022. Genotype to ecotype in niche environments: adaptation of Arthrobacter to carbon availability and environmental conditions. Isme Commun 2:1–10. https://doi.org/10.1038/s43705-022-00113-8.
22. Klazinga DR, Steelman CM, Cahill AG, Walton KM, Endres AL, Parker BL. 2019. Methane gas transport in unconfined aquifers: a numerical sensitivity study of a controlled release experiment at CFB Borden. J Contaminant Hydrology 225:103506. https://doi.org/10.1016/j.jconhyd.2019.103506.
23. Ma B, Charlet L, Fernandez-Martinez A, Kang ML, Made B. 2019. A review of the retention mechanisms of redox-sensitive radionuclides in multi-barrier systems. Applied Geochemistry 100:414–431. https://doi.org/10.1016/j.apgeochem.2018.12.001.
24. Hoffmann CC, Berg P, Dahl M, Larsen SE, Andersen HE, Andersen B. 2006. Groundwater flow and transport of nutrients through a riparian meadow - Field data and modelling. J Hydrology 331:315–335. https://doi.org/10.1016/j.jhydrol.2006.05.019.
25. Kim Y-J, Moon J-W, Roh Y, Brooks SC. 2009. Mineralogical characterization of saprolite at the FRC background site in Oak Ridge, Tennessee. Environ Geol 58:1301–1307. https://doi.org/10.1007/s00254-008-1633-1.
26. Fierer N, Schimel JP, Holden PA. 2003. Variations in microbial community composition through two soil depth profiles. Soil Biology & Biochemistry 35:167–176. https://doi.org/10.1016/S0038-0717(02)00251-1.
27. 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, Eren AM, Genome Standards Consortium et al. 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.
28. Wu X, Wu L, Liu Y, Zhang P, Li Q, Zhou J, Hess NJ, Hazen TC, Yang W, Chakraborty R. 2018. Microbial interactions with dissolved organic matter drive carbon dynamics and community succession. Front Microbiol 9. https://doi.org/10.3389/fmicb.2018.01234.
29. Guigue J, Leveque J, Mathieu O, Schmitt-Kopplin P, Lucio M, Arrouays D, Jolivet C, Dequiedt S, Prevost-Boure NC, Ranjard L. 2015. Water-extractable organic matter linked to soil physico-chemistry and microbiology, at the regional scale. Soil Biology & Biochemistry 84:158–167. https://doi.org/10.1016/j.soilbio.2015.02.016.
30. Kallenbach CM, Frey SD, Grandy AS. 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nat Commun 7. https://doi.org/10.1038/ncomms13630.
31. Beloin RM, Sinclair JL, Ghiorse WC. 1988. Distribution and activity of microorganisms in subsurface sediments of a pristine study site in Oklahoma. Microb Ecol 16:85–97. https://doi.org/10.1007/BF02097407.
32. Mary-Cathrine Leewis, Corey R. Lawrence, Marjorie S. Schulz, Malak M. Tfaily, Christian Orlando Ayala-Ortiz, Gilberto E. Flores, Rachel Mackelprang, Jack W. McFarland. 2022. The influence of soil development on the depth distribution and structure of soil microbial communities. Soil Biology and Biochemistry 174:108808. https://doi.org/10.1016/j.soilbio.2022.108808.
33. Xin L, Yuanquan C, Dalong M, Dandan S, Lin L. 2022. Vertical distribution of bacterial community diversity in the Greater Khingan Mountain permafrost region. Ecology and Evolution 12. https://doi.org/10.1002/ece3.9106.
34. Xiao C, Jiguang F, Zongju D, Mao T, Biao Z. 2022. Changes in soil total, microbial and enzymatic C-N-P contents and stoichiometry with depth and latitude in forest ecosystems. Science of The Total Environment 816:151583. https://doi.org/10.1016/j.scitotenv.2021.151583.
35. Bunger W, Jiang X, Muller J, Hurek T, Reinhold-Hurek B. 2020. Novel cultivated endophytic Verrucomicrobia reveal deep-rooting traits of bacteria to associate with plants. Sci Rep 10. https://doi.org/10.1038/s41598-020-65277-6.
36. Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA, Knight R, Fierer N. 2011. The under-recognized dominance of Verrucomicrobia in soil bacterial communities. Soil Biol Biochem 43:1450–1455. https://doi.org/10.1016/j.soilbio.2011.03.012.

37. Wu X, Spencer S, Gushgari-Doyle S, Yee M, Voriskova J, Li Y, Alm E, Chakraborty R. 2020. Culturing of "unculturable" subsurface microbes: natural organic carbon source fuels the growth of diverse and distinct bacteria from groundwater. Front Microbiol 11:610001. https://doi.org/10.3389/fmicb.2020.610001.
38. Zheng Y, Saitou A, Wang C-M, Toyoda A, Minakuchi Y, Sekiguchi Y, Ueda K, Takano H, Sakai Y, Abe K, Yokota A, Yabe S. and2019. Genome features and secondary metabolites biosynthetic potential of the class Ktedonobacteria. Front Microbiol 10. https://doi.org/10.3389/fmicb.2019.00893.
39. Collins G, O'Connor L, Mahony T, Gieseke A, de Beer D, O'Flaherty V. 2005. Distribution, localization, and phylogeny of abundant populations of Crenarchaeota in anaerobic granular sludge. Appl Environ Microbiol 71:7523–7527. https://doi.org/10.1128/AEM.71.11.7523-7527.2005.
40. Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, Tyson GW. 2016. Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. Nat Microbiol 1. https://doi.org/10.1038/nmicrobiol.2016.170.
41. Borrel G, Parisot N, Harris HM, Peyretaillade E, Gaci N, Tottey W, Bardot O, Raymann K, Gribaldo S, Peyret P, O'Toole PW, Brugère J-F. and2014. Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics 15. https://doi.org/10.1186/1471-2164-15-679.
42. Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437:543–546. https://doi.org/10.1038/nature03911.
43. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, Brochier-Armanet C, Chain PSG, Chan PP, Gollabgir A, Hemp J, Hügler M, Karr EA, Könneke M, Shin M, Lawton TJ, Lowe T, Martens-Habbena W, Sayavedra-Soto LA, Lang D, Sievert SM, Rosenzweig AC, Manning G, Stahl DA. and2010. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proc Natl Acad Sci U S A 107:8818–8823. https://doi.org/10.1073/pnas.0913533107.
44. Konneke M, Schubert DM, Brown PC, Hugler M, Standfest S, Schwander T, von Borzyskowski LS, Erb TJ, Stahl DA, Berg IA. 2014. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. Proc Natl Acad Sci U S A 111:8239–8244. https://doi.org/10.1073/pnas.1402028111.
45. Kampfer P, Denner EBM, Meyer S, Moore ERB, Busse HJ. 1997. Classification of "'Pseudomonas azotocolligans" Anderson 1955, 132, in the genus Sphingomonas as Sphingomonas trueperi sp nov. Int J Syst Bacteriol 47:577–583. https://doi.org/10.1099/00207713-47-2-577.
46. Cheng W, Yang X, Xue H, Huang D, Cai M, Huang F, Zheng L, Yu Z, Zhang J. 2022. Reproductive toxicity of furfural acetone in Meloidogyne incognita and Caenorhabditis elegans. Cells 11:401. https://doi.org/10.3390/cells11030401.
47. Fish JA, Chai BL, Wang Q, Sun YN, Brown CT, Tiedje JM, Cole JR. 2013. FunGene: the functional gene pipeline and repository. Front Microbiol 4. https://doi.org/10.3389/fmicb.2013.00291.
48. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, Campbell JH, Fortney JL, Mehlhorn TL, Lowe KA, Earles JE, Phillips J, Techtmann SM, Joyner DC, Elias DA, Bailey KL, Hurt RA, Preheim SP, Sanders MC, Yang J, Mueller MA, Brooks S, Watson DB, Zhang P, He Z, Dubinsky EA, Adams PD, Arkin AP, Fields MW, Zhou J, Alm EJ, Hazen TC. 2015. Natural bacterial communities serve as quantitative geochemical biosensors. mBio 6. https://doi.org/10.1128/mBio.00326-15.
49. Hug LA, Thomas BC, Brown CT, Frischkorn KR, Williams KH, Tringe SG, Banfield JF. 2015. Aquifer environment selects for microbial species cohorts in sediment and groundwater. ISME J 9:1846–1856. https://doi.org/10.1038/ismej.2015.2.
50. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, Thomas BC, Banfield JF. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. mBio 4. https://doi.org/10.1128/mBio.00708-13.
51. Zeugin JAH, James L. 1985. Ethanol precipitation of DNA. Focus 7:1–2.
52. Lui LM, Nielsen TN, Arkin AP. 2021. A method for achieving complete microbial genomes and improving bins from metagenomics data. PLoS Comput Biol 17:e1008972. https://doi.org/10.1371/journal.pcbi.1008972.
53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
54. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
55. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
56. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. Peerj 7:e7359. https://doi.org/10.7717/peerj.7359.
57. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
58. 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.
59. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927.
60. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29:2933–2935. https://doi.org/10.1093/bioinformatics/btt509.
61. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. Peerj 4:e2584. https://doi.org/10.7717/peerj.2584.
62. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9.
63. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glockner FO. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res 42:D643–D648. https://doi.org/10.1093/nar/gkt1209.
64. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825–5829. https://doi.org/10.1093/molbev/msab293.
65. Zhang H, Yohe T, Huang L, Entwistle S, Wu PZ, Yang ZL, Busk PK, Xu Y, Yin YB. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.
66. Rivas-Ubach A, Liu YN, Bianchi TS, Tolic N, Jansson C, Pasa-Tolic L. 2018. Moving beyond the van Krevelen diagram: a new stoichiometric approach for compound classification in organisms. Anal Chem 90:6152–6160. https://doi.org/10.1021/acs.analchem.8b00529.
67. Riera JL, Baldo L. 2020. Microbial co-occurrence networks of gut microbiota reveal community conservation and diet-associated shifts in cichlid fishes. Anim Microbiome 2:36. https://doi.org/10.1186/s42523-020-00054-4.
68. Zhu H-Z, Zhang Z-F, Zhou N, Jiang C-Y, Wang B-J, Cai L, Liu S-J. 2019. Diversity, distribution and co-occurrence patterns of bacterial communities in a karst cave system. Front Microbiol 10. https://doi.org/10.3389/fmicb.2019.01726.