BSSD 2021 Performance Metric Q2

Goal: Develop new omics-based techniques to understand microbiome function in environmental samples

Q2 Target: Describe progress on using omics-based techniques to infer activities among microbial communities in an environmental microbiome

Introduction

The LLNL "Microbes Persist" Soil Microbiome Scientific Focus Area (SFA) seeks to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues and formation of soil carbon (C). Our SFA research program is now four years old; it evolved from previously funded BSSD projects in the Firestone (UCB), Banfield (UCB), Sullivan (OSU) and Hungate (NAU) labs. We use stable isotope probing (SIP) in combination with 'omics analyses to measure how dynamic water regimes shape activity of individual microbial populations *in situ* and how their ecophysiological traits affect the fate of microbial and plant C. Using measures of population dynamics and microbiome-mineral interactions, we are working to synthesize both genome-scale and ecosystem-scale models of soil organic matter (SOM) turnover, to predict how soil microbiomes shape the fate of soil C. Here we focus on approaches (and examples) that advance our understanding of activity in complex soil microbiomes.

Environmental microbiology has long been consumed by questions of *Who is where,* and *What are they doing*? Classical DNA-based fingerprinting studies suggest there is enormous microbial diversity in soil, including functions that have no apparent tie to current conditions[4]. Yet SIP assays indicate only a fraction of these taxa are active at any particular moment or location—for example, we find only 30-50% of the surface soil community is active during the moisture-replete plant growing season in a Mediterranean climate annual grassland, and only 4% are active immediately following wet-up in the fall[5]. In addition, many taxa are simply too rare to be counted. Are inactive and uncounted OTUs merely non-viable relic DNA[6, 7], or perhaps dormant and poised to become active under different conditions[8]? These issues are fundamental to what maintains the enormous latent microbial diversity and functional potential in soils. We suspect the answer lies in a combination of both time, space, and biotic interactions —if snapshots of the active community were summed across enough timepoints and soil habitats, the active populations might equal the total 'present' community. Still, if an organism is present—but not active—does it represent meaningful potential for future activity? Or simply a relic of past climates or habitats? Our motivation for studying environmental microbiomes is linked to our desire to measure how they modify their local environment and affect system biogeochemistry. However, metanalyses have failed to find a strong correlation between microbial community composition and biogeochemical functions[9], or gene abundance and the corresponding process[10]. We argue that this disconnect is a result of studies that do not distinguish the more relevant populations from those that are dormant, i.e. a lack of focus on the *active microbiome.*
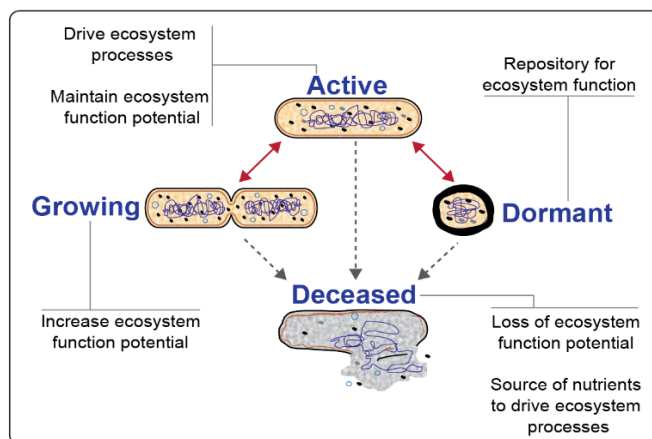

Figure 1. Microorganism metabolic states and their contribution to ecosystem functioning. Viable microorganisms exist in one of three general metabolic states that are all subject to mortality[2].

Environmental microbes exist in different metabolic states: growing, active, dormant, and recently deceased (**Fig 1**)[2]. These states correspond to the influence that populations have within their given habitat. To understand how microbiomes and their genome-scale functional properties drive system-level ecosystem functions, it is important to accurately associate microbial identity with <u>active</u> metabolic states.

Simultaneous identification of active microbes, their genomic capacity, and metabolic states has been a longstanding goal in microbial ecology, and methods to achieve this have accumulated in our molecular toolboxes—even for highly complex systems like soil. These methods include standard 'omics approaches (amplicon fingerprinting, metagenomics, viromics, metatranscriptomics, metaproteomics, metabolomics, lipidomics), and SIP-enabled versions of these 'omics tools[11-14]. In this report, we provide examples of how our SFA team has developed new understanding of microbiome activity using these tools.

'Omics techniques that infer activity

16S rRNA as a measure of activity

For over a decade, researchers have used ribosomal RNA gene surveys (rRNA genes) to identify environmental microorganisms present in a sample, and many studies measure ribosomal RNA (rRNA) as an indicator of growing or active taxa (e.g. [15-20]). However, in a 2013 metanalysis, Blazewicz et al. found conflicting patterns between rRNA content and growth rate[2]. Their work indicates that real-time activity and rRNA profiles in environmental samples do not always correlate due to differences in life histories, life strategies, and non-growth activities. Therefore, rRNA analysis may not be a reliable metric of currently active microbial communities.

Gene expression (metatranscriptomics)

Unlike amplicon or metagenomic data, which only provide information about the functional *potential* of a sample (not its current metabolic state), metatranscriptomic (mRNA) data comes one step closer to reflecting true microbial activity, albeit without providing direct evidence of translation or enzyme activity. Metatranscriptomics in soil remains a challenging approach—due in part to difficulty of extracting enough high-quality mRNA, successfully depleting rRNA, and having an ideal reference database.

Community mRNA response to carbon addition: Microbial traits can be observed at the genome level (in metagenomes), but alterations in gene-expression (reflected in metatranscriptomes) give microbes a flexibility that is much greater than what we can deduce from their genomic traits. These alterations in response to environment changes happen
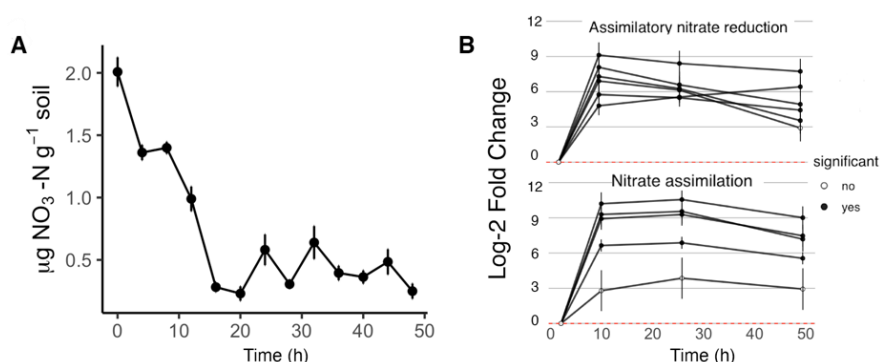

Figure 2. The concentration of nitrate (A) and the transcriptional response for inorganic nitrogen transporters (B) following the addition of labile carbon to an agricultural soil in a laboratory incubation.

quickly. We studied the short-term response of soil microbes to a change in C availability by adding glucose. We took samples before and 8, 24 and 48 h after glucose addition, extracted DNA and RNA, and sequenced and assembled the metagenomes and metatranscriptomes[22]. Within 8 h after glucose addition, very large changes in microbial N cycling occurred: transcript abundances for nitrate uptake, reduction to ammonium and assimilation increased 32 to 250-fold, while transcripts for the enzymes for nitrification and denitrification decreased (**Fig 2**)[23]. <u>This research shows that in combination, metagenomes and metatranscriptomes are powerful in helping understand microbial functions relevant to soil health at a foundational biochemical level.</u>

In a second study, we examined transcript abundances of inorganic nutrient transporters (N, P, K, Ca, Mg, S, Cl, Cu, Fe, Mn, Mo, Ni and Zn; Dijkstra et al in prep), again in combination with a soil glucose addition (**Fig 3**). Before glucose addition, transcripts for iron transport systems were particularly abundant, while after glucose addition, the ammonium, nitrate/nitrite and urea transporters increased dramatically, becoming the most abundant transporter transcripts. Transcript profiles of nutrient transporters before

glucose addition were similar to those measured in a range of soil, freshwater and marine ecosystems. Interestingly, while transcript abundances (mRNA) changed rapidly following the glucose addition, overall gene abundances (from DNA) remained the same. This research shows that alterations in gene-expression allow microbes to respond to their environment in the short-term by altering their ecophysiological traits and suggests that metatranscriptomes are a far more relevant read-out of the current activity occurring within a microbiome than surveys of taxonomic abundance (amplicons) or genomic potential (metagenomes).

Figure 3. Transcript abundances of inorganic nutrient transporters change within 8 h after a soil glucose addition, while the gene abundances remain constant. (Dijkstra et al., in prep.)

Using gene expression to identify rhizosphere decomposer guilds: Soil resource availability also changes quickly in the soil surrounding roots (rhizosphere). Stimulated by exudates and root decay, rhizosphere organisms interact to break down C derived from root tissues and move it into the surrounding soil, ultimately regulating how soil C is stabilized or lost. However, the microbial mechanisms that underpin rhizosphere C cycling are poorly understood. We analyzed time-resolved metatranscriptomes to compare microbial functions in rhizosphere, detritusphere, and combined rhizosphere-detritusphere habitats[3]. Aggregated population transcripts were binned using a custom reference database generated from soil isolate genomes, single cell amplified genomes, metagenomes, and SIP metagenomes[24]. Our results helped to identify distinct groups ('guilds') of microbes that express carbohydrate active enzyme (CAZy) genes in a spatially or temporally coherent manner. We show that these guilds have a cohesive metabolic niche within the rapidly changing rhizosphere environment. While the soil habitat significantly affected both community composition and overall gene expression, the succession of microbial functions (gene transcripts) occurred at a faster time scale than compositional changes (**Fig 4A,B**). Our study is one of the first in soils to explain how niche differentiation occurs, and to map transcripts back to individual genomes, generating population-resolved gene expression (**Fig 4C**). Genome-centric analyses like these allow us to track transcription in individual populations, which may be a more relevant approach than grouping transcripts across disparate classes or phyla.
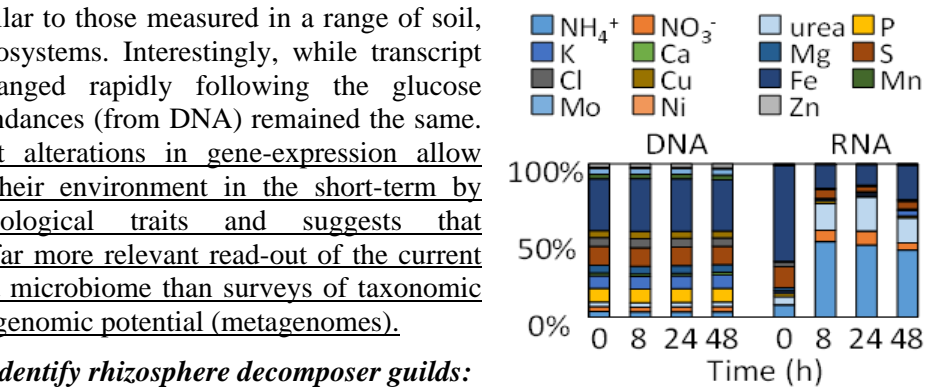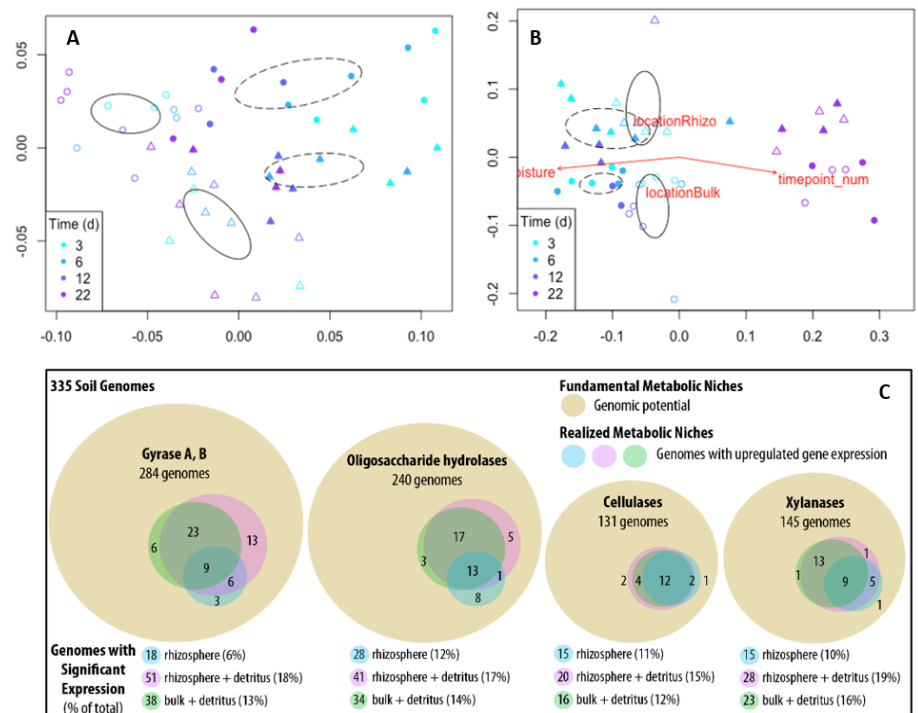

Figure 4. NMDS ordinations show the influence of time on 16S bacterial communities (A) and mRNA transcripts (B) during 22 days of plant root growth. (C) Number of functionally active taxa by soil habitat relative to the total metagenomic capacity. The outer circle (brown) indicates the number of unique genomes in our reference database with genomic potential for the specified gene class; inner circles reflect the number of taxa that differentially upregulated each gene class, per treatment.

3

Nitrogen cycle activities: In the same metatranscriptome dataset described above, we are also analyzing genes involved in N cycle processes. Most soil N is embedded in complex organic molecules such as proteins, chitin, DNA and lipids. However, plants can only take up N that is either inorganic or in small organic molecules (e.g. amino acids). Thus, degradation of complex N molecules by soil bacteria is a limiting step in plant N acquisition, but it is unknown which bacteria actualize their genomic potential to participate in N mineralization. We find extracellular proteases are among the most highly expressed N cycling genes in our soils (**Fig 5**).
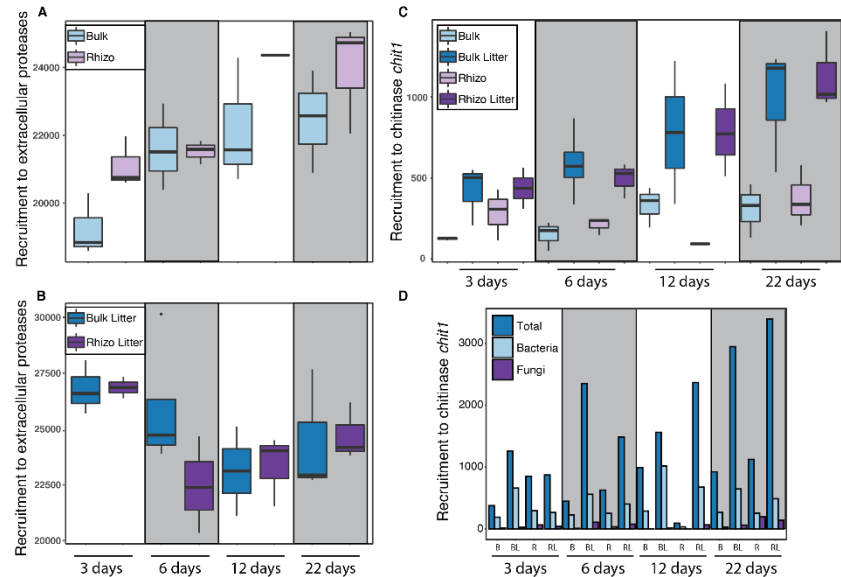

Figure 5. Expression of extracellular proteases in bulk and rhizosphere soils from common wild oat *Avena fatua* microcosms, grown without root litter amendment (A), or with litter amendment (B); expression of chitinase gene *chit1* (C) and mean counts of chitinases at the domain level (D).

The taxonomy of the bacteria expressing these proteases varies with time and soil habitat (rhizosphere, bulk soil, litter amended), suggesting that a succession of different bacteria affect plant-available N throughout plant growth. While chitinase was not as highly expressed as protease, it was significantly correlated with the presence of root litter, which may be because litter attracts fungi and other chitinous organisms.

In an inclusive analysis of all the expressed genes involved in organic and inorganic N cycle processes, we find that the presence of litter and/or living roots significantly alters the trajectory of N cycling gene expression (**Fig 6**). Expression (normalized to sequencing depth and gene length) was highest for ammonia monooxygenase subunits *amoAC*, nitrate reductase *nirK*, glutamate synthase *glnA* and extracellular proteases (exoprotease). Ammonium oxidation gene expression was almost exclusively performed by archaea and downregulated in the rhizosphere. Glutamate synthase expression patterns (a key indicator of ammonium assimilation via the GS/GOGAT pathway during N-limited conditions) suggest N limitation occurred during early rhizosphere development as well as in bulk soils that had been amended with root litter. Assimilatory nitrate reduction and ammonium assimilation were upregulated in the rhizosphere, whereas dissimilatory nitrate reduction and chitinase were upregulated in the detritusphere. Overall, these transcript expression patterns reveal dominant processes may be carried out by low-abundance microorganisms and N transformations patterns that were not reflected in community composition data.
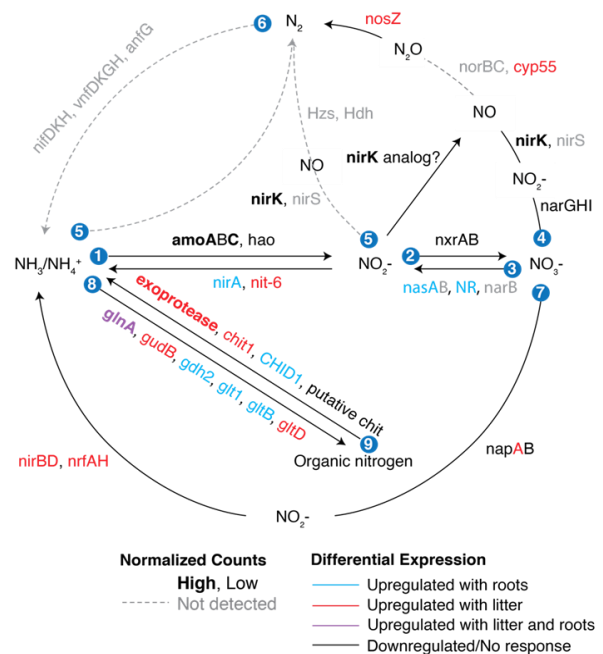

Figure 6. Expression level and upregulation of N cycling ORFs from assembled transcripts. Bold font indicates statistically significant expression. Numbers denote pathways: 1=ammonia oxidation, 2=nitrite oxidation, 3=assimilatory nitrate reduction, 4=denitrification, 5=Anammox, 6=nitrogen fixation, 7=DNRA, 8=ammonia assimilation (GDH and GS/GOGAT), 9=macromolecular N mineralization.

Metabolomics and Lipidomics

In soil, the diverse array of metabolites and microbial taxa make it challenging to directly assess the functional associations between metabolites and microorganisms under different conditions, unless individual isolate genomes are available[25]. Still, after transcriptomics and proteomics, metabolomics comes a step closer to reflecting the active metabolic state of a microbiome. If changes in functionally associated metabolites and microbial taxa are tightly bound, the alteration of one pool (e.g. in response to changing abiotic conditions) should be reflected in changes of the other. Soil metabolites can be measured several ways; here we describe results collected with LC-MS-MS and Fourier-transform ion cyclotron resonance-mass spectrometry (FTICR-MS).

<div style="border:1px solid">

Active Viruses

Amplicon and metagenome surveys include many soil organisms that are inactive. But are viral sequences different? Does viral DNA have an inherently shorter half-life? i.e., when we identify viral sequences, can we infer they are active? A serendipitous finding in our mRNA dataset is a diversity of RNA viral genomes[1] that we mapped to our site-specific metagenome[3]. Based on the phylogeny of the identified RNA viruses, it appears that most of them target fungal hosts.

</div>

Rhizosphere LC-MS-MS metabolites:

Elevated CO_2 (eCO_2) studies provide a unique opportunity to assess effects of altered root exudate (metabolite) patterns on microbial community succession and function, since eCO_2 stimulates higher rates of photosynthesis, increased belowground biomass production, and soil deposition of labile C[26, 27]. In our research on root-soil interfaces, we have found that plants grown under elevated (700 ppm) CO_2 increased both C allocated belowground and the amount of root-derived ^{13}C in the mineral-associated fraction[28]. Metabolites produced in early weeks of plant growth under eCO_2 conditions clustered distinctly from "later" produced metabolites (**Fig 7**). Since eCO_2 both increased and decreased specific exudate components (**Fig 7**), additional work is needed to parse how these changes map to the active microbiome in the soils.

Our team has also been working to substantiate the molecular formulae assigned and identify known metabolites using LC-MS/MS based approaches. We are using MS/MS based molecular network analysis to relate unknown components of soil organic matter to each other, and to known compounds, via the similarity of their fragmentation spectra. In this way, molecules can be grouped into more finely classified families where detailed structural differences can be detected and ultimately used to expand the number of known molecules in soils.

Figure 7. Heat maps of metabolites from an experiment where *A. barbata* was grown with eCO_2 for 9 weeks. A. Root exudate profiles for the 25 most common metabolites. Warm colors reflect higher abundance. B. Metabolite heat map for root exudates that were significantly different between eCO_2 and ambient CO_2 treatments.

Soil organic matter characterization with FTICR-MS:

Molecular formula assignment, made possible by the high-resolution mass accuracy of FTICR-MS, has brought this tool to the forefront for untargeted investigation of soil metabolites and dissolved organic carbon (DOC) composition. <u>The molecular composition of soil organic carbon (SOC) is a strong determinant for whether it is consumed, modified, or left untouched by microbial decomposers</u>. Hence, the information gathered from formula assignments can help identify the type and amount of microbial activity

in soil. For FTICR-MS, soils are typically sequentially extracted using water, methanol, then chloroform to capture polar and nonpolar soil metabolites[29] and then analyzed. We have used this approach for multiple studies, including one where tropical soil was incubated under static and oscillating soil redox conditions to determine the effect of O_2 availability on soil C pools. The FTICR-MS metabolite data suggest that redox conditions significantly impacted soil C composition (**Fig 8**). Specifically, fluctuating and oxic treatments had more dynamic metabolite profiles, while the anoxic
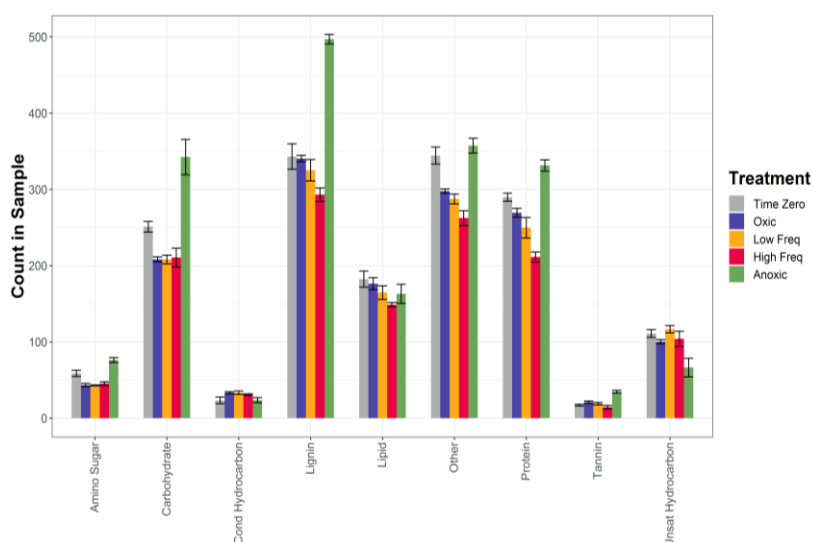

Figure 8. Impacts of redox conditions on water extractable soil metabolites in soils incubated under static oxic or anoxic conditions, or two oscillating treatments (low freq., high freq.) for 44 days (Campbell, Pett-Ridge et al. in prep)

treatment changed less. Soils kept under static anoxic conditions had more amino sugars, carbohydrates, lignin, proteins, and tannins, and less unsaturated hydrocarbons, than oxic or oscillating soils.

In a related study, we found that more labile, oxidized substrates released from fresh plant litter were equally depleted by both anaerobic and aerobic heterotrophs[30]. However, the conversion of more reduced, less thermodynamically favorable SOC to CO_2 primarily happened under aerobic conditions. This nuance in the behavior of microbes actively consuming different types of C under varying redox conditions will be an important addition to improve predictive models of C fluxes in tropical forest soils.

Soil lipidomics:
Minerals preserve the oldest most persistent soil C and play a critical role as a microbial habitat. In a recent study, we incubated four different mineral types (ferrihydrite, kaolinite, quartz, and density fractionated minerals) in the rhizosphere of ^{13}C-labeled plants. Distinct bacteria and fungi colonized the different mineral surfaces and had distinct capacities for rapid growth; their ribosomal copy number was significantly correlated with relative isotope enrichment on minerals[31]. We then used lipidomics (at EMSL) to trace plant- versus microbial-derived C onto minerals, characterizing total C, ^{13}C enrichment, and SOC chemistry[32]. Many of the mineral-associated lipids that we found were microbially-derived, including a large fraction of fungal lipids (**Fig 9**). Microbial assimilation of SOC compounds may result in transformation of those compounds to lipids as microbes synthesize their membranes and storage lipids. The lipidomics technique could prove to be a powerful connection between microbiome ecophysiology and SOC chemistry, especially if it becomes more possible to directly link specific lipid classes to microbial phylotypes.
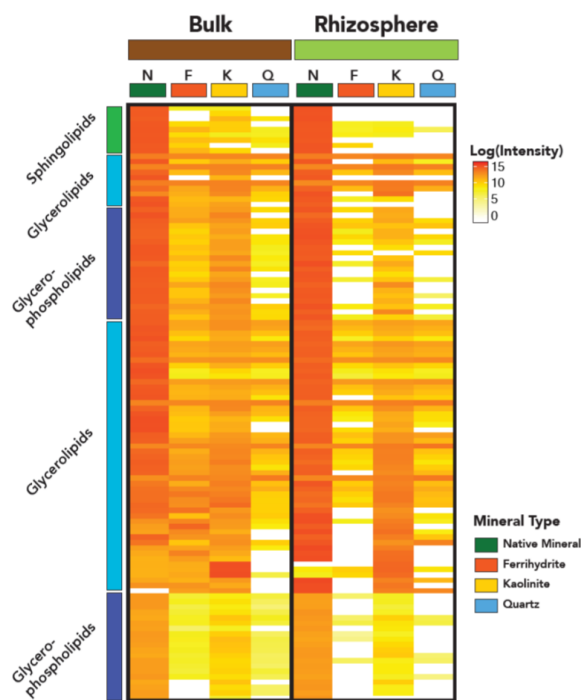

Figure 9. A heatmap comparing log$_{10}$ transformed lipid intensities for lipids identified in bulk and rhizosphere soil, by mineral type.

Stable Isotope Probing

Stable isotope probing (SIP) is one of the few approaches that can identify the ecophysiology of active microorganisms in their native environments, making it one of the most powerful techniques in microbial ecology. Broadly speaking, SIP refers to any technique where microorganisms that have actively consumed substrates enriched in rare stable isotopes (e.g. ^{13}C, ^{15}N, ^{18}O) are identified based on the resulting isotopic enrichment of their nucleic acids, proteins, and metabolites. Density gradient SIP is the culture-independent gold standard for directly linking sequence to function in complex microbial communities[33]. When a microorganism consumes a substrate enriched or "labeled" with a heavy isotope, the cellular components of that cell also become labeled in the heavy isotope. Density gradient SIP takes advantage of the increased density of microbial nucleic acids (due to assimilation of heavy isotopes), using a density gradient to separate the heavy (labeled) nucleic acids from lighter (unlabeled) ones. Isolated heavy nucleic acids can then be characterized to identify the organisms that actively assimilated substrates of interest.

Our SFA team has pioneered new SIP methods that quantify element fluxes with high taxonomic resolution. In particular, quantitative stable isotope probing or qSIP, developed at NAU, is the isopycnic separation of nucleic acids in cesium chloride combined with a mathematical model to quantify isotope enrichment [11, 34-37]. With qSIP we measure growth rates of individual taxa in complex soil communities using ^{18}O-labeled water as a universal substrate that is used by all actively growing organisms. qSIP is also applicable to metagenomic and viromics analyses, as we describe below.

16S rRNA gene targeted SIP

Taxon-specific growth following soil rewetting:
Microbial activity is stimulated by the rewetting of dry soils, resulting in a pulse of C mineralization and nutrient availability. While there has been much interest in the response of indigenous communities to soil wet-up, until our 2020 paper on the subject[5], no work had identified the specific microorganisms in semi-arid soils that grow in response to soil wet-up. We used heavy water ($H_2{}^{18}O$) DNA SIP coupled with high throughput sequencing of bacterial 16S rRNA genes to characterize taxon-specific growth activities following the rewetting of a seasonally dried California annual grassland soil. We hypothesized that after an abrupt change in soil water potential due to wetting of dry soil, new growth would follow a sequential response pattern due to differences in growth strategies. Using qSIP, bacterial growth was detectable within the first three hours after wet-up (**Fig 10**), and approximately 40% of the measured taxa were found to be growing during the 7-day incubation. At the genus level, bacteria exhibited different growth rate response

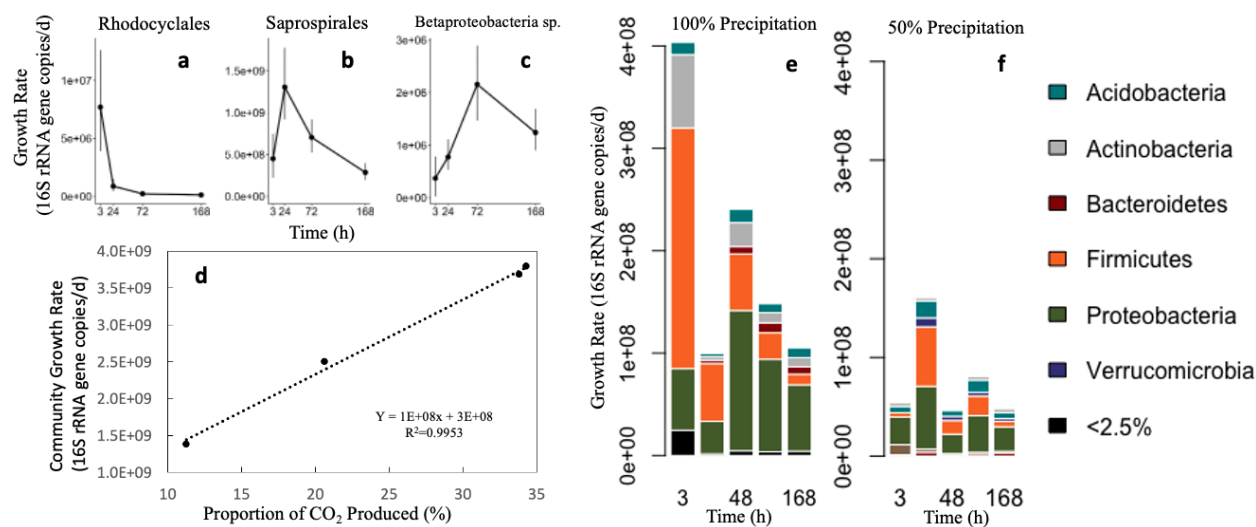**Figure 10. a-c) Examples of taxon-specific growth rate response patterns (a=primary, b=secondary, c=tertiary). Delayed responders are not shown because most of these taxa only grew at the final timepoint. d) Correlation between the portion of CO$_2$ produced vs. community growth rate. e,f) Phylum level growth rates for (e) 100% and (F) 50% precipitation treatments following wet-up, calculated with qSIP.**

patterns, with growth rate curves representing four potential responses (**Fig 10a-c**). Average community growth rate was well-correlated with the proportion of total CO_2 released at each time period (**Fig 10d**), indicating that even in complex microbial communities, population dynamics can translate quantitatively to ecosystem-scale biogeochemistry. Since the biological predictors of this CO_2 pulse are challenging to identify, it is exciting to observe a clear linear relationship between microbial growth rate and CO_2 efflux.

In the fall of 2018, we performed a second ^{18}O-water SIP wet-up experiment designed to ask how reduced spring rainfall affects the amount of C subsequently lost after re-wetting and phage-driven bacterial mortality. After the annual summer dry period, we wet soils with ^{18}O-water and harvested at six times (0, 3, 24, 48, 72, 168 h). Using 16S rRNA qSIP, we see that spring precipitation treatment differences do affect subsequent microbial growth rates during the fall wet-up (**Fig 10e,f**). Overall, growth began more rapidly and was sustained at higher rates in 100% precipitation soils. These results suggest a soil's historic rainfall patterns significantly impact microbial activities, turnover, and persistence.

16S qSIP highlights taxon-specific response to changing redox periodicity:

We have also used 16S rRNA qSIP to characterize the active soil microbiome in our tropical soil redox studies. In many wet tropical forest soils, fluctuating redox conditions support a metabolically diverse microbial community that mediates C turnover and critical ecosystem processes. However, global climate change is causing less regular rainfall at tropical latitudes and may lead to inconsistent soil redox periodicity in these systems. We conducted an incubation with ^{13}C-enriched plant litter to distinguish the microbial consumers of fresh plant litter versus native soil organic matter under different redox regimes in a tropical forest soil[30, 38]. [13]C qSIP provided evidence for phylogenetically conserved trends in microbial C assimilation from plant litter. Most notably, bacteria from the phylum Firmicutes assimilated proportionally more litter C under static anoxic conditions (**Fig 11a**). Several taxa derived a substantial portion of their nucleic acid C from litter regardless of soil redox status, indicating that facultative metabolisms are relatively common in this naturally O_2 dynamic environment. Surprisingly, many of the taxa that derived the highest proportion of their nucleic acid C from litter were not abundant, suggesting that 'rare' organisms may play an important role in terrestrial C cycling (**Fig 11b**). <u>[13]C qSIP revealed aspects of microbial activity and litter C assimilation that were not apparent from our traditional 16S rRNA community analyses.</u> Along with our biogeochemical measurements, these findings show that microbial response to shifting redox periodicity influences C cycling in tropical forest soils.
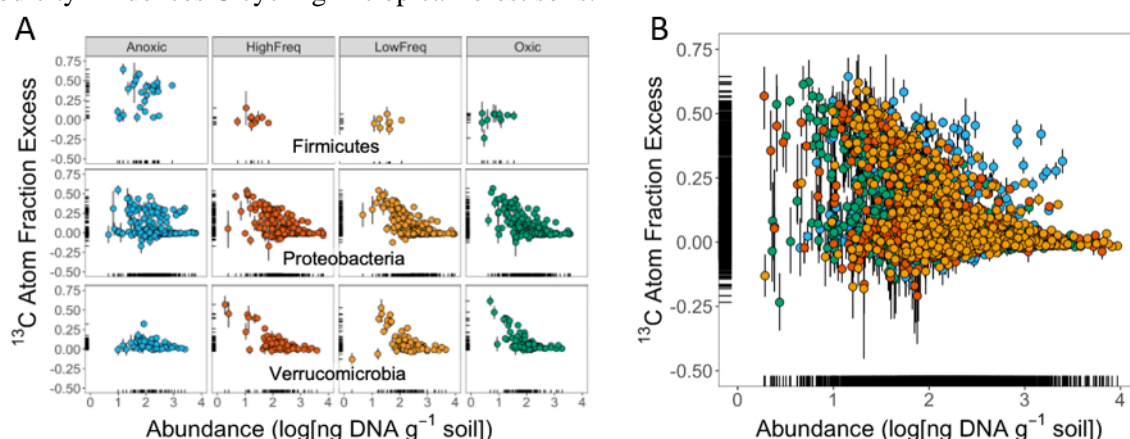

Figure 11. (A) Taxon-specific litter C assimilation of three phyla (shown as ^{13}C atom fraction excess) following 44 days of incubation under four different redox regimes. Each point represents the ^{13}C enrichment of a bacterial or archaeal ASV (amplicon sequence variant). (B) Taxon-specific litter C assimilation of all taxa identified as active by qSIP (shown as ^{13}C atom fraction excess) following 44 days of incubation under different redox regimes. Colors reflect the same phyla as in (A). (Hestrin, Pett-Ridge et al. in prep)

SIP-identified bacterial predators:

Stable isotope probing is ideally suited to studies of cross-kingdom interactions and food webs. Predation helps structure food webs and the flow of nutrients through systems. By studying environmental predators
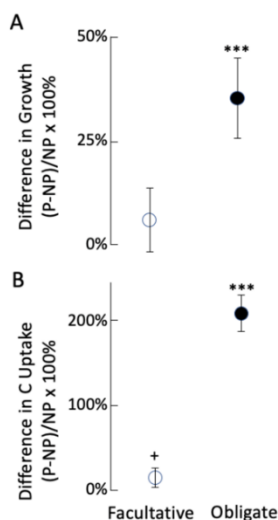
through the lens of stable isotope probing (SIP) we can begin to identify the impact of these predators. Analyzing the growth of soil bacteria with qSIP, we found that predatory bacteria grew more and faster than non-predatory bacteria (**Fig 12**). This indicated that they were consuming a large amount of C from non-predatory bacteria and could have a large influence on the ecology and nutrient flow through soil. After finding isotopically enriched *Bdellovibrio* sequences in several of our datasets, we hypothesized that bacterial predation is a significant driver of soil microbial mortality. Predatory bacteria are common in nature, yet their roles in soil microbial food webs are poorly known. We used meta-analysis to synthesize $H_2^{18}O$ SIP measurements from 81 experiments conducted at 14 sites--nearly 100,000 taxon-specific estimates of *in situ* bacterial growth rates[39]. Growth rates of predatory bacteria were 19% higher than for non-predatory bacteria, and especially high for obligate predators in the Vampirovibrionales and in the genus *Lysobacter*, which are capable of gliding motility and known to exhibit wolfpack hunting behavior[40-42]. Added C substrates disproportionately stimulated growth rates of obligate predators compared to non-predators and facultative predators. These qSIP findings support the ecological theory that higher productivity increases food chain length and suggest that the functional significance of bacterial predators increases with energy flow.

Figure 12. Relative difference in predator growth rate (A) and ^{13}C uptake rate (B) compared to non-predators. Values are shown separately for facultative (open symbols) and obligate (filled symbols) predators. Symbols are means ± standard errors.

Characterizing the total vs active soil microbiome with SIP:

We characterized the present (total) vs. growing (active) soil bacterial and archaeal communities in three California Mediterranean grasslands using heavy water SIP 16S rRNA amplicons. While the sites span a mean annual rainfall gradient (383 mm yr-1 to 2160 mm yr-1), we conducted our study during the spring season when soil water content was similar between sites and not limiting. Richness and Shannon's diversity index did not vary between sites for the present microbial communities (measured by 16S rRNA amplicon sequencing), but the diversity of actively growing microorganisms was higher in the intermediate site than the wettest site. Despite similar environmental conditions leading up to the incubation, mean relative growth rates were lowest at the driest site. The growth rate of a bacterial taxon in one site explained up to 57% of the variation in that same taxon's growth rate in another site (**Fig 13**), suggesting that cross-site variation in community mean growth rate is influenced by evolutionary history of the taxa at each site. Low relative growth rates in soils from the driest site may reflect a tradeoff between traits that allow microorganisms to survive chronic drought and maximum potential growth rate.
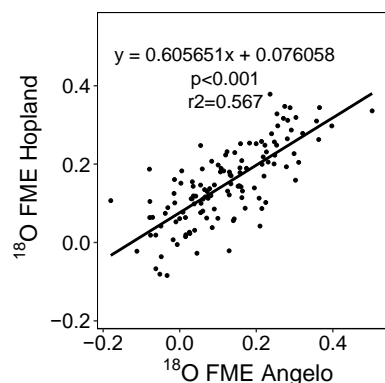

Figure 13. Linear regression of ^{18}O fraction maximum enrichment of microbial taxa co-occurring at our highest mean annual rainfall site (Angelo) vs. our intermediate mean annual rainfall site (Hopland). (Foley, Pett-Ridge, Hungate et al. in prep)

SIP metagenomics

California annual grassland rainfall gradient metagenomes:

Only a handful of studies have combined SIP with metagenomics or metatranscriptomics to investigate genomic potential and actively expressed genes by functional guild[43, 44]. In our SFA research, SIP-metagenome studies have allowed us to determine the core soil microbiome ecophysiologies corresponding to growth under water replete conditions. Our first challenge was assessing the best combination of the 180 bulk and SIP samples we shotgun sequenced (>1.3 Tbp), to yield the best sequence assemblies. After testing multiple approaches, bulk/SIP fractions from the same sample were pooled and co-assembled into 18 total assemblies, and every SIP-fraction library was mapped to each sample's metagenome co-assembly. These

coverages were used for differential-coverage binning, yielding 433 unique metagenome assembled genomes (MAGs). We modified the standard qSIP approach to calculate the atom fraction excess (AFE) (proportional to growth rate) of ^{18}O incorporated in each genome bin[13]. Broadly, the most isotopically enriched, and thus actively reproducing, microbial populations had lower abundance (consistent with our 16S qSIP results). Genomic capacity for aerobic respiration, methanol and carbon monoxide oxidation and carbohydrate decomposition was broadly distributed across both active and inactive bacteria in all three soils. Organisms with flagella were less isotopically enriched than those without flagella at the drier Sedgwick site, but were enriched at the wetter sites, suggesting that motility is an important response to water availability at the wetter sites.

*Arctic soil metagenomes***:**
In SIP datasets where only the heaviest fractions are sequenced (in our older datasets, this was necessary due to financial limitations), activity must be calculated using methods other than AFE. For our "ICE-SIP" study, we incubated permafrost active layer communities in ^{16}O and ^{18}O water in sub-freezing temperatures and detected incorporation of ^{18}O into DNA, thereby indicating the microbial taxa that were active in these frozen soils. Assembling and genome binning identified 153 total MAGs. Here, active organisms incorporating the heavy label can have two patterns, 1) MAG sequences are present in both the ^{16}O and ^{18}O samples and have a corresponding density shift, or 2) MAG sequences are present only in the ^{18}O samples but were below the density cutoff in the ^{16}O samples and no MAG sequences were present. With the first approach, 39 of 153 MAGs had significantly more reads in the ^{18}O vs. the ^{16}O samples and were thus considered active. For the second approach, a read-subtraction method was used to identify ^{18}O reads not present in a ^{16}O assembly, followed by assembly and binning steps of these ^{18}O-only reads. This approach yielded 32 total active MAGs. After a dereplication step between MAGs from the first and second approaches, 46 of 153 MAGs were considered active and growing, implicating these bacteria in biogeochemical processes even in anoxic sub-freezing temperature soils.

Targeting soil viruses and microbe-virus links with SIP metagenomes
Challenges associated with identifying virus 'activity':
Over the last decade, we have learned that viruses – mostly those that infect bacteria – have important ecological roles in diverse environments, including the oceans, the human gut, and soil. For soil, gaining an understanding of how microbes and their viruses are intertwined in the cycling of nutrients, including C and N, is a key goal for our SFA. Currently, we have limited data on the diversity and function of viruses, and it appears that each soil harbors a unique virus community.

The application of isotope tracers for direct assessment of activity is not as straightforward for viruses as it is for other microorganisms, because of different viral infection cycles, their lack of metabolism, and the many states in which they can be present[45]. The metabolic state of a virus depends on the infection cycle. Temperate viruses can undergo lytic infection, where activity is identified by progeny viruses, or lysogenic infection, where activity is difficult to assess. A temperate virus that is undergoing lysogenic infection (present as a provirus) during a heavy-water SIP incubation would become isotopically enriched and its abundance would depend on its host's division rate. Active proviruses undergoing lysogenic infection need to be distinguished from viruses undergoing lytic infection because the effect of proviruses on the host metabolism (and therefore ecosystem) will not be as pronounced. While SIP-enabled metagenomics alone cannot unequivocally identify a virus' state (e.g., lysogenic vs lytic), it does provide definitive evidence of de novo nucleic acid synthesis, a solid foundation for assessing activity.

Targeting viruses and their microbial hosts with ^{18}O-SIP:
One of the greatest benefits of SIP-metagenomics for viral ecology studies, is the ability to link active viruses with their active microbial hosts (using CRISPR and other approaches). In the California annual grassland rainfall gradient dataset described above, we used the ^{18}O SIP-metagenome data to determine 1) viral diversity, 2) the 'active' viral DNA, and 3) the microbial host range of virus communities in our three grassland soil sites. Across the three soil sites, we found >8,000 viral "species" (vOTUs), >90% of which were novel and comprised over 400 new taxonomic groups at the genus level (Fig 14). For about 50% of

these viruses, we were able to predict their microbial hosts; this showed that Actinobacteria (specifically Mycobacteria) were the dominant host taxa across all three soils. The identification of these viruses (one-third of which were greater than 10kb) greatly expands the known viral content of soils, with a 4.5-fold increase compared to previous bulk metagenome soil studies[46]. To explore the impact of potentially novel functions of these viruses in our soil systems, we curated a set of auxiliary metabolic genes (AMGs) using DRAM-v[47] and found they encode a variety of cellulose and xylan catabolism genes, plus a nitrogen regulator gene (PII) thought to enhance C sourcing within infected cells.

To further show that these viruses were 'active', we calculated AFE values for each vOTU. This showed that a majority (70%-90%) of the viruses were active, and 10% of these active viruses encoded AMGs. The number of active viruses varied greatly across the three soil sites and viral diversity varied inversely with soil moisture (**Fig 14**). Ranked SIP-activity level analyses indicate our intermediate-rainfall site's viruses had the highest activity levels and appear to target Actinobacteria, Proteobacteria and Bdellovibrio as hosts. We predicted high-confidence host linkages for 49-55% of the identified vOTUs. Further, in the driest soils, we discovered several viruses of Nitrososphaerales[48] (Crenarchaeota). To date, only a single other Thaumarcheota provirus has been identified in the literature[49]. <u>Taken together, these predictions suggest that viruses are actively preying upon key microbial taxa known to have fundamental roles in soil ecosystem function</u>.
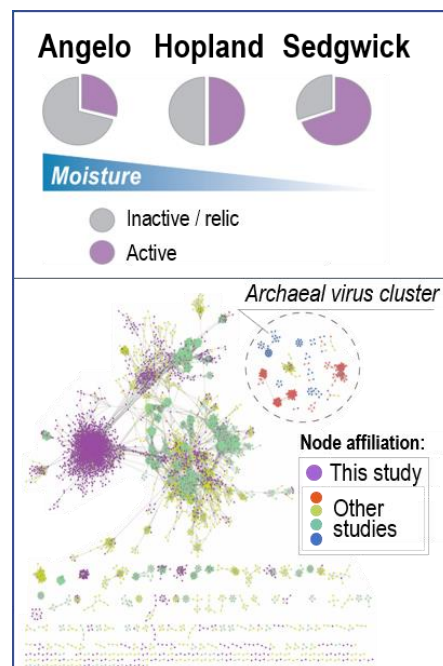

Figure 14: Virus metagenome analyses in a California annual grassland rainfall gradient. (A) Proportion of active viruses at each site along our rainfall gradient. (B) Gene-sharing network shows clustering of nonredundant vOTUs. (Zablocki et al. in prep.)

Viral activities in frozen soils:

Most ecosystem C models neglect winter months for arctic systems, since C losses (primarily driven by microbial decomposers) are assumed to be negligible at low temperatures. However, winter C loss in northern ecosystems may actually be greater than the average growing season C uptake. We used stable isotope probing (SIP) targeted metagenomics to assess the genomic potential of active soil microbial populations under winter conditions, with an emphasis on viruses and virus-host dynamics. Peat soils from the Bonanza Creek LTER site in Alaska were incubated under subzero anoxic conditions with $H_2^{18}O$ for 184 and 370 days. We identified 46 active bacterial populations (MAGs; spanning 9 bacterial phyla) and 243 active viral populations (vOTUs) that actively took up ^{18}O and produced significant CO_2 throughout the incubation (**Fig 15**). Active MAGs represented only a small portion of the detected microbial community and were capable of fermentation and organic matter degradation, whereas active vOTUs

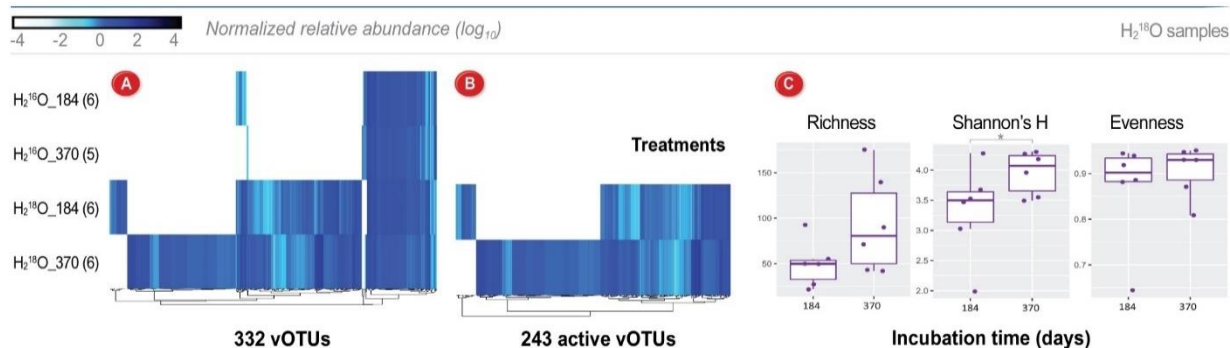were identified from 23 samples (natural abundance water "$H_2^{16}O$", and heavy water "$H_2^{18}O$", at 184 and 370 days). (B) Relative abundances of 243 vOTUs considered 'active' due to DNA ^{18}O enrichment patterns. (C) Diversity metrics for the 243 active vOTUs.

represented a large portion of the detected viral community and one-third were linked to active MAGs. We identified 86 auxiliary metabolic genes, the majority of which were carried by active vOTUs and had diverse functions such as C utilization and scavenging that could provide their host with a fitness advantage for utilizing much-needed C sources or to acquire essential nutrients. These results illustrate that substantial active virus-host interactions occur in sub-freezing anoxic conditions and highlight viruses as a major community-structuring agent that may modulate C loss in peat soils, even during winter months.

Summary

An ongoing challenge in microbial ecology is moving beyond what organisms are merely present in a system or even what their metabolic potential is, to inferring actual *activity* of organisms, particularly in complex communities. In the LLNL *Microbes Persist* Soil Microbiome SFA, we use a multi-domain approach to identify the microbial and viral inhabitants of soil ecosystems, designed to provide a comprehensive understanding of biotic interactions, ecophysiological traits, and the fate of microbiome biomass organic carbon. In both our empirical research and methods development, we are moving beyond traditional assessments of microbial communities by pairing stable isotope probing with metagenomic and metatranscriptomic surveys. This allows us to concentrate on the *active* and growing microorganisms that make up the soil microbiome, giving us an unprecedented picture of the most relevant taxa in soil ecosystems. Here, we have described a variety of methodologies we use to infer activity of organisms (including phage), from measuring rapid changes in gene expression (transcriptomics) and metabolic products (metabolomics and lipidomics), to the incorporation of isotopically labeled substrates into DNA. With an increase in temporal and spatial resolution, we aim to better capture the dynamics of ever-changing subpopulations of active microorganisms as their functional traits shift with environmental drivers. We also recognize the need to consider biotic interactions—with bacterial predators, RNA and DNA viruses—which play a large but uncharacterized role as regulators of microbial growth dynamics and microbial biogeochemistry. Combining our knowledge of the functions of active microorganisms with biogeochemical measurements at various scales will enable the implication of key microbial players in bulk-scale biogeochemical process rates. Further, studying how these microorganisms sense and respond to changing environments, and how those changes modulate their activities will help inform future predictive models.

References

1. Starr, E.P., E.E. Nuccio, J. Pett-Ridge, J.F. Banfield and M.K. Firestone, *Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil.* Proceedings of the National Academy of Sciences, 2019. **116**(51): p. 25900-25908.
2. Blazewicz, S.J., R.L. Barnard, R.A. Daly and M.K. Firestone, *Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses.* ISME Journal, 2013. **7**(11): p. 2061-2068.
3. Nuccio, E.E., E. Starr, U. Karaoz, E.L. Brodie, J. Zhou, S.G. Tringe, R.R. Malmstrom, T. Woyke, J.F. Banfield and M.K. Firestone, *Niche differentiation is spatially and temporally regulated in the rhizosphere.* The ISME Journal, 2020: p. 1-16.
4. Delgado-Baquerizo, M., A.M. Oliverio, T.E. Brewer, A. Benavent-González, D.J. Eldridge, R.D. Bardgett, F.T. Maestre, B.K. Singh and N. Fierer, *A global atlas of the dominant bacteria found in soil.* Science, 2018. **359**(6373): p. 320-325.
5. Blazewicz, S.J., B.A. Hungate, B.J. Koch, E.E. Nuccio, E. Morrissey, E.L. Brodie, E. Schwartz, J. Pett-Ridge and M.K. Firestone, *Taxon-specific microbial growth and mortality patterns reveal distinct temporal population responses to rewetting in a California grassland soil.* The ISME Journal, 2020. **doi.org/10.1038/s41396-020-0617-3**: p. 1-13.
6. Carini, P., P.J. Marsden, J.W. Leff, E.E. Morgan, M.S. Strickland and N. Fierer, *Relic DNA is abundant in soil and obscures estimates of soil microbial diversity.* Nature Microbiology, 2016. **2**: p. 16242.
7. Lennon, J.T., M.E. Muscarella, S.A. Placella and B.K. Lehmkuhl, *How, When, and Where Relic DNA Affects Microbial Diversity.*, in *MBio*. 2018, American Society for Microbiology.
8. Lennon, J.T. and S.E. Jones, *Microbial seed banks: the ecological and evolutionary implications of dormancy.* Nat Rev Micro, 2011. **9**: p. 119-130.
9. Graham, E.B., W.R. Wieder, J.W. Leff, S.R. Weintraub, A.R. Townsend, C.C. Cleveland, L. Philippot and D.R. Nemergut, *Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes.* Soil Biology and Biochemistry, 2014. **68**: p. 279-282.
10. Rocca, J.D., E.K. Hall, J.T. Lennon, S.E. Evans, M.P. Waldrop, J.B. Cotner, D.R. Nemergut, E.B. Graham and M.D. Wallenstein, *Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed.* ISME J, 2015. **9**(8): p. 1693-1699.
11. Hungate, B.A., R.L. Mau, E. Schwartz, J.G. Caporaso, P. Dijkstra, N. Van Gestel, B.J. Koch, C.M. Liu, T.A. McHugh, J.C. Marks, E. Morrissey and L.B. Price, *Quantitative Microbial Ecology Through Stable Isotope Probing.* Applied and Environmental Microbiology, 2015.
12. Koch, B.J., T.A. McHugh, M. Hayer, E. Schwartz, S.J. Blazewicz, P. Dijkstra, N. van Gestel, J.C. Marks, R.L. Mau, E.M. Morrissey, J. Pett-Ridge and B.A. Hungate, *Estimating taxon-specific population dynamics in diverse microbial communities.* Ecosphere, 2018. **9**(1): p. e02090-15.
13. Sieradzki, E.T., B.J. Koch, A. Greenlon, R. Sachdeva, R.R. Malmstrom, R.L. Mau, S.J. Blazewicz, M.K. Firestone, K. Hofmockel, E. Schwartz, B.A. Hungate and J. Pett-Ridge, *Measurement error and resolution in quantitative stable isotope probing: implications for experimental design, BioRxiv doi.org/10.1101/2020.02.25.965764.* mSystems, 2020. **5**: p. e00151-20. https://doi.org/10.1128/mSystems.00151-20.
14. Sachsenberg, T., F.A. Herbst, M. Taubert, R. Kermer, N. Jehmlich, M. von Bergen, J. Seifert and O. Kohlbacher, *MetaProSIP: Automated Inference of Stable Isotope Incorporation Rates in Proteins for Functional Metaproteomics.* Journal of Proteome Research, 2015. **14**(2): p. 619-627.
15. Egert, M., I. Schmidt, H.-M. Höhne, T. Lachnit, R.A. Schmitz and R. Breves, *Ribosomal RNA-based profiling of bacteria in the axilla of healthy males suggests right-left asymmetry in bacterial activity.* FEMS Microbiology Ecology, 2011. **77**(1): p. 146-153.
16. Jones, S.E. and J.T. Lennon, *Dormancy contributes to the maintenance of microbial diversity.*

Proceedings of the National Academy of Sciences, 2010. **107**(13): p. 5881-5886.

17.   Schippers, A., L.N. Neretin, J. Kallmeyer, T.G. Ferdelman, B.A. Cragg, R.J. Parkes and B.B. Jørgensen, *Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria.* Nature, 2005. **433**(7028): p. 861-864.

18.   Wüst, P.K., M.A. Horn and H.L. Drake, *Clostridiaceae and Enterobacteriaceae as active fermenters in earthworm gut content.* The ISME Journal, 2011. **5**(1): p. 92-106.

19.   DeAngelis, K.M., W.L. Silver, A.W. Thompson and M.K. Firestone, *Microbial communities acclimate to recurring changes in soil redox potential status.* Environmental Microbiology, 2010. **12**(12): p. 3137-3149.

20.   Mannisto, M.K., E. Kurhela, M. Tiirola and M.M. Haggblom, *Acidobacteria dominate the active bacterial communities of Arctic tundra with widely divergent winter-time snow accumulation and soil temperatures.* FEMS Microbiology Ecology, 2013. **84**(1), 47-59.

22.   Chuckran, P.F., M. Huntemann, A. Clum, B. Foster, S. Roux, K. Palaniappan, N. Varghese, S. Mukherjee, T.B.K. Reddy, C. Daum, A. Copeland, N.N. Ivanova, N.C. Kyrpides, T.G. del Rio, E.A. Eloe-Fadrosh, E.M. Morrissey, E. Schwartz, V. Fofanov, B. Hungate, and P. Dijkstra, *Metagenomes and metatranscriptomes of a glucose-amended agricultural soil* Microbiology Resource Announcements, 2020. **9**: p. e00895-20.

23.   Chuckran, P., V. Fofanov, B. Hungate, E. Morrissey, E. Schwartz, J. Walkup and P. Dijkstra, *Rapid response of nitrogen cycling gene transcription to labile carbon amendments in a soil microbial community.* bioRxiv https://www.biorxiv.org/content/10.1101/2021.03.26.437204v1 *mSystems* (in press), 2021.

24.   Starr, E., S. Shi, S. Blazewicz, B.J. Koch, A. Probst, B.A. Hungate, J. Pett-Ridge, M. Firestone and J. Banfield, *Stable isotope informed genome-resolved metagenomics uncovers potential trophic interactions in rhizosphere soil.* bioRxiv, 2020 **doi.org/10.1101/2020.08.21.262063**. (mSystems, in response to review)

25.   Zhalnina, K., K.B. Louie, Z. Hao, N. Mansoori, U.N. da Rocha, S. Shi, H. Cho, U. Karaoz, D. Loqué, B.P. Bowen, M.K. Firestone, T.R. Northen and E.L. Brodie, *Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly.* Nature Microbiology, 2018. **3**(4): p. 470-480.

26.   Hungate, B.A., *Ecosystem responses to rising atmospheric CO2 : Feedbacks through the nitrogen cycle*, in *Carbon Dioxide and Environmental Stress*, Y. Luo and H. Mooney, Editors. 1999, Academic Press: San Diego. p. 265-285.

27.   Phillips, R.P., A.C. Finzi and E.S. Bernhardt, *Enhanced root exudation induces microbial feedbacks to N cycling in a pine forest under long-term CO2 fumigation.* Ecology Letters, 2011. **14**(2): p. 187-194.

28.   Pett-Ridge, J., J. Shi, K.Y. Estera-Molina, E.E. Nuccio, M. Yuan, R. Rijkers, T. Swenson, K. Zhalnina, T. Northen, J. Zhou and M.K. Firestone, *Rhizosphere carbon turnover from cradle to grave: the role of microbe-plant interactions*, in *Rhizosphere Biology: Interactions with Plants*, V. Gupta and A.K. Sharma, Editors. 2020, Springer Nature Press.

29.   Tfaily, M.M., R.K. Chu, J. Toyoda, N. Tolić, E.W. Robinson, L. Paša-Tolić and N.J. Hess, *Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry.* Analytica Chimica Acta, 2017. **972**: p. 54-61.

30.   Lin, Y., A.N. Campbell, A. Bhattacharyya, N. DiDonato, A.M. Thompson, M.M. Tfaily, P.S. Nico, W.L. Silver and J. Pett-Ridge, *Differential effects of redox conditions on the decomposition of litter and soil organic matter.* bioRxiv 2020, https://doi.org/10.1101/2020.08.21.262246. Biogeochemistry Letters (in review)

31.   Whitman, T., R. Neurath, A. Perera, I. Chu-Jacoby, D. Ning, J. Zhou, P. Nico, J. Pett-Ridge and M. Firestone, *Microbial community assembly differs across minerals in a rhizosphere microcosm.* Environmental Microbiology, 2018. **20**(12): p. 4444-4460.

32.   Neurath, R., T. Whitman, I.X. Chu-Jacoby, A.M. Thompson, A. Lipton, M. Tfaily, J. Kyle, K.J. Bloodsworth, D.J. Herman, J. Pett-Ridge and M. Firestone, *Root carbon interaction with soil*

minerals is dynamic, leaving a legacy of microbially-derived lipids. bioRxiv https://biorxiv.org/cgi/content/short/2021.03.23.436628v1; Environmental Science & Technology, 2021.

33.   Murrell, J.C. and A.S. Whiteley, eds. *Stable Isotope Probing and Related Technologies.* 2011, ASM Press: Washington, D.C. 345.

34.   Hayer, M., E. Schwartz, J.C. Marks, B.J. Koch, E. Morrissey, A.A. Schutettenberg and B.A. Hungate, *Identification of Growing Bacteria During Litter Decomposition In Freshwater through H218O Quantitative Stable Isotope Probing.* Environmental Microbiology, in review.

35.   Koch, B.J., T.A. McHugh, E.M. Morrissey, E. Schwartz, N. van Gestel, P. Dijkstra and B.A. Hungate, *Estimating taxon-specific bacterial growth rates in intact soil communities.* Ecosphere, 2018, 9: e02090-15

36.   Morrissey, E.M., R.L. Mau, E. Schwartz, J.G. Caporaso, P. Dijkstra, N. van Gestel, B.J. Koch, C.M. Liu, M. Hayer, T. McHugh, J.C. Marks, L.B. Price and B.A. Hungate, *Phylogenetic Organization of Bacterial Activity.* ISME J, 2016.

37.   Schwartz, E., M. Hayer, B.A. Hungate, B.J. Koch, T.A. McHugh, W.G. Mercurio and K. Soldanova, *Stable isotope probing with 18O-water to investigate microbial growth and death in environmental samples* Current Opinion in Biotechnology, 2016, 41: 14-18.

38.   Bhattacharyya, A., A.N. Campbell, M.M. Tfaily, Y. Lin, R.K. Kukkadapu, W. Silver, P.S. Nico and J. Pett-Ridge, *Redox fluctuations control the coupled cycling of iron and carbon in tropical forest soils.* Environmental Science & Technology, 2018. **14**: 14129-14139.

39.   Hungate, B., J. Marks, M. Power, E. Schwartz, K. van Groenigen, S. Blazewicz, P. Chuckran, P. Dijkstra, B. Finley, M. Firestone, M. Foley, A. Greenlon, M. Hayer, K. Hofmockel, B. Koch, M. Mack, R. Mau, S. Miller, E. Morrissey, J. Propster, A. Purcell, E. Sieradzki, E. Starr, B. Stone, C. Terrer, and J. Pett-Ridge, *The functional significance of bacterial predators.* bioRxiv https://doi.org/10.1101/2021.02.22.432408; mBio, 2021 (in press).

40.   Muñoz-Dorado, J., F.J. Marcos-Torres, E. García-Bravo, A. Moraleda-Muñoz and J. Pérez, *Myxobacteria: moving, killing, feeding, and surviving together.* Frontiers in microbiology, 2016. **7**: p. 781.

41.   Seccareccia, I., C. Kost and M. Nett, *Quantitative analysis of Lysobacter predation.* Applied and Environmental Microbiology, 2015. **81**(20): p. 7098-7105.

42.   Wang, W., X. Luo, X. Ye, Y. Chen, H. Wang, L. Wang, Y. Wang, Y. Yang, Z. Li and H. Cao, *Predatory Myxococcales are widely distributed in and closely correlated with the bacterial community structure of agricultural land.* Applied Soil Ecology, 2020. **146**: p. 103365.

43.   Kalyuzhnaya, M.G., A. Lapidus, N. Ivanova, A.C. Copeland, A.C. McHardy, E. Szeto, A. Salamov, I.V. Grigoriev, D. Suciu and S.R. Levine, *High-resolution metagenomics targets specific functional types in complex microbial communities.* Nature Biotechnology, 2008. **26**(9): p. 1029-1034.

44.   Wilhelm, R.C., R. Singh, L.D. Eltis and W.W. Mohn, *Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing.* The ISME journal, 2019. **13**(2): p. 413-429.

45.   Trubl, G., H.B. Jang, S. Roux, J.B. Emerson, N. Solonenko, D.R. Vik, L. Solden, J. Ellenbogen, A.T. Runyon, B. Bolduc, B.J. Woodcroft, S.R. Saleska, G.W. Tyson, K.C. Wrighton, M.B. Sullivan, and V.I. Rich, *Soil viruses are underexplored players in ecosystem carbon processing.* mSystems, 2018. **3**.

46.   Emerson, J.B., S. Roux, J.R. Brum, B. Bolduc, B.J. Woodcroft, H.B. Jang, C.M. Singleton, L.M. Solden, A.E. Naas, J.A. Boyd, S.B. Hodgkins, R.M. Wilson, G. Trubl, C. Li, S. Frolking, P.B. Pope, K.C. Wrighton, P.M. Crill, J.P. Chanton, S.R. Saleska, G.W. Tyson, V.I. Rich, and M.B. Sullivan, *Host-linked soil viral ecology along a permafrost thaw gradient.* Nature Microbiology, 2018. **3**: p. 870-880.

47.   Shaffer, M., M.A. Borton, B.B. McGivern, A.A. Zayed, S.L. La Rosa, L.M. Solden, P. Liu, A.B. Narrowe, J. Rodríguez-Ramos, B. Bolduc, M.C. Gazitúa, R.A. Daly, G.J. Smith, D.R. Vik, P.B. Pope, M.B. Sullivan, S. Roux, and K.C. Wrighton, *DRAM for distilling microbial metabolism to*

automate the curation of microbiome function. Nucleic acids research, 2020. **48**(16): p. 8883-8900.

48. Krupovic, M., V. Cvirkaite-Krupovic, J. Iranzo, D. Prangishvili and E.V. Koonin, *Viruses of archaea: structural, functional, environmental and evolutionary genomics.* Virus Research, 2018. **244**: p. 181-193.

49. Krupovic, M., A. Spang, S. Gribaldo, P. Forterre and C. Schleper, *A thaumarchaeal provirus testifies for an ancient association of tailed viruses with archaea.* Biochemical Society Transactions, 2011. **39**(1): p. 82-88.