**1. JUSTIFICATION**

Soils contain more organic carbon (C) than the atmosphere and all vegetation combined, making them a critical sink for atmospheric CO2 (Ciais et al., 2013). However, the ability of soils to maintain—let alone increase—their C sequestration capacity is threatened by climate perturbation and land-use intensification (Lal, 2004). Two modes of SOC stabilization are thought to interactively govern the fate of new C inputs, specifically, whether those resources are mineralized to CO2 or retained belowground. The first is abiotic and includes factors such as climate, soil pedology, and physicochemical properties that facilitate C occlusion in aggregates or the availability of adsorption sites on mineral surfaces. The second mode is biotic, and includes the genomic diversity of the microbial community, their functional traits, and their aggregated metabolic efficiency.

Microbial community diversity is immense: in one gram of soil, 5x104 different bacterial species (Curtis et al., 2002) and 2x108 fungal hyphae (Leake et al., 2004) may coexist. These microbial communities are essential for maintaining ecosystem services and nutrient cycling (Wagg et al., 2014; Danczak et al., 2020; Hall et al., 2018) and their necromass and metabolic byproducts can account for more than half of SOC (Liang and Balser, 2011; Liang et al., 2019). However, their activity also represents the single largest source of heterotrophic respiration (Schlesinger and Andrews, 2000). The role of microorganisms as a de/stabilizing force acting upon SOC depends on environmental factors, substrate quality, and life history traits. Improving our understanding of conditions that select for different life-history traits (Fierer et al., 2012; Malik et al., 2019) and impact the ratio between catabolic respiration *and* anabolic growth (Liang et al., 2017a) is critical for improving IPCC and global C cycle models (Bradford et al., 2016; Crowther et al., 2016).

While the genomic revolution has increased our appreciation for the functional diversity of soil microorganisms, we have yet to characterize how their ecological strategies and life-history traits contribute to SOC formation and persistence. Untangling these processes is critical for understanding how microbial metabolism structures the chemical composition and longer-term fate of SOC (Doetterl et al., 2018). A central intermediary in the continuum between active microbial decomposer and persistent SOC is dissolved organic carbon (DOC). DOC is one of the world’s largest active C pools, containing a similar amount of C as all living biomass in terrestrial and marine environments combined (Hedges, 1992). However, the molecular composition of DOC—hereafter termed the *metabolome*—remains poorly defined.

Linking complexity within the microbiome and metabolome with SOC persistence is a ‘holy grail’ in terrestrial ecosystem science. Using a newly collected soil set—representing topsoil and subsoil samples collected from 73 sites across N. America (Figure 1)—we will critically evaluate the following question: **How do interactions among soil biogeochemical properties, microbial community structure, and metabolite composition, govern C-sequestration potentials across soil ecosystems?** Results from this one-year small grant will be used to develop a conceptual model—explicitly representing microbial community and metabolome diversity—to improve our understanding of mechanisms governing SOC formation and persistence.This information will help inform policy and decision-making processes, improve sustainable land management, and increase the adoption of ‘natural climate solutions’ (Bastin et al., 2019; Griscom et al., 2017) to help decarbonize the atmosphere.



**Figure 1:** Locations of samples taken at 73 sites across N. America in 2019. Background relief displays the aridity index, extracted from WorldClim 2.0. Samples (colored by biome) were collected from the A-horizon and C-horizon at each site using sterile sampling techniques and transported on ice to the laboratory where they were processed immediately.

**2. BACKGROUND**

*The genesis of soil organic carbon*

Microbial community composition is shaped by, and in turn shapes, soil properties (Kaiser et al., 2016; Fierer et al., 2016; Fierer and Jackson, 2006; Hall et al., 2018), but the consequences of specific microbial traits and niche partitioning on SOC formation and persistence remain unclear. Many predictive frameworks have been borrowed from plant ecology. For example, plant communities display a positive relationship between diversity and niche complementarity, such that greater plant diversity results in the occupation of greater niche space (Zuppinger-Dingley et al., 2014). These feedbacks can translate to higher functional efficiency of a community (Knelman and Nemergut, 2014). Similar correlations for microbial communities are difficult to explicitly test, given complex microbial and substrate mixtures present in soil environments. However, a study of desert biological soil crusts, which have fewer species and less complex metabolite pools, found that taxa specialized on the assimilation of specific substrates (Baran et al., 2015). The authors suggest this metabolite-based niche partitioning could be an important factor maintaining microbial diversity. Because the natural habitats of microbes typically have limited resources (where the concentration of any individual metabolite is low), selection may favor the development of communities with high resource use efficiency (Roller and Schmidt, 2015). These communities are also characterized by low SOC mineralization rates that may contribute to SOC formation (Liang et al., 2017).

Maximal growth rates are organism-dependent; the efficiency of metabolism aggregated across a community has important consequences for SOC persistence. Microbial communities with access to higher-quality substrates (i.e. nitrogen-rich plant materials, root exudates) are thought to divert more energy toward anabolic pathways and the production of biomass (Liang et al., 2017). Metabolic byproducts (enzymes, metabolites), necromass, and cellular debris have been shown to efficiently adsorb on soil mineral surfaces (Miltner et al., 2012; Swenson et al., 2015). The resulting mineral-associated organic matter (MAOM) (Cotrufo et al., 2013; Kaiser and Kalbitz, 2012) increases SOC formation and persistence (Lehmann et al., 2008; Miltner et al., 2012; Lehmann and Kleber, 2015). Our soil collection spans a broad range of SOC (0.1—24 % in mass), soil nitrogen (0.1—1.0%), and litter quality (from structurally rigid conifer needles, to leguminous forbs), allowing us to probe fundamental relationships between microbial metabolism, metabolite production, and consequences for SOC storage.

In addition to governing rates of SOC cycling and long-term persistence, microbial communities drive ecosystem metabolite transformations. Resulting metabolite assemblages act as both a substrate and a product of microbial metabolism (Danczak et al., 2020) and are immensely complex: a single sample can contain upwards of 20,000 molecular formulae with varying degrees of chemical reactivity and structural complexity (Dittmar and Stubbins, 2013; Zark and Dittmar, 2018). Dynamic feedbacks between the efficiency of microbial metabolism, and both the quality and quantity of individual compounds within the metabolite assemblage likely shapes the probability of SOC retention belowground. For example, biogeochemical hotspots can become enriched in nitrogenous metabolites related to microbial growth and biomass production (Graham et al., 2018). These N-rich metabolites have a relatively high capacity to interact with soil minerals to form MAOM (Kopittke et al., 2020; Rillig et al., 2007; Schmidt and Martínez, 2016). Metabolite assemblages are also influenced by non-microbial factors, including climate conditions, vegetation inputs, and mineralogy. Developing a framework integrating cross-continental microbial and metabolome assemblages is imperative for better calibrating Earth-System models and predicting the whole-system influence on SOC persistence (Danczak et al., 2020).

*Linking microbial functional traits with metabolite composition*

Multiple attempts have been made to classify the enormous taxonomic diversity of microorganisms into trait-based frameworks analogous to those successfully developed for plant communities. One of the most common frameworks is the copiotroph—oligotroph continuum, which separates taxa into two functional groups: copiotrophs inhabit nutrient-rich environments and are characterized by rapid growth and reproductive rates, while oligotrophs inhabit nutrient-poor environments and concentrate resources towards energy acquisition and survival (Koch, 2001). Field and lab-based inquiry suggest the two-pool continuum may be overly simplistic (Fierer et al., 2016), particularly when applied to fungal communities (Lustenhouwer et al., 2020). A framework has recently been developed that classifies microbial taxa into three life-history strategies—high yield, resource acquisition, and stress tolerance (Y-A-S)—across two primary axes of environmental variation (resource availability and abiotic stress) (Malik et al., 2019). Here, we propose organizing taxonomic diversity using the Y-A-S framework to probe relationships between edaphic properties, plant-litter quality, metabolite composition, and SOC persistence.

According to the Y-A-S framework, ecosystems with low levels of resource limitation (temperate grassland or deciduous forest rhizosphere soils), may be dominated by high-yield strategists (similar to copiotrophs), which invest strongly in cell division, RNA metabolism (Song et al., 2017), central metabolism, and associated assimilatory pathways, such as amino acid, nucleotide, and fatty acid synthesis—the precursors of various cellular components (Malik et al., 2019). Ecosystems with complex or scarce resources under environmental conditions that limit metabolism (acidic coniferous forests, subsoils) are thought to select for resource-acquiring strategists (Malik et al., 2019), which reduce growth rates in favor of resource acquisition. By producing extracellular enzymes, these communities gain access to what is assumed to be structurally protected substrates that are sparingly available to other taxa (Allison, 2005). Resource-acquiring taxa are described as being functionally similar to oligotrophs, where individuals occupy narrow niche widths, but overall diversity is high (Fierer et al., 2016). Similarly, oligotrophic and resource-acquiring taxa would then invest more strongly in defense, motility, energy acquisition, and carbohydrate metabolism pathways relative to taxa with other life-history traits (Song et al., 2017). Finally, dryland systems (arid grassland, desert) may select for microbial communities that have adapted stress tolerance strategies (Malik et al., 2020, 2019). By releasing osmolytes (i.e. trehaolse, glycine betaine, proline, ectoine) and polysaccharide-rich extracellular polymeric substrates, stress tolerators actively protect their cells from desiccation, helping them outcompete taxa with lower tolerance to drought, freeze-thaw cycling, metal contamination, and other abiotic stressors (Schimel et al., 2007).

The Y-A-S framework has been successfully applied to forest ecosystems with variable stand ages and soil nutrient availability (Shao et al., *in review*). Researchers found that mature forest litter was of lower quality (higher C:N) than secondary forests. Mature forest systems were dominated by resource-acquiring and stress tolerating taxa (*Rozellomycota* and *Verrucomicrobia*), that invested energy in the degradation of complex substrates and survival. These soils were also characterized by lower microbial biomass, minimal accumulation of microbial residues (amino sugars indicative of necromass), and higher rates of CO2 mineralization. In contrast, nutrient-rich secondary forests were dominated by high-yield strategists (*Ascomycota, Proteobacteria, Bacteroidetes*). These taxa were associated with greater biomass production and residue accrual in secondary forest soils. These results suggest the balance between microbial growth and SOC persistence is modulated by aggregated community traits and environmental factors that influence the microbial metabolic efficiencies.

*Correlations between microbiome and metabolome diversity*

Evidence from a 25-year drought experiment in California shrub and grassland ecosystems suggests abiotic factors broadly shape microbial taxonomic and functional profiles, while substrate quality mediates the composition of expressed metabolite assemblages (Malik et al., 2020). Chemically complex shrub litter harbored a more diverse microbial community than grassland litter and was dominated by a diverse coalition of resource-acquiring taxa. These taxa invested in carbohydrate and amino acid metabolism, increasing the degradation, uptake, and assimilation of shrub litter. Communities on shrub litter were functionally similar in drought and ambient shrublands, suggesting properties of the litter horizon governed metabolic efficiency to a greater extent than precipitation. It is less clear whether differences are a result of organic carbon composition or the ecosystem of an organic horizon. In contrast, drought-exposed grasslands were dominated by stress-tolerators that allocated more energy towards survival (production and uptake of osmolytes and extracellular polymeric substances to retain water) than communities growing under ambient precipitation. These findings suggest that (1) substrate quality and climate collectively govern microbial community and metabolite diversity, and (2) complex substrates degraded by diverse microbial communities increase metabolome heterogeneity. **We therefore suggest diversity is a self-reinforcing feedback that cascades across multiple levels of organization**—from plant litter composition, through microbial communities, and into metabolite assemblages that retain the chemical footprint of microbial metabolism (Figure 2, Baran et al., 2015).

Bacteria are thought to be more opportunistic than fungi due to differences in metabolic capacities and physiological traits (Maron et al., 2018): while growth rate was a primary determinant of bacterial success during soil colonization (favoring rapidly growing high-yield strategists), metabolic efficiency determined fungal success (favoring *Basidiomycota* over *Ascomycota*). Overall, systems with high microbial diversity were found to accelerate the decomposition of simple and complex C sources. This may result from niche compartmentalization; while fungi and bacteria directly compete for simple plant-derived substrates (a highly redundant function), mutualism is commonly observed during the decomposition of more complex substrates (e.g. cellulose and lignin, a weakly redundant function) (Boer et al., 2005; Maron et al., 2018). The underlying composition of microbial communities drives differences in respiration, suggesting microbial diversity is an important explanatory factor explaining differences in C cycling (both rates and direction) at broader ecosystem scales (Albright et al., 2020).

**

**Figure 2:** Conceptual diagram showing the relationship between soil microbial diversity and molecular diversity of the metabolome in relation to microbial traits (adapted from Malik et al., 2019).

Within a highly controlled chemostat environment, researchers found that functional gene richness was linearly related to OTU richness (R2 = 0.52, p < 0.001) across time and a range of nutrient concentrations (Song et al., 2017). Although resulting metabolite diversity was not measured, it is likely higher functional gene richness encodes a relatively more diverse metabolite profile, linking microbial taxonomic and metabolite diversity. Similarly, researchers found that directly manipulating microbial community diversity, using a dilution-to-extinction approach, reduced overall SOC decomposition rates when microbial diversity was low, but increased C turnover particularly in soils with high nutrient availability (Maron et al., 2018). Others found that the chemical diversity of DOC tends to decrease over time in the absence of new C inputs (Mentges et al., 2017), suggesting sequential turnover, metabolization, and mineralization of DOC by microbial communities reduces its chemical reactivity. Although it is hypothesized that functional redundancy within microbial communities reduces the importance of any individual taxa on SOC decomposition (Roger et al., 2016), specialized species, particularly those with enzymatic capacities required to degrade more complex C compounds, can play an outsize role in ecosystem function (Niemenmaa et al., 2008).

Although still debated, considerable evidence supports the view that species diversity is positively correlated with community and ecosystem stability due to greater functional diversity (McCann, 2000). Results from a litter-bag decomposition study suggest diverse communities may also influence SOC composition (Wallenstein et al., 2010). Different forest types (aspen, pine, spruce) were found to harbor unique soil microbial communities that decomposed a common substrate (aspen litter) at different rates. In turn, they also produced unique metabolomes (Wallenstein et al., 2010), suggesting the composition of microbial taxa within an ecosystem can influence SOC formation and sequestration.

Triangulating ecosystem biogeochemical properties, microbial community composition, and metabolite profiles, is critical for improving process-based predictive models of soil C cycling. After fully characterizing biological and edaphic properties across our sample set, we will leverage machine learning and multivariate statistical approaches to evaluate how microbial community and metabolite assemblage influence ecosystem function (Woolf and Lehmann, 2019). The analytical flexibility of our approach will enable us to identify empirically defined community traits that should be incorporated in global climate models.

**3. PROPOSED RESEARCH**

*Objective:* Determine whether aboveground litter quality influences microbial and/or metabolome diversity and SOC persistence.

*Hypothesis:* Complex and/or lower quality (high C/N) plant litter increases microbiome and metabolome diversity.

*Alternative Hypothesis:* Plant litter quality does not influence microbiome and metabolome diversity.

*Test:* During this one-year small grant, we intend to relate plant litter properties (e.g., C:N ratio, chemical composition) to the diversity of microbial and metabolite assemblages. For sites where litter has a higher C/N ratio and more structural tissue content, cumulative CO2 mineralization will be high, but initial increases in respiration will be delayed until enzymatic activity begins breaking apart plant structural material. When communities reach basal respiration, glucose amendment will result in less CO2 mineralization relative to sites dominated by more homogenous, high-yield communities (as determined by querying the EcoDB) that mineralize more than assimilate plant-derived C.

*Rationale:* Sites with lower quality plant litter are degraded by taxonomically diverse microbial communities (Malik et al., 2020). Resource-acquisition strategists increase DOC diversity by releasing extracellular enzymes that fragment plant material (Allison and Martiny, 2008; Liang et al., 2017b; Lustenhouwer et al., 2020; Maron et al., 2018), and by producing a diverse suite of metabolites related to energy acquisition and amino acid metabolism (Song et al., 2017). In the absence of new C inputs (Fontaine et al., 2007), available nutrients will be exhausted, resulting in a degraded DOC pool. Although molecular diversity remains high, compound concentrations and reactivity are low (Landa et al., 2014; Mentges et al., 2017), reducing SOC mineralization.

*Experimental design*

We will leverage an extensive sample and data set for this one-year small grant project. In 2019, paired surface and subsurface soil samples (to 1 m depth) and representative live plant material were collected from 73 sites across the continental US, Alaska, and northeastern Canada (Figure 1). A-horizon (approximately 0-0.3 m) and C-horizon (approximately 0.5-1.0 m) samples were collected from soil pits established at each site. Bulk density samples were collected using the ring cutting method and were measured gravimetrically from three replicate samples per soil horizon using known sample volume after drying at 105°C. Subsamples were either immediately frozen, air-dried, or processed, as appropriate. At each site, representative plant samples were collected from a 1 m radius surrounding the soil pit and plant cover was estimated using Fiji image processing (ImageJ, National Institutes of Health). Sampling sites were selected to represent wildland and grazed systems spanning ten of the twelve USDA taxonomy soil orders (Histosols and Oxisols were explicitly excluded) and major biomes across North America—tundra, temperate deciduous forest and coniferous forests, temperate and arid grasslands, shrublands, and deserts. Across the sample set, mean annual temperatures range between -10 and 17°C, mean annual precipitation between 105 and 3205 mm, and elevation between 4 and 2200 m (climate data extracted from WorldClim 2.0 (Fick and Hijmans, 2017)).

The value of our sample set consists not only in the fact that vegetation, surface, and subsurface soils were collected in the same manner by the same team across a vast climatic and edaphic expanse, but that soils were extracted fresh. Extracting fresh samples permits microbial community and metabolomics analyses to be conducted that are not possible with dry soils from national archives. We suggest datasets such as the one collected here are essential for catalyzing discovery and gaining mechanistic insight into fundamental rules governing SOC formation and persistence.

For each sample, we have already measured fundamental soil properties including carbon (C) and nitrogen (N), total dissolved organic C and N, microbial biomass C and N (using the direct chloroform fumigation method), cation exchange capacity, pH, electrical conductivity, bulk density, particle size distribution, and moisture (both gravimetric and hygroscopic). We have also measured several critical geochemical drivers of SOC sequestration. Using hydroxylamine hydrochloride, we extracted non-crystalline oxides (quantified by ICP-OES) and measured the amount of DOC released upon oxide dissolution (Chao and Zhou, 1983; Kramer and Chadwick, 2018). We also used an established size fractionation technique to assess how SOC is distributed between two ecologically unique pools: (1) mineral associated organic matter (MAOM), and (2) particulate organic matter (POM) (Cotrufo et al., 2019). POM is predominantly of plant origin and represents an important substrate for microbial metabolism, while MAOM is comprised primarily of microbial products and persists belowground through physical protection in aggregates and chemical bonding to minerals (Cotrufo et al., 2019; Kleber et al., 2015). **Together, this suite of biogeochemical properties allows us to probe questions central to C distribution in surface *and* subsurface soils that span a range of environmental conditions, but they do not allow us to assess how microbial communities regulate the chemical reactivity of DOC and SOC persistence.**

|  |  |
| --- | --- |
| **A close up of a map  Description automatically generated** | **Figure 3:** Canonical correspondence analysis elucidating relationships between fungal (ITS) and bacterial (SSU) assemblages and their environment (biome or soil order). Distances are based on Bray-Curtis dissimilarity calculated from the weighted abundance of individuals shared among samples. Extreme samples (Gelisols, Aridisols, tundra soils) were removed for clarity (unpublished preliminary data). |

To address questions pivotal to biotic control of SOC persistence, we characterized bacterial and fungal diversity by sequencing (MiSeq 2 x 250 bp) phylogenetic gene marker libraries for bacteria (v4 region of the 16S rRNA gene) and fungi (ITS region 2 of the 18S rRNA gene) from soil DNA (Figure 3). We recovered an average of 80,000 sequences per sample for bacteria and 100,000 sequences per sample for fungi. Overall, the composition of bacterial communities was significantly influenced by climate (PERMANOVA R2 = 0.27, p < 0.001) and to a lesser extent by soil order (R2 = 0.08, p < 0.001) and soil horizon (R2 = 0.04, p < 0.001). Similarly, fungal diversity was most influenced by climate (R2 = 0.31, p < 0.001) and soil order (R2 = 0.08), but not by soil horizon (R2 = 0.01, p = 0.2).

To better understand the functional attributes of these diverse soil ecosystems, we propose in this one-year project to use a new ecological database (‘ecoDB’, Wilhelm et al., *in prep*) developed by our colleagues at Cornell University, which can attribute growth rate, pH preference, substrate use, and stress tolerance to a breadth of soil organisms with unique roles in SOC cycling (Morrissey et al., 2019; Pepe-Ranney et al., 2016; Rocca et al., 2019). Based on the ecoDB, we will define high-yield (copiotrophic) phylotypes as organisms (i) exhibiting rapid growth and senescence, or (ii) those possessing high numbers of ribosomal gene copies, determined using the rrnDB (Stoddard et al., 2015). A subset of resource-acquiring taxa, phenolic acid degrading phylotypes, are particularly relevant to SOC turnover. These phylotypes will be defined, by (i) relatedness to genomes encoding peripheral pathways for aromatic degradation (Fuchs et al., 2011), including *pobA,* which contributes to soil priming (Wilhelm et al., *in prep*), and with (ii) the capacity to degrade phenolic acids (benzoic acid, 4-hydroxybenzoate, and coniferyl alcohol). Figure 4 displays an example of the ecoDB applied to a metabolomic study of litter decomposition (Lynch et al., unpublished data), where different plant litter samples were incubated with their native phylosphere communities (green network) or a mixture of phylosphere and added soil inoculum (brown networks represent phylotypes found only in soil inoculum, purple represents phylotypes found in both environments). Using the ecoDB we show that litter decomposition is driven by phylotypes assigned *copiotrophic* (high-yield) and *cellulolytic* status (starred symbols), which were active in soil but not phylosphere environments.

|  |  |
| --- | --- |
| A picture containing basketball, athletic game, sport  Description automatically generated | **Figure 4:** Co-occurrence patterns among taxa degrading different plant litters. Nodes correspond to individual taxa and edges define correlation patterns between taxa. Here, the ecoDB shows that litter decomposition is driven by phylotypes assigned *copiotrophic* (high yield) and *cellulolytic* status (starred symbols). These taxa were present in soil inoculum (brown), but not native phylosphere (green) or shared (soil + phylosphere, purple) environments (Lynch et al., unpublished data). |

*Metabolite composition and distribution*

Metabolites extracted from soils spanning a continental scale will encompass a vast array of molecular and biochemical classes, chemical properties, and concentration ranges. Understanding how metabolite composition influences SOC accrual will therefore require applying several analytical platforms in tandem and integrating computational pipelines (De Preter and Verbeke, 2013; Vernocchi et al., 2016). Mass spectrometry (MS) is an established technique for characterizing the molecular fingerprint of complex environmental samples (White III et al., 2017); with broad dynamic range, high sensitivity, and established compound libraries, MS-based analyses are well suited to our applications (Martins-de-Souza, 2014; Zhao and Hartung, 2015). In this one-year small grant, we aim to cover multiple metabolic pathways in both primary and secondary metabolism, and quantify the abundance and relative distribution of representative compound classes (i.e. carbohydrates, amino acids, amino sugars, organic acids, lipids/fatty acids, vitamins, etc.) (Sumner et al., 2003; Wercinski, 1999).

We will extract organic matter from each soil and analyze metabolite composition using three complementary MS approaches to profile polar, nonpolar, and volatile organic C compounds. Water-extractable OC will be analyzed by LC-MS/MS in positive mode (C18 column) and negative mode (HILIC column) to characterize non-volatile semi-polar and polar metabolites associated with primary metabolism (i.e. carbohydrates, organic acids, amino acids, and nucleobases). Non-polar OC will be analyzed by reverse phase LC-MS/MS (C18 column) to profile sterols, flavonoids, alkaloids, phenolic acids, lipids, and other secondary metabolites that are more species-specific than primary metabolites can help differentiate pathways of plant *versus* microbial metabolism (Salminen and Karonen, 2011; Zwetsloot et al., 2018). Volatile and semi-volatile compounds will be analyzed by electron impact gas chromatography MS (EI GC-MS) to profile fatty acids, carbohydrates, and organic acids associated with primary metabolism. Coupling different MS technologies will facilitate acquisition of extensive metabolomic fingerprints that are not achievable with a single technique (White III et al., 2017).

We will prepare samples for untargeted metabolite characterization following standard procedures (Swenson and Northen, 2019). Briefly, polar and semi-polar metabolites will be extracted by combining 8 g of lyophilized soil (previously sieved to < 2 mm) with 24 mL of LC-MS grade water. After shaking samples for 1 h at 200 rpm on a refrigerated orbital shaker, extracts will be centrifuged for 15 min at 4°C and 3220 *g*. Supernatants will be filtered through 0.45 μm Acrodisc® syringe filters fitted with gamma irradiated Supor® membranes (Pall Life Sciences). Lyophilized extracts will be resuspended in the appropriate carrier and with HILIC or C18 internal standard solutions and analyzed on a QE-HF LC-MS/MS system. Fully non-polar metabolites will be extracted as above but using 24 mL of nonpolar extractant (50% ethyl acetate, 50% water) rather than water to extract soils. After samples are shaken and centrifuged, the upper ethyl acetate layer will be separated and dried on a speedvac for 1 hour, then resuspended in C18 internal standard solution, centrifuged to remove salts, and analyzed by reverse phase LC-MS/MS as above. Finally, water-extractable molecules will be derivatized (using 50 μL N-methyl-N-trimethylsilyltrifluoroacetamide and 1% trimethylchlorosilane) following standard protocols (Lynch et al., 2019) followed by analysis of volatile compounds on a Trace GC Ultra coupled to a Thermo ISQ MS. All sample volumes will be normalized to ensure the same carbon concentration is injected for each sample (targeted concentration of 2 mg C mL-1). Samples will be analyzed at University of Idaho Mass Spectrometry Core Laboratory, with QA/QC samples injected between every six samples. Data will be collected in full MS/dd-MS2 mode and collision energies of 10-40 eV will be applied to fragment unknown compounds of interest.

All MS data will be processed using advanced open-source Molecular Networking (MN) software. MN is a powerful strategy to characterize molecular differences within and between complex samples, clustering molecules based on chemical relatedness, or the similarity of MS fragmentation patterns (Watrous et al., 2012). We will pre-process tandem mass spectrometry (MS/MS) and EI GC-MS data using MZmine2, which reliably dereplicates spectra by resolving isomeric compounds, annotating MN with predicted chemical formulas, and accounting for ion abundance to facilitate between-sample comparison (Olivon et al., 2017). Next, we will annotate molecular features of interest using the Global Natural Products Social Molecular Networking (GNPS) platform (Wang et al., 2016), and assign annotated molecules to a five-level chemical taxonomy (kingdom, superclass, class, subclass, and direct parent) using ClassyFire (Feunang et al., 2016). We will infer chemical relationships among samples using supervised machine learning, visualized within the newly developed Qemistree platform. Qemistree output will then be integrated with QIIME 2 to simultaneously assess metabolome and microbiome complexity (Tripathi et al., 2020). Finally, we will use MetaboAnalyst to probe statistical relationships among samples (Xia and Wishart, 2016) and to identify molecular features that separate samples across our primary scales of interest (biome, soil order, soil horizon). All raw spectral files will be published on the open-access GNPS platform.

Once we quantify carbon partitioning between particulate (typically plant-derived) and mineral-associated pools, we will test whether surface colonizing communities correlate with SOC accrual on mineral surfaces. We expect organic acids, particularly plant-derivatives (e.g. ferulic acid, benzoic acid, catechin, naringenin, cinnamic acid) will be enriched in surface soils where root exudation is high (Swenson et al., 2015; Zwetsloot et al., 2018), while amino acids and phosphorus-bearing compounds (e.g. guanine, lysine, thymine, xanthine, inosine monophosphate and 2’-deoxyadenosine monophosphate) will predominate in subsurface soils (Graham et al., 2018; Swenson et al., 2015) where metabolite—mineral interactions dominate (Kaiser and Guggenberger, 2003).

*Potential carbon mineralization*

Frozen soil samples will be slowly thawed over a 48-60 h period from -18 to 4°C in a refrigerator (Walz et al., 2017). Once samples are thawed, 50 g subsamples (pre-sieved to < 4 mm) will be transferred to sterile 1 L glass Mason jars fitted with gas-tight lids. Water contents will be adjusted to 60% water holding contents and monitored throughout the incubation to ensure constant weight. Because cumulative respiration (i.e. using standard KOH trap methods) cannot account for nuanced shifts in CO2 production over time that are associated with microbial dormancy, enzyme induction lags, or *in-situ* shifts in community-level carbon use efficiency (Allison, 2005), we will measure rates of CO2 production with high temporal resolution (every 10 hours). As soon as water is added, we will continuously monitor CO2 efflux using an in-house Picarro CO2 stable isotope analyzer (G2201-*i*, Santa Clara, CA, USA). This approach allows us to probe fundamental relationships between microbial community activity dynamics and DOC molecular complexity, as well as assessing SOC vulnerability to mineralization at each site.

Our Picarro system is fitted with a custom-built sequential-multiplexing manifold (DeCiucies et al., 2018) to monitor CO2 concentrations in 60 jars. During each sample collection period, headspace CO2 concentrations will be sampled for six minutes, purged for six minutes with CO2-free air, and sampled again to determine the starting CO2 concentration for the next measurement period. This cycle time has been calibrated to ensure headspace CO2 concentrations do not exceed 2%, thereby maintaining an oxic environment throughout the duration of the experiment (Lynch et al., 2018).

While we will monitor the initial flush of CO2 (including time to reach peak respiration, and the slope of the decline towards basal respiration) these dynamics are typically associated with a disturbance response to aggregate destabilization, among other processes. Therefore, we will also calculate steady-state (or basal) respiration (Creamer et al., 2014) and cumulative CO2 mineralization over a 30-day incubation period after the mineralization flush, for each site. After this period, we will amend soils with universally labeled, isotopically enriched 13C-glucose (increasing SOC by 1% for every soil, resulting in variable C additions). The linear increase in CO2 mineralization after glucose amendment is directly proportional to the concentration of *in situ* microbial biomass (Kaiser et al., 1992). This substrate-induced respiration (SIR) can also provide a useful index of the catabolic potential of the soil community (Degens, 1997). Using a two-endmember stable isotope mixing model approach, we will calculate the relative proportion of 13C-glucose *versus* native SOC assimilated by microbial communities or mineralized to 13C-CO2. This approach will also allow us to determine whether the addition of an easily assimilable substrate results in priming and turnover of SOC or retention of microbial-derived residues belowground (Lynch et al., 2018). We will then relate these findings to functional diversity within the soil microbiome.

CO2 mineralization dynamics (initial CO2 efflux, basal respiration, cumulative mineralization over a 30-day period, and SIR response curve) will be repeated at two temperatures, allowing us to account for the known differences in community adaptation to site-specific temperatures (Bradford et al., 2019). These data allow us to assess the potential vulnerability of SOC to mineralization at each site (calculation of a Q10 response is not the primary goal of this assessment, Crowther et al., 2016). Accordingly, soils will be incubated at 20°C to induce the maximum potential heterotrophic soil respiration rate (Doetterl et al., 2015; Lal et al., 2001) and at 10°C to represent mean July temperature conditions experienced by Arctic samples collected from northern Alaska (Lynch et al., 2018) and the mean annual temperature of the majority of our study sites. These combined mineralization studies will allow us to probe how soil efflux rates are related to the underlying complexity of DOC pools, microbial communities, and their functional attributes.

*Integrating potential mineralization with DOC functional complexity*

A total of twenty additional samples will be selected for detailed post-incubation analysis. Ten samples will be selected to capture pre-incubation metabolite diversity (highest and lowest diversity) and ten additional samples will be selected to capture post-incubation variability in CO2 mineralization (highest and lowest cumulative release of CO2, normalized to SOC). After the thirty days of incubation (and before addition of glucose) sample subsets will be extracted for bacterial and fungal sequencing and for DOC and its metabolome characterization, following the same protocols as outlined above. Together, this post-incubation sample set will allow us to assess how metabolome composition shifts over time and to relate the structural composition and reactivity of DOC with potential SOC mineralization. We will integrate these results with litter composition and edaphic properties to build a new conceptual framework that we will interrogate and revise in future experiments (Figure 5).

****

**Figure 5:** Conceptual diagram showing possible results that generate new insights into the relationship of microbial metabolism and DOC molecular diversity.

**4. BROADER IMPACTS**

*Imagination is more important than knowledge. Knowledge is limited.*

*Imagination encircles the world. –Albert Einstein*

In 1878, J.H. van ‘t Hoff gave an inaugural address at the University of Amsterdam titled *Verbeeldingskracht in de Wetenschap* (The power of imagination in Science), in which he concluded that the most prominent scientists are uniquely gifted with this quality (“Jacobus H. van ‘t Hoff – Nobel Lecture,” n.d.). J. H. van ‘t Hoff was later awarded the first Nobel Prize in Chemistry (1901) for his work on stereochemistry and the theory of dilute solutions, and is now considered a founder of physical chemistry.

Although instruction within the arts is thought to improve students’ creativity, spatial reasoning, observational acuity, objectivity, and perseverance (Gurnon et al., 2013), the humanities are often marginalized in science curricula (Root-Bernstein et al., 2008). We will develop course material and promote informal gatherings to foster interdisciplinary communication and collaboration between the arts and sciences at two land-grant institutions: Cornell University and the University of Idaho. Based on our previous work at Cornell University (Fig. 6), we propose to implement a transformative approach that utilizes artistic practices to explore scientific themes, broadens collaborations between disciplines, enhances educational tools in the sciences, and expands communication of scientific insights outside the university.



**Figure 6:** “Fifty Shades of Soil” exhibition exploring the diversity of soils collected for this proposed project. Traditional climate and biogeochemical data for each soil can be seen marching behind the figurines and up the walls of the installation (Tjaden Hall, Cornell University, November 2019; collaboration between the French artist Karine Bonneval (https://www.karinebonneval.com), the co-PI Laurel Lynch, and the PI Johannes Lehmann). Ms. Bonneval connected science and humanities disciplines during her month-long artist residency in the Lehmann Lab. An accompanying outdoor installation included a collection of interactive sculptures that encouraged passerby to listen to the sounds of soil organisms beneath their feet (‘ecouter la terre’). She also created a sonic rhizotron to assess whether the root systems and fungal symbionts of sunflowers and beans were attracted or repelled by various low frequency sounds (including frequencies selected by a plant neurobiologist to improve plant growth (Mancuso and Viola, 2015), an Antarctic soundscape, barred owls, Sandai waves, and predatory grubs eating plant roots).

Specifically, we propose three activities in collaboration with Karine Bonneval during this grant period. First, we will hold one-day workshops at each institution under the leadership of Ms. Bonneval to connect local biologists, ecologists, climate scientists, and environmental scientists with artists, curators, poets, and performance artists, to discuss connectivity between the sciences and environment. We are particularly interested in working with the Office of Tribal Relations (OTR at the University of Idaho) to encourage reciprocal collaboration with Nez Perce County and other local communities. Second, Ms. Bonneval will explore current and future science themes of our student groups and query the assumptions, perceptions, behavior, and conclusions from students’ research and educational experience. We will link this activity to the classroom in i) Environment and Sustainability (ES 20000) that PI Lehmann teaches at Cornell that includes sciences and humanities and to ii) Introduction to Soil Sciences that coPI Lynch will teach at the University of Idaho (typical enrollment includes ~80 first-year students from 12 different majors). The outcome of these co-lab experiences will generate an interactive display staged at a central location on each science campus and an educational module tailored for high school students. Third, we will bring the educational module to a local community center that serves underrepresented minorities, including black and Hispanic communities (Cornell) and tribes (U of ID). Our goal is to bring female middle and high school girls closer to science through hands-on activities, and to increase the number of girls choosing a STEM career. Our intent is to make the sciences accessible by conveying the exciting and creative nature of scientific exploration.

We will make a considerable effort to attract and promote staff and students from underrepresented minorities through this grant. One graduate student that we intend to support is Rachelle LaCroix, who has been funded for the first two years of her PhD studies through a Cornell diversity fellowship. The second fellowship will be used to attract a minority candidate to the University of Idaho, where coPI Dr. Laurel Lynch will start her tenure-track faculty position in Fall 2020. We will work with the University of Idaho NSF Louis Stokes Alliances for Minority Participation (NSF-LSAMP) program and Office of Tribal Affairs to attract Native American graduate students and with the Cornell Diversity and Inclusion Recruitment Office to attract students from Historically Black College and Universities and Hispanic Serving Institutions. We will also leverage our existing college recruitment weekends to identify suitable candidates.

Through these activities, we intend to create an institutional environment that explicitly values and promotes diversity and transformative creativity in science, both of which are central to scientific pursuit. We will make use of artistic creativity, which is defined as ideas and actions that transform laws, principals, materials, and thoughts both of the artist and the audiences (Lehmann and Gaskins, 2019). With this approach, we are particularly interested in linking to indigenous concepts of knowledge and place in our regions.