Title. Disentangling the role of microbial interactions and spatial structure in soil organic matter decomposition

Call Topic. ETI 2

Specific Aims. The study of the soil microbiome is critical to understanding the ecological functions carried out by this ecosystem, particularly with respect to the cycling of carbon and nitrogen nutrient sources. Despite the importance of this community, the direct analysis of the soil microbiome *in situ* is difficult due to the incredible complexity of this environment with thousands of species and millions of potential interactions. Understanding is further complicated by the intrinsic heterogeneity of soil where the multi-step pathways related to soil organic matter (OM) decomposition are often focused in certain "hotspots" (areas of high nutrient concentration with increased bacterial biomass and metabolic activity) while other areas of the same soil site are more muted in their expression of decomposition pathways. A better understanding of OM decomposition processes is also hampered by the lack of fundamental knowledge of the genomic and phenotypic characteristics of the major microbial players. Despite these hurdles, OM decomposition is a critical component of natural soil systems and therefore its understanding is key to improving the mechanistic underpinnings of genome-enabled models, carbon and nitrogen cycling in soil, and interactions between species surrounding hotspots. In the experiments proposed here, we will fill these knowledge gaps by using *previously developed project resources including a Model Soil Consortium (MSC) for OM decomposition (a chitin-degrading microbiome naturally evolved from the soil microbiome) and newly developed analytical devices* in combination with EMSL resources to derive novel insights into the molecular controls that drive interaction networks related to carbon and nitrogen cycling and how these interaction networks are driven by hotspots of OM decomposition. Specifically, we will address the following **hypotheses**:

1) Within a community of microbial species with widespread chitin degrading potential, chitin metabolism is driven only by a small subset of microorganisms expressing chitinase. Subsequent loss of these primary chitin degraders will promote the expression of chitinase by alternative species to fill this metabolic gap, resulting in minimal loss of overall community growth.

2) Chitin patches added to a structured environment will drive a spatial hierarchy of species with chitinase expressers (primary degraders) directly adjacent to a chitin source (defined as a chitin hotspot) and secondary degraders occupying the periphery and expressing less expensive metabolic processes.

Mission Relevance. This work directly addresses the BSS mission to predictively understand intra- and inter-cellular microbial systems using molecular approaches and computational analyses to reveal interactions that underlie OM decomposition. We also specifically address EMSL's mission to advance and integrate process-level understanding of complex systems across spatial scales, develop new technology tools couple observations, experiments, and theory with modeling and simulation.

Background. The incredible amount of soil microbial diversity is a major challenge to understanding OM decomposition in soil, as only a fraction of soil biochemical pathways is currently categorized in annotation databases [1] and the complexity of interactions make data analysis and interpretation daunting. In addition, OM decomposition processes occur in a heterogeneous soil environment that facilitates hot spots of decompositions rather than a spatially homogenous breakdown of complex carbon and nitrogen sources. Understanding decomposition hotspots is therefore a key challenge in soil science.

Here, we use chitin as a model substrate to investigate the microbial interactions that mediate decomposition of organic substrates rich in C and N. Chitin is the second most abundant structural polysaccharide on the planet after cellulose [2], and, because it is insoluble in water, chitin dispersion is non-uniform at the microbial scale making it an ideal substrate to study OM hotspots. In addition, the breakdown of chitin is likely the result of interactions among species [3]. Together, the ecological relevance and prevalence of

chitin decomposition, along with broad genomic characterization makes this functional process a strong model for investigating how microbial interactions regulate OM decomposition.

To address these scientific questions, a taxonomically diverse, chitin-degrading consortium was developed built from eight species isolated from a naturally evolved soil microbial community [4]. This consortium is termed Microbial Soil Consortium 2 (**MSC-2**) and has been used to understand how coordinated microbial phenotypes can facilitate chitin decomposition. All eight members of MSC-2 have fully sequenced genomes that can be used for interpreting data generated by this proposal. Many of the MSC-2 species have been labeled with fluorescent tags, facilitating spatiotemporal analysis of MSC-2. Using monoculture and co-culture growth assays it has been shown that chitin degradation is carried out only by a subset of species within MSC-2, but that non-chitin degraders (e.g. Streptomyces) comprise a significant portion of the total community biomass (**Figure 1**). These results suggest several putative interspecies interactions critical to OM decomposition.

A

Organism	Chitin Growth
Streptomyces venezuelae	No
Neorhizobium galegaae	No
Dyadobacter fermentans	Yes
Sphingopyxis fribergensis	No
Ensifer adhaerens	Yes
Variovorax paradoxus	Yes
Sinorhizobium meliloti	Yes
Rhodococcus	Yes

B

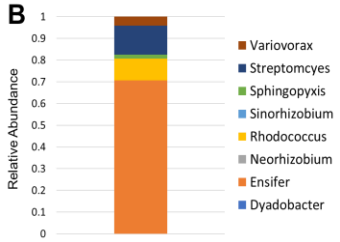

Figure 1. MSC-2 Chitin growth. A) Table showing the eight members of MSC-2 and their ability to grow on chitin in monoculture. **B)** Relative abundance of MSC-2 species when grown on chitin as a co-culture.

This consortium will be used in a series of integrated experiments and models to address a set of related questions surrounding our hypotheses including: How do organisms coordinate metabolism during chitin decomposition? Which interactions lead to microbial hotspots of decomposition? Which taxa are the primary decomposers of chitin and how do other taxa not directly involved with chitin metabolism support and/or derive nutrients from the primary decomposers, especially in a spatially resolved system representative of soil?

Approach or Work Plan. This work will leverage existing capabilities developed in collaboration with EMSL as described below.

Our proposal is divided into two Aims, each of which will answer one of the hypotheses proposed above.

Aim 1. Delineate the role of key microbial taxa in MSC-2 during active chitin degradation. Because of the enormous physicochemical variation in soils, we will begin our analysis of species interactions driving chitin degradation using a liquid system. Use of a liquid system will allow us to easily modify the size of cultures if necessary, to sample in a non-destructive manner, and to collect sufficient cell biomass to meet the needs of the -omics technologies that we will use in collaboration with EMSL. As chitin is non-soluble, there will exist in this liquid system hotspots at the microbial scale (chitin particles) and we propose to study such hotspots in greater detail in the context of spatial structure in **Aim 2**. A great deal of data has been gathered on the MSC-2 consortium when grown with chitin in liquid culture conditions supporting the feasibility of our scope and have also developed modeling and analysis pipelines [5]. Our first task within this proposal will be to grow the complete MSC-2 community (all eight species) under conditions where chitin is the only source of carbon. Growth will take place in 100 mL liquid cultures of M9 supplemented with 100 ppm chitin with each species of MSC-2 added from an axenic stock and all eight species started at identical cell numbers. We will then collect samples at three timepoints (0, 72 and 120 hours after start of growth) with five replicates each (n = 15). These timepoints were chosen based on previous analysis of MSC-2 and what is predicted to be high points in chitin degradation kinetics. Samples will be collected for amplicon, transcriptomic, proteomic, lipidomic and (extracellular) metabolomic analyses. Amplicon data will be examined with generalized Lotka-Volterra (gLV) models to identify microbial interaction networks of MSC-2 species

during chitin metabolism [6]. Then, we will use an advanced mixed-integer linear programming-based regression algorithm (available as an in-house code) to explore a set of candidate gLV models that each equally fit the species abundance data. The resulting interaction networks will be subsequently evaluated using metabolic models to select the most plausible/accurate network of microbial interactions during chitin decomposition, which will be used to inform experiments below. In collaboration with KBase, we will cross-reference the gLV with genome scale metabolic models of the eight MSC-2 members (already constructed; https://narrative.kbase.us/narrative/60446) to predict interspecies metabolic interactions [7]. Transcriptomic and proteomic data will be used to determine which species of MSC-2 are expressing chitin degrading processes and which other species express processes to feed of these chitin degraders. Metabolomic and lipidomic data generated in collaboration with EMSL will enable us to investigate microbial expression patterns that are not observable with genomic and transcriptomic data alone, thereby greatly expanding our ability to test our hypotheses. Metabolic data will also be integrated with transcriptomic data using the previously developed MEMPIS algorithm [8] to define community level processes taking place. Outcomes from these experiments will be the identification of which species are involved with chitin degradation processes, what functional pathways secondary consumers express and which processes are carried out at the community level vs. at the individual species level.

Next, we propose to use leave-out experiments to define the role of certain constituent members of MSC-2 in more detail in driving the community process of chitin degradation and community growth. We will choose two species of MSC-2 that, based on the data collected above, we predict to have the greatest effect on community growth. Above experiments will be repeated two times, each time lacking one of the selected members with five replicates each (leading to 15 additional samples for each leave out experiment and a total of 45 samples for Aim 1). Similar data analysis will take place and gLV and MEMPIS models will be compared between the complete MSC-2 community and MSC-2 sub-communities lacking one member to identify how the lack of certain driver species affects chitin degradation by the whole community. A better understanding of how the community is organized with and without certain driver members will help answer H1: *Within a community of microbial species with widespread chitin degrading potential, chitin metabolism is driven only by a small subset of microorganisms expressing chitinase. Subsequent loss of these primary chitin degraders will promote the expression of chitinase by alternative species to fill this metabolic gap, resulting in minimal loss of overall community growth.*. This Aim will require 45 16S rRNA gene amplicon analyses (external), 45 transcriptomic samples (external), *45 proteomic (EMSL), 45 lipidomic (EMSL), and 45 metabolomic (EMSL) samples* and result in a *community metabolic model* describing interactions among microbes in relation to chitin decomposition.

Aim 2. Map the spatiotemporal structure and activity of MSC-2 members in the context of OM hotspots in a synthetic soil analog. Understanding community interactions in the context of a structured soil-like system will lead to more realistic and transferable results for understanding the soil microbiome. In Aim 2, we will build on the chitin incubation experiments described in Aim 1, with three critical changes. First, incubations will take place in a soil-like fabricated ecosystem (Soil EcoFAB). The Soil EcoFAB is a microfluidic device that mimics soil porosities and is constructed through a combination of deep reactive ion etching and soft lithography technologies (https://eco-fab.org/). Our soil EcoFAB consists of different porosities (50 µm to 20 µm) to represent porosities of different soil systems. The microscopy that can be applied to Soil EcoFABs and the fact that they are soil analogs means that we can use them to better understand the spatial structure of microbial communities centered on chitin hotspots. MSC-2 has already been used in Soil EcoFAB devices, a success that was driven by the incorporation of several fluorescent tags into the genomes of MSC-2 species (**Figure 2**) allowing for visualization of the spatial arrangement and the growth rate of individual MSC-2 species within the device. The second change will be the introduction of chitin patches (areas of high chitin concentration added to pre-determined coordinates within the EcoFAB), an approach that has already been used with success in previous work [9]. This approach will help us to define where hotspots are and make imaging and analysis of these hotspots much easier. The third change will be the incorporation of a chitinase specific

probe (**Figure 2**) that, through fluorescent microscopy, will confirm that chitin patches represent true microbiological hotspots of chitinase activity and also reveal which species are degrading chitin. Preliminary work already completed using these strains, chitinase probes and Soil EcoFAB devices speaks to the feasibility of these experiments and we expect to have all samples sent to EMSL by April 2022.

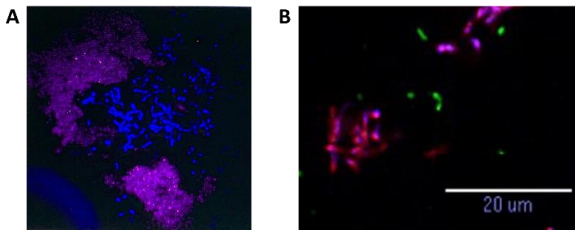

Figure 2. Fluorescent microscopy of MSC-2 chitin degradation in structured system. A) Co-incubation of labeled Rhodococcus (blue) and Ensifer (purple). **B)** Co-incubation of Rhodococcus (red) with Sinorhizobium (green) and chitinase probe (purple) associated with Rhodococcus, showing that this species expresses this enzyme.

Mirroring the experimental design in Aim 1 in a more spatially-complex environment, MSC-2 species will be incubated in Soil EcoFAB devices with M9 and chitin patches. We propose to use two sets of four species each for incubation in Soil EcoFABs (5 replicates per set, n=10). We will choose species based on the outcomes of Aim 1 which will reveal keystone members and we will also be guided by preliminary data showing which strains grow well together in the Soil EcoFAB. Imaging of Soil EcoFAB chitin hotspots will take place at 3, 7, and 14 days post inoculation using a Yokogawa Spinning Disk confocal mounted on a Leica DMi8 microscope at EMSL. These longer time courses are a result of observed slower growth in Soil EcoFAB devices compared to liquid cultivation. One day prior to each imaging day chitinase probe will be added to the devices so that species-specific expression of chitinase can be determined. Imaging will be used to determine 1) Which species are located adjacent to the chitin source, 2) which species are co-located with these chitin-adjacent species and 3) which species are expressing active chitinases. Imaging will only be applied to 2 replicates per set with the additional three replicates included for statistical power in our -omics analysis. At the conclusion of the experiment, we will destructively sample to collect bulk amplicon, transcriptomic, proteomic, lipidomic and (extracellular) metabolomic data. Image data will be analyzed together with agent-based and reaction-diffusion microbial interaction models, which enable predicting microbial assemblies and interactions at μm and cm scales, respectively. These models contain interspecies interactions as key parameters, which will be determined such that predicted spatial patterns are consistent with image data. Multi-omic data will reveal species abundance levels, expressed pathways and metabolic/lipidomic environments but with the added knowledge of who is an active chitinase expresser and who is a secondary consumer (based on our probe results). This work will answer H2: *Chitin patches added to a structured environment will drive a spatial hierarchy of species with chitinase expressers (primary degraders) directly adjacent to a chitin source (defined as a chitin hotspot) and secondary degraders occupying the periphery and expressing less expensive metabolic processes*. This Aim will require 10 16S amplicon analyses (external), 10 transcriptomic samples (external), *10 proteomic (EMSL), 10 lipidomic (EMSL), and 10 metabolomic (EMSL) samples.* Outcomes from Aim 2 will be the delineation of which species assemble adjacent to chitin patches (primary chitin degraders) and which assembly in periphery positions (secondary consumers of breakdown products), which species express active chitinases and what processes and pathways are being expressed by primary and secondary species.

Data generated by this proposal will be assimilated into the National Microbiome Data Collaborative (NMDC) database and will be made publicly available in accordance with User Facility and DOE guidelines. Models developed as part of this proposal will leverage existing pipelines through (KBase); and we will generate additional applications and narratives as needed. Together our experiments will demonstrate how Soil EcoFAB and MSC-2 community resources can serve the EMSL user community in future experiments and user calls. Collectively, we expect outcomes of this proposal to greatly expand our knowledge of interaction networks and community phenotypes that drive the consumption of complex carbon and nitrogen nutrient sources. This knowledge will be gained in the context of a spatially structured environment driven by hotspots of decomposition meaning that our conclusions will directly inform our knowledge of native soil processes. In addition to the scientific gain, the successful completion of these experiments will greatly advance BER and EMSL mission goals.

Computing Approach.
This proposal is requesting 25,000 node hours on Tahoma for inference of species interaction models and network and modeling analysis of multi-omic (amplicon, transcriptomic, proteomic, lipidomic and metabolomic) data.

1) **Species interaction network modeling**. This project will infer interaction models using an advanced mixed-integer linear programming-based regression algorithm (available as an in-house code) to explore a set of candidate Lotka-Volterra models that each equally fit the species abundance data. We expect to be able to deploy this code on Tahoma to facilitate determination of these interaction networks, which will be used, in part, to integrate and interpret the other kinds of data generated under this proposal.
2) **Multi-omic data processing**. We have been collaborating with IRP Lead Jay Bardhan to implement softwares for the analysis of sequence and molecular data on Tahoma. We will leverage this progress to complete the proposed analytics.

☒ By checking this box, I am confirming that participants on this proposal will NOT process, store, or transmit sensitive data (e.g. Personally Identifiable Information, Official Use Only, etc.) on Tahoma, Cascade or Aurora.

Total CPU Hours Request for first year of proposal: 25,000					
Total GPGPU Hours Request for first year of proposal: 0					
Total Data Archive Request for first year of proposal: 2 TB					
Software Details	**Node Request (CPUs or GPGPUs)**	**Estimated # of jobs**	**Estimated Node Hours**	**Expertise of your investigators for these requests**	**EMSL Support Requested** *Specific Needs (e.g., compiling code, libraries needed, help running jobs, etc.)*
gLV/MEMPIS	CPU	40	5,000	Expert	Minimal Support Required
Burrows Wheeler Aligner/HTSeq	CPU	60	10,000	Expert	Minimal Support Required
Mass spectrometry - omics analysis	CPU	40	10,000	Expert	Minimal Support Required

Appendix 1: List of References

[1]

Active Collaborators List:

Name	Affiliation	Collaborator	Advisor/Advisee