LA-UR-19-30124
Approved for public release; distribution is unlimited.

Title: BSSD 2019 Performance Metric – End-of-year summary report

Author(s): Dunbar, John Martin

Intended for: Report

Issued: 2019-10-07

BSSD 2019 Performance Metric – End-of-year summary report

Goal: Develop metagenomics approaches to assess the functioning of microbial communities in the environment.

Summary:

The LANL Science Focus Area program in Terrestrial Microbial Carbon Cycling aims to inform climate modeling and enable carbon management in terrestrial ecosystems. To achieve these aims, the program develops and uses community genomics approaches to discover widespread biological processes that control carbon storage and release in temperate biome soils. The first nine years of the program focused on developing and applying community genomics approaches to examine the composition, spatiotemporal variability, and response of soil microbial communities in two biomes to environmental changes that alter carbon cycling (e.g., elevated atmospheric CO_2, nitrogen fertilization, and drought). Owing to persistent inability to distinguish cosmetic from functionally consequential responses, the SFA recently shifted research strategy. The new research strategy focuses on soil microbial communities that drive differences in carbon cycling *within* the same environment. With this approach, there is an inherent cause-effect relationship between microbial community composition and ecosystem function. This approach is revealing core micro-organisms and interactions at different scales (molecular to ecosystem-level) that drive contrasting patterns of carbon cycling. Elaborating processes that shape the composition and function of soil microbial communities driving terrestrial carbon cycling will accelerate progress toward more accurate climate models and strategies for terrestrial carbon sequestration.

This report summarizes the material found in previous reports describing the use of genomics techniques and supporting computational approaches to decipher the functioning of soil communities driving variation in carbon cycling. Previous reports focused on the following:

- Progress in the use of metagenomic techniques to detect and describe the composition of microbial communities in complex environmental samples. The LANL SFA applied diverse metagenomic techniques to increase knowledge of the structure and composition of fungal and bacterial communities in 13 terrestrial ecosystems across the U.S. responding to drivers of ecosystem changes in carbon cycling. Insights from these efforts enabled development of a more efficient strategy to discover microbial processes that can drive variation in ecosystem function.

- Concerted use of metagenomics and metatranscriptomic analyses to elucidate microbial community function in environmental samples. Metagenomics and metatranscriptomics provide contrasting views of the microbial community function. Consequently, combined use of these techniques is routinely applied in the LANL SFA to accelerate understanding of the genes, organisms, physiological processes, and interactions that underpin community function. Effective methods for soil shotgun metatranscriptomics that provide inter-kingdom (e.g. prokaryotes and eukaryotes) coverage emerged only in the past 5 years, including the technique published by the SFA in 2015.

- 'Omics'-based techniques used to describe microbial activity in the environment. Genomic, metagenomic, transcriptomic, proteomic, and metabolomic techniques have been applied in various ways to investigate aspects of microbial activity linked to carbon cycling. The suite of techniques enabled functional characterization of reference organisms—a stepping stone to understanding community functioning. The techniques also enabled ranking environment factors by their capacity to perturb soil community composition. A suite of these techniques is central to current efforts deciphering the aspects of community composition that drive carbon fate in soils.

- The latest computational approaches to analyze large complex 'omics' datasets to describe microbial community function in environmental samples. Terrestrial carbon cycling involves complex interactions at many spatial and temporal scales. Discovering biological drivers of carbon cycle variation requires integration of increasingly powerful 'omics capabilities with advanced computing. To address this complexity, the LANL SFA has been establishing a computational approach that combines machine learning, exascale computing, and multi-scale ecosystem modeling. This approach provides tight integration of experimentation and modeling, accelerating efforts to inform climate modeling and enable carbon management in terrestrial ecosystems.

Highlights of key efforts and insights from previous reports are synthesized below.

The LANL SFA research program began 10 years ago when shotgun metagenomics for complex environmental samples was first emerging, and use of metagenomics for comparison of ecosystems was mostly conceptual. Analysis pipelines were nascent. Even the concept of sequencing replicate samples to enable statistical analyses of experimental treatments—a standard of rigorous research—was novel because of the cost and sequencing capacity needed. A central objective of the project at that stage was to begin developing soil metagenomics as a research tool and apply this unique capability across ecosystems to gain insight into soil microbial responses to future climate regimes.

Figure 1. Major biomes examined. Approximate locations of ecosystem field experiments, ●.

The SFA initially used metagenomics techniques to examine microbial community (i.e. bacteria and fungi) responses to elevated atmospheric CO_2 in six DOE climate change field experiments led by external collaborators. This was the first-of-its-kind attempt to find common microbial responses to climate manipulations in multiple ecosystem experiments by "large-scale" sequencing [1]. The central hypothesis was that common microbial responses occur among different ecosystems exposed to climate forcing. Identification of common responses and the underlying microbial processes could potentially inform climate models projecting carbon cycling over centennial time scales. Our studies showed that the fungal and bacterial communities differed significantly among the ecosystems [1-4], strongly influenced by local soil geochemistry and soil depth [1-3, 5, 6].

Responses to elevated CO_2 were mostly ecosystem-specific and were smaller in magnitude compared to responses to other experimental factors (Figure 2), such as inorganic nitrogen fertilization, which has increased globally over the past century as an anthropogenic perturbation of natural terrestrial ecosystems.

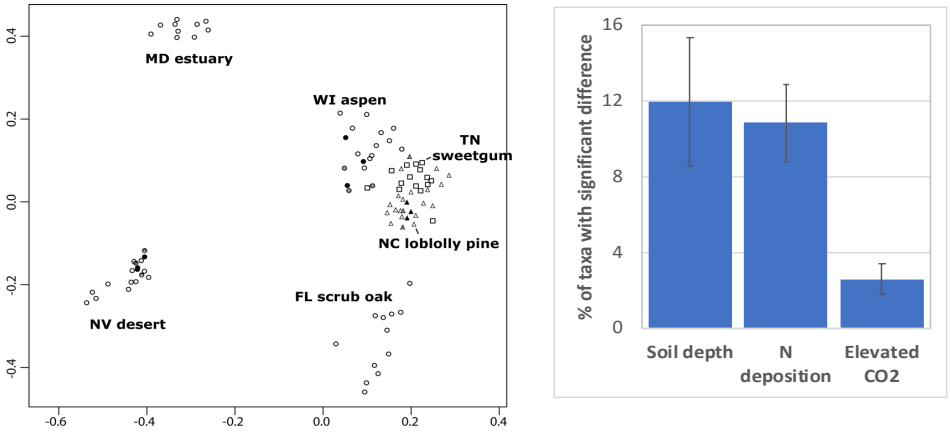

Figure 2. Use of metagenomics techniques to assess community variability and activity (degree of response) in field experiments testing the impact of elevated atmospheric CO_2 and other factors in six ecosystems. Left panel - Nonmetric multidimensional scaling of soil bacterial community dissimilarities (from [1]). Similar results were obtained with fungal community data (unpublished) and with cellulolytic fungal community data [4]. Right panel – Ranking of soil depth, N fertilization, and elevated atmospheric CO_2 as perturbations of soil bacterial communities [1]. Error bars are 95% confidence intervals.

Because N fertilization tends to alter soil microbial communities more strongly than elevated atmospheric CO_2, the SFA subsequently focused on common microbial responses to N fertilization in 8 ecosystem studies (5 forest and 3 aridland ecosystems) led by external collaborators. This shift in focus involved SFA development of a shotgun metatransriptome technique for soils to broadly examine gene expression [7]. Prior to developing this approach, metatranscriptome studies of soil microbial communities had been difficult and rarely reported, largely because the high abundance of rRNA overwhelmed mixed-template sequencing. Our approach significantly improved upon earlier studies by (a) removing the majority of rRNA from an environmental RNA sample without relying on polyA enrichment, and thus (b) retaining the bacterial, archaeal, and eukaryotic components. This approach facilitates studies of interactive metabolism and signaling between prokaryotes and eukaryotes—for example between plants, their mycorrhizal fungi, and rhizosphere bacteria [7].

Application of metatranscriptomics to forest and arid grassland ecosystems provided insights that were not feasible using metagenomic techniques alone ([7-9], unpublished data). Integrating metagenomics and shotgun metatranscriptomics is important because each technique provides a sharply different view of soil microbial communities. Shotgun metagenomes are often dominated by bacterial sequences because in soil, bacteria typically outnumber fungi by 100 to 1000-fold. In contrast, our metatranscriptome studies of forest soils have been dominated by fungal gene expression. For example in a study of two Maple forest ecosystems subjected to chronic nitrogen fertilization treatments for 21 years, we found twice as many carbon cycling gene transcripts attributed to fungal than to bacteria [7]. Soil metatranscriptomics is more effective than metagenomics in documenting inter-kingdom interactions that underpin community function.

Soil metatranscriptomes can also reveal sets of genes that are potentially co-regulated or interacting [7]. Granted, the interpretation of co-expression patterns in metatranscriptome samples is challenging owing to the large number of genes and organisms represented in the data. Nonetheless, transcript correlation can be used to formulate hypotheses about potentially interacting transcripts or organisms. In Maple forest soil samples for example, we found strong clusters of co-expressed fungal carbon cycling gene transcripts (Figure 3; [7]). In one illustrative case, a cluster included many Carbohydrate Active enZyme (CAZyme) families mostly known for fungal lignocellulolytic activities, suggesting the genes/organisms were acting in concert to decompose lignocellulosic plant litter. This cluster of genes was negatively correlated with nearly all of the abundant bacterial CAZyme transcripts indicating possible competition or other antagonistic interactions between fungal and bacterial decomposers [7]. This highlights the power of metatranscriptomics to illuminate possible mechanistic interactions contributing to terrestrial carbon cycling or other processes of interest in other application areas such as bioenergy production.

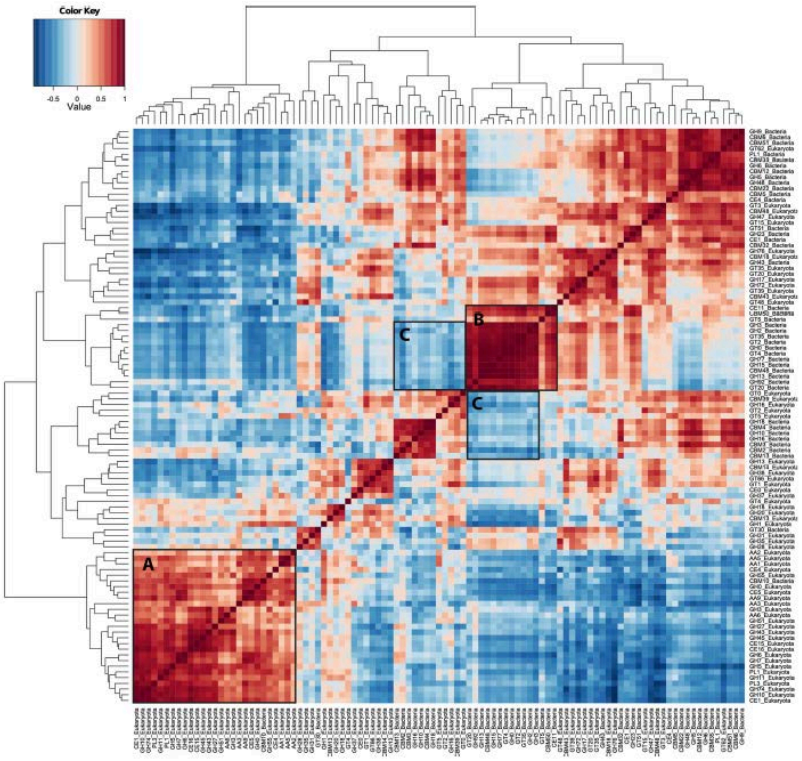

Figure 3. Carbohydrate Active enZyme transcript abundance correlation heatmap. The co-expression of many CAZymes in the metatranscriptomes illustrates that this approach is able to detect complex patterns of co-expression among carbohydrate-active genes. Stronger correlations, either positive (red) or negative (blue), are illustrated as darker shades. Box A indicates a group of eukaryote CAZyme families (with one bacterial family) that are co-expressed. Box B shows a group of bacterial CAZyme families that are also co-expressed, while the C Boxes include bacterial and fungal CAZyme families that are anti-correlated with those in Boxes A and B. From [7].

Through the compendium of field studies, the SFA used metagenomics not only to understand ecosystem responses but also to guide isolation of priority bacteria and fungi widely sought by the international research community as key reference organisms [10-15]. From our isolate collections, we obtained genome sequences for the dominant photosynthetic cyanobacterium *Microcoleus vaginatus* in arid grassland biocrusts of the US southwest [12], heterotrophic bacterial partners that interact with *M. vaginatus*, Actinomycetales bacteria that are prevalent in arid land soils but under-represented in genome databases [15], and various

Mucoromycotina fungi in forest soils including *Bifiguratus adelaidae*—one of the international research community's 'top 50 most wanted' fungi [3, 13]. Genomes from under-represented organisms like these are a vital contribution to the genome databases (e.g., the DOE Integrated Microbial Genomics database) that are a foundation for analysis of meta-'omics data from complex communities.

Whereas the aim of studying multiple ecosystems under N fertilization was to find common soil microbial responses among ecosystems, these studies revealed the same obstacles as in prior studies of responses to elevated CO_2. The numerous microbial responses to N fertilization were highly variable and location specific within and among ecosystems [3, 11, 16-18].

The difficulty of finding common responses to perturbations among ecosystems, combined with the inability to determine which responses are cosmetic versus functionally significant revealed a major obstacle in the SFA research strategy to inform climate models. We used this insight to improve the SFA research strategy.

Shift in research strategy. The new SFA strategy emphasizes discovery of microbial communities that represent substantially different (and measurable) functional states *under the same environmental conditions,* not perturbed versus unperturbed conditions. This approach shifts the focus from discovering *responsive* populations to discovering populations and processes that directly *cause* substantial variation in carbon cycling. Our revised strategy requires sampling a large number of microbial communities and measuring the function of each one under the same environmental condition in order to identify cohorts of communities representing functional states of interest (Figure 4). For the SFA, functional states of interest are patterns of carbon cycling with contrasting implications for climate feedbacks and for soil carbon sequestration. Once the functional states are delineated, metagenomic and other 'omics techniques enable the search for the common features in each cohort of communities that represent a specified functional state. This strategy leverages technology advances that have increased metagenomic sequencing capacity 100 to 1000-fold.

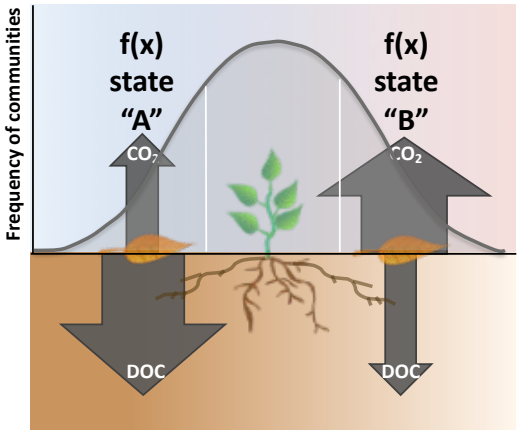

Figure 4. SFA research strategy concept. An observed distribution of functional outcomes enables delineation of contrasting functional states, each represented by a cohort of communities.

With the strategic shift, the SFA is examining thousands of decomposer microbial communities to identify processes driving variation in carbon flow. The SFA is examining variation in carbon flow from plant leaf litter at the soil surface, from root litter in the subsurface, and from mineral-bound organic matter. For example, targeted metagenomic techniques applied to communities decomposing different types of leaf litter under the same environmental conditions show, as expected, that the average composition of the decomposer communities varies by plant litter type, yet substantial similarities occur (Figure 5).

Our interest is the common taxonomic and/or metabolic features among litter types that underpin carbon cycling functional states. In our studies thus far, the functional states are

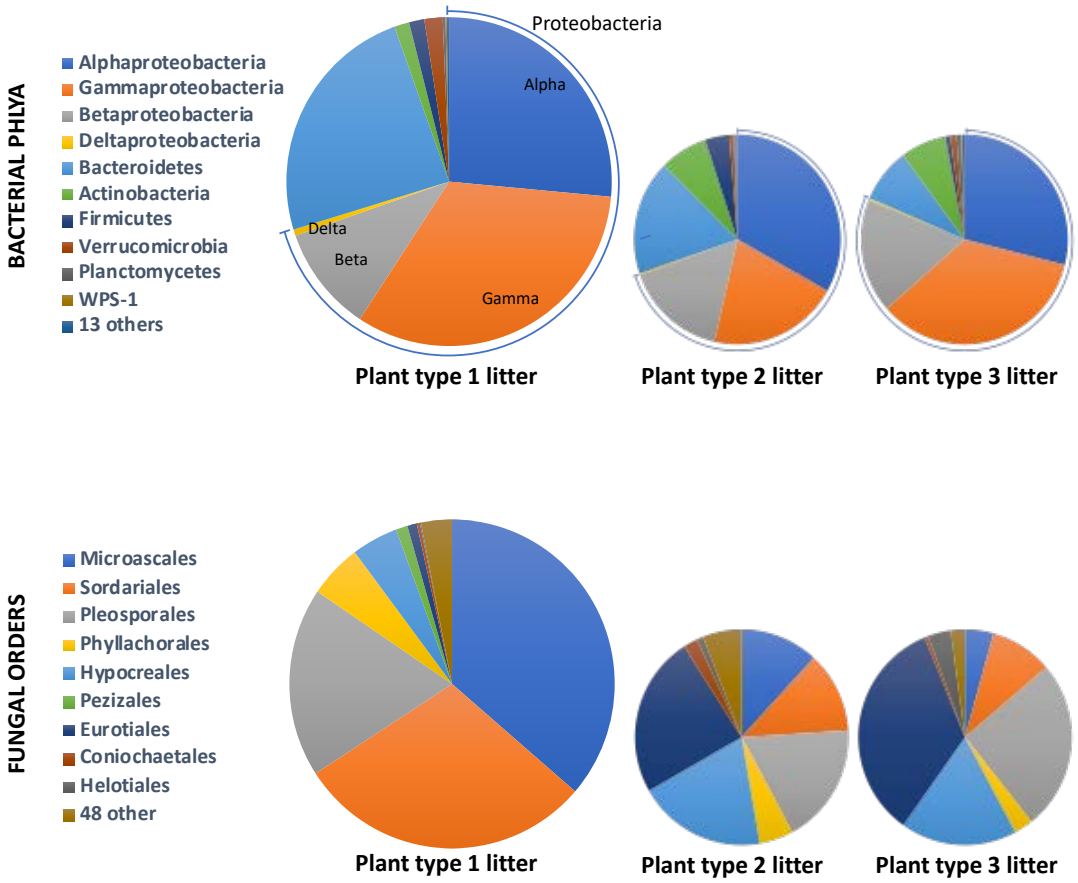

Figure 5. The average composition of hundreds of microbial decomposer communities that assembled on different plant litter types in the early phase of decomposition.

delineated based on the abundance of dissolved organic carbon (DOC) from the decomposition interval. We have found a substantial fraction of taxa exhibit similar shifts in relative abundance between the "high" verus "low" DOC functional states among litter types [19]. These data are a promising indication that the revised strategy will provide insights to biological processes broadly relevant to terrestrial carbon cycling.

To support the new research strategy, the SFA is establishing for routine use a computational approach that integrates exascale metagenomic computing with multi-scale ecosystem modeling. In this approach, the SFA is developing, applying, and distributing to the research community computational tools that use machine learning techniques to reduce the dimensionality of larger 'omics datasets that arise from the SFA's research approach [20]. These techniques yield a subset of core taxa (or other community features) that best predict functional outcomes from microbial community activity [20]. Related computational techniques are being used to infer interactions between core taxa and metabolic products (Figure 6A)[19] and to infer differences in physiological processes that may exist between carbon-cycling functional states (Figure 6B)[21]. These inferences provide specific research targets for targeted mechanistic studies.

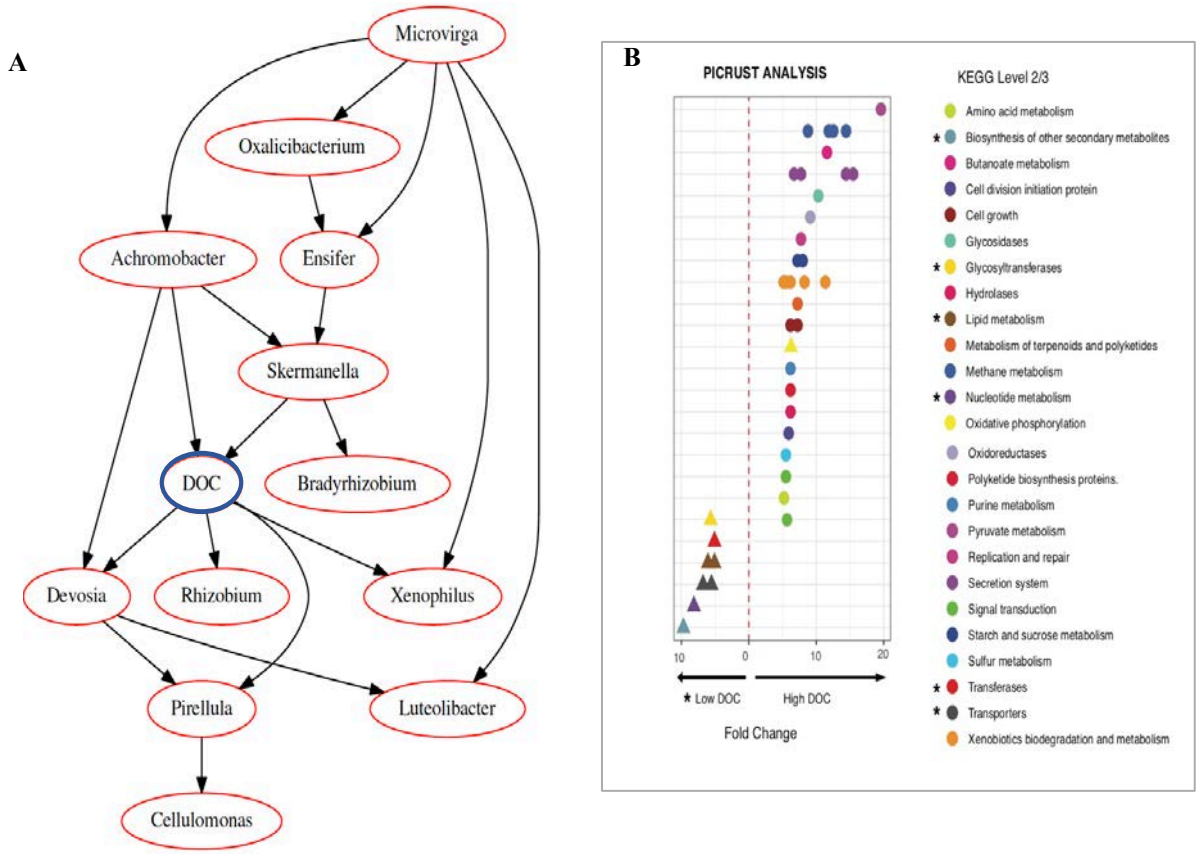

Figure 6. Computational inference of interactions among taxa driving dissolved organic carbon abundance and associated physiological differences between community cohorts representing contrasting carbon cycling patterns. Panel A - Interaction network for DOC and a set of 12 selected genera. Panel B – Physiological differences from predicted metagenomes for contrasting community cohorts. [19, 21]

The SFA research strategy continues to use metatranscriptomics to validate and augment functional inferences derived from metagenomics data. Metatranscriptome measurements have confirmed that distinctive physiological processes underpin the observed carbon cycling functional states (Figure 7). Metatranscriptome data show significant changes in the number and expression levels of carbon cycling genes (CAZymes) associated with the high and low DOC functional states.

A general challenge of metatranscriptomics is that data interpretation is weakened by the routine absence of an assembled reference metagenome to which gene transcripts can be accurately mapped. Assembled reference metagenomes are usually absent because the vast amount of sequencing data required to build them is costly to collect. To address this challenge, the SFA is using exascale computing through DOE's user facility for high performance computing (NERSC) to build ecosystem metagenome assemblies that can be used and sequentially improved by the entire research community as reference metagenomes. This approach involves aggregating all ecosystem-relevant metagenomic and metatranscriptomic data from every available source (different experiments, research labs, collection dates).

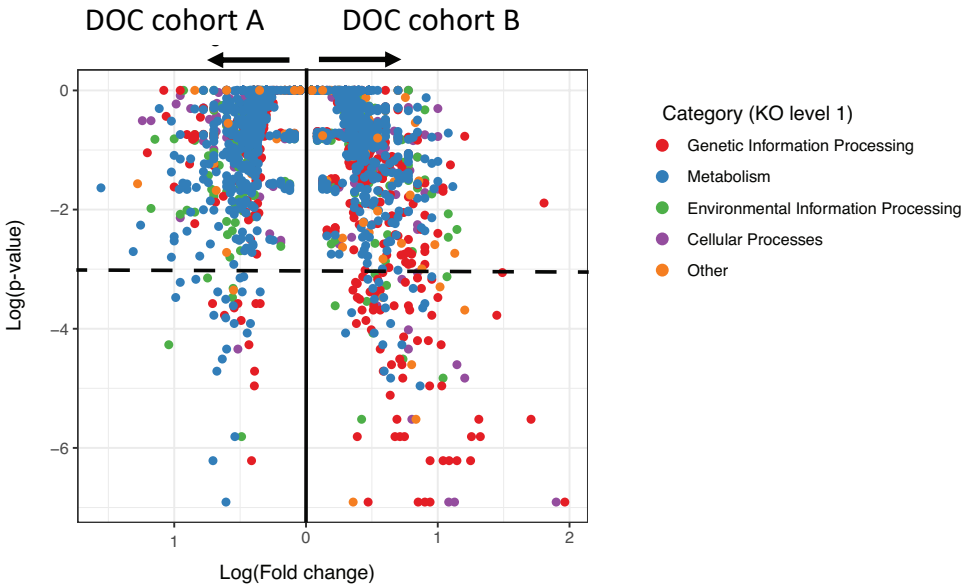

Figure 7. Volcano plot showing differential gene expression of decomposer communities (n=9 per cohort) that represent distinct patterns of carbon flow (measured as dissolved organic carbon abundance) from leaf litter decomposition in laboratory microcosms. The dashed line is the p-value threshold. Points below the dashed horizontal line represent gene families with significant differences in expression. From [21].

Within the new SFA strategy, metagenomic and metatranscriptomic techniques remain central to monitoring microbial activity. However, the SFA is now integrating these techniques tightly with additional functional measurements to accelerate discovery of microbial processes driving variation in carbon cycling. For example, capabilities at DOE's Environmental Molecular Science Laboratory user facility are being leveraged to acquire different types of metabolomic data for integration with metagenomic and metatranscriptomic data. The metabolite composition of dissolved organic carbon from plant litter decomposition shows large differences in the composition of organic carbon from microbial communities that represent contrasting functional states (Figure 8) [22]. The SFA's machine learning pipeline and other analytical techniques indicate relatively small groups of microbial taxa control the abundance of specific compounds, providing powerful new insights into underlying mechanisms.

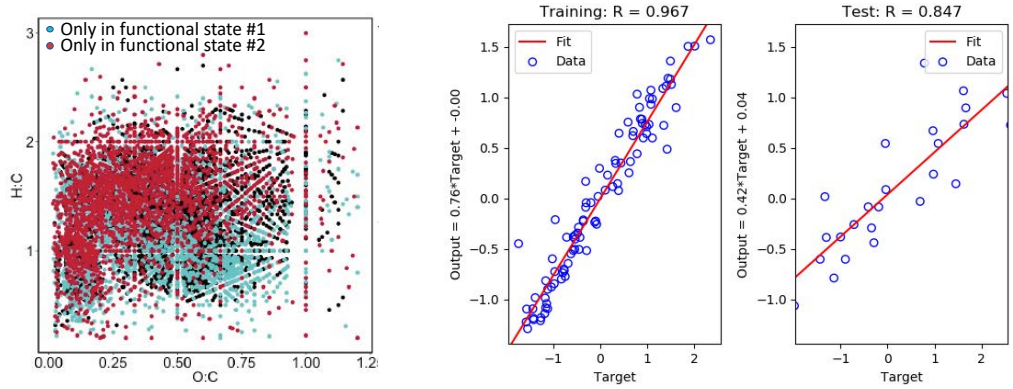

Figure 8. Metabolomic data from >300 plant litter decomposer communities. Left panel – Van Krevlen plot of chemical composition. Right panels – Machine learning model using metagenomics data (the abundance of core taxa) to predict the abundance of a particular chemical component [22].

A distinctive facet of the SFA's recent shift in research strategy has been to incorporate soil carbon modeling and simulation [23] as an overarching framework to guide research priorities, experimental design, and to improve data interpretation (e.g. [24]). This is a capability and strategic refinement added within the past year. The SFA is using SOMic, a microbially-based SOC model recently developed by Cornell collaborators [23]. SOMic captures the recent paradigm shift in concepts about soil organic matter persistence, wherein persistence depends on microbe-mineral-organic matter interactions, instead of recalcitrance of the organic matter.

The combination of the SFA's improved research strategy, development and use of advanced computational tools, and tight integration of modeling (SOMic) and experimentation promises is accelerating the pace and effectiveness of the SFA's research program to inform climate modeling and enable soil carbon management.

Bibliography

1. Dunbar, J., S.A. Eichorst, L. Gallegos-Graves, S. Silva, G. Xie, N. Hengartner, B.A. Hungate, R.B. Jackson, D.R. Zak, R. Vilgalys, R.D. Evans, C.W. Schadt, J.P. Megonigal, and C.R. Kuske, *Common bacterial responses in six ecosystems exposed to ten years of elevated atmospheric carbon dioxide.* Environ Microbiol, 2012.
2. Berthrong, S.T., C.M. Yeager, L. Gallegos-Graves, B. Steven, S.A. Eichorst, R.B. Jackson, and C.R. Kuske, *Nitrogen fertilization has a stronger effect on soil nitrogen-fixing bacterial communities than elevated atmospheric CO2.* Appl Environ Microbiol, 2014. **80**(10): p. 3103-12.
3. Weber, C.F., R. Vilgalys, and C.R. Kuske, *Changes in Fungal Community Composition in Response to Elevated Atmospheric CO2 and Nitrogen Fertilization Varies with Soil Horizon.* Front Microbiol, 2013. **4**: p. 78.
4. Weber, C.F., D.R. Zak, B.A. Hungate, R.B. Jackson, R. Vilgalys, R.D. Evans, C.W. Schadt, J.P. Megonigal, and C.R. Kuske, *Responses of soil cellulolytic fungal communities to elevated atmospheric CO(2) are complex and variable across five ecosystems.* Environ Microbiol, 2011. **13**(10): p. 2778-93.
5. Steven, B., L.V. Gallegos-Graves, S.R. Starkenburg, P.S. Chain, and C.R. Kuske, *Targeted and shotgun metagenomic approaches provide different descriptions of dryland soil microbial communities in a manipulated field study.* Environ Microbiol Rep, 2012. **4**(2): p. 248-56.
6. Steven B, L.G.-G., C Yeager, J Belnap, CR Kuske, *Common and distinguishing features of the bacterial and fungal communities in biological soil crusts and shrub root zone soils.* Soil Biol Biochem, 2013.
7. Hesse, C., M.R. C, M. Vuyisich, L. Gallegos-Graves, C.D. Gleasner, D.R. Zak, and K.C. R, *Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests.* Front Microbiol, 2015. **6**: p. 337.
8. Mueller, R.C., L. Gallegos-Graves, D.R. Zak, and C.R. Kuske, *Assembly of Active Bacterial and Fungal Communities Along a Natural Environmental Gradient.* Microb Ecol, 2016. **71**(1): p. 57-67.
9. Steven, B., J. Belnap, and K.C. R, *Chronic physical disturbance substantially alters the response of biological soil crusts to a wetting pulse, as characterized by metatranscriptomic sequencing.* Front Microbiol, 2018. **9**: p. 2382.

10. Eichorst, S.A., C.R. Kuske, and T.M. Schmidt, *Influence of plant polymers on the distribution and cultivation of bacteria in the phylum Acidobacteria.* Appl Environ Microbiol, 2011. **77**(2): p. 586-96.
11. Hesse, C.N., T.J. Torres-Cruz, T.B. Tobias, M. Al-Matruk, A. Porras-Alfaro, and C.R. Kuske, *Ribosomal RNA gene detection and targeted culture of novel nitrogen-responsive fungal taxa from temperate pine forest soil.* Mycologia, 2016. **108**(6): p. 1082-1090.
12. Starkenburg, S.R., K.G. Reitenga, T. Freitas, S. Johnson, P.S. Chain, F. Garcia-Pichel, and C.R. Kuske, *Genome of the cyanobacterium Microcoleus vaginatus FGP-2, a photosynthetic ecosystem engineer of arid land soil biocrusts worldwide.* J Bacteriol, 2011. **193**(17): p. 4569-70.
13. Torres-Cruz, T.J., T.L. Billingsley Tobias, M. Almatruk, C.N. Hesse, C.R. Kuske, A. Desiro, G.M.N. Benucci, G. Bonito, J.E. Stajich, C. Dunlap, A.E. Arnold, and A. Porras-Alfaro, *Bifiguratus adelaidae, gen. et sp. nov., a new member of Mucoromycotina in endophytic and soil-dwelling habitats.* Mycologia, 2017. **109**(3): p. 363-378.
14. Torres-Cruz TJ, C.H., CR Kuske, A Porras-Alfaro, *Presence and distribution of heavy metal tolerant fungi in surface soils of a temperate pine forest.* Applied Soil Ecol, 2018.
15. Yeager, C.M., V. Gallegos-Graves, J. Dunbar, C.N. Hesse, H. Daligault, and C.R. Kuske, *Polysaccharide Degradation Capability of Actinomycetales Soil Isolates from a Semiarid Grassland of the Colorado Plateau.* Appl Environ Microbiol, 2017. **83**(6).
16. Kuske, C.R., C.M. Yeager, S. Johnson, L.O. Ticknor, and J. Belnap, *Response and resilience of soil biocrust bacterial communities to chronic physical disturbance in arid shrublands.* Isme j, 2012. **6**(4): p. 886-97.
17. Weber, C.F., M.M. Balasch, Z. Gossage, A. Porras-Alfaro, and C.R. Kuske, *Soil fungal cellobiohydrolase I gene (cbhI) composition and expression in a loblolly pine plantation under conditions of elevated atmospheric CO2 and nitrogen fertilization.* Appl Environ Microbiol, 2012. **78**(11): p. 3950-7.
18. Sinsabaugh, R.L., J. Belnap, J. Rutgers, C.R. Kuske, N. Martinez, and D. Sandquist, *Soil microbial responses to nitrogen addition in aridland ecosystems.* Front Microbiol, 2015.
19. Thompson, J., N. Lubbers, J. Dunbar, and B. Munsky, *Applying Bayesian network structure learning to plant decomposer microbiomes to predict functional outcomes* PLoS Computational Biology, 2019. **in prep**.
20. Thompson, J., R. Johansen, J. Dunbar, and B. Munsky, *Machine learning to predict microbial community functions: An analysis of dissolved organic carbon from litter decomposition.* PLoS ONE, 2019. **14**: p. e0215502.
21. Albright, M.B.N., J. Thompson, R. Johansen, D.E.M. Ulrich, L.V. Gallegos-Graves, B. Munsky, and J. Dunbar, *Microbial physiology linked to divergent carbon flow from litter decomposition.* Frontiers in Microbiology, 2019. **Submitted**.
22. Ulrich, D.E.M., M. Tfaily, J. Thompson, B. Munsky, M.B.N. Albright, J. Tyoda, and J. Dunbar, *Microbial drivers of protein and dissolved organic carbon abundance in early phase of pine litter decomposition.* in revision, 2019.
23. Woolf, D. and J. Lehmann, *Microbial models with minimal mineral protection can explain long-term soil organic carbon persistence.* Scientific Reports, 2019. **9**: p. 6522.
24. Albright, M.B.N., A. Runde, D. Lopez, J. Gans, S. Sevanto, D. Woolf, and J. Dunbar, *Initial microbial biomass abundance is a weak driver of variation in CO2 flux during plant litter decomposition.* PloS ONE, 2019. **submitted**.