Session 1

Decoding the Structure-Function Secrets of Environmental Microbiomes

Posters alphabetical by first name of presenting author*

EMSL's Rhizo-Critical Campaign: Critical Minerals Biogeochemistry in the Rhizosphere – Ultramafic Soils

Amir H. Ahkami* (amir.ahkami@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

Amir H. Ahkami, Jayde Aufrecht, Arunima Bhattacharjee, Vimal Balasubramanian, Paul Piehowski, Tom Wietsma, Kaizad Patel, Tanya Winkler, Cayden Perdue, Sarah Leichty

Environmental Molecular Sciences Laboratory (EMSL), Pacific Northwest National Laboratory (PNNL), Richland, WA 99352 USA

The growing demand for critical materials and minerals (CMM) in the United States has intensified interest in ultramafic soils, which contain valuable metals such as nickel (Ni), chromium (Cr), manganese (Mn), cobalt (Co), and copper (Cu)—essential for batteries, magnets, wind turbines, and stainless steel technologies. Hyperaccumulator plants grown in these soils can extract economically meaningful concentrations of CMMs through phytomining. However, the fundamental mechanisms of metal uptake, transport, and sequestration—particularly those shaped by microbial communities, root exudation, rhizosphere chemistry, and root–microbe interactions—remain poorly understood.

To address these gaps, EMSL established the Rhizo-Critical Campaign, a multi institutional, community science effort aimed at overcoming challenges such as insufficient mechanistic knowledge of root–microbe–mineral interactions, limited characterization of Ni speciation and uptake pathways, and the absence of standardized sampling and molecular workflows. Leveraging EMSL's advanced capabilities—including integrated imaging, omics, computational modeling, and synthetic soil habitats (TerraForms)—the campaign will: (i) build mechanistic understanding of how metabolites and microbiomes regulate Ni mobilization and uptake in hyperaccumulators versus non hyperaccumulators; (ii) create an integrated field–lab framework with shared reference sites and standardized rhizosphere sampling; and (iii) develop coordinated workflows using model systems, advanced omics/imaging, and shared data platforms for reproducible, scalable studies.

The resulting molecular datasets will be feature-rich, mechanistically grounded, and stored in structured, searchable formats—aligned with the Genesis Mission vision. These data will power AI-driven modeling and predictive analytics, enabling next-generation insights into rhizosphere dynamics and advancing DOE-BER priorities for critical mineral supply chain resilience.

Developing a Functional Screening Platform for Environmental Microbiomes

Beth Davenport* (beth.davenport@ubc.ca)
University of British Columbia

Beth Davenport, Wendy Wu, John Collins, Andras Szeitz, Rosalie Zhou, and Steven J. Hallam

Microbial communities also known as microbiomes provide a network of genetic operating systems useful in developing sustainable bioprocesses with wide-ranging industrial applications including strategic metal recovery. Advances in high-throughput sequencing have revealed extensive, yet underexplored, genetic potential in mining microbiomes, including metalloenzymes and metal-binding domains useful in bioprocess development. However, realizing this potential requires functional screening paradigms that effectively bridge gene prediction and discovery with functional validation and biological engineering. Here, we present a modular platform developed to enable high-throughput functional screening of microbiome-derived DNA or cDNA libraries on the cell surface of Caulobacter vibrioides CB2A JS4038. Using JGI's genomic, transcriptomic, and DNA synthesis capabilities, an existing surface display platform was refactored into a flexible vector architecture supporting heterologous gene expression. This architecture enables selection for in-frame insertion sequences significantly improving screening efficiency and is compatible with downstream fluorescence-activated cell sorting (FACS) and microtiter plate–based functional assays. As an initial application, we explore functional screening on Caulobacter cell surfaces using mine tailings samples to identify proteins mediating selective copper binding, a function of growing importance for future energy supply. Taken together, this work establishes a novel screening paradigm based on cell surface interactions to identify genes or transcripts of interest that can be applied to wide ranging targets or substrates including but not limited to strategic metals, lignocellulosic biomass, and chemicals of concern.

Seasonal Bloom Dynamics and Urbanization Gradients Reshape Dissolved Organic Matter and Microbial Networks across the St. Louis River Estuary adjacent to Lake Superior

Christian Ayala-Ortiz* (christian.ayalaortiz@pnnl.gov)
Pacific Northwest National Laboratory

Christian Ayala-Ortiz, Hannah Nicklay, Chris Filstrup, Euan Reavie, Peter Birschbach, Rosalie Chu, Will Kew, Caroline Simonsen, William Peterson, Izabel Stohel, Andrew Townsend, Emily Graham

Lakes are the largest reservoirs of freshwater on Earth supporting key ecological functions, including earth system regulation, carbon cycling, and biodiversity maintenance. Yet human activities increasingly modify surrounding watersheds, altering water quality, organic matter (OM) composition, and microbiota. Here, we address how urbanization shapes aquatic biogeochemistry by studying multiple sites across the St. Louis River Estuary adjacent to Lake Superior, with a specific focus on periods of harmful algal blooms. We integrated ultrahigh resolution FTICR MS characterization of OM, and amplicon sequencing of bacterial/archaeal and algal assemblages, alongside measurements of nutrients, ions, turbidity, temperature, and land-use. Our results showed that water column variations in OM profiles and algal communities were primarily organized along a seasonal temperature gradient, while bacterial/archaeal communities were more tightly linked to dissolved ions and nutrients concentrations. Conversely, sediment OM and microbial composition were strongly influenced by urbanization gradients, with highly urbanized sediments showing distinct OM profiles and copiotrophic, eutrophication tolerant taxa compared with low urbanization sites. During high algal blooms, surface waters were enriched in labile, nitrogen rich OM and specific algal and cyanobacterial lineages that represent potential markers for bloom monitoring and early warning. Additionally, identified differences between organic matter and microbial composition between sediments of high- and low-urbanization sites can be used to provide candidate indicators for watershed assessment and regulatory monitoring. Collectively, these results reveal water column drivers of bloom phenology as well as a persistent effect of urbanization on the benthic zones, that can be used for tracking harmful algal bloom risk and land use impacts in estuarine ecosystems.

Leveraging metagenomic and metatranscriptomic time series data to identify biomarkers of community mortality during a harmful algal bloom

Claire Elise Elbon* (celbon2@uw.edu)
University of Washington

Claire Elbon1, Emma Timmins-Schiffman1, Rachel Schauer1, Miranda Mudge1, Gabriella Chebli2, Tatiana Rynerson3, Julia Kubeneck2, Brook Nunn1

Harmful algal blooms (HABs) are disruptive events in marine ecosystems that significantly alter the structure and abundance of the co-occurring microbial community. It remains difficult to predict the beginning and end of HABs, which is critical information for public health forecasting and modeling of biogeochemical cycles (i.e. the carbon pump). Importantly, the interactions and mechanisms driving mortality of the microbial community and HAB itself remain uncharacterized. Understanding drivers of mortality, including viral lysis/infection and programmed cell death would allow for prediction of HAB termination. This study utilizes metagenomic and metatranscriptomic data sequenced by JGI from a marine time series that captured the duration of an initial HAB and the start of a second HAB during May-June 2021 off Orcas Island in Washington. We identified an increase and fall in estimated gene copies of the giant algal virus, Phycodnaviridae, before and during the initial HAB in our metagenomic data, as well as significant transcription of this dsDNA virus during the chlorophyll peak. Conversely, dsDNA bacteriophages, Caudoviricetes, abundance trends are inversely proportional to the chlorophyll peak; estimated gene copies drop and rise ~10-fold before and after the initial HAB. We also found transcriptional evidence of Cathespin X (CPX), a conserved algal protein implicated in viral infection-induced programmed cell death and spore formation. CPX abundance was elevated during the initial HAB through its termination. We will utilize this 'omic data to guide targeted proteomics of mortality for bloom forecasting, and to investigate potential diel cycles of mortality in the microbial community.

Methane-Derived Melanin Supports Heat and Drought Resilience of Plants and Soil

Courtney Cameron* (ccameron3333@sdsu.edu)
San Diego State University

Courtney Cameron, William Gebbie, Deborah Jaffey, Edward Baidoo, Emine Akyuz Turumtay

(Semi)arid ecosystems cover one-third of the Earth's surface and contributors to terrestrial cycling of atmospheric gases. Understanding arid biome interactions (soil microbiota and plants) can uncover mechanisms of resilience toward environmental stressors. Methylocaldum spp. is the predominant methanotroph in arid soils and possess genetic signatures typically linked to symbiotic interactions with plants and microbes. We present insight into genetic elements required to produce photoactive redox polymers and the characterization of signaling networks that shape interactions with soil community members. "Melanin" is a diverse range polymers known for improving abiotic stress resilience. Genome comparison between the M. gracile wild-type strain and spontaneous white mutant revealed that the lack of melanin biosynthesis can be associated with an insertion element (IS) transposition. Genetic evidence suggested the presence of copper-controlled regulation, so we speculate that methanotrophs possess a yet-to-be identified biosynthesis pathway which similarly to eumelanin is initiated by tyrosinase. Our findings indicate that methanotrophic bacteria in arid soils produce a previously uncharacterized eumelanin-like polymer. We show that methanotrophic melanin can protect from UV and heat stress, benefiting themselves and associated biota-microbes and plants. Experiments also show that specific soil microbiota can influence methanotroph activity and growth. Collectively, these results suggest that members of arid soil microbial communities have evolved specialized mechanisms that enhance their resilience to drought and thermal stress. Additional research efforts are underway to elucidate the molecular structure and biosynthetic pathways of C1-melanin.

Genomic dark matter drives metabolic innovation in high-altitude tropical wetland

Daniela Betancurt-Anzola* (dkb5890@psu.edu)
Penn State

Daniela Betancurt-Anzola, Luisa Robles Zaragoza, Sarah Glass, Sophie Jurgensen, Kelly Wrighton, Johanna Santamaría-Vanegas, Estelle Couradeau

Wetlands are the single largest natural source of atmospheric methane, currently threatened by climate change and expected to contributing significantly to global warming. However, current climate prediction models rely heavily on data from northern latitudes, leaving tropical wetlands, like the Andean Páramos, critically underrepresented. This data limitation creates a significant blind spot in our ability to accurately predict methane feedback loops in a warming world. To address this gap, we investigated the soil metagenomes of the Chingaza National Natural Park (Colombia) across three distinct plant assembled driven ecosites: Espeletia, Chusquea, and peatland systems.

Our analysis reveals that microbial community structure is significantly driven by ecosite type, with peatland soils exhibiting the most distinct composition. We also recovered 109 high-quality Metagenome-Assembled Genomes (MAGs), 41 representing unclassified microbial dark matter. Metabolic potential reconstruction (DRAM2) coupled with physiological inference (MicroTrait) revealed that these Dark Matter lineages are specialized ecosystem engineers rather than functionally redundant opportunists. They harbor critical pathways for methanogenesis, nitrogen fixation, and sulfate reduction that vary by vegetation type. Furthermore, life-history analysis indicated a convergent persistence strategy among these novel taxa where they prioritize stress tolerance and low yield over resource acquisition. This genomic investment explains their success in the extreme fluctuating weather conditions of the high-altitude Andes. These findings establish the Chingaza páramo as a reservoir of potentially endemic, high-abundance microbial novelty that drives biogeochemical cycling but remains invisible to standard reference databases.

Linking microscopy with genome sequencing of uncultured microbes

Danielle Goudeau* (dgoudeau@lbl.gov)
Joint Genome Institute (JGI)

Danielle Goudeau, Daniel R. Utter, Trina Mcmahon, Rex R. Malmstrom

Fluorescence-activated cell sorting (FACS) followed by whole genome amplification enables genome sequencing of individual microbial cells. Although this method has a high throughput and can be coupled with interesting functional assays, it cannot accurately measure basic structural info such as cell size and morphology. FACS is also limited to cells and small microbial aggregates <50um in size.

Laser Microdissection (LMD) is an alternative approach for isolating cells and large aggregates (e.g. <= 250um) that relies on using a laser to cut the substrate around the particle and transfer the cut section into a collection vessel. More importantly, LMD enables imaging of cells and aggregates by brightfield, fluorescence, and other modalities before cell recovery and whole genome amplification, thus linking microscopy with genome sequencing. Here, we discuss two applications of LMD followed by whole genome sequencing.

Linking microbial community structure to function through strain-resolved metaproteomics

Delaney Beals* (bealsdg@ornl.gov)
Oak Ridge National Laboratory

Delaney G. Beals[1], Jonghyun Kim[1], Robert L. Hettich[1], Dale A. Pelletier[1], Mitchel J. Doktycz[1] [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Functional redundancy is frequently invoked to explain the stability of microbial communities, yet it is often inferred from shared genomic potential rather than directly measured from realized activity. Consequently, it remains unclear whether closely related strains that encode similar functions are functionally interchangeable within multispecies communities. Here, we use strain resolved metaproteomics to assess functional redundancy among closely related *Pseudomonas* strains embedded within a defined bacterial community derived from the *Populus* root microbiome. We constructed a ten-member synthetic community in which a single *Pseudomonas* representative was systematically replaced with one of nine alternative isolates while holding all other community members and growth conditions constant. By quantifying protein expression across the entire community before and after stabilization, we evaluated functional overlap ranging from individual proteins to emergent physiological states.

Despite substantial conservation of genomic content among *Pseudomonas* isolates, metaproteomic profiles resolved these strains into two statistically distinct clusters. These clusters were characterized by differential allocation toward biosynthetic, growth-associated functions versus stress-associated maintenance functions. These differences were driven primarily by strain specific regulation of conserved core pathways rather than by accessory gene content. Additionally, variation in *Pseudomonas* functional state propagated to the broader community, affecting the relative abundance and proteomic responses of non-*Pseudomonas* members. Together, these results demonstrate that functional redundancy among closely related bacteria is context- and scale-dependent, emerging from shared functional capacity while remaining constrained by regulatory and physiological divergence. This work highlights the importance of measuring realized activity to understand how diversity within a genus shapes community function and stability.

Metabolic specialization and phenotypic heterogeneity in natural microbiome

Dmitry Sutormin* (dsutormi@systemsbiology.org)
Institute for Systems Biology

Dmitry Sutormin, Karl Gaisser, Crystal Perez, Christian Diener, Alex Carr, Sean M. Gibbons, Anna Kuchina

In complex natural communities, individual bacteria encounter a diverse array of environmental niches, as well as signals from other microbiome members. As a result, a multitude of community-wide behaviors emerge from a multiplication of variable phenotypic states and responses to highly heterogeneous microenvironment. Bulk metagenomic and metatranscriptomic studies due to inherited averaging, however, are unable to resolve phenotypic heterogeneity of bacterial populations. As a result, this hides functional subpopulations and hinders mechanistic understanding of bacterial ecology in complex consortia.

To fill this gap and reveal single-cell transcriptional heterogeneity at a microbiome-scale, we developed mmSPLiT, a scalable single-cell RNA sequencing approach optimized for microbiomes. We successfully applied mmSPLiT to two synthetic microbiomes, comprised of 7 and 15 taxonomically diverse species, respectively. Using these data, we developed a computational pipeline that identifies the microbiome composition and decomposes the initial multi-species scRNA-Seq dataset into species-specific datasets with high accuracy and sensitivity.

Next, we profiled fecal microbiomes obtained from three healthy human donors with mmSPLiT, capturing over 19,000 high-quality single-cell transcriptomes from across >25 bacterial species from each microbiome. We revealed extensive transcriptional heterogeneity in 16 species, including the health-related gut residents Segatella copri, Dorea longicatena, and Anaerostipes hadrus. Specifically, we observed ubiquitous heterogeneous expression of metabolic pathways, stress pathways, and core housekeeping genes. Further, we detected multiple transcripts from mobile genetic elements and assigned them to hosts from different genera belonging to Eubacteriales, Bacteroidales, and Burkholderiales. Finally, we performed mmSPLiT on marine microbiome samples from Puget Sound, uncovering multiple microbial species and strains.

Advancing Metagenomics and Metatranscriptomics with Ultra-Accurate ONT Long-Read Sequencing

Dori Deng* (dorid@lnl.gov)
Joint Genome Institute (JGI)

Dori Deng, Tomas Tyml, Annika Danneman, Frederik Schulz

Environmental sample sequencing is a powerful tool for understanding microbial ecosystems and unlocking their genomic potential, with significant implications for biogeochemical process understanding, novel biotechnological resource discovery, and environmental health monitoring. However, working with environmental samples poses challenges, including limited biomass, high complexity of microbial communities, and degraded DNA/RNA. Currently, short-read sequencing is widely used, but it has limitations, including high costs, long turnaround times, and difficulties in resolving conserved sequences for diverse species and structural variants. Long-read sequencing technologies, such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences(PacBio), offer a promising alternative. However, these long-read sequencing methods have been hindered by either high error rates or high costs. Here, we have developed an ultra-accurate long-read sequencing method using ONT sequencing. By combining Rolling Circle Amplification and Consensus calling with Unique Molecular Identifiers (R2C2+UMI), we achieved per-base accuracy of 99.97% for low-input environmental samples, suitable for strain-resolved sequencing. This method allows us to generate high-quality, high-throughput sequencing data from diverse environmental samples, facilitating comprehensive research in metagenomics, metatranscriptomics, and amplicon sequencing, and driving scientific discovery.

Using AlphaFold3 to Model and Edit Ice Nucleation Proteins and Understand their Role in Bacterial Survival in Adverse Atmospheric Conditions

Eleah Flockhart* (eleah.flockhart@colostate.edu)
BROADN, CSU

Eleah Flockhart, Sam Golon, Kevin Barry, Tom Hill, Marina Nieto-Caballero, Pankaj Trivedi, Sonia Kreidenweis, Brad Borlee.

The microorganisms that live in the air are an understudied contributor to natural and agricultural ecosystems. The ability of some of these organisms to produce Ice Nucleation Proteins (INPs), which allow ice to form at higher temperatures, is a potentially integral part of their resistance to environmental stresses and their effect on plant health. INPs function as a virulence factor, allowing bacteria to damage plant tissues and infect the plants. However, studies have not looked at how INPs contribute to bacterial survival during aerosolization and desiccation. This project aims to investigate the physiological role of INPs in the fitness of airborne bacteria via whole genome sequencing and assembly, protein modeling, and an allelic exchange vector to create mutants of INP expressing strains of Pantoea ananatis. The INPs have been modeled in Alphafold3 based on the genome assembly. The mutant strains will have regions of the INP gene deleted via an allelic exchange process. These mutants will then be used to evaluate how the structure and presence of the protein affect key phenotypes such as freeze tolerance, UV tolerance, and desiccation. If INPs provide increased fitness under any of these conditions, it would demonstrate that INPs increase bacterial survival in addition to acting as a virulence factor.

Comparative Functional Genomics of Metagenome-Assembled Genomes from Heavy-Metal-Impacted Non-Marine Environments

Elijah Bring Horvath* (eli.bringhorvath@pharm.utah.edu)
University of Utah

Elijah R. Bring Horvath, Jaclyn M. Winter

Metagenome-assembled genomes (MAGS) provide a powerful lens for examining how microbial communities, including hard-to-culture species, adapt to inhospitable environments shaped by diverse environmental stressors. Here, we present a comparative genomic analysis of MAGs recovered from multiple non-marine environmental regions characterized by extreme salinity and elevated heavy-metal burdens. Using a combination of taxonomic profiling and functional annotation, we investigate whether microbial communities inhabiting these environments converge on shared genomic strategies for metal tolerance and homeostasis. Across sites, we observe both conserved and environment-specific patterns of metal-associated functional potential. While uptake systems for essential metals such as iron, zinc, and manganese are broadly represented, environments with higher inferred metal stress show enrichment of efflux systems, detoxification enzyme, and regulatory pathways associated with metal resistance. These trends suggest that microbial survival in metal-impacted environments is not driven solely by taxonomic composition, but by differential investment in specific functional strategies. Together, this work highlights the value of function-centric comparative genomics for interpreting microbial adaptation in extreme environments and provides a framework for linking environmental chemistry to genomic analysis.

Subsurface Microbial Community Dynamics Following a Deuterium Injection in Antrim Shale

Elliott Barnhart* (epbarnhart@usgs.gov)
U.S. Geological Survey

Elliott P. Barnhart Hannah Schweitzer, Matthew Varonka, Elizabeth Tomaszewski, Anna Martini, Elisha Moore and Matthew W. Fields

Microbial methane (CH4) accumulations in terrestrial organic-rich subsurface reservoirs are prevalent worldwide, but the timing and ongoing nature of deep subsurface CH4 generation remains uncertain. In this study, downhole sampling and an in situ injection of deuterated water (D2O) were used to investigate the microbial communities and methanogenic activity in Antrim Shale within the Michigan Basin, USA. The Antrim Shale stands out for its secondary biogenic CH4 production from depths of 150 to 600 meters. Three production wells were initially sampled across the northern portion of the Michigan Basin with diffusive microbial samplers to examine the microbial ecology and to inoculate laboratory microcosms. Methanogens were detected and CH4 production was evident in microcosms from all three wells. Subsequently, one well underwent a 2384-liter injection of 0.084% D2O to isotopically label newly generated CH4 (i.e., CH3D, CD4 etc.). Gas samples were collected at 21, 50 and 71 days postinjection at the wellhead to monitor any changes in gas isotopes. At 85 days postinjection, 3179 liters of water was pumped to collect final gas and water samples. The water samples were filtered, and the DNA was extracted for metagenomic and metaproteomic analysis. The results provided valuable insights into methanogenic microbial communities and their activity in subsurface environments, with broad implications for CH4 resource assessments and environmental management. Understanding the spatiotemporal dynamics of CH4 generation in the subsurface is crucial for informing and optimizing resource exploitation strategies.

Time Series Modeling of a Marine Algal Bloom-Associated Microbiome

Emma Timmins-Schiffman* (emmats@uw.edu)
University of Washington

Emma Timmins-Schiffman, Claire Elbon, Anna Leipertz, Sean Perez, Miranda Mudge, Gabi Chebli, Rachel Schauer, Tasman Nunn, Mark Scheuerell, Tatiana Rynearson, Julia Kubanek, Brook Nunn

Aquatic microbiomes encompass complex ecological interactions among prokaryotes, eukaryotes, and environmental factors. These interactions are chemically mediated, leading to shifts in relative abundances of microbes and affecting ecosystem dynamics. Bacterial influence over harmful algal blooms is of particular interest in the field of cross-kingdom microbial interactions due to the ecological and economic impacts of blooms. We captured a high-resolution time series, sampling every four hours over 21 days, of the bacterial community off Orcas Island, Washington, USA during a period of high nutrient input followed by an algal bloom and the start of a second bloom. The 0.2 um size fraction of bacterioplankton was sequenced by JGI and the resulting estimated gene copy numbers for taxonomy and gene functional categories were used as input into time series models to distill patterns in the bacterial microbiome over the course of these large and transient environmental changes. Both bacterial community and function were impacted by the algal blooms. The first bloom, coinciding with a large nutrient influx, saw an increase in classes in the Chloroflexi phylum, while the second bloom was dominated by classes in the PVC superphylum. Bacterial KO functional categories also differed between the environmentally distinct bloom periods, suggesting contrasting functional profiles of the bacterial community. These findings uncover some of the complexity of bacterial responses to ecological disturbances and how they may influence algal bloom dynamics.

eBERlight: Enabling Multimodal Imaging and Structural Biology for BER Science at the Advanced Photon Source

Gyorgy Babnigg* (gbabnigg@anl.gov)
Argonne National Laboratory

Gyorgy Babnigg2, Xiaoyang Liu1, Malgorzata Korbas1, Debora Meira1, Alex Lavens1, Qiaoling Jin3, James O'Sullivan1, Priyanka Gade1, Kemin Tan1, Changsoo Chang1, Youngchang Kim1, Michael Endres1, Karolina Michalska1

The eBERlight program at the Advanced Photon Source (APS) enables biological and environmental research by providing integrated access, scientific support, and multimodal X-ray capabilities aligned with the U.S. Department of Energy Office of Biological and Environmental Research (BER) mission. Following the APS Upgrade, ultrabright, highly coherent X-ray beams now support experiments spanning angstrom-to-millimeter length scales and microsecond-to-hour time domains, enabling new classes of measurements on complex biological and environmental systems.

eBERlight connects BER researchers to advanced synchrotron techniques including macromolecular crystallography, X-ray fluorescence microscopy, X-ray absorption spectroscopy, X-ray computed tomography, coherent imaging, and scattering methods. These approaches enable structural, chemical, and spatial characterization of proteins, enzymes, microbes, plant tissues, soils, and bio-derived materials under realistic and dynamic conditions. Applications include enzyme and pathway engineering, microbial and plant systems biology, metal and nutrient mapping, biomass and rhizosphere studies, and characterization of engineered and natural biogeochemical processes.

Beyond beamline access, eBERlight provides end-to-end user support spanning experimental design, proposal development, sample preparation strategies, multimodal workflows, and data analysis. The program actively promotes coordinated use of BER user facilities by supporting cross-platform experiments that combine synchrotron measurements with genomics, multi-omics, and molecular characterization resources available through JGI and EMSL. By lowering technical barriers and enabling integrated experimental strategies, eBERlight accelerates discovery and supports scalable, reproducible BER science across diverse research communities.

Biochar-induced Root Exudates Rewire the Rhizosphere Microbiome and Its Functionality

Hanyue Yang* (hyang5@esf.edu) **presenting for Yaqi You** (yyou@esf.edu)
SUNY College of Environmental Science and Forestry

Biochar has multifaceted benefits on the soil-plant system; yet the underlying mechanism remains obscure. Integrating multi-omics and using wheat as a model plant, we demonstrate that biochar induces differential root exudation which rewires the rhizosphere microbiome and its functionality. Complex molecules in biochar-induced root exudates, particularly plant secondary metabolites and phytohormones, evoke a plant-beneficial rhizosphere microbiome centered by diverse plant growth promoting rhizobacteria. The restructured microbiome exhibits shifted nitrogen metabolism, characterized by enhanced nitrogen fixation, nitrification, and complete denitrification with reduced N_2O emission potential. Biochar suppresses methane production through orchestrated microbial functional shifts: Stimulated methionine salvage that deprives the methanogenesis precursor methanethiol; suppressed biosynthesis of coenzyme M, a cofactor required for the methane-forming step; and enhanced methane oxidation through enrichment of methanotrophs. The restructured microbiome also contributes to the transformation of phytohormones and biosynthesis of redox-active quinone compounds. Our work provides new insights into biochar's profound impact on the soil plant system and highlights the promise of engineering the rhizosphere through reshaping root-microbe interactions using biochar or root exudate cocktails. Since fire derived black carbon shares similar chemistry with biochar and increased frequency of wildfire could cause more black carbon deposition, this research also spotlights the overlooked impact of fire-derived black carbon on belowground carbon and nitrogen dynamics through intricate rhizosphere interactions.

The Strain Engineering Platform at the Joint Genome Institute (JGI)

Hualan Liu* (hualanliu@lbl.gov)
Joint Genome Institute (JGI)

Hualan Liu*, Anandharaj Marimuthu, David Robinson, Zhiying Zhao, Ian Blaby

The Joint Genome Institute's strain engineering platform facilitates the functional assessment of genomic sequences by developing heterologous expression hosts. This is achieved through Chassis-independent Recombinase-Assisted Genome Engineering (CRAGE), a technology capable of integrating large metabolic pathways directly into diverse bacterial chromosomes. The CRAGE workflow utilizes transposon-mediated integration to install a "landing pad" equipped with a Cre-lox recombineering system. To date, we have successfully engineered CRAGE hosts across four bacterial phyla—including Proteobacteria, Firmicutes, Bacteroidetes, and Actinobacteria—and quantified their cassette exchange efficiencies. Furthermore, we have adapted CRAGE for select model yeasts and are currently expanding the platform to include a broader range of phylogenetically diverse microbial hosts.
Concurrently, we are optimizing conjugation-based DNA transformation protocols to leverage high-throughput robotic systems, reducing the time and cost associated with pathway integration in CRAGE hosts. Furthermore, we are establishing a systematic workflow for comprehensive vector engineering. This initiative aims to produce standardized plasmid toolkits and robust transformation methods to facilitate precise genetic manipulation. Together, these advancements provide the foundational tools necessary to characterize and engineer emerging non-model microbes.

Free access to Cryo-EM for BER science

James Evans* (james.evans@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

A Krios cryogenic transmission electron microscope is available for free user access at the Environmental Molecular Sciences Laboratory (EMSL) to advance DOE BER user research in protein/small molecule structural biology and whole cell ultrastructure. The operation of the Krios G3i instrument is a joint funding venture between EMSL and BER. The microscope is available to the general EMSL user community and BER researchers in a 50/50 split allocation.
EMSL users can access this instrument free of charge via the normal EMSL user proposal calls, which permit combining cryo-EM with other EMSL capabilities such as mass spectrometry or super-resolution fluorescence microscopy. Access is to cryo-EM only resources is available for BER users and this access mechanism allows for an expedited submission and review process.

The Krios G3i is fully operational and has been applied to multiple EMSL and BER User projects. The microscope has complete screening, data collection, and image processing workflows for (1) microelectron diffraction of small molecule or protein crystals, (2) single-particle analysis of soluble and membrane protein complexes, and (3) electron tomography of whole cells or isolated organelles. It is equipped with a K3 direct electron detector, Ceta-D camera, phase plate, and Bioquantum energy filter. In addition to semiautomated data collection, the facility has installed automated image processing workflows for real-time monitoring feedback of session quality and full 3D reconstruction of all workflows. To date, the facility has demonstrated sub-Å resolution microelectron diffraction, sub-2 Å resolution from 3D single-particle protein structure determination, and sub-nanometer resolution for whole-cell tomography. While the facility provides rapid access for samples that arrive frozen on clipped and prescreened grids, users can also begin with samples that arrive in buffer and require all steps of the cryo-EM workflow. In a subset of cases, users can also start from a provided gene of interest and employ the cell-free expression system to produce enough protein for structural characterization. We will highlight several recent user results as well as an example of going from cell-free expression through cryo-EM structure determination in less than 24 hours.

Exploring Microbial Responses to Sediment Amendments in Coastal Wetland Restoration

Jing Yuan* (Jing.Yuan@ct.gov)
Connecticut Agricultural Experiment Station

Jing Yuan, Beth Lawrence, Ashley Helton, Chris Elphick, Itamar Shabtai, Blaire Steven

Coastal wetlands are increasingly threatened by human activities, including salt marsh squeeze driven by sea-level rise and coastal development. As lower marshes become submerged and landward vegetation migration is constrained, bird habitat is lost or degraded. Along Long Island Sound (LIS), up to 97% of marsh habitat is predicted to be lost by 2100, underscoring the urgency of effective restoration strategies to sustain ecosystem function. Sediment amendments using locally sourced dredged material are widely used to elevate marsh surfaces, promote vegetation recovery, and support wildlife habitat. However, their biogeochemical consequences remain unknown, particularly regarding greenhouse gas emissions (GHGs) and bacterial oxidation of reduced inorganic sulfides in dredged sediments that form acid sulfate soils and inhibit plant recovery. To address this gap, we evaluated sediment amendments across multiple LIS marshes, including Great Meadows Marsh (CT). GHGs varied interannually (2022–2024), specifically CO_2 fluxes were positively associated with vegetation cover, indicating restoration efforts influence GHG cycling. Initial 16S rRNA gene sequencing revealed significant shifts in bacterial community composition associated with vegetation cover. Preliminary experiments also suggest that altered flooding regimes can produce acid sulfate soils, directly linking microbial activity to restoration outcomes. Together, these results highlight the need to resolve microbial processes connecting sediment biogeochemistry, vegetation recovery, and ecosystem function. Building on this foundation, we were recently awarded a JGI New Investigator project to apply functionally resolved genomic approaches to link microbial activity with sediment amendments and biogeochemical outcomes, providing a mechanistic framework for resilient salt marsh restoration.

A High-Throughput Discovery Engine for Carbon-Sequestering Enzymes

John Collins * (jcolli11@student.ubc.ca)
University of British Columbia

John Collins, Sebastian Hyland, Beth Davenport, Steven J. Hallam

The microcosmos pervades virtually every habitat, thrives under powerful extremes, and sustains conditions for life on Earth. High-throughput sequencing (HTS) has transformed our perception of the microcosmos, linking microorganisms to biogeochemical processes and providing a blueprint for sustainable biotechnology innovation. Despite its promise, a paucity to convert sequence information into validated enzyme functions limits knowledge creation and translation. Here we describe a workflow for enzyme discovery and validation, that links together phylogenetic mapping of functional anchor genes using the Tree-based Sensitive and Accurate Phylogenetic Profiler (TreeSAPP) with supervised and semi-supervised HTS using a recently developed cell surface display system deployed in Caulobacter vibrioides CB2A JS4038. We focus our proof of concept on carbon capture and conversion using a reference package for alpha-carbonic anhydrase (aCA) which catalyzes the reversible hydration of CO2 to bicarbonate (HCO3). Carbonic anhydrase plays important roles in photosynthetic carbon fixation and facilitating microbial induced calcium carbonate precipitation (MICP) for long-term carbon sequestration. The aCA reference package was deployed across NCBI genomes, GTDB (Genome Taxonomy Database) MAGs, and SRA contigs (Logan) to map the global distribution of aCAs across diverse taxonomic and biogeographic zones, identifying known and novel aCAs. Selected aCAs were synthesized and inserted into a display vector for functional validation on JS4038 cell surfaces. In parallel, a plate-based HTS paradigm was developed for unsupervised recovery of aCAs by adapting the JS4038 surface display system for environmental cDNA expression. This discovery and validation engine is extensible to a wide range of protein functions and bioprocess applications.

Decoding viral impacts in plant-rhizobacterial interactions

Jonelle Basso * (jbasso@lbl.gov)
Joint Genome Institute (JGI)

Jonelle Basso, Thai Dao, Vlastimil Novak, Maureen Berg, Nicolas Grosjean, Saray Maeda, Zhiying Zhao, Mária Džunková, Nathalie Elisabeth, Meghana Faltane, Mimi Lynde, Katherine Louie, Benjamin Bowen, Trent Northen, Axel Visel

The plant microbiome includes beneficial microbiota that provide the plant protection, increased nutrient availability and stress resilience. While previous and ongoing studies of the rhizosphere microbiome have been critical for assessing the impact of specific plant-microbe interactions, their focus has overwhelmingly targeted bacterial and fungal members of the microbiome. Viruses are ubiquitous, outnumbering all other biological entities on the planet, yet they are remarkably understudied in the rhizosphere. Prior transposon-sequencing data analysis conducted in our group identified functional roles for hundreds of genes in the plant growth promoting rhizobacterium Pseudomonas simiae WCS417 that are important for its colonization of the rhizosphere. Two of these genes that cause reduced fitness in the rhizosphere when mutated are components of a latent bacteriophage and are present among two phage loci ranging in size from 15-65kbp. To better understand this phenomenon, we used a loss of function approach to generate fluorescently-labeled phage gene deletion mutants and subsequently conducted experimental characterization studies such as root colonization assays and phenotypic comparative assessments. We co-cultivated the GFP-labeled wild-type (WT) and prophage absent (Δphage) strains with Arabidopsis thaliana and Brachypodium distachyon to assess the prophage's impact on both bacterial host and plant physiology. Our results demonstrate that prophage presence correlated with change in plant root architecture, root exudate metabolic profiles, and shoot biomass under thermal stress without affecting bacterial colonization ability. Collectively, these findings support a triadic plant-bacterium-phage axis in the rhizosphere, underscoring the presumptive critical role of phages in shaping community dynamics and plant-microbiome interactions.

Targeted multi-omics of metabolically active microbial populations in anaerobic digesters bioaugmented with carbon-based conductive materials.

Juan Santana * (juancamilo.santanamartinez@ubc.ca)
University of British Columbia-Hallam Lab

Juan Santana, Steven Hallam

Anaerobic digestion (AD) is a cornerstone technology in the bioenergy sector, bridging renewable natural gas (RNG) production with sustainable waste management. However, municipal-scale digesters often operate far below their theoretical efficiency due to poorly constrained microbial interactions governing electron transfer and methanogenesis. Enhancing direct interspecies electron transfer (DIET) between syntrophic bacteria and methanogenic archaea has therefore emerged as a promising strategy to improve the conversion of complex organic matter into methane-rich biogas, with evidence showing systems amended with conductive carbon (CC) materials enhance RNG production. Building on prior demonstrations, we tested a novel side-stream incubator reactor incorporating conductive carbon cloth to selectively enrich slow-growing, methanogen-associated consortia for bioaugmentation of conventional digesters resulting in improved process stability and biogas quality. To constrain microbial interactions driving these process improvements we applied biorthogonal noncanonical amino acid tagging (BONCAT), stable isotope probing (SIP), and fluorescence-activated cell sorting (FACS), coupled with microdroplet processing in one pot for trace samples (μPOTS) to recover proteomes, metabolomes and transcriptomes of active microbial populations assimilating labeled short-chain fatty acids that serve as key electron donors for methane production in the CC-enhanced AD milieu. These data were mapped onto high-quality reconstructed microbial genomes derived from short- and long-read metagenomic datasets and single-cell amplified genomic datasets to directly link microbial identity, function, and activity. In addition, data were used to construct activity-resolved genome correlation networks to identify syntrophic interactions including DIET and syntrophic acetate oxidation. Taken together these results provide the first direct, *in situ* validation of syntrophic networks in CC-enhanced ADs, identifying key microbial players and metabolic interactions with the potential to inform more robust, efficient biogas production outcomes at municipal scales.

A simple but hardy bacterial community detoxifies chlorite and breaks down benzene in the cooling water of the world's largest supercomputer for AI

Justin Podowski* (jpodowski@anl.gov)
Argonne National Laboratory

Justin Podowski, Ali Fox, Sarah Owens, Mitchell McClellan, Susan Coghlan, James J Davis, Rick L Stevens

Microbial growth is a potentially serious concern in industrial cooling water systems, as it can lead to biofilm development which can impede water movement and reduce cooling capacity. We used deep metagenomic and metatranscriptomic sequencing of microbial communities in the cooling water of Argonne National Laboratory's Aurora Supercomputer, to understand the metabolic potential and activity of these microbes. We found a relatively simple community dominated by bacteria in Azospira, Rhizorhabdus and Pseudomonas, among several others. An average of nearly 50% of the total community was made up of the top 8 most abundant species, which were predicted to be aerobic, heterotrophic, and mesophilic. Transcriptional activity demonstrates that energy saving glyoxylate shunt and beta-oxidation were highly active, suggesting energy starvation and scavenging. Xenobiotic degradation of benzotriazole – a corrosion resistance additive – was evident in transcription of the full degradation pathway of benzene to acetyl-CoA and pyruvate, which likely feeds into the highly active glyoxylate shunt. Response of bacteria to the biocide chlorine dioxide is also evident from the abundant transcription of chlorite dismutase and the presence of chlorite dismutase in several minor and abundant MAGs. Autotrophic and photosynthetic pathways were not present or at very low abundance, suggesting a system which is likely dependent overall on allochthonous organic carbon. Our work provides guidance for future HPC systems management as water cooling becomes dominant, as well as insight into an emergingly prominent microbial ecosystem in the built environment which contains metabolically flexible bacteria.

Populus trichocarpa selects unique microbial communities under different climate-induced abiotic stress conditions

Kathryn Bazany* (bazanyke@ornl.gov)
Oak Ridge National Laboratory

Kathryn E. Bazany, William Argiroff, Alyssa A. Carrell1, Kelsey R. Carter, Nancy L. Engle, Dawn M. Klingeman, Justus Smith, Ian Morris, Amy Schaefer, John Lagergren, Daniel Jacobson, Stanton L. Martin, Melissa A. Cregger, David J. Weston, Dale A. Pelletier, Christopher W. Schadt

Abiotic factors like temperature, water, and nutrients directly influence both plants and their associated microbial communities. Abiotic stress conditions elicit targeted microbial community responses which can have a considerable influence on broader ecosystem processes like nutrient cycling, and on plant fitness, as plants under stress form advantageous partnerships with microbes possessing stress relieving functionality. Using amplicon sequencing, we examined the bacterial and fungal communities of bulk soil, rhizosphere soil, and root endosphere of Populus trichocarpa trees from 11 sites along a naturally occurring temperature and rainfall gradient with variable soil conditions over the Cascade Mountains in Washington. We found that climate factors, especially average temperature and rainfall, correlated with both bacterial and fungal community composition in all of the observed niche compartments. We also observed that microbial communities broadly clustered into three main compositional groups driven by climate condition: a high precipitation, warmer group along the lower Cowlitz River on the Western side, a colder high elevation community at the crest of the pass and along the upper Tieton River, and a lower precipitation, hotter condition community along the lower Tieton River on the Eastern side. We are currently investigating the functional potential of Populus trichocarpa rhizosphere and endosphere microbial communities under these varying climate and soil conditions using metagenomic approaches. We hypothesize that the host trees will select for microbes with the functional potential to alleviate specific types of abiotic stress and that the functional potential of these communities will cluster based on these same broad climate categories.

Storage impact on IMPROVE filters for microbial communities and ice nucleating particles

Kevin Barry* (kevin.barry@colostate.edu)
Colorado State University

Kevin Barry, Amy Sullivan, Taylor Melling, Bret Schichtel, Faris Jawda, Lawrence Tsai, Yongjing Zhao, Tetsuya Kawamoto, Nicole Hyslop, Ann Dillner, Sarah Gering, and Sonia Kreidenweis

The air contains many microorganisms, and their components, collectively termed bioaerosols. They are important to detect and monitor as they can provide a baseline and spot changes in an ecosystem, especially within a changing climate. Linking to established networks may be a cost-effective and valuable way to deepen understanding of bioaerosols with increased spatial and temporal resolution. The Interagency Monitoring of PROtected Visual Environments (IMPROVE) network currently has about 160 sites and samples every 3 days dating back decades. One of the filters collected is currently only used to report PM10 gravimetric mass concentrations. These filters could be leveraged for bioaerosol analyses, however, they are stored at ambient conditions before and after analysis. Thus, quantifying the losses of DNA as a function of storage time would assist with establishing appropriate uses of this archive. In this project, we sought to understand how storage of IMPROVE PM10 filters affects recovery of bioaerosol composition and total DNA concentration. We set up three IMPROVE PM10 Modules next to an established site and collected samples with the same protocols and frequency. Initial results point to broad similarities of microbial community composition for bacteria and fungi within a week but changes out to 3 months. We also tested storage impact on a special type of aerosol, ice nucleating particles (INPs), which have a large impact on cloud composition and lifetime and can also indicate microbial damage. Ice nucleating particle concentrations decreased after 1 month of room temperature storage, indicating some degradation over this time period.

Soil Moisture Modulates Microbial Carbon Processing in a Combined Rhizosphere– Detritusphere Habitat.

Linnea K. Hernandez* (hernandez163@llnl.gov)
Lawrence Livermore National Laboratory

Linnea K. Hernandez* (hernandez163@llnl.gov), Noah W. Sokol, Megan M. Foley, Steven J. Blazewicz, Nicole Didonato, Ljiljana Pasa-Tolic, William Kew, Katerina Estera-Molina, Mary Firestone, Alex Greenlon, Bruce A. Hungate, Jennifer Pett-Ridge

Shifts in precipitation regimes are expected to impact microbial transformations of soil organic matter (SOM) in semi-arid California grasslands. While drought has been shown to alter plant rhizosphere (live root) and detritusphere (litter)-derived carbon transformations, the microbial mechanisms driving these processes across mixed habitats, where roots interact with decomposing litter, remain poorly understood. Here, we used stable isotope probing and multi-omics to examine how drought influences microbial activity and carbon flow across three experimental soil habitats in mesocosms: the rhizosphere, the detritusphere, and the combined rhizosphere–detritusphere. We tracked isotopically labeled carbon from plant material into mineral-associated organic matter (MAOM) and measured microbial gene transcript abundance (metatranscriptomics) and SOM chemical composition (FTICR-MS). We found that the presence of living roots suppressed detritus-derived MAOM formation while simultaneously driving microbial gene expression toward a rhizosphere-like functional state under ambient moisture conditions; however, under drought conditions the presence of living roots had a less pronounced impact on detritus-derived MAOM formation, gene expression, and SOM chemistry. Under ambient conditions, root presence reduced transcript abundances of plant polymer degradation genes while increasing expression of simple sugar metabolism genes. These functional shifts corresponded with enrichment of lignin-like compounds in the detritusphere, indicative of decomposed plant material, and condensed hydrocarbon-, protein-, lipid-, and amino sugar–like compounds in rhizosphere habitats, indicative of microbial biomass and root exudates. Together, these results highlight how interactions among soil moisture,

habitat, and microbial ecophysiology regulate the pathways through which plant and microbially derived carbon contributes to SOM.

Mapping of Great Salt Lake Prophage Diversity Across Salinity and Heavy Metal Gradients Using Bacterial Isolates and Metagenomes.

Lydia Dresler* (Lydia.dresler@pharm.utah.edu)
University of Utah

Lydia K. Dresler, Jaclyn M. Winter

Bacteriophages influence microbial community structure and adaptation by mobilizing genes that alter host fitness, metabolism, and stress tolerance. In extreme environments, such as hypersaline, metal-contaminated, and effluent wastewater systems, phage-mediated gene exchange may be especially important for microbial survival, serving as a reservoir for genetic functions. Despite studies of bacterial diversity in Great Salt Lake, phage and prophage diversity in this hypersaline, heavy metal-rich environment remains poorly characterized. Here, we profile prophage regions recovered from a collection of cultured bacterial isolates and metagenomic datasets using GeNomad. From the datasets, we mapped and recovered prophage sequences to lake sampling locations spanning salinity gradients (8-28%) and elevated concentrations of toxic metals, including arsenic, lead, and mercury. Using CheckV to ensure confident results, 172 phage regions were identified from this dataset. Comparative analyses identified clustering of related prophage regions and identified patterns of phage distribution associated with environmental conditions and microbial community composition. Genome similarity, measured by the Jaccard index, revealed 36 clusters with each region sharing greater than 99% similarity, indicating conserved phage species across hosts and sampling sites. We further conducted a preliminary survey of auxiliary genes encoded within prophage regions, including candidate genes linked to stress tolerance, antimicrobial resistance, and adaptive functions in hypersaline and metal-rich environments. Together, these data establish a foundation for understanding how prophage mobilization may contribute to microbial ecology and adaptation in Great Salt Lake.

Antimicrobial Discovery from Soil-Derived Bacteria Using a Genome-Level Approach

Martel DenHartog* (denhartog2@wisc.edu)
University of Wisconsin–Madison

Martel L. DenHartog, Murrel Saldanha, Neha Kashyap, Kaetlyn T. Ryan, Sarah Miller, Marc G. Chevrette, Jo Handelsman

Resistance to pesticides and drugs is increasing plant and animal diseases, threatening production of bioenergy crops and other agricultural industries. Soil bacteria are a rich source of many natural products used as antimicrobials. The challenge is to minimize rediscovery of known compounds to detect new ones. In classrooms at 486 institutions across the United States, college students enrolled in a research course (Tiny Earth) isolate antimicrobial-producing bacteria from the soil and send them to a collection of 4,400+ isolates at the University of Wisconsin–Madison where scientists extract DNA for PacBio sequencing at the Joint Genome Institute. To find novel antimicrobial compounds, we take a comparative genomics approach using antiSMASH to mine the genome sequences for biosynthetic gene clusters (BGCs) that encode the machinery for production of secondary metabolites with antimicrobial potential. Taxonomic classification conducted via GTDB-tk and OrthoFinder is used to perform a multi-locus sequence analysis and phylogenetic tree construction. An annotated phylogeny with mapped BGC data provides researchers with genetic insight into the biosynthetic potential and evolutionary patterns of BGCs for soil bacteria across geographical samples. This analysis of 470+ genomes reveals the antimicrobial-producing potential of soil-dwelling bacteria and provides a basis for prioritizing isolates for chemical analysis of bioactive compounds. Since 2018, we have identified and solved structures of 20 bioactive molecules, including a nematicidal indole compound from a bacterial strain active against several pathogens that cause plant diseases. Genomic analysis provides a new basis for sorting through the chemical cornucopia of the soil microbiome.

A decadal, continental-scale view into the drivers of soil microbiome ecology and function

Matthew Kellom* (mkellom@lbl.gov)
Joint Genome Institute (JGI)

Matthew Kellom, Simon Roux, Alicia Clum, Hugh Cross, Neha J. Varghese, Sujay Patil, Christopher J. Mungall, Montana Smith, Julia M. Kelliher, Robert Riley, Kurt LaButti, Robert Egan, T.B.K. Reddy, Marcel Huntemann, Natalia N. Ivanova, Nikos C. Kyrpides, Axel Visel, Nigel J. Mouncey, Tijana Glavina del Rio, Igor V. Grigoriev, Ulas Karaoz, Eoin Brodie, Steven Hofmeyr, Leonid Oliker, Katherine Yelick, Paula M. Mabee, Katherine M. Thibault, Emiley A. Eloe-Fadrosh

Microbial communities drive Earth's biogeochemical cycles, yet our understanding of their function and ecosystem services remains limited. The U.S. National Science Foundation's National Ecological Observatory Network (NEON), in partnership with the U.S. Department of Energy's Joint Genome Institute (JGI) and National Microbiome Data Collaborative (NMDC), offers an opportunity to study soil microbiomes across space and time. NEON's continental-scale design systematically samples soil microbiomes and environmental variability using standardized methods with over 4,000 samples collected over the span of a decade.

Using a combination of sequencing read analysis and genome-resolved metagenomics, we relate microbial community taxonomy and function to the environmental conditions recorded across NEON stations. A total of 23 Tb of metagenome data has been generated from ~2,000 metagenomes, enabling sequencing read-based analysis of environmental microbiome composition and relationships with sample environmental data.

Additionally, 4.8 Tb of data were assembled to provide a consensus assembly that enabled reconstruction of hundreds of metagenome-assembled genomes as well as predicted viral genomes. This unprecedented dataset provides a unique view of soil microbiome composition, metabolism, ecological trends, and a look into the future of soil microbiome research. This data is continuing to be generated and is publicly available for further analysis.

Targeted multi-omics tool for genomic and functional profiling of biosynthetic gene clusters within natural and synthetic microbiomes

Minxi Jiang* (mj2770@lbl.gov)
Lawrence Berkeley National Laboratory

Minxi Jiang, Frederik Schulz, Lydia Rachbauer, Qingyun Dan, Dori Deng

Biosynthetic gene clusters (BGCs) encode diverse natural products with strong relevance to energy independence, biomanufacturing, and human health. Microbial communities harbor vast BGC biodiversity, yet much of it remains inaccessible due to community complexity and low-abundance functional taxa. Metagenomic assemblies frequently fail to recover complete BGCs from these taxa, while functional characterization is hindered by limited analytical sensitivity and strong background interference in environmental samples. In addition, how BGC regulation interferes with growth environments at the community level remains poorly understood, limiting scalable application.

To address these gaps, we will develop a new BGC function-oriented multi-omics panel with targeted hybridization probes that sensitively characterize the genomics and regulation of BGCs and pathway-related enzymes across diverse organisms within microbial communities, particularly low-abundance organisms. We will use metallophores, especially lanthanophores as critical raw materials with high supply risk, as proof-of-principle to validate the tool, which has vast extension potential to target numerous other BGCs.

We will leverage existing BGC databases and run complementary bioinformatics screening of selected metagenomic datasets to identify conserved genes and homolog enzymes related to lanthanide metabolism, and design capture probes matching sequence heterogeneity across defined species. Combined with targeted long-read DNA and RNA sequencing, and downstream bioinformatic analysis, this approach will enable sensitive BGC detection and expression profiling across diverse organisms within microbiomes, with minimal reliance on full metagenomic assembly. We will first validate these tools using several cultured strains (e.g., Methylobacterium) from controlled metal-feeding gas fermenters, and benchmark their performance against existing multi-omics approaches. We will then apply these tools to natural and synthetic microbiomes across diverse environmental and high-throughput operational conditions to reveal unexplored lanthanophore BGCs on community-level. Such a lanthanide utilization profile can serve as reference for downstream protein functional prediction and validation, genome editing, and process engineering toward scalable production of diverse natural products.

Addiction by Design: Detecting Phage Toxin–Antitoxin Systems and Assessing Their Impact on Host Metabolism

Moshe Steyn* (moshe.steyn@pnnl.gov)
University of Arizona, PNNL

M.D. Steyn, Ruonan Wu, David A. Baltrus, Jason McDermott

Addiction cassettes are paired protein systems composed of a stable toxin and a labile antitoxin. Once viewed as parasitic, they are now known to play diverse roles in prokaryotes, including viral defense, plasmid stability, stress responses, antimicrobial resistance, and virulence. Although addiction cassettes have been detected in bacteriophage, their function in these systems remains poorly understood. We developed TAfinder3D, an improved addiction cassette detector that incorporates structural similarity via Foldseek, substantially expanding the detectable diversity of these systems. This approach nearly doubles the number of toxin–antitoxin systems identified in distantly related bacteriophage from metagenomic data. Using TAfinder3D, we find addiction cassettes in over 15% of temperate phage but rarely in virulent phage, supporting their use as a lifestyle marker and enabling downstream enrichment analyses. To investigate whether phage-encoded addiction cassettes function beyond genome stability, we apply a multi-omic framework combining Hi-C viral–host linkage with metatranscriptomics. By integrating emerging machine-learning methods to predict and match prokaryotic transcription factors with their binding sites, we link phage gene content with host metabolic regulation. This framework enables interrogation of host physiological state during lysogeny and provides a path toward understanding how phage-encoded addiction cassettes reshape metabolic networks. Together, these approaches leverage recent advances in addiction cassette discovery and computational biology to characterize their prevalence, function, and ecological impact across diverse systems.

The role of plant signaling molecules in the tripartite interaction between plants, protists, and bacteria

Ravikumar Patel* (ravikumar.patel@ct.gov)
Connecticut Agricultural Experiment Station

Ravikumar Patel, Blaire Steven, Lindsay Triplett

Microbial communities in the rhizosphere play essential roles in plant health and ecosystem function, yet the roles of heterotrophic protists, active regulators of microbial community structure, remain underexplored. Beyond simple predator–prey relationships, protists form stable associations with bacteria in the rhizosphere, but the mechanisms of these interactions are poorly understood. Understanding protist–bacteria interactions may therefore provide new insights into processes influencing plant health. We investigated bacterial associates of ten heterotrophic protists isolated from the maize rhizosphere using a combined metagenomic and culture-based approach. Metagenome-assembled genomes revealed that each protist harbors a distinct but relatively simple bacterial community enriched in genes associated with phytohormone biosynthesis. Among phytohormones, indole-3-acetic acid (IAA) was the most prevalent, and 68% of protist-associated bacteria produced IAA in vitro. We next tested whether this auxin hormone influences protist growth. Exogenous IAA significantly increased the growth of all ten protists. Transcriptomic analysis of the ciliate Colpoda sp. identified more than 1,700 genes that were differentially expressed in response to IAA, including homologs of proteins involved in auxin transport, metabolism, and cell-cycle regulation. These results provide evidence that auxin can regulate growth and gene expression in non-photosynthetic eukaryotes. Building on these findings, our JGI New Investigator project aims to define the molecular and metabolic mechanisms underlying tripartite signaling among plants, auxin-producing bacteria, and auxin-responsive protists. Using EcoFAB devices combined with metabolomics and metatranscriptomics, we will determine how plant-derived signals shape bacterial phytohormone production, protist responses, and plant gene expression in controlled rhizosphere systems.

Two Decades of Environmental Microbiome (e-biome) Changes at a Superfund Site, 2006 – 2025 and the Fate of a Biodegrading Inoculum

Rebecca Reiss* (rebecca.reiss@nmt.edu)
New Mexico Tech

Rebecca Reiss (rebecca.reiss@nmt.edu) (New Mexico Tech, Socorro NM, USA), Peter Guerra, Peter Guerra (NM Bureau of Geology & Mineral Resources, Socorro, NM, USA)

The North Railroad Avenue Plume (NRAP) iss the result of over 30 years of leakage of tetrachloroethene (PCE) from a dry-cleaning facility into the sole-source potable water aquifer for the city of Española New Mexico, the Santa Clara Pueblo, and surrounding rural populations. The toxic plume was discovered in 1989 and extended 1.2 km from the source and reached a depth of 80 m below the ground surface. NRAP was declared an EPA Superfund Site in 1999. In 2006 quantitative PCR results indicated that indigenous chlorinated solvent-degrading microbes were present making biostimulation to encourage enhanced reductive dechlorination (ERD) the remedy of choice. We previously reported on short-read whole-metagenome sequencing (WMS) and 16S amplicon data to characterize shifts in the e-biome from baseline, 0.5, 2.0 and 3.5 years after remedy application. As of 2020, over 90% of the PCE was eliminated in the shallow aquifer. However, the significant levels of contamination remaining in the deep zone necessitated the application of bioaugmentation. In 2024 a commercial biodegrading inoculum was injected into the aquifer. E-biome samples were collected in 2023, 2024 (post-injection) and 2025, 18 to 20 years since remediation began. WMS data from baseline to 20 years as well as the inoculum were run using the latest workflows and reference sequence databases. The two-decade timeline of e-biome changes includes tracking the non-indigenous microbes used for bioaugmentation and determining their impact on chloroethene biodegradation. Efforts are underway to apply proximity-ligation (Hi-C) to the 2025 e-biome and compare the results to WMS.

Microbial Community, Activity, and Decomposition Processes are Influenced by Temperature and Moisture Interactions in Peatland Soil

Regan McDearis* (mcdearisrw@ornl.gov)
Oak Ridge National Lab

Regan McDearis, Christopher W Schadt, Melanie A Mayes, Joshua Birkebak, Natalie Griffiths, Randy K Kolka, Matthew Craig, Jeff Chanton, and Beth Holmes

Northern peatlands play a vital role in the global carbon cycle due to extensive organic carbon storage, but are experiencing increasing pressures from warmer temperatures, which may significantly impact carbon cycling processes. Within the ORNL SPRUCE experiment, warming treatments aimed at understanding these effects have also resulted in increased surface drying, complicating efforts to understand their individual effects on soil processes. To address this, we conducted a lab-based incubation to disentangle the effects of temperature and moisture on peat decomposition.

Peat from 20–30 cm depths was incubated factorially for 90 days under three temperature treatments (14ºC, 18.5ºC, 23ºC) and three moisture levels (95% VWC, 75% VWC, 55% VWC). Carbon dioxide and methane fluxes were continuously monitored throughout the experiment, revealing a significant interactive effect between temperature and moisture treatments. Warmer conditions increased CO2 respiration rates, but the highest CO2 flux was in the 75% VWC treatment. Methane flux was also strongly influenced by treatment conditions: samples incubated at higher temperatures initially produced CH4, which was subsequently consumed, implying a shift in microbial activity as the moisture in the jars decreased.
We are currently performing DNA- and RNA-based 16S rRNA amplicon sequencing and Fourier-Transform Infrared (FTIR) spectroscopy to assess microbial community composition, identify active populations, and characterize warming- and drying-induced changes in peat carbon compounds and metabolites. Finally we would like to work toward the incorporation of metagenomic and metatranscriptomic analyses, which would enable more in-depth understanding of the microbial mechanisms driving the observed gas flux and decomposition changes.

Aerobiome from a central grassland is shaped by time of day

Rocio Rodriguez* (rocio.rodriguezdelasllagas@colostate.edu)
Colorado State University

Rocio Rodriguez, Avinash Dhar, Marina Nieto-Caballero, Beth Hayes, Kristen Otto, Thomas Hill, Paul DeMott, Lily Jones, Ashley Milller, Jane E. Stewart, Noelle Bryan, Mark T. Hernandez, Sonia Kreidenweis, Jan E. Leach, Pankaj Trivedi

The aerobiome is fundamental to environmental and human health, influencing cloud and ice formation and pathogen transmission processes. Despite significant advancements in understanding other environmental microbiomes, how air microbial communities are composed and structured and what drivers affect these still need to be better understood. Moreover, we need a more comprehensive understanding of the aerobiome in natural environments such as grasslands that are particularly impacted by climate change. In this study, we explored the diurnal and vertical variability of the composition and diversity of the aerobiome from a central midwestern grassland site. We collected air samples from the NEON tower located at the Central Plains Experimental Range (CPER) from two different heights and two 12-hour high-volume filter samples per day (AM and PM) for 33 days in Spring 2023 (May-June). We profiled bacterial and fungal communities by amplicon sequencing using Earth Microbiome Primers (EMP). Our results showed non-significant impacts of sampling height on the diversity and composition of the aerobiome. Ordination analysis showed a clear separation between the daytime and overnight samples for bacteria and fungi, indicating the strong impact of time of day on the aerobiome community composition. The diversity of the overnight samples was significantly ($p = 0.001$) higher than the daytime samples. This research provides valuable insights into the factors shaping air microbial communities in natural areas such as grasslands and exposes the need to further investigate this field.

Uncovering the gut microbial biosynthesis of human health related metabolite delta-valerobetaine

Sam Syberg* (samsyber@uw.edu)
University of Washington

Sam Syberg, Timothy James Hall, Nolan Clayton, Lauren Rajakovich

Delta-valerobetaine (dVB) is a gut microbial metabolite with multiple, yet contradictory, links to human health. It has strong associations with various metabolic diseases, including obesity and liver disease, as well as cognitive decline. Conversely, it is also reported to have health benefits, such as promoting neural development and improving gut barrier function. However, these correlations either lack evidence of causality or rely on direct administration of dVB rather than in situ generation to establish causal effects. A better understanding of this metabolite and its production by the gut microbiota is thus needed. We have addressed major gaps in knowledge of microbial dVB biosynthesis in the human gut. Through screening a panel of gut bacterial strains and isolation of colonies from fecal samples, we discovered a select group of obligate anaerobes from the Lachnospiraceae family of Clostridia capable of synthesizing dVB. We then elucidated the three-step biosynthetic pathway originating from trimethyllysine (TML) and identified a small molecule targeting a key enzyme that can modulate dVB production. Using comparative genomics and transcriptomics, we identified the gene cluster that encodes this metabolism and showed that it serves as a biomarker for dVB production. We are now using these biomarkers to evaluate multi-omic datasets in order to assess correlations with human disease. Lastly, we demonstrated that germ-free mice monocolonized with a dVB-producing strain have increased levels of serum and liver dVB. With this new knowledge of microbial dVB biosynthesis, we are now investigating the physiological effects of this microbial metabolism on host biology.

Evaluating absolute metagenome quantification using synthetic spike-in controls

Samuel Koehler* (sikoehler@lanl.gov)
Los Alamos National Laboratory

Samuel I. Koehler, Erik R. Hanschen, Jennifer T. Pentz

Metagenomic analyses commonly use relative abundance proportions that identify some important microbial community properties but fail to capture quantitative absolute abundance that is necessary for many important applications like pathogen monitoring, microbial therapeutics, and microbiome profiling. Here we share ongoing work developing methods using Sequins spike-in DNA standards to calculate absolute abundance of low-input de novo assembled metagenomes. A first major challenge in quantifying metagenome samples is identifying what species are present, especially low abundance taxa. We have developed a metagenome pipeline that uses standard tools to perform read quality control, exclude Sequins sequences, assemble contigs, bin and classify contigs, merge similar bins, and download reference genomes for identified species. Co-assembly and read-base classification methods are currently being developed to improve performance of low abundance samples and rare taxa. Metagenomic quantification of per-species input cell counts is achieved through a second pipeline using linear regression to model the relationship between known sequins input quantity and resulting sequencing coverage. To date, two methodological experiments have been performed, yielding promising results. Our first experiment used 90 commercial ZymoBIOMICS 7-species metagenomes to explore input quantity thresholds, successfully quantifying Sequins spike-ins to predict input species absolute cell counts. Our second experiment used 30 custom in vitro 9-species metagenomes of known cell counts to ground-truth validate absolute quantification modeling. Future work will use real environmental metagenome samples to explore the minimum sample input required for detection.

From tropics to tundra: expanding microbiome analytics across saturated soils

Sophie Jurgensen* (sophie.jurgensen@colostate.edu)
Colorado State University

Sophie K. Jurgensen, Jamie Irby, Alexandra M. Wettengel, P. Summer Turner, Laura Mason, Jared B. Ellenbogen, Jason Shabaga, Alan Gorchov-Negron, Else Radeloff, Elisha M. Wood-Charlson, Tom Douglas, Tamara Harms, Merritt Turetsky, Mikayla A. Borton, Kelly C. Wrighton

Saturated soils, defined by rich organic matter and low porewater oxygen concentrations, are vulnerable to anoxic microbial decomposition of stored carbon resulting in greenhouse gas production. Despite this climate relevance, the underlying microbial mechanisms and drivers of carbon decomposition across ecosystems remain poorly resolved. To address this knowledge gap, we leveraged a JGI Community Science Program (CSP) award to create the Multi-omics for Understanding Climate Change (MUCC) database. This continuously growing resource contains identity, distribution, and functional information from over 26,000 microbial genomes paired to gas fluxes and geochemistry from freshwater wetlands across the United States. Here we to expand MUCC to include international sites spanning tropical to arctic ecosystems. Using a distributed science approach, our team has collected samples from rice fields, alpine peatlands, saltwater wetlands and permafrost soils in the United States, Colombia and Sweden for metagenomic and metatranscriptomic sequencing. A significant advantage of the highly-curated MUCC database is it can elevate amplicon-based studies. In one project focused on microbial mechanisms underpinning permafrost thaw in Alaska, we linked 16S rRNA amplicons to microbial genomes in MUCC. This revealed the importance of iron redox metabolisms in soils experiencing abrupt permafrost thaw, and informed sample choice for metagenomic sequencing through this CSP. All genomes, annotations, metadata, and workflows are being released through open community resources including KBase to enable reuse, reproducibility, and collaboration. Ultimately, this work provides a comprehensive perspective on the microorganisms and metabolisms governing carbon cycling across saturated soils, facilitating their translation and integration into climate-scale models.

Mobile genetic elements as an engine of adaptation in the plant microbiome

Stephanie Porter* (stephanie.porter@wsu.edu)
Washington State University, Vancouver

In host-associated microbiomes, horizontal transfer of mobile genetic elements (MGEs) is considered rampant and critical for adaptation. Here, adaptive loci, ranging from heavy metal-resistance to pathogen- or symbiont-infectivity, may move between genomes to expand microbial function, without discernible shifts in taxonomic composition. However, MGEs can be costly and pursue selfish fitness interests, which may constrain microbial adaptation. To connect theoretical dynamics to empirical patterns of MGE transmission and function in the plant microbiome, we compare two adaptive MGEs in rhizobia bacteria: 1) symbiosis islands (SIs), which enable formation of nitrogen-fixing root nodules on a legume, and 2) nickel-resistance islands (NRIs), which confer resistance to toxic levels of this critical mineral. Across 300 strains from wild microbiomes, we find surprisingly congruent phylogenetic histories for these MGEs and their host chromosomes, despite their potential to transmit horizontally. Consistent with selection for co-adapted MGE-chromosome combinations, historic MGE transfer has been largely restricted to closely related strains, especially for the larger SI. However, rare MGE transmission among divergent strains has played a key role in niche expansion over deeper timescales. Subsequent mating experiments, in which these MGEs may transfer among host genomes experimentally, show that while transmission of these selfish elements is promiscuous, MGE function is higher if donor and recipient genomes are more closely related. By providing genomic and functional evidence of how MGE transfer both potentiates and constrains adaptation in plant microbiomes, we shed light on the evolvability of microbial traits and the potential to use MGEs to engineer beneficial symbioses.

Competitors, Collaborators, and Frenemies: Elucidating the Mechanisms Governing Microbes Divide Labor in Polysaccharide Degradation

Stephen Lindemann* (lindemann@purdue.edu)
Perdue University

Toko Hisano, Marcelo Guerrero, Victoria Gutierrez, Sajal Bhattarai, Jack Russo, LaRae Hudson, Aaron Wright, Steve Lindemann

Although much is known about how microbes compete for simple substrates (i.e., those that can be directly transported through the cell envelope, such as tryptophan)[1], correspondingly little is known about the ecology governing how they compete for complex ones (i.e., those that need to be externally processed before being imported into the cell). Because the consumption of complex substrates requires external activity, such as hydrolysis, community degradation of these substrates inherently requires resource allocation (typically, synthesis of enzymes) outside the cell, where they potentially become "public goods" from which other members may benefit[2]. Therefore, growth on complex substrates (either soluble, such as a polysaccharide, or insoluble, such as lignocellulosic biomass (LCB)) invokes fitness trade-offs to these resource allocation decisions, which generates a complex web of potential interactions among microbes that can be either cooperative, competitive, or a mix of the two. Using anaerobic communities fermenting sorghum hemicellulose and aerobic communities degrading cyanobacterial exopolysaccharide, we reveal that microbial carbohydrate-active enzyme-encoding gene expression is neighbor-dependent. Further, we demonstrate that even low-abundance partners alter the feeding behavior of their more dominant neighbors and support increases in community productivity. Finally, cooperative and competitive interactions respond to phylogenetic distance among partners, suggesting the hypothesis that microbes may be "hard-wired" for interactions with specific (typically distant) taxa in degrading consortia. We are presently working to elucidate the mechanisms by which these microbial interactions arise, with the goal of developing theory to describe ecological interactions in complex substrate degradation.

1. Hansen, S. R. & Hubbell, S. P. Single-nutrient microbial competition: qualitative agreement between experimental and theoretically forecast outcomes. *Science* **207**, 1491–1493 (1980). 2. Konopka, A., Lindemann, S. & Fredrickson, J. Dynamics in microbial communities: unraveling mechanisms to identify principles. *ISME J* **9**, 1488–1495 (2015).

4-Hydroxybenzoic Acid, an Algal Exometabolite, Impacts Algal-Bacterial Interactions and Carbon Flow

Vanessa Brisson* (brisson2@llnl.gov)
Lawrence Livermore National Laboratory

Vanessa Brisson, Amrit S. Johal, Felizardo Salazar, Kristina Rolison, W. Camden McIntire, Nathan McCall, Peter K. Weber, Xavier Mayali, Rhona K. Stuart

Algal microbiome interactions have significant implications for global biogeochemical cycling and industrial algal processes, and algal and bacterial exometabolites are important mediators of these interactions. We previously found that 4-hydroxybenzoic acid (4-HBA), an algal exometabolite, influences the microbiome composition of the phycosphere, the region surrounding algal cells. Here we used 13C labeled 4-HBA and high-resolution imaging mass spectrometry (NanoSIMS) to investigate this metabolite's role in interactions between the model diatom Phaeodactylum tricornutum and four representative bacterial isolates from its phycosphere microbiome: Devosia B7WZ, Thalassospira 13M, Marinobacter 3-2, and Alcanivorax EA2. We found that 4-HBA acted as a selective carbon substrate for the growth of some bacterial strains (Devosia B7WZ and Thalassospira 13M1) but not others (Marinobacter 3-2 and Alcanivorax EA2). The level of 13C incorporation from 4-HBA by bacterial cells depended not only on the bacterial strain, but also on the presence/absence of the algal partner and/or algal exudates. Interestingly, crossfeeding of 13C from 4-HBA to P. tricornutum was observed with some bacterial partners, including one (Marinobacter 3-2) that did not use 4-HBA as its own carbon substrate. We have ongoing efforts to characterize the mechanism behind this carbon exchange, which may involve extracellular particles observed in the P. tricornutum-Marinobacter co-cultures. This research highlights distinct influences of a P. tricornutum exometabolite on bacterial taxa in its microbiome. By linking this exchange directly to carbon flow, we can improve predictions of factors governing algal productivity in these systems.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL microBiospheres SFA, FWP SCW1039. We thank the Livermore Lab Foundation for supporting A.S. Johal's and F. Salazar's work on this project through the LLF undergraduate fellowship program.

Thermal stress reshapes the chemical interface between seagrasses and their microbiomes

Vivian Rojas* (vivianrojas@ucmerced.edu)
UC Merced

Vivian K. Rojas, Gina M. Chaput, Karolina L. Zabinski, Diane-Marie Branche-Smith, Andres M Caraballo-Rodriguez, Jonathan A. Eisen, John J. Stachowicz, E. Maggie Sogin

Seagrasses rely on their associated microbial communities to sustain metabolism in coastal environments that are nutrient-limited. Unfortunately, rising ocean temperatures threaten seagrasses by imposing physiological stress that alters carbon allocation. It is unclear how temperature reorganizes the plant's metabolome and how these changes affect the bioavailability of plant-released organic matter (OM) that supports microbial growth. Here, we hypothesized that warming induces compartment-specific reorganization of seagrass metabolomes, altering root exudate composition and microbial access to plant-produced carbon. To test our hypothesis, we applied a shoot-to-root analytical framework using untargeted metabolomic analysis to quantify temperature
driven changes in seagrass metabolite composition under ambient and marine heatwaves (MHW) conditions. MHWs reduced metabolite diversity in shoots, rhizomes, and sediments, while diversity increased in roots and root exudates. The shoot metabolome was enriched in lipids and reduced carbon metabolites relative to ambient temperatures. This shift indicated that the plants underwent structural and physiological reorganization in aboveground tissues. In contrast, root exudates were enriched in oxidized carbon and nitrogen-rich metabolites, with shifts in energetic state and elemental stoichiometry compared to controls. Metabolites enriched in the root exudates in plants held at elevated temperatures include those involved in dampening inflammatory responses and mitigating oxidative stress. These changes suggest reduced microbial use of energetically favorable dissolved OM and restructuring of the root-associated microbiome during plant stress responses. Together, these findings demonstrate that MHW drives divergent above- and below-ground metabolic strategies in seagrasses, reshaping plant-microbe chemical exchange with implications for microbial function, nutrient cycling, and carbon feedbacks under thermal stress.

Discovery and Characterization of Critical-Metal Binding Small Molecules and Proteins

William Kew* (william.kew@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

William Kew, Paul Piehowski, Alex Beliaev, James Evans, Scott Baker, Christopher Anderton, Erin Bredeweg, Rene Boiteau, Kristin Burnum-Johnson

The "Metal-Binding Biomolecules" Campaign brings together academic and National Laboratory researchers in coordinated individual and multi-PI projects focused on shared scientific themes. In its inaugural year, the Campaign engaged the BER-funded CMM research community to identify pressing scientific knowledge gaps and opportunities where EMSL can enable new scientific outcomes for our Users. The inaugural year focuses on the theme of metal selectivity across scales and how biomolecules, individual cells, and consortia, selectively bind, transport, and utilize critical metals across diverse environments. EMSL's high-throughput proteomics, lipidomics, and metabolomics capabilities will be used for phenotypic characterization of these biological systems. These capabilities are extended using EMSL's recently upgraded ultra-high-resolution hybrid Orbitrap–21 Tesla FTICR mass spectrometer for discovery and characterization of metal-binding metabolites. Native-like online liquid chromatography coupled to 21T FTICR MS enables direct detection of intact metal-binding metabolites, with parallel ion fragmentation supporting structural characterization directly from complex samples. Complementary native top-down proteomics using high-field Orbitrap mass spectrometry provides critical data related to metalloprotein structure and metal-binding properties. Augmenting these capabilities are magnetic resonance spectroscopy, optical and electron microscopy, and nanoSIMS, which will provide additional molecular and spatial insights into metal uptake and accumulation by biological systems. Initial data will be presented from fungal (Trichoderma harzianum) and bacterial (Pseudomonas putida) systems grown under metal-limited and metal-replete conditions.

Elucidating the microbiome dynamics on Sorghum leaves under different watering regimes

Yorgos Kepesidis* (gkepesi@sandia.gov)
Sandia National Labs

Yorgos Kepesidis, Daniella Victoria Martinez, Christopher M De Ben, Daniel H Putnam, Kimberly Butler, Jeri Timlin, Raga Krishnakumar

Understanding the microbial communities on plant leaves is essential for enhancing crop fitness, particularly in relation to stress adaptation and nutrient uptake. This research is pivotal for developing predictive and detection tools in agriculture. Sorghum, known for its high drought tolerance, serves as an ideal model for studying these dynamics, enabling sustainable field studies under varying water conditions. While much of the existing literature focuses on the microbial communities in Sorghum's rhizosphere and soil, there is a significant gap in our knowledge regarding the epiphytic (leaf surface) and endophytic (leaf interior) communities. Just as their rhizosphere counterparts, these leaf-associated microbes play a crucial role in plant health and resilience, making their study vital for effective detection strategies, especially given their accessibility for optical detection.

In this poster, we will share the results of a comprehensive comparative metagenomics analysis of the epiphytic community on Sorghum leaves from three commercially significant strains, grown under different water availability conditions. Our findings reveal the environmental influences on these microbial communities and provide insights into the physiological interactions between microbes and plants. The bacterial genus of Bacilliaceae, as well as the fungal genera of Cladosporium and Vishniacozyma are found to be predominant in the epiphytic phylosphere of all three species and their presence is linked with the growth conditions (drought, heat, humidity) and the plants' physiological responses to those. Ultimately, we aim to utilize this data for advanced spectroscopy applications for the timely detection and characterization of biological and chemical responses of plants to environmental variability.

Genome-resolved expansion of giant viruses reveals new diversity and functional potential.

Yumary Vasquez* (ymvasquez@lbl.gov)
Joint Genome Institute (JGI)

Yumary M. Vasquez, Tiago Nardi, Tanja Woyke, Frederik Schulz

Nucleocytoplasmic large DNA viruses and viruses of the newly proposed Duplodnaviria phylum, Mirusviricota, exhibit taxonomic richness which continues to expand due to metagenomic sequencing of Earth's biomes. Giant viruses contain complex genomes encoding genes of both viral and cellular origin, representing a reservoir of unexplored biological functions with potential implications for ecology, evolution, and biotechnology. Here, we present the largest curated database of giant virus metagenome-assembled genomes (GVMAGs V2), comprising 8,508 species-level clusters inferred from 18,727 genomes, originating from marine, freshwater, anthropogenic and terrestrial environments. Phylogenomic analysis revealed 712 novel genera, 13 previously unknown viral families and a new proposed order, tentatively named Mycodnavirales. We improved gene calling of 12% of giant virus genomes by accounting for alternative and custom genetic codes, enabling more accurate identification of protein-coding genes. Orthologous clustering of 2.5 million proteins identified 135,998 orthogroups representing comprehensive metabolic capabilities, such as enrichment of genes involved in aromatic compound degradation (commonly associated with bioremediation) in Algavirales genomes. Furthermore, we detected widespread biosynthetic gene clusters underpinning antimicrobial-like activity and antibiotic resistance-like activity, suggesting roles of giant viruses in host defense. Conversely, 67% of orthogroups have unknown functions, underscoring a substantial unexplored potential. This comprehensive publicly available database provides a critical resource for the giant virus research community and a foundation for expanding giant virus diversity, uncovering virus-host interactions and, exploring viral evolution, and identifying reservoirs for novel enzymes with the potential to advance biotechnological applications.

Session 2

Genomic Ingenuity: Crafting the Future of Biodesign

Posters alphabetical by first name of presenting author*

A Modular Promoter Engineering Framework for Nonmodel Microbes

Amanda Robert* (amro5809@uw.edu)
University of Washington

Amanda M. Robert, Michael Guzman, Diego Alba Burbano, Jackson Comes, Nam Pham, Yasuo Yoshikuni, James M. Carothers

Fossil fuel–based chemical manufacturing drives greenhouse gas emissions and environmentally harmful waste streams, motivating the development of sustainable, carbon-efficient alternatives. Microbial biomanufacturing offers a self-renewing platform to convert waste carbon and atmospheric CO_2 into valuable products, but its adoption is limited by the genetic intractability of many metabolically promising nonmodel organisms. Purple nonsulfur bacteria (PNSB) are attractive engineering chassis due to their metabolic flexibility and diverse growth modes. However, the regulatory genetic parts required to tune gene expression for bioproduction, particularly promoters, remain poorly characterized in nonmodel species. Here, we developed a modular promoter engineering workflow for rapid construction and screening of regulatory elements using conjugative genomic integration. Leveraging the CRAGE-duet genome integration platform, we screened over 180 promoter variants in the marine nonmodel PNSB Rhodovulum sulfidophilum, overcoming transformation barriers commonly encountered in alphaproteobacteria. To further extend the scope of this approach, we performed cross-species screening of the Anderson promoter collection across multiple PNSB, revealing substantial variability in promoter activity between hosts and underscoring the need for portable, host-agnostic regulatory part characterization. Our workflow identified promoters spanning a >60-fold dynamic range that enable tunable and stable gene expression from a single-copy genomic context, bypassing plasmid instability and more accurately reflecting native transcriptional control. Together, this workflow and promoter toolkit provide a foundation for scalable biodesign in PNSB, enabling rational pathway tuning and expanding the genetic accessibility of nonmodel microbial chassis for sustainable bioproduction of industrially relevant biocommodities.

Automating the Domestication of Non-Model Bacteria

Andrea Garza Elizondo* (garzaelizoam@ornl.gov)
Oak Ridge National Laboratory

Andrea Garza Elizondo, Jiwoo Kim, Michael Melesse Vergara, William Alexander, Adam Guss, Carrie Eckert

Modern genomics has shown us that nature is rife with diverse bacterial phenotypes that can potentially be leveraged for a variety of applications - biodegradation of wastes, synthesis of petrochemicals and pharmaceuticals, and accumulation of valuable elements. Replicating these phenotypes in our model organisms isn't always feasible or preferred, thus, as our tools improve, engineering non-model bacteria is becoming a better strategy. However, the bulk of genetic parts, editors, and other synthetic biology technologies were designed in and for model organisms, necessitating novel tool development, and complex phenotypes aren't always a result of a single bacteria, requiring solutions that are translatable. Here, I present how we've used physical automation to accelerate the domestication of non-model bacteria – specifically, in isolates from a plastic-fed mealworm's gut microbiome. The semi-automation of electroporation protocol optimization is shown, a task that is usually laborious. Furthermore, a modular genetic element library designed for rapid prototyping through automation is presented, allowing quick discovery of functional genetic parts. Overall, I discuss how our workflows are facilitating our ability to both study and engineer the bacteria's potential for plastic degradation.

Using comparative genomics to identify and functionally validate conserved cis-regulatory elements in grass

Bliss Beernink* (bmbeernink@lbl.gov)
Joint Genome Institute (JGI)

Bliss M. Beernink[1] , Ronan O'Malley[1,2] , Dan Rokhsar[1,3], Axel Visel[1] , John Vogel[1] , Li Lei[1,4]
[1] U.S. Department of Energy Joint Genome Institute, Lawerence Berkeley National Laboratory, Berkeley, CA,
USA
[2] The University of Chicago, Chicago, IL, USA
[3] University of California Berkeley, Berkeley, CA, USA
[4] Syngenta, Greensboro, NC, USA

Gene expression changes, governed by cis-regulatory elements (CREs) including enhancers, are a major source of plant phenotypic variation, and understanding the mechanisms of gene expression changes is critical for domesticating and improving grasses as biomass crops. Enhancers are regions of noncoding DNA that regulate transcription and tend to be evolutionarily conserved due to their critical function in orchestrating gene expression. While comparative genomics approaches have revealed thousands of putative enhancers in mammals, far fewer have been identified and functionally characterized in plants whose large complex genomes pose greater challenges.

Here, we propose integrating comparative genomics and gene expression approaches to identify conserved enhancers in the genomes of DOE flagship grasses. Polyploid plant genomes offer a unique opportunity to leverage intra-species comparisons across ancestral subgenomes and their homeologous genes to identify functionally relevant mutations affecting enhancers. Leveraging comparative genomics, we identified regions of conserved genomes among a phylogenetically diverse set of nine diploid grasses. To further refine this list, we identified homeologous genes associated with the ranked CRE elements in the switchgrass and Zea mays genomes to identify cases where the deletion of a CNS element is correlated with a change in expression pattern. To validate this CRE identification approach, we will functionally validate up to a dozen enhancers via transgenic Brachypodium distachyon.

Scalable CRISPR Gene Regulation and Genome-Wide Programming in Pseudomonas putida

Brian Darst* (bhdarst@uw.edu)
University of Washington

Brian H. Darst, Cholpisit Kiattisewee, Tommy Primo, Kira Olander, Ian D. Faulkner, Yejun Kim, Ryan Cardiff, Allan Scott, Joshua Elmore, Hector Garcia Martin, Jacob Rapp, Patrick C. Kinnunen, Tijana Radivojevic, Jesse G. Zalatan, James M. Carothers

Synthetic sensing, memory, and dynamic control have transformed eukaryotic systems by enabling systems-level analysis and programmable regulation of multi-gene networks; here, we extend these capabilities into a bacterial chassis optimized for industrial biomanufacturing. Pseudomonas putida offers a metabolically versatile platform, but realizing its full potential requires genetic programs that are scalable, multiplexable, and genetically durable. We develop next-generation CRISPR activation and inhibition (CRISPRa/i) architectures that expand the size and complexity of multi-gene programs deployable in P. putida, supporting coordinated control of metabolic pathways at unprecedented scale. Through iterative design-build-test-learn (DBTL) cycles, we construct large, multi-node CRISPRa/i programs and streamline their genomic integration using an enhanced serine-recombinase toolkit. Looking ahead, we are integrating CRISPRa/i outputs with next-generation biosensor circuits and programmable base editors to create input-responsive, memory-enabled control layers that emulate the dynamic regulatory capabilities demonstrated in advanced eukaryotic systems. These modules are poised to enable adaptive metabolic control in P. putida and further expand the size of genetic programs that can be engineered. As these capabilities scale, we are building expansive strain libraries and generating high-dimensional datasets suitable for AI/ML-driven discovery. Such datasets will support the development of predictive models of bacterial metabolism and regulation, accelerating the transition toward truly programmable biosystems design.

Expanding the engineering and discovery toolkit for Methanosarcina acetivorans

Darla Brennan* (dbrennan2@nebraska.edu)
University of Nebraska – Lincoln

Darla Brennan, Connor Hines, Dirk Anderson, Nicole Buan

This page intentionally left blank

Identification of Functional Promoters in Anaerobic Gut Fungi

Elaine Kirschke* (ekirschke@ucsb.edu)
University of California, Santa Barbara

Elaine Kirschke, Sarah Seagrave, Leo Baumgart, Ian Blaby, and Michelle O'Malley

Lignocellulose is the most abundant renewable carbon source on Earth, yet its recalcitrance remains a major bottleneck in biofuel and bioproduct production. Anaerobic gut fungi (AGF; phylum Neocallimastigomycota) initiate lignocellulose deconstruction in herbivore gut microbiomes under anaerobic conditions. Although they represent a minority of the community, AGF encode expansive repertoires of carbohydrate-active enzymes with broad substrate specificity, making them promising platforms for lignocellulose bioprocessing. However, functional characterization and engineering of AGF remain limited by genetic intractability. Highly AT-rich and repetitive genomes, horizontal gene transfer, complex life cycles, and transformation barriers have hindered molecular tool development, and robust promoters for heterologous expression are lacking.

To address this gap, we are pursuing two complementary strategies leveraging JGI construct generation and sequencing capabilities. First, we are developing a high-throughput promoter screening pipeline targeting ~200 candidate promoters identified from upstream regions of highly expressed AGF genes across transcriptomic datasets. JGI-generated promoter–reporter constructs will be evaluated in Saccharomyces cerevisiae and AGF using fluorescent reporters to quantify promoter strength across heterologous and native contexts.

Second, we are applying DAP-seq to characterize AGF transcription factors and identify high-confidence binding motifs within candidate promoter regions. Of 176 TFs analyzed (92 from Caecomyces churrovis and 84 from Neocallimastix californiae), ~16% passed quality thresholds (FRiP > 0.5), with six datasets exhibiting strong enrichment (>100 peaks per genome). While not sufficient for regulatory network reconstruction, these data enable motif discovery and prioritization of candidate promoters for functional validation. Together, these efforts establish a promoter toolkit to enable AGF genetic engineering for bioenergy and biomanufacturing applications.

Monitoring lytic phage host range in complex communities using an autonomous RNA barcoder

Elizabeth Zeng* (ez30@rice.edu)
Rice University

Elizabeth J. Zeng, Zach LaTurner, Jonathan (Joff) Silberg, and Lauren B. Stadler

Bacteriophage play a crucial role in the environment by regulating microbial behaviors and mediating the flow of genetic information across microbial communities. Their highly specific and diverse host range enable precise perturbations of complex communities, representing an underexplored approach for scalable microbiome manipulations. However, our understanding of lytic phage host range and transduction rules remains limited, due to the lack of tools for rapid measurement of community-level infection. RNA-addressable modification (RAM) is a new tool that is poised to overcome these challenges by enabling phage host range detection directly within microbial communities. With RAM, a synthetic ribozyme is coded into phage genomes, which splices a barcode onto the host 16S ribosomal RNA (rRNA) upon gene transfer, allowing host identification via targeted sequencing. We will describe how we encoded RAM into the T7 genome and showed that RAM yields a detectable signal in diverse gram-negative bacteria infected by T7 phage, including both lysing and non-lysing infections. Next, we will describe the application of barcoded T7 to a wastewater community and report the detection of over one hundred unique T7 ASV hosts, including Enterobacterales and Aeromonadales. By simplifying lytic phage host-range analysis, RAM converts phage–host mapping into a scalable, automation-compatible assay suitable for high-throughput data generation and computational analysis. This approach enables systematic exploration of how variables such as phage-phage interactions, community composition, and environmental conditions influence transduction in complex microbiomes, addressing fundamental questions about phage host range and guiding lytic phage engineering for therapeutics and microbial community programming.

DBTL cycles for engineering aromatic bioproduction in Pseudomonas putida with multi-node CRISPRa/i gene regulatory programs

Ian Faulkner* (idf@uw.edu)
University of Washington

Ian Faulkner, Cholpisit Kiattisewee, Brian Darst, Tommy G. Primo, Allan Scott, Tijana Radivojevic, Patrick C. Kinnunen, Jacob Rapp, Alex Beliaev, Hector Garcia Martin, Jesse G. Zalatan, and James M. Carothers

Design and improvement of microbial strains engineered for bioproduction can be accelerated by combining tools, like CRISPR gene regulation and multi-omics data analysis, with strategies like iterative incorporation of machine learning. CRISPR systems are well-suited to controlling the broad circuits of multi-gene expression programs, with part of the circuit controlling heterologous pathway expression and the rest of the circuit redirecting endogenous metabolic flux toward that pathway. While we have recently profiled bioproduction within a three-dimensional design space of enzyme stoichiometry, larger circuits encompassing endogenous metabolic perturbations yield impractically numerous design combinations. Here, we expand the gRNA array to include a number of perturbations to central metabolism and amino acid anabolism. These perturbations are intended to boost metabolic flux toward 4-aminocinnamic acid, a chorismite-derived precursor for advanced polymer, but it is difficult to rationally predict the most effective combinations of perturbation targets. To address this design problem, we are developing an iterative design-build-test-learn framework using multi-omics data analysis and ML-directed target recommendations to increase 4-ACA production through incremental addition of new perturbations and exploration of perturbation combinations. We find wide production variation across combinations, with modest improvements above baseline only when multiple endogenous genes are perturbed. We also suggest an upper limit of control circuit size when the circuit is fully controlled by a single CRISPR system. By developing new data-driven, model-guided approaches to strain optimization with accelerated DBTL cycles, we uncover productive multi-guide CRISPRa/i programs with endogenous and heterologous targets, and offer a broadly applicable route to early-stage strain development.

Probing Metabolic Division of Labor in Phototrophic Bacteria Using MicroSPLiT

Jackson Comes* (jscomes@uw.edu)
University of Washington

Jackson Comes, Margaret Cook, Yujia Huang, Jackie Haring, Anna Kuchina, James Carothers

Purple non-sulfur bacteria (PNSB) are phototrophic microbes often capable of metabolizing lignin-derived aromatic compounds, yet the transcriptional regulation and metabolic interactions underlying these pathways remain poorly understood at single-cell resolution. Understanding population-level heterogeneity and metabolic specialization is critical for both genome-to-function studies and for engineering PNSB as lignin-valorizing production hosts. Such heterogeneity may reflect metabolic interactions and division of labor that govern aromatic flux, coordinate photosynthetic and redox metabolism, and ultimately shape product formation.
We are developing MicroSPLiT-based single-cell transcriptomic workflows for PNSB to profile aromatic metabolism and engineered pathway responses across individual cells. Building on emerging split-pool frameworks such as MAP-SPLiT, we aim to couple single-cell profiling with planned CRISPRi perturbation targets to elucidate regulatory mechanisms controlling aromatic catabolism and carbon flux partitioning. By combining genetic integration strategies with ongoing optimization of permeabilization and barcode delivery, this platform aims to reveal metabolic subpopulations and interactions that drive or limit bioconversion. Ultimately, these efforts will establish new functional genomics capabilities for phototrophic non-model bacteria and enable more systematic metabolic engineering for sustainable chemical production.

mapSPLiT Enables Single-cell CRISPRa/i Screens for Predictive Functional Genomics in Bacteria at Scale

Jacob Brandner* (brandner@uw.edu)
University of Washington

Jacob R. Brandner, Quoc Tran, Yujia Huang, Dmitry Sutormin, Karl D. Gaisser, Georg Seelig, Jesse G. Zalatan, James M. Carothers, Anna Kuchina

Advancing microbial engineering requires scalable methods to probe gene function, map regulatory networks, and resolve genetic interactions across diverse bacterial hosts. CRISPR–Cas systems offer programmable perturbations compatible with high-throughput workflows, and pooled CRISPR screens paired with single-cell RNA sequencing (scRNA-seq) have transformed functional genomics in eukaryotes. However, sparse transcript capture in bacterial scRNA-seq has prevented analogous approaches, limiting the generation of structured, high-dimensional datasets needed for predictive design.

We introduce mapSPLiT (Microbial Analysis of Perturbations using Split-Pool Ligation Transcriptomics), the first platform to integrate CRISPR activation/interference (CRISPRa/i) with single-cell transcriptomics for pooled functional screens in bacteria. mapSPLiT links each perturbation to its transcriptional phenotype through a guide-barcode construct (GBC) co-captured with endogenous transcripts. A PCR-based enrichment strategy increases GBC detection and a sequencing-based rRNA depletion protocol improves recovery of informative reads, enabling pooled single-cell CRISPRa/i screens in prokaryotic systems.

In a single experiment, we profiled >100 perturbations targeting 52 known or putative E. coli transcription factors across 76,000 cells. This dataset revealed eight previously uncharacterized regulatory interactions and assigned functional roles to twelve regulators lacking prior expression data. Combinatorial perturbations further exposed how promoters integrate multiple regulatory inputs to shape transcriptional outcomes. We ported mapSPLiT to P. putida, uncovering new regulatory functions for the carbon-metabolism regulator gclR with implications for expanding substrate utilization.

By generating scalable, high-resolution perturbation–phenotype maps across model and non-model bacteria, mapSPLiT provides a foundation for mechanistic modeling and AI/ML-driven inference of regulatory logic, supporting the development of predictive, design-ready microbial systems.

Whole genome sequencing of Populus deltoides natural variants to enable genomic prediction and genome-wide associated studies

Jay Chen* (chenj@ornl.gov)
Oak Ridge National Laboratory

Jay Chen, Heidi Renninger, David Weston, Chris Schadt

Populus trichocarpa is the first tree species whose genome was fully sequenced. Furthermore, whole-genome resequencing of the P. trichocarpa natural variants has enabled genome-wide association studies (GWAS) that have led to numerous discoveries of genetic control of biomass production, cell wall chemical properties, and stress responses, revolutionizing tree genomics and biology research. Among the approximately a dozen Populus species in the United States, P. deltoides is the dominant species in the Southeastern region. Overwhelming field and laboratory evidence indicates that P. deltoides is physiologically and genetically distinct from P. trichocarpa, with species-specific genetic regulation of key traits influencing biomass quantity and quality, and plant-environment interactions. Due to its ecological prevalence and potential as a bioenergy feedstock, a deeper, species-specific understanding of P. deltoides population genetics is essential to support widespread deployment of Populus for sustainable bioenergy and biomaterial production across the southeastern U.S. To address this need, we propose whole-genome resequencing of 500 P. deltoides natural variants to enable genomic prediction and GWAS. These 500 individuals will be primarily collected from across the southeastern US to capture geographic, genetic and physiological diversities. The phenotypic diversity of this collection will be further evaluated under controlled conditions with high throughput phenotyping in the Advanced Plant Phenotyping Laboratory at Oak Ridge National Laboratory, and in common gardens at multiple locations. These data will enable genomic prediction and GWAS to identify genetic determinants of relevant bioenergy and bioproduct traits as well as traits associated with plant-environment interactions (including plant-microbe interactions) in P. deltoides.

Brachypodium distachyon Transformation, a New Capability for Plant Functional Genomics at JGI

John Vogel* (jpvogel@lbl.gov)
Joint Genome Institute (JGI)

Yezhang Ding, Yi Zhai, Tomáš Brůna, Sharon Greenblum, Li Lei, Leo A. Baumgart, Peng Wang, Vlastimil Novak, Mingqin Shao, Scott J. Lee, Samuel P. Hazen, Suzanne M. Kosina, Trent R. Northen, John P. Vogel

Transformation of many of the plant species DOE researchers study is challenging and beyond the capability of all but a few labs. This bottleneck has been recognized by DOE*. Hopefully, future investments might lead to a DOE Plant Transformation Capability that can satisfy the needs of the DOE research community. In the meantime, JGI has begun offering users access to efficient transformation and genome editing in the model grass Brachypodium distachyon with several pilot projects underway. This poster presents results from one such project that highlights the power of integrating plant transformation with existing JGI capabilities. Using a multi-omics approach that combined metabolomics, transcriptomics, population genomics, single-cell transcriptomics, and heterologous expression, we identified candidate genes involved in dopamine biosynthesis in plants. We then used plant transformation to create single and double knockouts of these putative biosynthetic genes. We discovered that Brachypodium roots produce and exude very high levels of dopamine. We further show that dopamine biosynthesis in roots and leaves is mediated by genetically redundant, tissue-specific enzymes that differ from the dominant enzymes used by animals. It is noteworthy that due to the tandem nature of the redundant genes, it is unlikely that traditional forward genetics or single-gene mutations could have been used to elucidate this pathway. This work underscores the potential of combining JGI capabilities with plant transformation to help users go from initial gene discovery to functional validation. Please contact us if you are interested in applying Brachypodium transformation and/or multi-omics analyses to your research.

*https://www.genomicscience.energy.gov/plant-transformation/

Integrating DAP-seq and iModulon analysis to map the transcriptional regulatory network of Streptomyces albidoflavus J1074 for improved strain engineering

Laia Meseguer-Monfort* (lamemo@dtu.dk)
BRIGHT, Technical University of Denmark

Laia Meseguer Monfort, Mathias Jönsson, Lei Yang, Emre Özdemir

Rational engineering of microbial chassis strains requires understanding their transcriptional regulatory networks. Streptomyces albidoflavus J1074 is commonly used for expressing biosynthetic gene clusters, yet comprehensive regulatory knowledge remains limited. Transcriptomic approaches such as iModulon analysis can identify co-regulated gene modules, but require extensive RNA-seq datasets and capture both direct and indirect regulatory effects. DNA affinity purification sequencing (DAP-seq) offers a complementary approach by identifying direct transcription factor (TF) binding sites in vitro, potentially providing a more cost-effective route to regulon definition.

We performed DAP-seq on 182 TFs from S. albidoflavus J1074, obtaining reliable binding profiles for 108 TFs. High-confidence target genes were defined as those with DAP-seq peaks in their promoter regions (-500 to +150 bp from the transcriptional start site) exceeding 5% of maximum peak intensity for each TF. Target sets were expanded to include co-transcribed operon members, reflecting prokaryotic transcriptional organization.

We compared these predicted regulons against published S. albidoflavus J1074 iModulons to assess correspondence between direct binding and transcriptional co-regulation.
Approximately 40% of TFs showed significant enrichment for genes within specific iModulons, indicating direct correspondence between binding and co-expression for this subset. Where GO and KEGG pathway enrichment was also observed, functional annotations were generally consistent with the associated iModulon, validating the biological coherence of DAP-seq-defined regulons.

Our results indicate that DAP-seq and iModulon analysis provide complementary views of the S. albidoflavus J1074 regulatory network. Integrating both approaches offers a practical framework for mapping transcriptional regulation in non-model industrial organisms, supporting more informed strain design for bioproduction applications.

Multi-omics of elongating internodes suggests multiple regulatory mechanisms for grass development and cell wall synthesis

Laura Bartley* (laura.bartley@wsu.edu)
Washington State University

Laura E. Bartley, Niharika N. Chandrakanth, Fan Lin, Duolin Wang, Skyler Kramer, Samuel Purvine, Steven Callister, Nicolas Gaitan, Matthew Mcgowan, Ruifeng He, Anna Lipzen, Vasanth Singan, Vivian Ng, Chris Daum, Yuko Yoshinaga, Kyle Palos, Dong Xu, Andrew Nelson, Jorge Duitama, Stephen Ficklin

Shifts in precipitation regimes are expected to impact microbial transformations of soil organic matter (SOM) in semi-arid California grasslands. While drought has been shown to alter plant rhizosphere (live root) and detritusphere (litter)-derived carbon transformations, the microbial mechanisms driving these processes across mixed habitats, where roots interact with decomposing litter, remain poorly understood. Here, we used stable isotope probing and multi-omics to examine how drought influences microbial activity and carbon flow across three experimental soil habitats in mesocosms: the rhizosphere, the detritusphere, and the combined rhizosphere–detritusphere. We tracked isotopically labeled carbon from plant material into mineral-associated organic matter (MAOM) and measured microbial gene transcript abundance (metatranscriptomics) and SOM chemical composition (FTICR-MS). We found that the presence of living roots suppressed detritus-derived MAOM formation while simultaneously driving microbial gene expression toward a rhizosphere-like functional state under ambient moisture conditions; however, under drought conditions the presence of living roots had a less pronounced impact on detritus-derived MAOM formation, gene expression, and SOM chemistry. Under ambient conditions, root presence reduced transcript abundances of plant polymer degradation genes while increasing expression of simple sugar metabolism genes. These functional shifts corresponded with enrichment of lignin-like compounds in the detritusphere, indicative of decomposed plant material, and condensed hydrocarbon-, protein-, lipid-, and amino sugar–like compounds in rhizosphere habitats, indicative of microbial biomass and root exudates. Together, these results highlight how interactions among soil moisture, habitat, and microbial ecophysiology regulate the pathways through which plant and microbially derived carbon contributes to SOM.

Conjugation-based Serine Recombinase-Assisted Genome Engineering (cSAGE) Enables Bioproduction in Non-model Bacteria

Michael Guzman* (msguzman@uw.edu)
University of Washington

Michael S. Guzman, Cholpisit Kiattisewee, Jackson Comes, Amanda M. Robert, Allan Scott, Ryan A.L. Cardiff, Diego Alba Burbano, Margaret Cook, Stella Anastasakis, Sarah Grube, Brian Darst, Jesse G. Zalatan, Joshua Elmore, Alex Beliaev, James M. Carothers

Non-model bacteria are promising platforms for bioproduction from waste-derived and C1 feedstocks, yet their limited genetic accessibility has constrained engineering efforts. Purple nonsulfur bacteria (PNSB), in particular, use low-energy infrared light to catabolize diverse waste carbon feedstocks, including lignin-derived aromatics and syngas, making them attractive chassis for carbon-conserving biomanufacturing. To overcome the limitations of genetic intractability of these and other non-model microbes, we developed Conjugation-based Serine Recombinase–Assisted Genome Engineering (cSAGE), a modular platform that combines conjugative DNA delivery with serine recombinase–mediated integration to enable predictable and stable genome editing in transformation-resistant bacteria. Using a panel of five PNSB, we systematically benchmarked eight serine integrases and identified Bxb1, R4, and TG1 as high-fidelity enzymes supporting efficient and on-target integration across diverse hosts. We further implemented a ΦC31-based marker-curing system and an orthogonal multi-payload integration toolkit that facilitates sequential strain engineering. To demonstrate its utility for bioproduction, we applied cSAGE in Rhodopseudomonas palustris, one of the most widely studied PNSB, to integrate a phenolic acid decarboxylase (BaPAD), achieving photosynthetic conversion of a lignin-derived aromatic (p-coumarate) to the thermoplastic precursor p-vinylphenol (p-VP). Finally, we demonstrate that cSAGE operates reliably across Gram-negative and Gram-positive bacteria by testing additional non-model bacteria, including two Pseudomonas species and Rhodococcus jostii. With its programmability and broad-host range applicability, cSAGE provides a versatile genome engineering approach for previously inaccessible microbes, supporting systematic exploration and acceleration of carbon-conserving bioproduction.

Identifying single-cell heterogeneities in an engineered strain of Pseudomonas putida for the bioproduction of itaconic acid

Nicholas Reichart* (nicholas.reichart@pnnl.gov)
Pacific Northwest National Laboratory

Nicholas J. Reichart, Soujanya Akella, Allan Scott, Jacob Brandner, James Carothers, Joshua Elmore, Alex Beliaev

Understanding how cell phenotypes drive functional variations can provide useful information for cellular metabolic engineering to improve bioproduction. For instance, heterogeneity in bioproduction cultures can impact titers, production rates, and yields of high-value chemicals. Recently, single-cell methods have been developed to aid in the understanding of the molecular basis of individual cell populations within well mixed bioreactor conditions. Applying microbial split-pool ligation transcriptomics to itaconic acid producing engineered Pseudomonas putida, we determined the functional roles of subpopulations of cells that are actively contributing to biochemical production of itaconic acid and cells that serve ancillary roles within a well-mixed system. Through screening of 10,368 single cells across multiple times points of bioproduction, we tracked the expression profiles of a subset of cells expressing itaconic acid related genes compared to subsets of cells that did not. Additional clustering of transcriptomic expression profiles revealed subpopulations of cells affected by redox stress, nutrient imbalances, and other regulatory systems. Single-cell transcriptomics provides a higher resolution of cellular activity than bulk RNAseq for the functional expression of bioreactor systems and identifies subpopulations that can be used in iterative design build test learn cycles to improve output of biochemical production.

CRISPRi-mediated gene function analysis in diverse cyanobacteria

Nicolas Grosjean* (ngrosjean@lbl.gov)
Joint Genome Institute (JGI)

Nicolas Grosjean, Daniel Peterson, Mike Sharkey, Lisa Simirenko, Yuko Yoshinaga, Matthew Blow, Ian K. Blaby

Genome-wide loss-of-function phenotypic screens using CRISPR guide RNA (gRNA) libraries provide powerful approaches to systematically interrogate gene function. By knocking down gene expression across a population using CRISPRi, we can identify genes enriched or depleted in selected versus control populations. Screens can be highly parallelized to rapidly obtain large gene- and condition-specific datasets, allowing researchers to assess gene essentiality across diverse growth conditions.
At the JGI, we developed a comprehensive workflow to streamline CRISPR library screening, from gRNA design and assembly through sequencing and data analysis. We also developed CLiP (CRISPR Library Portal) to advance community access to data and facilitate physical distribution of CRISPR libraries. As proof of concept, we applied this workflow to screen CRISPRi libraries in multiple cyanobacterial species. By screening libraries under 30 environmental conditions, including trophic conditions, light intensity, pH, and micronutrient deficiency and excess, we aimed to assign function to ~45% of unannotated genes and provide additional evidence for characterized genes in the model organism Synechocystis sp. PCC 6803. Our analysis uncovered 1,210 of 3,670 genes with significant effects in at least one condition. Notably, we identified 130 genes with putative functional roles during growth with lanthanum, including transporters representing interesting targets for critical mineral recovery.
The successful implementation in cyanobacteria demonstrates potential for broader application in photosynthetic organisms and other industrially relevant microbes. This workflow reduces technical barriers, making comparative functional genomics grounded in experimental data more accessible to the community working on diverse microbial systems.

Transcriptome-Informed Genome Engineering in Rhodobacter sphaeroides Enables Predictable Expression from Neutral Sites

Sarah Grube* (sgrube@uw.edu)
University of Washington

Sarah Grube, Michael Guzman, James Carothers

Microbial bioproduction offers a promising alternative approach to petrochemical-based production of value chemicals. Non-model bacteria such as purple non-sulfur bacteria (PNSB) represent a promising bioproduction chassis. In order to overcome the lack of robust tools for genomic modification in these non-model microbes, high-efficiency tools such as SAGE (Serine recombinase-Assisted Genome Engineering) have been developed. Approaches such as SAGE rely on integration of a so-called "landing pad"into a stable genomic location. These landing pads are located at neutral genomic sites that enable stable expression of heterologous genes without disrupting host physiology. However, neutral genomic sites are poorly characterized in Rhodobacter sphaeroides. Therefore, we employed a genome-wide transcriptomics-guided approach to identify candidate landing pad locations. We further validated these sites using cSAGE-mediated genomic integration of a fluorescent reporter (to assess expression) and growth assays (to evaluate fitness effects). Application of this workflow identified nine high-confidence sites across chromosomes and endogenous plasmids, providing a diverse set of genomic contexts for potential engineering. Overall, this work establishes a foundation for predictable, high-efficiency genome engineering in PNSB and supports their broader usage as chassis for C1/C2-based biomanufacturing.

Unraveling Non-coding RNA Responses to CRISPR Perturbations With Bacterial Single-Cell RNA Sequencing

Stephen Fedak* (sfedak@uw.edu)
University of Washington

Stephen Fedak, Yujia Huang, Jacob Brandner, James Carothers, Anna Kuchina

Bacteria use post-transcriptional gene expression control via non-coding small RNA (sRNA) in addition to transcriptional regulators to shape cellular responses to environmental change. Functional genomics of sRNA typically uses short-fragment RNA-seq, which does not retain information about coexpressed genes, or overexpression and knockout studies that are limited to a single sRNA regulon. Comparatively little is known about how global sRNA expression is affected by perturbations in one part of the regulatory network. Single-cell transcriptomics enables observation of sRNA at scale and in context with coding genes, providing insights into the cues that trigger expression of sRNA and the genes they regulate that bulk methods cannot. In this work, we profiled sRNA expression changes in the context of CRISPRi of three transcription factors in Pseudomonas putida using the bacterial single-cell CRISPR screening platform mapSPLiT. We detected 203 annotated and putative sRNA, with 91% of cells expressing at least one. Of these, 21 sRNA showed perturbation-specific expression changes, while 30 were differentially expressed in all CRISPRi conditions. Finally, we clustered the sRNA using their coexpressed genes and identified functional groups related to motility and stress responses. This work demonstrates the utility of bacterial single-cell RNA sequencing for studying sRNA regulation and highlights the broader impacts of targeted expression changes.

Engineered tunable CRISPRa synthetic promoters for rapid biosynthetic profiling

Tommy G. Primo* (tprimo@uw.edu)
University of Washington

Tommy G. Primo, James M. Carothers

Engineering microbial strains for high-titer, scalable bioproduction of value-added chemicals remains a central challenge in metabolic engineering. Limitations in tunable tools for programming multi-gene variations hinder optimal strain development required for iterative design-build-test-learn (DBTL) cycles. CRISPRa-based biosynthetic profiling offers a programmable framework for exploring high dimensional genetic design spaces through systematic modulation of gene expression.

Here we develop and characterize a suite of tunable synthetic CRISPRa-responsive promoters in Pseudomonas putida and deploy them to control both a three-gene and a refactored five-gene heterologous pathway for production of the target compound 4-aminocinnamic acid (4-ACA). We demonstrate that these engineered promoters enable robust expression within the expanded pathway architecture and support comparable production levels of 4-ACA.

To accelerate strain optimization, we are integrating these promoters into an ML-guided recommendation workflow and rationally select a minimal subset of combinatorial designs for each iterative DBTL cycle. The goal of this approach is reducing the experimental search space, increasing the efficiency of pathway exploration, and enabling rapid identification of high-performing strains.

RNA Aptamer-based biosensing Platform for Tracking Organism Responses to microenvironments (RAPTOR)

Tushar Aggarwal* (tushar.aggarwal@pnnl.gov)
Department of Energy, United States Government

Tushar Aggarwal, Clifford S Morrison, Alex S Beliaev

The RNA Aptamer-based biosensing Platform for Tracking Organism Responses to microenvironments (RAPTOR) introduces a platform for real-time monitoring of cellular heterogeneity within non-model organisms used in industrial bioprocesses and synthetic biology. Conventional fluorescent protein-based biosensors are limited by delayed signal response, high metabolic burden, oxygen dependency, phototoxicity, and narrow dynamic range. RAPTOR overcomes these challenges by integrating genetically encodable fluorescent RNA aptamers with metabolite-sensing riboswitches—either naturally occurring or in vitro selected—into a modular, adaptable system.

Unlike protein-based sensors, RAPTOR relies solely on transcription, enabling rapid biosensor synthesis and immediate responses to environmental changes while conserving cellular resources. Its architecture can allow precise detection of diverse targets, including small molecules, proteins, metal ions, and nucleic acids, with high sensitivity and specificity. Oxygen-independent fluorescence and reduced phototoxicity make RAPTOR suitable for anaerobic or stress-prone environments, ensuring robust performance across industrial conditions. By leveraging aptamer flexibility and riboswitch-mediated ligand recognition, RAPTOR achieves efficient signal transduction without complex multi-step pathways. This platform also provides spatiotemporal resolution for dynamic visualization of metabolic states and microenvironmental responses in organisms such as Pseudomonas putida and Clostridium tyrobutyricum, which are critical for producing biochemicals and sustainable aviation fuel precursors.

RAPTOR addresses key limitations of conventional biosensing methods, offering a versatile, cost-effective solution for optimizing bioprocesses and advancing synthetic biology. By enabling real-time insights into cellular heterogeneity, RAPTOR sets a new standard for biosensing technologies, driving innovation in industrial biotechnology.

CRISPRa/i circuits for downstream transcription factor-based biosensor control

Yejun Kim* (yejunkim@uw.edu)
University of Washington

Yejun Kim, James Carothers

Allosteric transcription factors (aTFs) are broadly used as tools for output-based bacterial strain screening or selection. However, native biosensors display a wide range of sensitivity and ligand specificity that are often mismatched to application and difficult to optimize through protein engineering. Here, we propose a downstream CRISPR-based system to functionalize native biosensor response curves. Multi-layer CRISPR activation and interference (CRISPRa/i) circuits offer a largely tunable and customizable space through characterized components and designs. Using a previously established LysR-type transcriptional regulator (LTTR) for itaconic acid detection, we demonstrate functionality of ligand dependent CRISPRa/i nodes. We rationally determine the most promising circuit designs towards linear input-output relationships through dynamic computational CRISPR equation optimizations. Multi-layer CRISPR systems are prototyped to display divergent relationships from the native logarithmic response curves. Precise control and adaptability of CRISPRa/i tools enable downstream signal engineering that is generalizable to other biosensor proteins and inputs, encouraging further sensing applications and accelerated multiplexed regulation.

Revealing functional genomics at high resolution and scale across diverse bacteria using a pooled single-cell CRISPR screening platform

Yujia Huang* (yujhuang@uw.edu)
University of Washington

Yujia Huang, Jacob R. Brandner, Stephen J. Fedak, Anna Kuchina and James M. Carothers

Engineering bacterial species as robust biotechnological chassis requires systematic tools to connect targeted genetic perturbations with system-level phenotypes. However, most organisms lack scalable platforms for high-throughput perturbation and transcriptomic readout. Here, we extend the mapSPLiT platform—originally developed in E. coli—to two bacteria widely used in industrial and environmental biotechnology: Pseudomonas putida and Bacillus subtilis. mapSPLiT combines CRISPR activation/interference (CRISPRa/i) with microbial split-pool ligation transcriptomics (microSPLiT) single-cell RNA sequencing to profile transcriptome-wide responses to gene perturbations at high resolution and scale. In this platform, a unique guide barcode construct (GBC) associates each CRISPR perturbation with its corresponding transcriptional response. To benchmark our ability to uncover functional genomics, we targeted known transcriptional regulators with CRISPR-based perturbations in both organisms. In P. putida, we profiled 15,004 cells with a median of 85 transcripts per cell, and 57.4% were correctly assigned to specific perturbations. In B. subtilis, we achieved a perturbation assignment accuracy of 90.6% across 90,702 cells. With this high assignment efficiency, we captured robust transcriptional responses and recapitulated numerous established regulatory interactions. In addition, we uncovered previously uncharacterized gene expression responses that are consistent with predicted regulatory functions and prior proteomic data. These results demonstrate the scalability and versatility of mapSPLiT across diverse bacterial systems, establishing a powerful framework for genome-wide interrogation of bacterial functional genomics.

Session 3

AI in Action: Automating Experiments for Genomic Exploration

Posters alphabetical by first name of presenting author*

A Minimal Set of Tools For Autonomous Microbial Experimentation

Angela Cintolesi* (angela.cintolesi@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

Angela Cintolesi, Todd Edwards, Riley Maltos, Moses Obiri, Erin Bredeweg, Olga Shishkov, Niaz, Chowdhury, Scott Baker, Jaydeep Bardhan, Justin Teeguarden, Kristin Burnum-Johnson, Douglas Mans

Autonomous experimental workflows hold great promise for accelerating biological discovery and optimization, particularly in industrial biotechnology. Here we present a prototype workflow and software toolset, Autonomously Repeating Experimentation Leveraging Integrated AI (ARELIA), designed to execute closed-loop microbial experimentation with minimal human intervention. ARELIA is a modular, minimum viable system built to enable autonomous experimentation today while being readily extensible to future advanced automation at PNNL, specifically the Anaerobic Microbial Phenotyping Platform (AMP2).

In this workflow, the scientist specifies the biological question, design space, and constraints in a spreadsheet describing the parameters of a microbial growth experiment. ARELIA parses this input and coordinates an iterative cycle of experiment design, automated execution, data acquisition, and AI-driven analysis. In each loop, ARELIA selects a subset of experimental conditions; once initial experiments have been executed, it employs a machine learning algorithm to select subsequent conditions. A Tecan Fluent platform executes the experiments, and optical density (OD) measurements are collected as a proxy for microbial growth. The resulting data are aggregated and analyzed by an AI-based decision module, which proposes the next set of experimental conditions. This process continues until a predefined exit criterion is met, at which point ARELIA generates a report summarizing the optimal conditions and the path taken to identify them. To demonstrate feasibility and impact, we applied the workflow to maximize cell growth of an industrially relevant yeast, Yarrowia lipolytica, used for production of fuels, chemicals, and nutritional products.

Exploration of growth frontiers of Pseudomonas putida using reinforcement learning

Bram Stone* (bram.stone@pnnl.gov)
Pacific Northwest National Laboratory

Bram Stone, Joonhoon Kim, Andrew Frank, Abby Reynolds, Joshua Elmore

The emerging need for AI-ready data in biology demands extensive exploration of large combinatorial parameter space. Automated design and exploration using reinforcement learning patterns offers a promising avenue to build such data. We modified an existing reinforcement learning tool, BacterAI, to explore a multivariate continuous parameter space in order to understand the conditions that allow for the growth of Pseudomonas putida AG5577 (a derivative strain of the bioproduction chassis KT2440). We tested five carbon compounds (D-glucose, citrate, octanoate, acetate, benzoate), two nitrogen sources (ammonium, urea), three pH settings (5, 7, and 9), and two inorganic salts (potassium chloride, sodium chloride) intended to induce cellular stress. Experimental additions were made with an Echo automatic liquid handler. Beginning with a Monte Carlo initialization scheme, we undertook two rounds of Bayesian optimization. BacterAI was successfully able to identify more challenging growth conditions by lowering concentrations of nutrients and essential trace minerals and by increasing concentrations of salts. Between the second and third round, the summed multivariate distance from ideal growth conditions increased from 2.18 ± 1.42 to 3.57 ± 1.35 (mean ± 1 SD). Least absolute shrinkage and selection operator (LASSO) regression performed on the resulting growth data indicated that pH was the most important variable, with lower growth (based on final optical density, OD600) observed at pH 5. However, BacterAI was still able to identify conditions yielding higher growth despite more challenging growth conditions such as fewer nutrients, more salt, and non-optimal pH.

Toward Autonomous Biosystems Engineering: AI-Driven Protein and Strain Design with Automated Evolution

Christopher Henry* (chenry@anl.gov)
Argonne National Laboratory

Chantel Duscent-Maitland, Paul Hanke, Xinran Lian, Alexander Brace, Brian Hsu, Nicholas Chia, Filipe Liu, Casey Stone, Isabella Rose Zempel, Natascha Spahr, Sai Karanam, Zahmeeth Sakkaff, Paul Abraham, Yusuke Otani, Yasuo Yoshikuni, Gyorgy Babbing, Adam P. Arkin, Ellen Neidle, Nidhi Gupta, Arvind Ramanathan and Christopher S. Henry

We are combining AI, automated labs, and mechanistic models to build a platform for autonomous engineering of bacterial proteins, pathways, and strains. We use Acinetobacter baylyi ADP1 as a chassis because it enables rapid genome rewriting via efficient natural transformation with linear DNA. The platform executes two coupled reinforcement-learning loops: one for protein design and one for strain design.
In the protein-design loop, protein/DNA language models (e.g., GenSLM-family models) and agentic reasoning workflows propose candidate enzyme sequences. Designs are scored through a computational evaluation pipeline (foldability, activity, and promiscuity proxies) and iteratively improved using reinforcement-style optimization. We apply this loop to two enzyme families: (i) engineered aldolase variants (DgoA*) that replace the canonical DAHP synthase step by condensing erythrose-4-phosphate with pyruvate, creating a redesigned entry into the shikimate pathway; and (ii) O-demethylases with expanded specificity for catabolism of non-native lignin-derived aromatics. Top-ranked variants are validated in vivo and in vitro, closing the loop with measured functional outcomes.

In the strain-design loop, AI-driven experimental planning—grounded in mechanistic and causal models of ADP1 metabolism—selects genetic interventions and adaptive laboratory evolution (ALE) conditions to maximize adaptive success. Robotics-enabled, plate-based ALE performs serial transfer, growth monitoring, and competitive selection, while sequencing identifies designed and emergent beneficial mutations affecting central metabolism, transport, and regulation. These outcomes provide reward signals that refine strain-design policies and improve predictive models for subsequent rounds.
Together, these reinforcement loops turn automated evolution into a learning engine for AI, enabling rapid, scalable, and increasingly autonomous protein and strain engineering.

Cultivating Solutions: Targeted Isolation and Characterization of Zostera marina Associated Bacteria to Enhance Seagrass Resilience and Support Restoration Efforts

Diane-Marie Brache-Smith* (dbrachesmith@gmail.com)
University of California, Merced

Diane-Marie Brache-Smith, Jaquelyn Badillo, Saray Maeda, Maggie Sogin

Seagrass meadows serve as nurseries for fisheries, stabilize coasts, improve water quality, and sequester carbon. Unfortunately, eutrophication and rising ocean temperatures are causing rapid decline in seagrass meadows worldwide. To enhance restoration efforts, innovative solutions are needed to mitigate these stressors. Like terrestrial plants, seagrasses form partnerships with diverse microbe that provide then with nitrogen, phosphorus, sulfur, and phytohormones. One strategy to enhance seagrass conservation efforts is to isolate, characterize and identify plant growth promoting bacteria (PGPB) from seagrass tissues. Here, we used a novel culture-dependent approach to isolate putatively PGPB from Zostera marina's root compartments. We recovered 195 isolates spanning 26 families, 40 genera, and 88 species. Of the 104 isolates tested for probiotic potential using in vitro assays, 66 mobilized nitrogen, 13 detoxified sulfides, 38 solubilized phosphorus, and 26 produced indole-3-acetic acid. Of the 61 isolates send for whole genome sequencing, 17 represented genomes absent from current databases (ANI 80-94%), including putative sulfur-oxidizing Roseibium and nitrogen-fixing Agarivorans. The vast majority of sequenced isolates had genes for nitrogen mobilization, including nitrogen fixation (10%), denitrification (51%), dissimilatory nitrate reduction to ammonia (71%), and C-N bond cleavage (83%). Additionally, 52% of the genomes had genes for sulfur/thiosulfate oxidation, 88.5% for phosphorus solubilization, and 60.5% for IAA production. Based on these functional profiles, we designed a minimal bacterial community capable of producing essential plant growth compounds including nitrogen-containing metabolites, auxin indole-3-acetic acid, and sulfur detoxification products, highlighting key microbial species with potential to aid in seagrass restoration efforts.

Process Analytical Technologies for Autonomous Bioreactor Operation

James Collett* (james.collett@pnnl.gov)
Pacific Northwest National Laboratory

Isaiah J. Lemmon, A. Brooke Remington, Erik R. Hawley, Teresa L. Lemmon, Marie S. Swita, Shuang Deng, Ziyu Dai, James R. Collett

Rapid, quantitative phenotyping can be a major bottleneck in translating microbial genome engineering into predictable industrial bioreactor performance, especially under realistic process conditions. Crude, cost-advantaged feedstocks such as agricultural and forestry residues, waste oils and fats, or municipal and organic wet waste for bioproduct manufacturing often contain high amounts of suspended particles and complex chemical mixtures that increase turbidity and background signals and confound conventional bioprocess data acquisition and control systems. Yet these harsh, variable media are often where strain-to-strain differences in carbon utilization, inhibitor tolerance, and byproduct formation are most pronounced—making them essential environments for genome-to-phenotype studies in bioreactors. To meet this challenge, we integrated commercial, off-the-shelf liquid autosampling systems with dielectric spectroscopy, near-infrared spectroscopy, and Raman spectroscopy systems to build multivariate, chemometric models for real-time tracking of multiple sugar substrates and organic acid products and for fed-batch control of bioreactor cultivations of oleaginous yeast and filamentous fungi in volumes up to 30 L. Further integration of spectral data with those from online sensors for cell mass, density, and viability along with other process data, process event logs and experiment metadata into database schemas presents a rich opportunity for AI-driven process optimization and phenotype dissection using LLMs with retrieval-augmented, tool-calling agents to query the structured data, generate mechanistic hypotheses, propose next experiments, and develop interpretable control policies linking genotypes, conditions, and phenotypes.

Automation and standardization of biogeochemical measurements for AI analysis and modeling

John Barger* (john.bargar@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

Swarup China, Todd Edwards, Emily Graham, Sarah Leichty, Kaizad Patel, and Greg Vandergrift

Investments in laboratory automation and generation of high-quality datasets are critical to DOE's Genesis Mission and the Nation's AI Action Plan. EMSL's Molecular Observation Network (MONet) is advancing these priorities by developing automated sample handling and processing systems to deliver high-throughput, AI-ready biogeochemical molecular data from sediments and soils. These capabilities support EMSL's community science campaigns while providing unprecedented new opportunities to link process-rich molecular data with regional-scale perspectives.

Standardization of sample handling, data acquisition, and processing workflows through automation improves reproducibility and reliability while reducing error and measurement bias. Such data are foundational for robust AI utilization and predictive Earth system simulations. Moreover, molecular data offers unprecedented opportunities to understand biogeochemical function at reservoir and regional scales, reduce model uncertainties, and inform strategies for recovering critical minerals from unconventional sources such as soils and mine tailings. These insights are possible because molecular signatures provide quantifiable proxies for physical and biological processes that collectively govern soil function.

Our automation efforts span laboratory and imaging workflows—from soil aliquoting and particle size measurements to TOC/TON analyses, water extractions for high-resolution mass spectrometry, and automated data processing for imaging platforms including XCT and SEM-EDS. This poster will present an overview of our automation plans, current capabilities, and strategic investments in data automation to accelerate scientific discovery.

Proteus: an integrated automation platform for accelerating bioconversion discovery

Katarina Aranguiz* (aranguiz@wisc.edu)
Great Lakes Bioenergy Research Center
University of Wisconsin- Madison, Madison, WI

Katarina Aranguiz, Jeffrey Yao, Samantha Kelly, Adela Arrington, Daniel Noguera, Michael Botts

Proteus is a laboratory automation platform deployed by the Great Lakes Bioenergy Research Center (GLBRC) to accelerate the discovery of innovative bioconversion technologies. This platform consists of a staging/prep chamber, an automation chamber, a plate reader, an incubator, a liquid handler, and a storage carousel - all of which are integrated with advanced scheduling software, allowing us to interleave processes and run multiple experiments simultaneously.

Linking automated data collection with an existing data repository and high throughput compute cluster, we have 1.) laid the groundwork for analytical pipeline development where experimental metadata can be combined with plate reader measurement data to explore input-output relationships at a high throughput, and 2.) created an automated adaptive laboratory evolution (ALE) passaging technique where passage timing can be determined by a variety of criteria, requiring real time data intake and analysis. These criteria include: phase of growth, max OD, doubling time, and even colorimetric absorbance measurements for traits not coupled with growth.

We demonstrate proof-of-concept data from a microbial growth screen and automated ALE experiment, the metadata schemas being used to organize sample information for downstream applications, and a step-by-step example of how a current GLBRC ALE project is being expedited using this platform. Ultimately, leveraging automation and rational data management will accelerate the delivery of next-generation microbial cell factories.

Development of automated approaches for pooled microbial library screening to advance bioconversion chassis engineering

Michael Botts* (botts@wisc.edu)
Great Lakes Bioenergy Research Center
University of Wisconsin- Madison, Madison, WI

Pratham Lotia, Katarina Aranguiz, Adela Arrington, Michael R. Botts

Improving the efficiency of bioconversion of lignocellulosic feedstocks to high value products remains a key challenge for improving the profitability of the bioeconomy. Laboratory automation provides enormous value for accelerating and increasing the throughput of microbial engineering efforts, ultimately decreasing the time required to develop and deploy impactful bioconversion technologies. One microbial engineering pipeline in development at the Great Lakes Bioenergy Research Center (GLBRC) is marrying pooled library screening with high throughput laboratory automation. Pooled gene perturbation libraries are a powerful approach to rapidly identify high value engineering targets across a broad swath of bioproduction relevant media and environmental conditions. We are using rapid prototyping approaches to develop automated workflows that allow for reproducible growth kinetics for multiple chassis and a wide range of culture volumes, supporting many classes of genome-scale libraries. Using this emerging platform, we have begun performing pooled library selections in the alphaproteobacterium Zymomonas mobilis and budding yeast Saccharomyces cerevisiae across lignocellulosic feedstocks and feedstock-related stressors. We aim to leverage productionized automated pooled library screening to accelerate microbial engineering and deliver optimized bioconversion chassis.

A generalized platform for artificial intelligence-powered autonomous enzyme engineering

Nilmani Singh* (nsingh16@illinois.edu)
University of Illinois Urbana-Champaign

Nilmani Singh, Stephan Lane, Tianhao Yu, Jingxia Lu, Adrianna Ramos, Haiyang Cui, Huimin Zhao

Proteins are essential molecular machines with broad applications from therapeutics to sustainable chemistry. However, engineering proteins with desired functions remains slow, costly, and resource intensive. We designed a generalized platform for autonomous protein engineering that integrates AI/ML models and robotic workflows to minimize the need for human intervention and domain expertise. This work advances a self-driving laboratory framework for accelerated and time-efficient DBTL cycles in protein engineering.

The process begins by designing a mutant library using a protein LLM and epistasis model. The library is constructed and screened at Illinois biofoundry (iBioFAB) using optimized modular robotic workflows. The assay data from each cycle is used for training a ML model to predict fitness for subsequent iterations. With only an input protein sequence and a quantifiable fitness assay, this system can rapidly engineer diverse proteins. As a proof of concept, we applied the platform to engineer Arabidopsis thaliana halide methyltransferase (AtHMT) and Yersinia mollaretii phytase (YmPhytase). In just four rounds, requiring construction and characterization of fewer than 500 variants per enzyme, we achieved a 90-fold improvement in AtHMT substrate preference, a 16-fold increase in ethyltransferase activity, and a 26-fold improvement in YmPhytase activity at neutral pH. These advances were enabled by high-quality mutant library design using unsupervised protein language models, automated library construction and screening using iBioFAB robotic platform, and low-N machine learning models for iterative fitness prediction.
This work establishes a roadmap for generalized autonomous experimentation in synthetic biology, highlighting its potential to accelerate biotechnological innovation and discovery.

Biogeochemical Controls on Critical Mineral Recovery from Unconventional Sources

Odeta Qafoku* (odeta.qafoku@pnnl.gov)
Pacific Northwest National Laboratory

Odeta Qafoku*, Arunima Bhattacharjee, Sandy LaBonte, Nurun Nahar Lata, Young Song, Zihua Zhu, Jeremy Bougoure, Shuttha Shutthanandan, Brian O'Callahan, Jay Cheng, Dušan Veličković

Unconventional sources of critical minerals and materials (CMMs), including mine waste, treatment systems, and other non-ore Earth materials, represent underutilized domestic reservoirs that are becoming increasingly important as conventional CMM deposits become depleted. Although typically lower in CMM content, these systems offer opportunities for microbe-driven "biomining" approaches that reduce energy use and environmental impact. However, biomining recovery is constrained by slow leaching kinetics, limited understanding of host phases driven by the chemical and mineralogical heterogeneity of these sources, and a lack of standardized, multiscale workflows.

Autonomous JGI Sequencing Analysis Software

Sophia Koehn* (sakoehn@uw.edu)
University of Washington

Sophia Koehn, Claire Elbon, Michael Riffle, Julia Kubanek, Tatiana Rynearson, Emma Timmons-Schiffman, Brook Nunn

The contextualization of larger meta-omics projects in space, time and condition is an important analysis tool for chemists, biologists and students. Currently, JGI's pipeline projects provide distinct sequencing datasets spanning space or time. To reach maximum usability, these projects need to be traversable via a term (e.g., taxonomic or functional annotation) and ordered by space, time or a given condition. JGI's pipeline outputs functional terms (e.g., KEGG, GO, COG, TIG) and NIH taxonomic classifications. However, currently no autonomous program exists to organize these annotated terms in a user-friendly environment.

This platform will combine JGI data with scraped functional and taxonomic annotations to construct python hierarchical data structures and analytical tools, simplifying the analysis of gene abundance. Once completed, a user can enter a key term, phrase, term id, or gene id and the software will automatically return the abundance of that searched item and hierarchical information for every identified JGI project. It will also provide visualizations to enhance user understanding. Further development of this platform has potential application beyond environmental studies to areas such as human health.

This platform not only decreases a researcher's analysis time, but allows unskilled users to interact with meta-omic sequencing projects to increase understanding for their studies and projects. The goal is for this platform to keep JGI datasets from stagnating as the amount of data JGI produces increases in step with the continued rise of genomic research.

Applying Genomic Language Models to Advance DOE Plant Science

Tomas Bruna* (tbruna@lbl.gov)
Joint Genome Institute (JGI)

Nicolas Haas, Sumaira Zaman, Tomas Bruna

Understanding functional DNA elements remains a significant challenge in plant genomics, yet is crucial for advancing virtually all areas of genomics research. Current in silico approaches require vast amounts of experimentally-verified training data available only for select model plants, and the complex genome architecture of plants complicates knowledge transfer between species. Genomic language models (gLMs), pre-trained on large amounts of unlabeled sequences, offer a promising alternative that can generalize across species with minimal additional training. We present a project to systematically evaluate and develop gLMs for plant genomes. Our initial benchmarking reveals that model scale and training data breadth do not guarantee performance: smaller, domain-specialized models outperform larger general-purpose models. We demonstrate a practical application by training a model to predict tissue-specific gene expression from gLM embeddings of promoter sequences in Arabidopsis, showing that predicted expression levels enable clear delineation of tissue types from held-out genes. We outline a path toward accelerating plant genomics research through carefully validated AI methods, leveraging JGI's experimental capabilities for rigorous benchmarking.

Leveraging multi-condition and -organism omics data for ML/AI applications

Yannick Mahlich* (yannick.mahlich@pnnl.gov)
Pacific Northwest National Laboratory

Yannick Mahlich, Ben Drucker, Kate Schulz, Jeremy Jacobson, Dylan Ross, Sara Gosline and Jason McDermott

Determining protein function is a crucial component in elucidating the biochemical potential of microorganisms and the microbial communities they participate in. For example, this knowledge of protein function is essential to establish control points important for optimization problems in biomanufacturing applications. Traditional methods to detect and establish protein function mostly employ sequence driven approaches (e.g. annotation transfer by homology), their limiting factor being the dependence on previously established functional knowledge (e.g. via experimental determination). Large parts of the "functionome" have yet to be uncovered, with experimentally derived abundance profiles from multi-omics studies potentially containing information that can be leveraged to establish the molecular function of proteins.

We have demonstrated that our VaLPAS (Variation-Leveraged Phenomic Association Study) framework which uses multiple omics modalities derived from mass-spectrometry experiments under different experimental conditions (e.g. proteomics and metabolomics abundances of the same organism cultured under different environmental conditions) can successfully establish associations between features using unsupervised machine learning to enable insight into previously unannotated protein function.

As part of the new DOE Genesis Mission we are building on this work, to explore how the approach can be modified to accommodate and utilize data of more than a single organism bridging the gap to multi-organism communities as well as leveraging this information to train a ML/AI framework to answer how functional potentials can inform knowledge about environmental condition properties for example in order to identify organisms or conditions of interest for biological applications.

Session 4

Integrating Data and Modeling for Systemic Insights

Posters alphabetical by first name of presenting author*

Multi-omics Data Access via the National Microbiome Data Collaborative's Data Portal and API

Alicia Clum* (aclum@lbl.gov)
Lawrence Berkeley National Laboratory

Alicia Clum1, Bea Meluch2, Cameron Giberson2, Chris Mungall1, Cody O'Donnell1, Cristina Stone Pedraza2, Eric Cavanna1, James Tessmer2, Jasmyn Pangilinan1, Jonelle Basso1, Katherine Heal2, Mark Miller1, Montana Smith2, Olivia Hess2, Patrick Kalita1, Paul Piehowski2, Sam Purvine2, Samantha Miller2, Samantha Obermiller2, Sanjay Janarthanan2, Set Sarrafan1, Shreyas Cholia1, Sierra Moxon1, Simon Roux1, Sujay Patil1, Yuri Corilo2, Emiley Eloe-Fadrosh1 1Lawrence Berkeley National Laboratory, Berkeley, California 2Pacific Northwest National Laboratory, Richland, Washington

Contact: aclum@lbl.gov
Project Lead Principal Investigator (PI): Emiley Eloe-Fadrosh BER Program: BSSD Project Website: https://microbiomedata.org/ Project

The volume and variety of microbiome data being generated continues to grow substantially, creating a significant data resource for researchers addressing critical challenges in environmental science. The National Microbiome Data Collaborative's (NMDC) guiding principles are to make microbiome data findable, accessible, interoperable, and reusable (FAIR), to embrace open science, and to democratize data access through community engagement. NMDC's Data Portal enables search, access, and download of consistently processed and integrated multi-omics data via a web interface and an Application Programming Interface (API) for programmatic data access. These tools are built on a foundation of community standards and a robust data model designed to advance the creation, use, and reuse of microbiome data. The NMDC currently hosts multi-omics datasets from 36 studies including individual investigator projects, Science Focus Area (SFA) and Bioenergy Research Center (BRC) projects, and collaborative efforts including the National Ecological Observatory Network. In the Data Portal users can use faceted search to drill down on studies, samples, and datasets, driven by curated metadata, environmental context, geolocation, and gene function. Once datasets have been identified, standardized workflow results can be downloaded in bulk via the website, API, or Globus endpoints. NMDC coordinates across other BER funded projects via Extract, Transform, Load (ETL) scripts, support for external identifiers and working groups. Planned improvements for this year include extending ETL scripts to pull from additional data repositories and ensuring that data exported from NMDC are compatible with formats required by Biological and enviRonmental Infrastructure for Data manaGement and Exploration (BRIDGE) effort.

SRP Analytics: A FAIR Data Pipeline for Multi-Source Integration and Systemic Analytics

Christine Chang* (christine.chang@pnnl.gov)
Pacific Northwest National Laboratory

Christine Chang, David Degnan, Sara Gosline, Michael Barton, Katrina Waters

The ever-increasing amount of available experimental data introduces several challenges in data integration for downstream analysis and extraction of systemic insights. For instance, heterogeneous datasets acquired over extended periods of time may be in different formats, lack standardized schemas, and require complex processing steps integrating multiple analytical tools. To address these integration challenges, we developed a comprehensive pipeline for data acquired for the Superfund Research Project (SRP). The SRP Analytics pipeline ingests and transforms over a decade's worth of data into a single unified, web-accessible platform that supports both downstream exploratory analysis and systematic modeling. We also designed a FAIR-compliant database schema to ensure reproducibility and standardization of each database build. Our approach demonstrates how reproducible pipeline design can effectively facilitate analytical workflows for systemic understanding of complex systems.

Why is sharing your scientific workflows still so hard? Building FAIR workflows with JAWS

Daniela Cassol* (dcassol@lbl.gov)
Joint Genome Institute (JGI)

Daniela Cassol, Mario Melara, Seung-Jin Sul, Ramani Kothadia, Nick Tyler, Elais Player, Ludovico Bianchi, Joshua Boverhof, Setareh Sarrafan, Kjiersten Fagnan

Sharing a scientific workflow should be simple. In practice, it rarely is. Workflows that run smoothly on a laptop or at one facility often fail or become time-consuming when handed to collaborators, moved to a different HPC site, or applied to new datasets. JAWS (JGI Analysis Workflow Service) was built for this reality: portability across DOE computing environments without reinventing the wheel. It is a carefully integrated set of established tools that teams already trust. WDL (Workflow Description Language) provides the "write once" layer, separating scientific intent from infrastructure. JAWS delivers the "run anywhere" experience by executing the same workflow across sites using containers (Shifter/Apptainer), Cromwell orchestration, HTCondor scaling, and Globus data staging and transfer. The result is more consistent execution with provenance captured for traceability and reuse. The challenges JAWS targets are familiar: dependencies break, paths diverge, schedulers and policies differ, data staging becomes manual, and provenance gets lost. These friction points slow collaboration and make portability, reproducibility, and FAIR reuse harder than it needs to be. The JAWS team also helps scientists migrate legacy pipelines that were often designed for a single system and required substantial custom code for task execution. Over time, these pipelines can become difficult to maintain, with ad hoc handling and duplicated effort. In our production experience, a careful migration, rather than a simple line-by-line conversion, improves maintainability and shareability and can also deliver major performance gains. In some migrations, we have observed up to 70% faster runtime and fewer batch-scheduler jobs.

This poster focuses on JAWS and what we have learned from migrating manual or ad hoc pipelines into WDL workflows executed by JAWS across DOE computing environments.
We present practical best practices and common anti-patterns: separate workflow logic from dataset-specific inputs, break complex pipelines into reusable modules, remove site-specific assumptions like hard-coded paths, pin container versions, and make data staging and integrity checks explicit. The goal is to help teams avoid migration traps and build workflows that are portable, reproducible, and easier to maintain.

Stop by for a portability checklist, migration patterns and anti-patterns, and a conversation about how JAWS could support your workflows.

MAP-Chat: AI Chatbot for the Multi-Omics Analysis Portal

David Degnan* (david.degnan@pnnl.gov)
Pacific Northwest National Laboratory

The Environmental Molecular Sciences Laboratory's (EMSL) Multi-omics Analysis Portal (MAP) is a powerful suite of online web applications supporting omics data analysis, integration, and visualization (https://map.emsl.pnnl.gov/). Though MAP has documentation and tutorial videos, users often have specific questions which can be difficult to answer without expertise or knowledge of available resources. To address this challenge an AI-powered chatbot, MAP-chat, was developed and built to specifically address MAP and omics questions. The engine of the bot is gpt-4.0, and context is provided to the bot using a Retrieval-Augmented Generation (RAG) system with R's ragnar package. Context provided to the bot includes 20+ years of published omics research from EMSL, application documentation, and transcripts from videos. The bot has been developed to always provide a citation for all answers, whether that be a link to a publication, video timestamp, etc. MAP-chat also limits the scope of questions to those that are MAP and omics-related. MAP-chat empowers users to ask questions and provides real-time references for its answers to ensure continued improvement and accuracy. MAP-chat is in its alpha phase and will be released publicly this year.

Demystifying AI and data discovery for 'omic researchers, scientists, and engineers

Deanna Beatty* (dsbeatty@lbl.gov)
Joint Genome Institute (JGI)

Deanna Beatty*, Daniela Cassol*, Ellen Dow*

The rapidly evolving AI landscape brings new opportunities and challenges for researchers across disciplines. Staff at the Joint Genome Institute (JGI), DOE Systems Biology Knowledgebase (KBase), and our partners at the National Microbiome Data Collective (NMDC) and Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) have come together; sourcing topics of common interest across genomics, computational biology, AI, data management, and high performance computing in order to support internal community learning and knowledge sharing through seminars and workshops: the Demystify AI and Data Management Series. This series is a forum to learn new tools, demonstrate best practices and approaches to working with data at scale, and share the latest advances in AI for biology. As we build new data infrastructure and integrate 'omic datasets for our user communities through the Biological and Environmental Research Program, we aim to lower barriers to large-scale data discovery and analysis across domains, continents, environments, and data types. To meet our user community where they are, we are extending an invitation to scientists and developers to share their work from tools they have developed to interdisciplinary approaches to tackle their research questions with us and take part in discussing challenges and applications. Join our community of practice invested in lowering barriers to data access, large-scale scientific analysis, and data management best practices.

Data-driven integration of multimodal data to illuminate the dark metabolome

Dylan Ross* (dylan.ross@pnnl.gov)
Pacific Northwest National Laboratory

Dylan Ross, Yannick Mahlich, Lummy Monteiro, Sutanay Choudhury, Thomas Metz, and Jason McDermott

Assigning function to unknown proteins and annotations to unknown metabolites are central objectives in the exploration of biological dark matter. Typical sequence homology-based approaches to protein functional prediction fail without sufficient homology to known proteins, and they are not applicable to small molecule metabolites. We instead set out to leverage patterns in multiomic data to extract meaningful functional annotations for unknown metabolites and proteins using statistical and machine learning approaches via our VaLPAS (Variation-Leveraged Phenomic Association Screen) framework. Metabolite annotation presents a major barrier in this data-driven approach to functional prediction: only a small fraction of observed features in a typical metabolomics experiment are able to be assigned a molecular annotation, limiting the utility of metabolomics data for functional elucidation. The complexity of small molecule chemical space makes unambiguous annotation of metabolites based solely on their observable properties difficult, necessitating the development of innovative approaches.
We hypothesize that unknown metabolite annotations can be inferred based on integration of observable properties, metabolite-metabolite and protein-metabolite associations, and phenotype-level information in a data-driven fashion from complex multiomics data. Using multiomics data from oleaginous yeast grown under different conditions, we demonstrate the effectiveness of this novel approach for annotating unknown metabolites. Our results so far show that unknown metabolites have strong associations with proteins and phenotype-level information, providing many leads for potential annotations. We have additionally developed utilities for robust tracking of unknown metabolites across separate experiments, enabling propagation of novel metabolite annotations to benefit both existing and prospective multiomics studies.

Using "metadata" to contextualize microbial isolates: Can understanding their native genetic potential better inform downstream engineering?

Elisha Wood-Charlson* (elishawc@lbl.gov)
KBase / Lawrence Berkeley National Laboratory

Elisha M Wood-Charlson[1], Jonelle Basso[2], Alicia Clum[2], Montana Smith[3], Deanna Beatty[2], Emiley Eloe-Fadrosh[2], Adam P Arkin[1]

Affiliations:
[1] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA
[2] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA
[3] Pacific Northwest National Laboratory, Richland, WA

Microbial isolates are often the foundation of genotype-to-phenotype studies and bioengineering activities. These cultivation-centric processes often focus on the downstream implications of genetic variability and variation, often with direct modification to test theories of function. However, an original isolate is often derived from an environmental sample, which provides insight into their function as part of a community and ecosystem.

KBase and the National Microbiome Data Collaborative (NMDC) are teaming up to help you reconnect your microbial isolates back with their environmental context! Future improvements to our infrastructure would benefit from your input!

- NMDC sample submission portal will include fields to characterize isolates grown in culture. Help us define what those fields should be.
- KBase is testing out a new Data Lakehouse environment that can help characterize everything we know about your isolate genomes: its pangenome, gene annotation by a variety of tools (including where they agree and where they don't), metabolic and machine-learning models that predict phenotype, and any measured phenotype information currently available. We're looking for input on what results would be most helpful and how to best deliver them.

The goal of this project is to provide additional characterization of microbial isolates that could inform bioengineering projects. By exploring an isolate's genetic potential

across the pangenome and predictive models of metabolism or function, we aim to provide additional information regarding the potential for biodesign work.

Soil Organic Indicators at Large Scale for Artificial Intelligence (SOILS-AI) Campaign

Emily Graham* (emily.graham@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

Emily B. Graham, Sakthi Kumaran, Arjun Chakrawal, Christian Ayala-Ortiz, Cheng Shi, John Bargar, Odeta Qafoku, Sarah Leitchy, and the MONet Team

High-quality AI-ready data are critical to achieving BER's mission to develop fundamental process understanding of biological systems across scale. Accurately predicting biogeochemical dynamics at regional to continental scales is essential for national energy security and remains a major challenge for Earth system models, in part because of extreme soil heterogeneity and limited molecular-level data for globally important soil taxa. To address this, the SOILS-AI campaign integrates AI/ML methods to translate molecular measurements into scalable predictors of biogeochemical processes, improving their integration into predictive modeling frameworks. This gap currently limits estimates of carbon use efficiency, soil respiration, and belowground responses to environmental change, reducing confidence in projections relevant to energy infrastructure, land-use planning, and national energy security. The campaign conducts distributed sampling to capture spatiotemporal variability in microbial activity and soil respiration across diverse ecosystems. SOILS-AI leverages EMSL's MONet database and SSURGO open-source soil data to identify and prioritize soil taxa that currently limit model performance. Feature extraction pipelines are being developed to identify key soil organic matter molecular signatures and microbial traits linked to CUE and SOC dynamics, representing early steps toward translating high-dimensional datasets into tractable inputs for ESMs and developing workflows to be shared with the broader EMSL and JGI user communities. By linking molecular-scale observations to regional and continental predictions, SOILS-AI advances community efforts to improve representation of belowground processes in ESMs and supports more robust forecasts of soil carbon dynamics critical for national energy security under changing environmental conditions.

KBase BER AI-Native Lakehouse: From Data Integration to Agentic Discovery

Gazi Mahmud* (gazimahmud@lbl.gov)
Lawrence Berkeley National Laboratory

Gazi Mahmud, Tianhao Gu, Gavin Price, Boris Sadkhin, David Lyon, Ahmed Khan, AJ Ireland, Ellen G. Dow, Elisha Wood-Charlson, Yue Wang, Joseph Bezouska, Mikaela Cashman McDevitt, John-Marc Chandonia, Paramvir S. Dehal, Janaka N. Edirisinghe, José P. Faria, Prachi Gupta, Marcin P. Joachimiak, Roy Kamimura, Keith Keller, Dileep Kishore, Filipe Lui, Katherine O'Grady, Chris Neely, Priya Ranjan, William Riehl, Samuel Seaver, Alan Seleman, Gwyneth Terry, Shinjae Yoo, Robert Cottingham, Chris Henry, Adam P. Arkin

The KBase BER Data Lakehouse is a next-generation data and AI platform designed to accelerate scientific discovery across the Department of Energy's Biological and Environmental Research (BER) data ecosystem. Built on a Lakehouse architecture, the platform provides a unified data and AI foundation that combines scalable storage, high-performance analytics, and consistent governance for diverse biological, environmental, and multi-omics datasets. By integrating data from multiple DOE laboratories and research programs, the KBase Data Lakehouse enables cross-domain interoperability while supporting secure, multi-tenant collaboration across independent scientific domains.

A defining feature of the KBase BER Data Lakehouse is its built-in, multi-tenant governance model that applies uniformly across both data and AI assets, laying a critical foundation for AI-readiness of BER data. Rather than treating governance as an external control layer, the platform embeds fine-grained access control, stewardship boundaries, and policy enforcement directly into the data and AI lifecycle. This unified governance framework manages data access, compliance, and reproducibility while extending naturally to AI-driven workflows, models, and agentic systems. The platform provides native capabilities for tracking AI agent interactions, including reasoning trace retention, auditability of decisions, and full lineage of AI-generated data products, ensuring transparency and accountability in AI-assisted scientific discovery.

With these capabilities, the KBase BER Data Lakehouse enables trustworthy multi-agent AI systems to reason over governed, high-quality datasets, generate reproducible insights, and support collaborative hypothesis generation across tenants without compromising autonomy or compliance. By bridging data silos through standardized interfaces and enforcing governance at scale, the platform creates an AI-ready environment aligned with DOE's open science principles while maintaining rigorous standards for security, traceability, and responsible AI use. Ultimately, the KBase Data Lakehouse serves as a model for modernizing

DOE's data infrastructure by unifying data and AI governance, enabling explainable and auditable AI, and supporting scalable, multi-tenant analytics to advance systems biology, environmental modeling, and synthetic biology.

Modeling Standards for Accelerating Research - Progress Report

Herbert M Sauro* (hsauro@uw.edu)
University of Washington

Herbert M Sauro, Lucian Smith

Mechanistic modeling has become a central tool for studying metabolism in bacterial systems, but its impact depends critically on the use of robust, community-driven standards. This progress report highlights recent advances centered on SBML, SED-ML, and the ongoing curation of the BioModels database. SBML has become the de facto standard for representing biochemical network models, including genome-scale metabolic models, enabling the creation of large, interoperable model repositories and facilitating model exchange, reuse, and comparison. Importantly, SBML has also ensured the long-term persistence of models, allowing them to remain usable even as modeling platforms and software ecosystems evolve. SED-ML has complemented this by standardizing the description of simulation experiments, improving reproducibility of computational studies. These standards are supported and extended by platforms such as our Python-based environment, Tellurium, which emphasizes transparency and reproducible workflows. Looking ahead, we report on ongoing efforts to develop a next-generation simulation description language, SED2, in collaboration with the NIH Center for Reproducibility, DARPA, and other partners. Together with SBML, SED2 aims to support the persistence of both models and computational experiments over much longer timescales than is currently possible, further accelerating reproducible research in systems biology.

Integrating Metabolomics and Phenomics to Understand the Response of C4 Crops to Water Deficit

Ivan Baxter* (ibaxter@danforthcenter.org)
Donald Danforth Plant Science Center

P.BHANDARI, C.LUEBBERT, A. HUBBARD. L. CONNELLY, P.OZERSKY, J.BARRETT, G. ZIEGLER, T.C.MOCKLER, I. BAXTER

Water use efficiency (WUE), defined as the ratio of biomass gained to water used, is crucial for the economic viability and agricultural sustainability of bioenergy feedstocks. However, traditional methods for genetic improvement of WUE have proven challenging due to low throughput and environmental heterogeneity in field settings, while molecular responses to water deficit at the population scale remain understudied. To address these limitations, we implemented a multi-pronged approach: integrating phenomics, genetics, transcriptomics and metabolomics to comprehensively characterize water deficit responses in C4 plants. We used high-throughput phenotyping to quantify plant size and water use across more than 11,000 plants of the C4 species Setaria viridis and Sorghum bicolor under contrasting water availability conditions. For physiological traits, we conducted 890 genome-wide association (GWA) mappings of WUE across 10 experiments, integrating the results through Bayesian meta-analysis and synteny. Additionally, we analyzed over 3,800 samples with untargeted metabolomics to characterize the biochemical response of the plants. Using novel software to process the metabolomics data, we were able identify and quantify >30,000 mass features for GWA analysis. By integrating these metabolomic traits with genomic variation and transcriptome data, we identified and characterized biochemical pathways conserved across both species. This integrated approach will enable a molecular understanding for targeted improvement of WUE in bioenergy Sorghum, with potential implications for many C4 species.

SOILARIUM: Automated synchrotron workflows for continental-scale soil characterization

James O'Sullivan* (osullivanj@anl.gov)
Argonne National Laboratory

James B. O'Sullivan[1], Debora M. Meira[1], Xiaoyang Liu[1], Jason G. Toyoda[2], Andrew T. Townsend[2], Kenneth M. Kemner[3], Maxim I. Boyanov[3,4], Lucie Stetten[3,5], Alex A. Lavens[1], Alan Kastengren[1], Gyorgy Babnigg[3], Yuri E. Corilo[2], Tamas Varga[2], John Bargar[2], Karolina Michalska[1]

[1]X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, United States
[2]Pacific Northwest National Laboratory, Richland, WA 99352, United States
[3]Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, United States
[4]Institute of Chemical Engineering, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria
[5]Radioactive Waste Disposal and Environmental Remediation National Nuclear User Facility and Williamson Research Centre, The University of Manchester, Manchester M13 9PL, United Kingdom

Soil pore structure and Fe speciation are fundamental controls on biogeochemical processes. Iron oxides interact with organic matter and nutrients, while pore networks regulate O2 diffusion that drives reduction of Fe(III) oxides to Fe(II) in anoxic microsites. Characterizing these pore structure and Fe speciation across diverse soils requires high-throughput methods that preserve redox sensitive features while resolving micron-scale structure. The SOILARIUM project, a collaboration between EMSL's Molecular Observation Network and the Advanced Photon Source eBERlight initiative, developed automated workflows at beamlines 20-BM and 7-BM to generate open-access X-ray absorption fine structure (XAFS) and micro-computed tomography (CT) datasets from soils collected in Kentucky, Michigan and Washington. CT scans intact cores using robotic sample exchange for high-throughput scanning of 18 cores without human intervention, completing scans in under 3 minutes at 3.45 µm resolution with phase contrast imaging to resolve organic matter and pore structures. For XAFS measurements, cores were destructively sampled in a glovebox then measured under cryogenic conditions to preserve Fe oxidation state. Sites differed in their proportions of reduced and

oxidized Fe and in their Fe mineral composition, with extended X-ray absorption fine structure (EXAFS) analysis identifying minerals such as ferrihydrite and goethite. These data feed into the MONet database works towards

---

A novel AI-enabled framework to confidently assign metabolic function to unknown genes, proteins, and metabolites using multi-omics data

Jason McDermott* (Jason.McDermott@pnnl.gov)
Pacific Northwest National Laboratory

Jason McDermott, Yannick Mahlich, Dylan Ross, Sutanay Choudhury, Lummy Monteiro

Despite continuing advances in sequencing and computational function determination, large parts of the studied gene, protein and metabolite space remain functionally undetermined. Most function assignment is driven by homology searches and annotation transfer from known and extensively studied molecules, which limits function prediction for genes and proteins when homology can't be established, and is impossible for metabolites and lipids. Alternate function prediction methods have been developed but do not make use of available experimental omics data generated via technologies like mass-spectrometry, which represents a rich source of unrealized functional information.

The VaLPAS (Variation-Leveraged Phenomic Association Screen) framework is an approach to shed light on the functional dark matter of molecular space by elucidating previously unknown functions of proteins and metabolites via statistical association metrics (correlation, mutual information) and novel unsupervised ML approaches to better capture non-linear associations between molecules. Furthermore, we have implemented supervised methods if known relationships are available, and these are used to calculate confidence metrics for predictions in a report format.

We benchmark VaLPAS on multiomic data from oleaginous yeasts and bacterial datasets using KEGG module and pathway membership. Our results show that VaLPAS is able to assign KEGG module membership with high confidence for a large number of proteins and metabolites that have no known functional labels. In this study we show the applicability of using experimental abundance data from detectable metabolites and proteins (extendable to other modes of experimental data) to infer protein functionality and metabolite annotation for as of yet unannotated data.

Endophytes induce systemic spatial reprogramming of metabolome in poplar roots under drought

Jayde Aufrecht* (jayde.aufrecht@pnnl.gov)
Pacific Northwest National Laboratory

Jayde Aufrecht; Dusan Veličković; Robert Tournay; Sneha Couvillion; Vimal Balasubramanian; Tanya Winkler; Daisy Herrera; Robert Stanley; Sharon Doty; Amir Ahkami

Beneficial microorganisms that reside in the rhizosphere and plant endosphere profoundly influence their host's health and ability to thrive in suboptimal environments including droughted conditions. One school of thought suggests that molecular signals produced by the beneficial microorganisms could prime the plant's immune response and prepare it for stressful conditions like drought, biotic stresses, or nutrient limitations. However, little is known about how microbial induced changes to the plant's metabolism at the cellular scale could propagate to the whole organism systems biology scale. Despite numerous studies which report a changing plant metabolome under stress, these metabolomic data have so far had limited application in predicting plant status or physiology.

In this work, we employed a high-resolution chemical imaging approach to map metabolic changes at the root zone and cell type levels. We integrated metabolomics data from MALDI mass spectrometry chemical imaging and bulk-tissue LCMS with ddPCR to correlate microbial abundance to metabolites within plant root tissues. We found that a 9-strain consortium of beneficial endophytes differentially altered the metabolome of droughted root tissues according to cell types and locations along the root system architecture. Using machine learning (ML) models, we identified root metabolites and exudates that have power to predict drought and endophyte inoculation status. We calculated the correlation between each endophyte and metabolite and found that this relationship shifts under drought conditions, indicating the dynamic role endophytes play in a plant's microbiome and metabolism in response to environmental changes.

Phytozome 14: New Capabilities and Expanded Genomic Data

Jeremy Phillips* (jlphillips@lbl.gov)
Joint Genome Institute (JGI)

Jeremy Phillips, Shengqiang Shu, Joseph Carlson, Patrick Davidson, Sumaira Zaman, Nathan Berkowitz, David Goodstein, Tomáš Brůna

Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant genomes, as well as selected genomes and datasets that have been sequenced elsewhere. By integrating this large collection of plant genomes into a single resource and performing comprehensive and uniform annotation and analyses, Phytozome facilitates accurate and insightful comparative genomics studies via interactive exploration and customized data access and download.

Pangenomes of complex organisms increase power in environmental modeling

Joanna Rifkin* (jrifkin@hudsonalpha.org)
HudsonAlpha Institute for Biotechnology

Joanna Rifkin, Avril Harder, Chloee McLaughlin, Adam Healey, Paul Grabowski, Lillian Padgitt-Cobb, Avinash Sreedasyam, John Lovell

Reference-based genetic strategies limit insights into functional variation by ignoring structural variation. However, until recently the size and complexity of plant genomes has limited implementation of pangenome approaches. We present recent advances in integrating novel pangenomic and pangenome-enabled approaches to genotype-environment association approaches that describe novel variation to support breeding programs for biomass and biofuel production.

CoreMS: An AI-ready Open-Source Toolkit for End-to-End Mass Spectrometry Data Processing Across Multi-Omics Workflows

Katherine Heal* (katherine.heal@pnnl.gov)
Pacific Northwest National Laboratory

Katherine Heal, Will Kew, Bea Meluch, , Danielle Ciesielski, Montana Smith, Yuri Corilo

CoreMS is an open-source, Python-based framework developed at EMSL to provide a foundation for processing small-molecule mass spectrometry data. Originally designed for FT-ICR MS natural organic matter analysis, CoreMS now serves as the backbone for end-to-end workflows in the Molecular Observation Network (MONet) and the National Microbiome Data Collaborative (NMDC), supporting LC-MS metabolomics, LC-MS lipidomics, GC-MS metabolomics, and FT-ICR MS NOM.

Unlike rigid pipelines, CoreMS offers a modular architecture that enables researchers to design workflows tailored to specific scientific questions. Built on a common infrastructure and designed with AI-ready architecture, CoreMS facilitates both standardized and custom approaches. This shared infrastructure facilitates integration across data types and workflows, forming a foundation for multi-omics interoperability. The ability to generate new workflows with AI assistance is a significant step towards autonomous and adaptable data processing pipelines.

Recent projects demonstrate this flexibility: An LC-MS metabolomics project required monitoring specific compounds alongside untargeted analysis. Leveraging CoreMS's infrastructure and an AI agent, we built a workflow combining targeted internal standard searches with untargeted profiling that delivered outputs in a harmonized, interoperable manner. Another project tracked $^{2}H_{3}$-labeled compounds through complex environmental samples. Building on the hybrid workflow, we added functionality to detect isotopic enrichment in both known and unknown molecules, enabling insights into metabolic pathways.

These examples illustrate how CoreMS extends beyond its established workflows to meet emerging research needs. By combining automation, modularity, and openness, CoreMS accelerates development of new approaches while maintaining reproducibility and scalability for high-resolution MS data processing.

Integrating genome-scale metabolic modeling to predict and engineer redox fluxes in Rhodopseudomonas palustris

Margaret Cook* (mkcook@uw.edu)
University of Washington Molecular Engineering & Sciences Institute

Margaret Cook, Michael Guzman, James Carothers

Lignin is one of the most abundant renewable carbon sources on Earth, yet its aromatic breakdown products remain difficult to funnel into efficient bioproduction pathways. Purple non-sulfur bacteria offer a promising alternative due to their environmental tolerance and metabolic versatility. Particularly, Rhodopseudomonas palustris is able to catabolize lignin-derived compounds such as p-coumarate while growing photosynthetically, enabling carbon-conserving bioproduction. However, photoheterotrophic growth on these reduced aromatic compounds generates excess reducing equivalents, which must be dissipated through native redox sinks such as the Calvin cycle, ultimately limiting productive carbon flux and thus growth. To systematically interrogate this, we use a genome-scale metabolic model of R. palustris to computationally quantify substrate-dependent redox stress and simulate how native redox-balancing pathways respond to increasing reducing pressure under targeted knockdowns of native sinks. These simulations reveal redox imbalance as a primary bottleneck to effective p-coumarate utilization. By integrating flux simulations across substrates with rational and machine-learning guided analyses, we identified genetic perturbations predicted to shift electron flow away from native sinks and toward engineered redox-consuming pathways, thereby coupling redox relief to growth and motivating heterologous chemical production. We demonstrate this in silico design framework through three modeled bioproduction case studies: xylitol, 4-ethylphenol, and butanol, which represent orthogonal, feedstock-coupled, and co-feeding strategies, respectively, for coupling redox balance to chemical production. Together, this work establishes a predictive framework for metabolic engineering in photosynthetic bacteria, linking genome-scale modeling, AI-assisted hypothesis generation, and modular genome engineering to rapidly prototype strains for lignin-derived bioproduction.

Chip2Flow: An AI-driven model-data-experiment integration workflows for lab-on-chip applications

Maruti Mudunuru* (maruti@pnnl.gov)
Pacific Northwest National Laboratory

Maruti Mudunuru, Lal Mamud, Rajesh Singh, Aramy Truong, Sameera Nalin Venkat, Bryan He, Arunima Bhattacharjee, Tamas Varga, Odeta Qafoku, Arjun Chakrawal, Erin Rooney, Kalyana Nakshatrala, Behzad Ghanbarian, Glenn Hammond, Satish Karra

Chip2Flow is an AI-driven, scalable, model-data-experiment integration workflow that accelerates lab-on-chip studies by tightly coupling multimodal imaging, pore-scale physics simulations, and deep-learning surrogates into a closed-loop, calibration-ready pipeline. Lab-on-chip platforms such as TerraForms provide a powerful way to interrogate how pore geometry and connectivity control flow, transport, and reactive processes (e.g., mineral dissolution/precipitation, nutrient scavenging, and critical mineral mobilization). But extracting quantitative insight remains challenging because (i) structural imaging and chemical maps are heterogeneous and difficult to fuse, (ii) high-fidelity pore-scale simulations are computationally expensive, and (iii) scaling results to core-scale process models requires robust upscaling of structure and properties. Chip2Flow addresses these gaps through three integrated steps. (1) Imaging-to-geometry and chemistry fusion: 3D X-ray computed tomography (XCT) is combined with 2D chemical maps (e.g., SEM-EDX from thin sections) to produce geometry- and chemistry-informed digital representations of porous media using CHIMERA framework. (2) Pore-resolved multi-physics simulation and experimental alignment: the workflow supports chip design generation (via Pore2Chip), high-fidelity flow and tracer/reactive transport simulations (e.g., OpenFOAM, PFLOTRAN), and systematic calibration/verification against lab-on-chip observations, including identification of dominant high-velocity channels (advection-controlled, early breakthrough) versus low-velocity secondary zones (diffusion-controlled, long residence time, enhanced local reactions). Chip2Flow is engineered to handle 2D and large 3D domains and property fields (including meshes scaling to ~125M cells where needed) and to run ensembles of chip designs in parallel to generate training and validation datasets. (3) AI acceleration and upscaling to PFLOTRAN: geometry-aware operator learning (e.g., GeONet) provides fast, high-fidelity emulation of flow and transport fields in unseen

chip geometries, while AI-based upscaling transforms imaging-derived structure and property information (e.g., permeability-porosity transforms and learned mappings) into PFLOTRAN-ready core-scale representations for breakthrough prediction and process-model calibration. As a result, Chip2Flow enables rapid hypothesis testing, uncertainty-aware parameter inference, and reusable community workflows that turn static imaging into actionable, predictive lab-on-chip science.

Meta-virus resource (MetaVR): expanding the frontiers of viral diversity with 24 million uncultivated virus genomes

Mateus Fiamenghi* (mbfiamenghi@lbl.gov)
Joint Genome Institute (JGI)

Mateus B Fiamenghi, Antonio Pedro Camargo, Iro N Chasapi, Fotis A Baltoumas, Simon Roux, Artyom A Egorov, Eleni Aplakidou, Eric Olo Ndela, Yumary M Vasquez, I-Min A Chen, Krishna Palaniappan, T B K Reddy, Supratim Mukherje, Natalia N Ivanova, Frederik Schulz, Tanja Woyke, Emiley A Eloe-Fadrosh, Georgios A Pavlopoulos, Nikos C Kyrpides

Viruses are ubiquitous across environments and impact host metabolism, evolution, and ecology. The rapid accumulation of genomic and metagenomic datasets has resulted into an explosion of uncultivated virus genomes (UViGs) and the creation of specialized databases to curate and catalogue this viral diversity. However, many existing resources lack comprehensive integration or scalability.
Here, we introduce the Meta-virus resource (MetaVR), the successor to the IMG/VR database, designed to address previous limitations regarding large-scale querying and programmatic access. By leveraging the vast increase in publicly available genomes and metagenomes, MetaVR significantly expands known viral diversity, now comprising over 24 million UViGs, a 57.6% increase from its predecessor, organized into over 12 million viral operational taxonomic units (vOTUs).
Key enhancements in MetaVR include the integration of curated eukaryotic host information, protein clusters with predicted structures for comparative studies, and a dedicated API for programmatic data access. Furthermore, the resource features an updated taxonomic framework based on ICTV release 39, assignment to Baltimore classes, and enhanced host assignment through novel computational tools such as iPHoP. These advancements position MetaVR as a unique and comprehensive resource for researchers exploring viral diversity, evolution, and host interactions across diverse ecosystems. MetaVR is freely accessible at https://www.meta-virome.org/.

Evidential deep learning for accurate and trustworthy prediction of enzyme functions

Mingon Kang* (mingon.kang@unlv.edu)
University of Nevada Las Vegas

Mingon Kang and Jeffery Shen

Characterizing enzymes is critical for understanding enzyme functions and their roles in biological processes to catalyze a wide range of commercial processes, such as pharmaceutical biosynthesis, food production and bioremediation. However, it is reported that only 33% of unknown proteins are matched with such well-characterized enzymes in reference databases, which indicates that there are a huge number of proteins whose biological functions are not yet known. Furthermore, most related computational annotation tools are predominantly trained with relatively few species of largely annotated datasets (e.g., humans), which subsequently causes underperformance with under-annotated species. Current state-of-the-art AI models have significantly improved predictive performance by automatically recognizing class-specific patterns of protein sequences from a large scale of labeled databases. However, despite the success, there is still significant room to improve in terms of predictive performance, model interpretation, and trustworthiness. In this talk, we address innovative deep learning-based models to predict enzyme commission (EC) numbers and bacteria protein and compound interaction predictions. First, our evidential deep learning models offer trustworthy predictions of functional annotation by incorporating prior biological knowledge and identifying significant amino acids to the prediction, which may correspond to known motif sites of a given protein. Second, we introduce a novel positive learning strategy that allows deep learning-based component-protein interaction (CPI) predictive models to train without negative interaction data. Our new solution, Positive-Unlabeled (PU) model addresses this limitation by generating pseudo-positive and negative labels from known positive interaction data and enabling effective training of deep learning models for CPI prediction.

AI-Ready Microbiome Data: Enabling FAIR Integration via NMDC and BRIDGE

Montana Smith* (montana.smith@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

Montana Smith[2]*, Alicia Clum[1], Bea Meluch[2], Cameron Giberson[2], Chris Mungall[1], Cody O'Donnell[1], Cristina Stone Pedraza[2], Eric Cavanna[1], James Tessmer[2], Jasmyn Pangilinan[1], Jonelle Basso[1], Katherine Heal[2], Mark Miller[1], Olivia Hess[2], Patrick Kalita[1], Paul Piehowski[2], Sam Purvine[2], Samantha Miller[2], Samantha Obermiller[2], Sanjay Janarthanan[2], Set Sarrafan[1], Shreyas Cholia[1], Sierra Moxon[1], Simon Roux[1], Sujay Patil[1], Yuri Corilo[2], Emiley Eloe-Fadrosh[1]

[1] Lawrence Berkeley National Laboratory, Berkeley, California
[2] Pacific Northwest National Laboratory, Richland, Washington

There is rapidly growing interest in leveraging microbiome datasets for AI-driven discovery, large-scale data integration, and predictive modeling. Realizing this potential depends critically on standardized, rich metadata, harmonized data formats and structures, and reliable, programmatic access. These elements are essential for accurate integration and reuse of microbiome resources in advanced machine learning workflows and BER-wide platforms such as BRIDGE, the data lakehouse for AI-enabled discovery. The NMDC Submission Portal offers a researcher-oriented interface for capturing metadata that adhere to community standards and ontologies and align with shareable, interoperable schema, thereby increasing the usability of microbiome datasets for integrated analysis and downstream modeling.
Since the release of the Submission Portal in April 2022, the NMDC has facilitated the metadata capture for over 100 studies, including submissions from large BER programs, and user facility projects from the JGI and EMSL. The NMDC Submission Portal supports metadata capture following the Genomic Standards Consortium's environmental extensions and provides tools for simplifying capture and validation for both metadata and study information. Current tools for metadata capture include suggesting new metadata values based on existing fields and external reference resources. Tools are in development to use AI to infer and recommend appropriate metadata fields and entries directly from study text, thereby improving and simplifying the user experience. By providing rigorously standardized, machine-readable microbiome metadata, programmatic access, and an expanded set of curated study descriptors, the NMDC supplies interoperable, analysis-ready inputs

that can be readily shared with BRIDGE to support cross-project data integration and AI-driven modeling.

Making heterogeneous data AI analysis ready tle

Roelof Versteeg* (roelof.versteeg@subsurfaceinsights.com)
Subsurface Insights LLC

Roelof Versteeg, Amir Ahkami, John Bargar, Sean Kacur, Arjun Chakrwal, Doug Johnson, Mariefel Olarte, Young Song, Aramy Truong and Tyler Turner

A large part of the value of heterogeneous datasets lies in system-level insights that can be obtained by examining relationships between variables—for example, understanding how precipitation and temperature influence microbial genotype and phenotype. While tools for generating these insights are increasingly available, their effectiveness depends on the ability to link diverse data together. Linking data from a single analysis type and institute is relatively straightforward, but integrating heterogeneous data from multiple providers is substantially more complex, yet essential for comprehensive analysis.
Our BSSD-funded SBIR project aims to develop software that enables this data integration, curation and analysis and allows researchers to extract maximum value from the growing wealth of environmental and geoscience data.
This software uses pipelines for data ingestion and fusion that generate AI analysis-ready, cloud-based databases. In addition, it includes software pipelines for data imputation and analysis. This work is being implemented as an open-source capability that builds on and extends our current data management product (odmx.org).
To ensure our tools are responsive to user needs, we are working with diverse biological and environmental datasets from DOE-funded projects. This collaborative and iterative approach helps us identify and address real-world data integration challenges across different data types, collection methods, research contexts, and user types.
In this presentation, we will detail our approach and share results from our work to date. We will also provide information on how interested users can get involved with our efforts and begin using our tools to make their own heterogeneous datasets analysis-ready.

Uncovering New Players in Photosynthesis by Multi-omics of Chlamydomonas reinhardtii Mutants

Sara Calhoun* (secalhoun@lbl.gov)
Joint Genome Institute (JGI)

Vy T. Duong, Cailyn Sakurai, Michela Lee, Sam Purvine, Anna Lipzen, Katherine Louie, Ben Bowen, Trent Northen, Mary Lipton, Krishna Niyogi, Setsuko Wakao, Sara Calhoun

Photosynthetic organisms play a significant role in global carbon cycling through the fixation of organic carbon from atmospheric carbon dioxide. The unicellular green alga *Chlamydomonas reinhardtii* is a predecessor of land plants and has been a powerful model organism for studying photosynthesis for more than half a century. Classical genetic analysis and recent phylogenomic approaches have hinted that there are hundreds of genes required for optimal photosynthesis in *C. reinhardtii*, including many genes that have little or no functional annotation. Towards this goal to gain a functional understanding of these unknown genes, we characterized photosynthesis-deficient *C. reinhardtii* mutants through multi-omics profiling, producing a large-scale dataset of transcriptomics, proteomics, and metabolomics measurements under different light conditions. We leveraged a manually curated set of high-confidence candidate genes identified from mutants generated through insertional mutagenesis, focusing primarily on the plasmid-disrupted candidate genes in 38 mutants. By clustering transcriptome profiles across 18 mutants associated with unknown genes, 20 mutants associated with known genes, and wild-type samples, we identified groups with similar expression patterns. Mutants with the most severe photosynthetic efficiency phenotypes show similar expression profiles, reflected in the downregulation of photosynthetic complexes and related metabolic pathways. Integrated correlation analysis reveals novel co-expression relationships between unannotated candidate genes and genes with photosynthesis-related functions. By integrating multi-omics along with phenotypic data, we establish testable hypotheses for gene functions that mediate photosynthetic responses to light stress. This work generates critical multi-omics resources for the algal research community while demonstrating a scalable framework for functional gene discovery.

An interactive, user-facing tool for JGI metabolomics and transcriptomics data integration

Sharon Greenblum* (greensi@lbl.gov)
Joint Genome Institute (JGI)

Brandon Kieft, Sharon Greenblum, Matt Blow

Many cellular and organismal processes involve cascades of information flow across multiple molecular and metabolic layers. At JGI, we offer assays to profile samples at multiple levels, such as genomic variation, regulatory mechanisms, gene expression, and metabolite production. Using statistical tools to integrate such multi-omics datasets can yield a deeper holistic understanding of organisms and their environment. In this pilot workflow, we integrate transcriptomics data, which quantifies RNA abundance, and metabolomics data, which quantifies the end products of metabolic pathways. By combining these two layers of biological information across the same samples, we capture networks of dependencies between gene expression and metabolite production, and reveal biochemical pathways that underlie complex phenotypes or experimental perturbations.

Myco-Ed: A National Platform for Scaling Fungal Genomics and Predictive Biology

Stephen Mondo* (sjmondo@lbl.gov)
Joint Genome Institute (JGI)

Sara Branco, Amy Honan, Sara Gremillion, Lotus Lofgren, Andrew W. Wilson, Peter Avis, Kerrie Barry, Scott Bates, Gerald M. Cobián, Ellen G. Dow, Chinyere Knight, Kurt LaButti, C. Alisha Quandt, Jane Stewart, Jayson Talag, Stephen J. Mondo

The fungal kingdom hosts an incredible amount of phenotypic and phylogenetic diversity, yet trait characterization and genomic representation for many taxa is still lacking. In a collaborative effort, mycologists across the United States are developing the Mycological Curriculum for Education and Discovery (Myco-Ed) to train students in fungal biology and comparative genomics while simultaneously enriching genomic representation and trait discovery across the kingdom. Students in mycology courses isolate environmental fungi, determine their taxonomic placement and perform structured assays to document specific phenotypes. Metadata are stored in a public iNaturalist repository for future use by the community while trait and taxonomic information are used to select taxa for genome sequencing at the Joint Genome Institute. Since the Myco-Ed pilot launched two and a half years ago, it has been successfully incorporated into 23 classrooms, reaching over 400 students. As a result, Myco-Ed has recorded over 1,100 observations in iNaturalist and isolated >50 novel fungal species for genome sequencing that fill important knowledge gaps. The Myco-Ed program was recently approved for sequencing of an additional 150 fungal genomes at the JGI to understand the genetic underpinnings of fungal drought tolerance - a critical area where knowledge is limited. Leveraging the wide ranging network and scale provided through collaboration with teaching labs, the Myco-Ed framework allows for rapid generation of thousands of data points which, when linked to genomic results using AI-based approaches, will allow prediction of genes associated with drought response (and other traits) across Fungi.

Graph-Based Predictive Phenomics: Linking Gene Fitness to Media Composition and Function with Knowledge Graphs and Heterogeneous Graph Inference Using GIMME

Winston Anthony* (winston.anthony@pnnl.gov)
Pacific Northwest National Laboratory

Winston Anthony, Sam Donald, Sumit Purohit, Kaizad Patel, Robert Egbert

Predicting microbial fitness across diverse conditions is a core challenge for predictive phenomics and autonomous experimentation. High-throughput assays generate genotype– phenotype data, but these are hard to integrate with experimental metadata and biological function in a way that supports mechanistic reasoning and experiment design. Knowledge graphs o@er semantically grounded frameworks to unify these modalities and enable context-aware inference.

GIMME (Graph Inference for Microbial Metabolism Exploration) is a heterogeneous knowledge graph that unifies gene fitness measurements with experimental metadata, growth media composition, and a functional biological hierarchy (gene–protein–reaction– EC). Media are decomposed into chemical components with concentrations, and experiments are linked to genes and natural-language descriptions. The graph supports: (1) symbolic traversal to surface candidate gene–environment and gene–chemical associations, and (2) learned inference via heterogeneous GNNs that propagate information across biological and environmental neighborhoods. Fitness prediction is formulated as link regression over (gene, media, experiment) tuples, combining learned gene representations with pretrained text embeddings of media names and experimental descriptions. We augment a baseline MLP with message-passing encoders (GraphSAGE/GAT) over gene–protein–function and media–chemical subgraphs, integrated via gated residual mechanism. In a case study across 10 Pseudomonas species, we achieve strong agreement with held-out fitness measurements ($r \approx 0.72$) while revealing challenges in positive-fitness regimes and specific enzyme classes. Aggregated GAT attention coe@icients provide interpretable estimates of biological and environmental relations driving predictions.

GIMME serves as a "context graph" that makes experimental conditions computable, enable explainable evidence retrieval, and support autonomous workflows for experiment prioritization

Multi-Omics Integration Reveals the Molecular Rewiring of Carbon-Dependent Ultrafast Growth in Picochlorum celery

Yasin Torres-Tiji* (yasin.torres-tiji@mines.edu)
Colorado School of Mines

Yasin Torres-Tiji, Alaina Lapanse, Devin Karns, Kent Bloodsworth, Marina Gritsenko, Lye Meng Markillie, Hugh Mitchell, Nathalie Munoz Munoz, Ernesto Nakayasu, Paul Piehowski, Camilo Posso, Nicholas Reichart, Alex Beliaev, Matthew Posewitz

Picochlorum celeri is a high-productivity marine microalga, showcasing sustained biomass productivities exceeding 25 g m^{-2} d^{-1} in outdoor ponds. However, it exhibits a striking dependence on elevated CO_2 with photoautotrophic doubling times approaching ~2 h under CO_2 supplementation but slowing to over 24 h at ambient air CO_2, suggesting the absence of a functioning carbon concentrating mechanism. To investigate the molecular basis of this CO_2-responsive phenotype, we cultivated wild-type P. celeri in climate-simulating photobioreactors under ambient air and CO_2-supplemented conditions, observing a ten-fold difference in areal biomass productivity. In parallel, we cultivated four carbonic anhydrase (CA) knockout strains under CO_2-supplemented conditions, where they exhibited no measurable productivity defects relative to wild type, challenging the assumption that these enzymes drive bulk carbon uptake.

We present an integrated multi-omics framework, including transcriptomics, global proteomics with targeted post-translational modification analysis, lipidomics, and metabolomics, to resolve the extensive metabolic and regulatory rewiring associated with the fast growth under high CO_2 availability. Comparative analyses highlight large-scale shifts between air- and CO_2 conditions, consistent with the dramatic phenotypic divergence. Notably, genomic context analysis reveals that CA genes co-localize with DNA repair and ubiquitin-associated genes. Consequently, we are integrating omics comparisons between wild-type and CA mutant strains to elucidate whether these enzymes serve non-canonical roles in carbon-responsive signaling or genome maintenance rather than direct carbon fixation. This study provides insight into the mechanisms enabling ultrafast photosynthesis under favorable conditions and its system-wide molecular consequences, informing future metabolic engineering efforts to enhance biomass productivity in photosynthetic organisms.

Current and future approaches of exploring microbial soil organic matter decomposition using the MONet metagenomics data

Young Song* (young.song@pnnl.gov)
Environmental Molecular Sciences Laboratory (EMSL)

Young C. Song, Cheng Shi, Kelly G. Stratton, Christian Ayala-Ortiz, Izabel Stohel, Viviana Freire-Zapata, Malak M. Tfaily, Emiley Eloe-Fadrosh, Arjun Chakrawal, Gihan U. Panapitiya and Emily B. Graham

Soil organic matter (SOM) decomposition by microorganisms represents one of the largest uncertainties in predicting terrestrial carbon–atmosphere feedbacks. Yet we lack a systematic understanding of the microbial taxa that drive decomposition of chemically distinct carbon pools and of the metabolic pathways they employ across environmental gradients. The Molecular Observation Network (MONet) at the Environmental Molecular Sciences Laboratory (EMSL) aims to address these challenges by integrating multi-omics data derived from soil, including shotgun metagenomics (through collaboration with the Joint Genome Institute, JGI) and molecular profiles measured using Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS). Here, we explore three examples of leveraging the MONet data to investigate various facets of microbial SOM depolymerization. First, we examine the integration of 828 metagenome-assembled genomes (MAGs) and 66,727 distinct SOM molecules derived from 47 standardized soil cores collected across the United States, revealing complementary metabolic specialization between bacterial and archaeal genera within the abundant orders *Rhizobiales*, *Chthoniobacterales*, and *Nitrososphaerales*. Second, we discuss inference of the carbon use efficiency (CUE) of the microbes represented by the MAGs assembled from these soils, in an effort to establish a robust model for inferring microbial interactions with soil carbon. This effort stems from previous studies suggesting that copiotrophs show lower CUE than oligotrophs, and that shifts toward more copiotrophic communities under increased substrate availability can lower community-level CUE. Consistent CUE measurements across diverse microbial taxa are essential for linking variation in resource-use physiology to CUE. We developed a KBase narrative to construct metabolic models derived from annotated genomes to estimate taxon specific fluxes, such as biomass growth and carbon uptake. Finally, we briefly explore the potential benefits and existing challenges of ongoing efforts at EMSL to implement a large language model–based platform for querying MONet data to address fundamental questions about taxonomic diversity in soils and the presence of common or unique metabolic pathways.