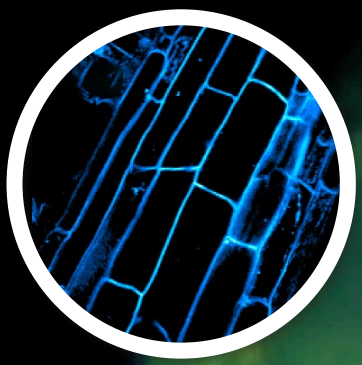
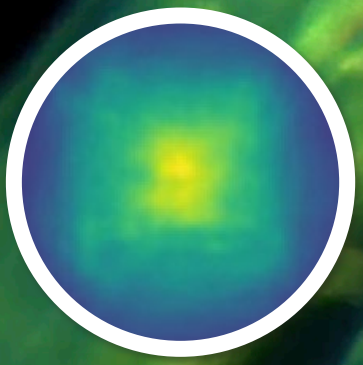
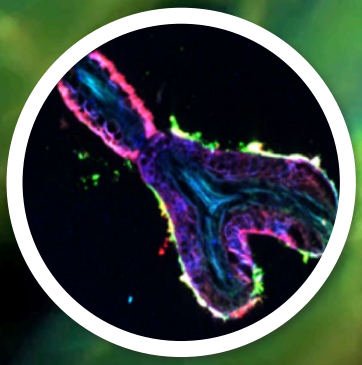
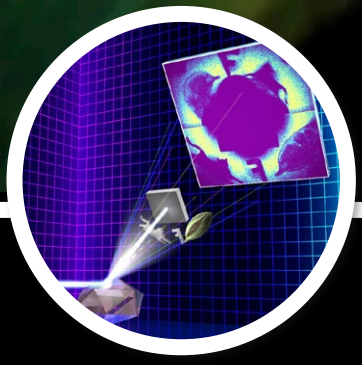








Enabling Capabilities and Resources

2024 Principal Investigator Meeting Proceedings

U.S. DEPARTMENT OF ENERGY | Office of Science

Biological and Environmental Research Program

Enabling Capabilities and Resources

2024 Principal Investigator Meeting

U.S. Department of Energy
Biological and Environmental Research Program

April 2–4, 2024
Bethesda, MD

DOE Program Managers

Paul Sammak
Biological and Environmental Research Program
Office of Science
U.S. Department of Energy

Amy Swain
Biological and Environmental Research Program
Office of Science
U.S. Department of Energy

Resham Kulkarni
Biological and Environmental Research Program
Office of Science
U.S. Department of Energy

Ramana Madupu
Biological and Environmental Research Program
Office of Science
U.S. Department of Energy

About BER

The Biological and Environmental Research (BER) program advances fundamental research and scientific user facilities to support U.S. Department of Energy missions in scientific discovery and innovation, energy security, and environmental responsibility. BER seeks to understand U.S. biological, biogeochemical, and physical principles needed to predict a continuum of processes occurring across scales, from molecular and genomics-controlled mechanisms to environmental and Earth system change. BER advances understanding of how Earth's dynamic, physical, and biogeochemical systems (atmosphere, land, oceans, sea ice, and subsurface) interact and affect future Earth system and environmental change. This research improves Earth system model predictions and provides valuable information for energy and resource planning.

Cover Images

Circle 1: Label-free structural imaging of live *Brachypodium distachyon* roots using third-harmonic generation microscopy, see p. 48; **Circle 2:** A cannula directs excitation light into the plant and collects emitted fluorescence, see p. 52; **Circle 3:** X-ray fluorescence indicating the uptake of iron, zinc, calcium, and copper in a mycorrhizal root tip of *Pinus taeda*, inoculated by *Suillus cothurnatus*, p. 7; **Circle 4:** Nonlinear X-ray diffraction spontaneous parametric down conversion (SPDC) imaging setup: 15 kiloelectron volt pump X-rays from a synchrotron light source generate SPDC biphotons from a diamond crystal, p. 12; **Background image:** Fluorescence lifetime light-sheet microscopy of *Arabidopsis* root labeled with green fluorescent protein (maximum intensity projection), see p. 51.

Suggested Citation

U.S. DOE. 2024. *Enabling Capabilities and Resources 2024 Principal Investigator Meeting Proceedings*, U.S. Department of Energy Office of Science. https://doi.org/10.2172/2440216.

Enabling Capabilities and Resources

2024 Principal Investigator Meeting Proceedings

Published October 2024

Biological and Environmental Research Program

Prepared for the
U.S. Department of Energy
Office of Science
Biological and Environmental Research Program
Germantown, MD 20874-1290

Prepared by
Biological and Environmental Research Information System
Oak Ridge National Laboratory
Oak Ridge, TN 37830
Managed by UT-Battelle, LLC
For the U.S. Department of Energy
Under contract DE-AC05-00OR22725

Contents

Executive Summary

As a major supporter of basic genome-enabled research, BER's Biological Systems Science Division (BSSD) fosters scientific discovery by funding fundamental biological research across disciplines in conjunction with enabling investigational tools and computational capabilities that include world-class user facilities. The overarching goal of BSSD is to provide the necessary fundamental science to understand, predict, manipulate, and design biological systems that underpin innovations for bioenergy and bioproduct production and enhance understanding of natural, DOE-relevant environmental processes (*Biological Systems Science Division Strategic Plan*, 2021). To accelerate the U.S. bioeconomy, BSSD pursues innovative science underpinning advances in sustainable biofuels and bioproducts and the development of next-generation technologies and computational resources for systems biology research.

The 2024 BSSD Enabling Capabilities and Resources (ECR) Principal Investigator (PI) meeting brought together PIs across the BSSD ECR portfolio to confer on shared interests and opportunities. The meeting was held concurrently with the Genomic Science program (GSP) PI meeting to optimize collaboration on research to advance bioenergy and the bioeconomy. Rick Stevens of Argonne National Laboratory gave a keynote on How Generative Artificial Intelligence Can Impact Biological Research (see Keynote: How Generative Artificial Intelligence Can Impact Biological Research, this page). Plenary presentations included several joint sessions that illuminated the integration and understanding of the larger BSSD mission. GSP's objective is to provide systems-level understanding of plants, microbes, and their communities through its Bioenergy Research, Biosystems Design, and Environmental Microbiome Research portfolios. The objective of the ECR portfolio is to support development of computational and instrumental platforms to advance fundamental GSP research—and BER more broadly—toward the overall goal of understanding the functional principles of living systems and their response to environmental challenges.

BSSD Enabling Capabilities and Resources

The ECR portfolio encompasses Computational Biology, User Facilities, and Biomolecular Characterization and Imaging Science (BCIS). Components include

- BCIS-sponsored basic research on new imaging instrumentation and measurement technologies for characterizing molecules and cells in living systems.
- BCIS's world-class resources for structural biology and imaging of biological and environmental systems prioritized for BER researchers.
- Quantum information science–supported quantum-enhanced light sources, detectors, microscope imaging and direct environmental quantum sensing of the metabolome, cells, and biological systems.
- BER awards in the Office of Science Biopreparedness Research Virtual Environment program (BRaVE) that provide foundational science and computing to prepare for future threats to the bioeconomy.

Keynote: How Generative Artificial Intelligence Can Impact Biological Research

Rick Stevens (stevens@anl.gov), Argonne National Laboratory

Stevens discussed the variety of approaches groups are taking to build domain-specific foundation models for biological problems (e.g., protein design, single cell analysis), current results, and future directions. Stevens discussed the use of general purpose large language models in biological use cases, including the automated analysis of tens of thousands of scientific papers, methods for high-throughput knowledge extraction, distillation, and synthesis and how frontier artificial intelligence systems can be used to explore the landscape of biological hypothesis and drive experimental design.

How Hypothesis-Driven Science Compares to Tool-Driven Science

"The effect of a concept-driven scientific revolution is to explain old things in new ways."

– The Structure of Scientific Revolutions, Thomas Kuhn, 2012

"The effect of a tool-driven revolution is to discover new things that have to be explained."

– The Sun, The Genome, and The Internet: Tools of Scientific Revolution, New York Public Library Lectures in Humanities. Freeman J. Dyson, 2000

- Computational biology research–supported development of cyberinfrastructure and computational tools, artificial intelligence, machine learning, and systems biology computations and enablement of the integration of large, disparate data types that are being advanced and leveraged for the BER research portfolio.
- BER User Facilities and Community Resources:
  - The Joint Genome Institute (JGI) is a national scientific user facility that advances genomics research by providing the research community with access to genome sequencing, DNA synthesis, metabolomics, and large-scale data analysis capabilities.
  - The Systems Biology Knowledgebase (KBase) is a cyberinfrastructure for the community to perform large-scale biological analyses, facilitate modeling and design, and reproducibly and collaboratively generate, test, and share new hypotheses.
  - Structural Biology and Imaging Resources are beamline-based and cryo-electron microscopy (cryo-EM) capabilities at DOE synchrotron light and neutron sources that provide access to and support for applying advanced structural biology and imaging techniques to BER research questions.
  - The National Microbiome Data Collaborative (NMDC) empowers the research community to harness microbiome data exploration and discovery through a collaborative and integrative data science ecosystem. NMDC enables findable, accessible, interoperable, and reusable data practices.
  - The National Energy Research Scientific Computing user facility provides high-performance computing to BSSD researchers through the Office of Science Advanced Scientific Computing Research program. BSSD allocates over 570k CPU and 300k GPU node hours every year to DOE and other research groups.
  - The Environmental Molecular Sciences Laboratory (EMSL) is supported by BER's Earth and Environmental Systems Science Division (EESSD). It enables molecular-scale, experimental, and predictive modeling research on aerosol chemistry, biogeochemistry, and subsurface science.

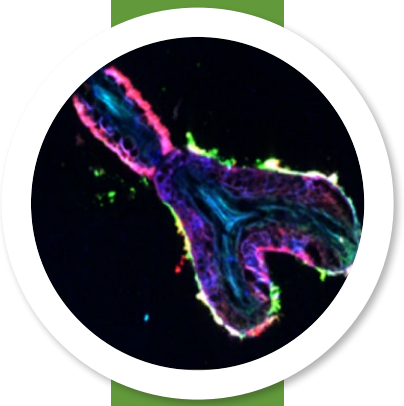
These computational tools, user resources, data sources, and bioimaging technologies in the ECR portfolio enable researchers to model and visualize the structural, spatial, and temporal relationships of key metabolic processes and the critical biomolecules governing phenotypic expression in plants and microbes under environmental- and bioenergy-relevant conditions important for bioenergy and environmental missions. The objective of measuring biosystem mechanisms and integration of data sources is to evaluate phenotypic function and the response of living systems to environmental change.

Evolution of the ECR PI Meeting

The ECR PI meeting has grown from its origins in the Bioimaging Science PI meeting to this year's expansion to encompass all enabling capabilities supporting BSSD science. The change is driven by an understanding that investigational tools are best conceived and developed with an end use in mind beginning at their conception, by a recent expansion of relevant measurement tools and computational methods, and by the integration of these methods and tools into data-driven modes of inquiry. Research to develop new tools works in concert with hypothesis-driven biological research and provides complementary advances to fundamental science. It is expected that providing more opportunity for combined sessions and networking among PIs across the GSP and ECR portfolios will enrich their potential to advance the understanding, prediction, manipulation, and design of biological systems and environmental processes.

PART 1

Oral Presentations



CHAPTER 1

Multidisciplinary Science Enabled by Facility Collaboration

Great breakthrough opportunities happen at the intersection of multidisciplinary science when experts from different domains come together to tackle grand challenges central to BER mission science. Multiple analytical approaches and techniques must be brought to bear on these complex problems. This chapter highlights an array of BER science that has been made possible through access to advanced capabilities beyond those available at many institutions. DOE enables multidisciplinary science by providing access to such capabilities through national scientific user facilities and computational resources. BER's facilities and user resources engage collaboratively to assist researchers in using complementary capabilities at multiple facilities. Facility experts can train users and help design experiments, take measurements, and, in some cases, help interpret user data. These facilities provide users with free access to the capabilities through proposal processes.

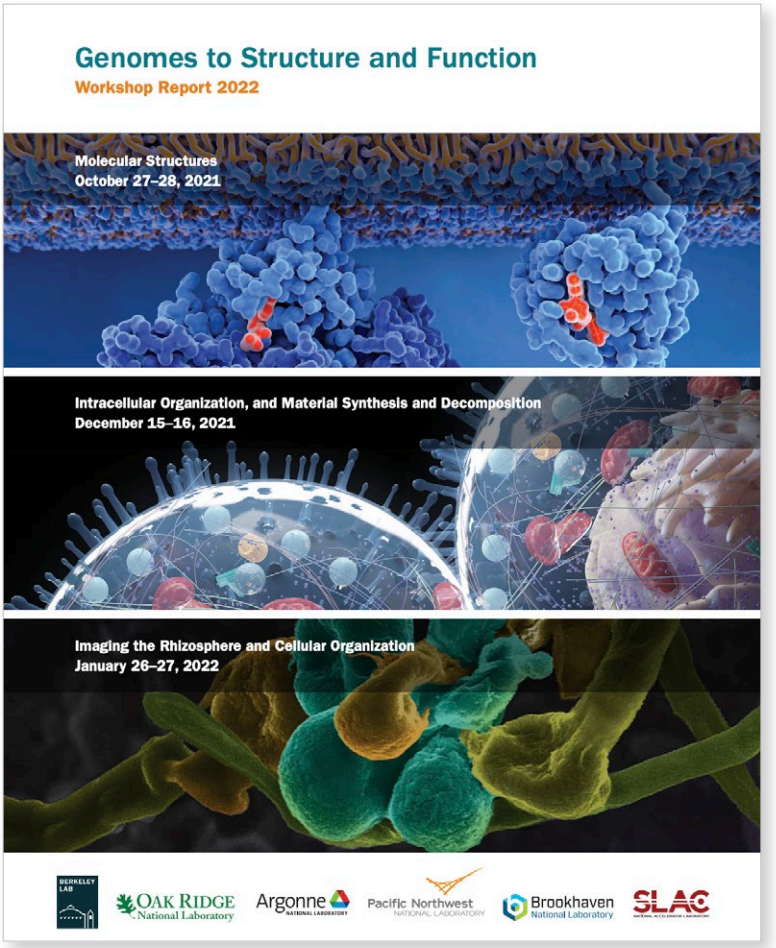
Opportunities for Team Science Enabled by Integrated User Facilities

Yasuo Yoshikuni* (yyoshikuni@lbl.gov, PI)

DOE Joint Genome Institute

Project Goals: Ongoing efforts among DOE's user facilities offer their unique and large-scale capabilities collaboratively and interactively to users to significantly advance their science programs.

The goal of the Biological and Environmental Research program is to achieve a predictive understanding of complex biological, Earth, and environmental systems with the aim of advancing the nation's energy and infrastructure security. To pursue this goal, team science—collaborations among experts in diverse research areas that lead to multidisciplinary projects—is indispensable. The roles of DOE user facilities, which offer unique and powerful resources for such research projects, are evolving, and expectations for the facilities are increasing. To respond to users' needs, the DOE Joint Genome Institute (JGI) and Environmental Molecular Sciences Laboratory (EMSL) initiated the Facilities Integrating Collaborations for User Science (FICUS) program in 2014. This collaboration has grown into a successful program, advancing more than 100 multidisciplinary projects to date. Similarly, new interfacility collaborations are becoming essential for



Genomes to Structure and Function Workshop Report. [Courtesy U.S. Department of Energy]

*Presenter

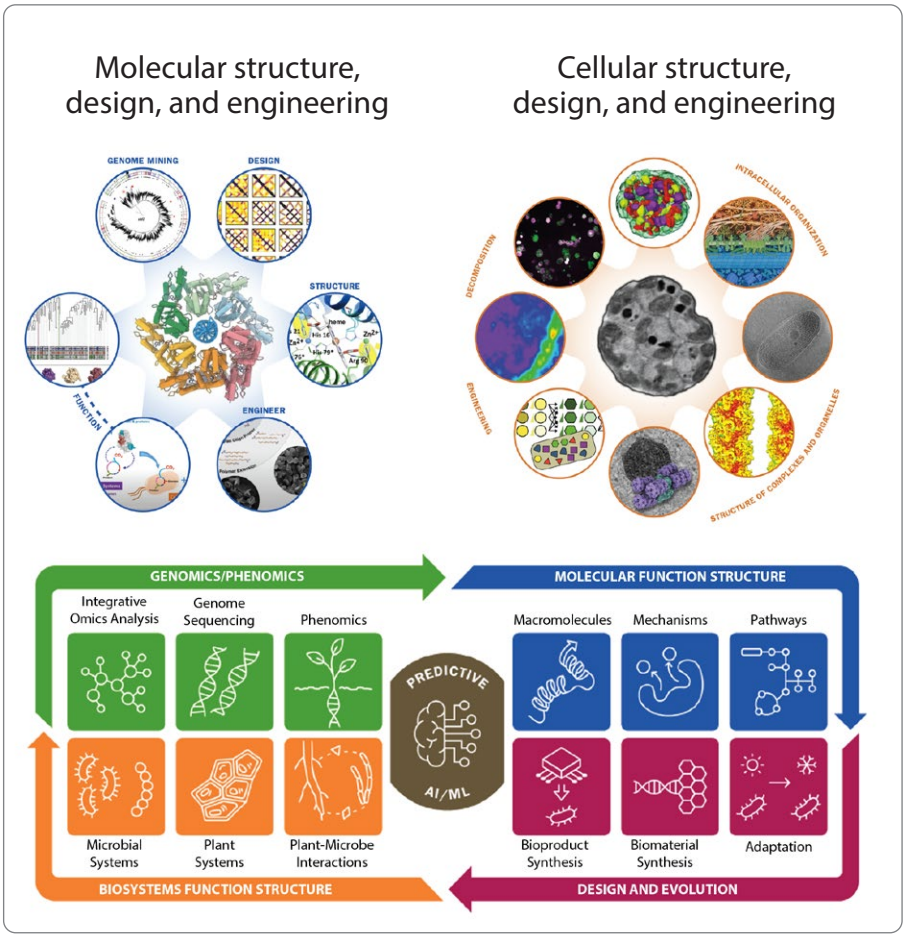
cutting-edge transdisciplinary science. These collaborations include not only JGI and EMSL but also user resources for (1) BER structural biology and imaging at synchrotron and neutron facilities supported by DOE's Basic Energy Sciences program and for (2) computing capabilities at the Advanced Scientific Computing Research program's high-performance computing facilities.

To explore the need for the BER research community to combine genomic, functional, systems, biosystems design, structural, and computation approaches to advance their research, researchers performed various pilot programs to test the feasibility of offering the integrative capabilities to users and hosted workshops to better understand users' needs. This report describes ongoing efforts among DOE user facilities to enable user research.

Funding Statement: Lawrence Berkeley National Laboratory is managed by the Regents of the University of California for DOE under contract DE-AC02-05CH1123.

Reference

U.S. DOE. 2022. *Genomes to Structure and Function Workshop Report 2022*, OSTI ID 1959294. U.S. Department of Energy Office of Science. DOI:10.2172/1959294.



Collaboration Between DOE User Facilities. Top left: Structural proteomics workflows. Discovering and engineering proteins with novel or desired functions is one opportunity and challenge identified during the breakout session. **Top right:** Building interactive and holistic cell models. Holistic cell models were identified as one challenge and opportunity during the breakout sessions. **Bottom**: Structure-informed discovery and design cycle. [Courtesy DOE Joint Genome Institute]

TerraForms User Program – Capability Development, User Science, Collaborations, and Deployment

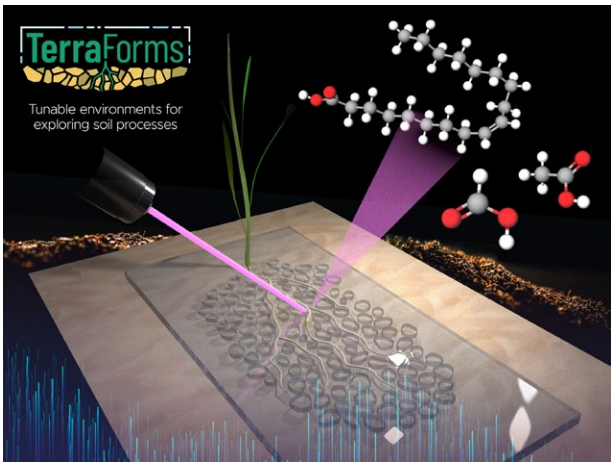
Arunima Bhattacharjee[1]* (arunimab@pnnl.gov, PI), Jayde Aufrecht[1], Jocelyn Richardson[2], Vimal Balasubramanian[1], Natalie Sadler[1], Nicholas Reichart[1], Ryan McClure[1], Christopher Anderton[1], Erin Nuccio[3], Ritimukta Sarangi[2], Jennifer Pett-Ridge[3], Kirsten Hofmockel[1]

[1]Pacific Northwest National Laboratory; [2]SLAC National Accelerator Laboratory; [3]Lawrence Livermore National Laboratory

Soils are highly heterogeneous and opaque and a challenging medium for spatial and *in situ* characterization. Yet spatial characterization of soils is essential to comprehend the pore scale biogeochemistry that contributes to ecosystem-scale events. Reduced complexity platforms, such as soil micromodels that emulate soil properties, have demonstrated the potential for *in situ* visualization of microbial and plant growth dynamics. Inspired by previous studies, this project developed "TerraForms", a platform that replicates specific physical and chemical properties of soils, reducing the complexity of analysis for soil-driven processes. TerraForms support plant and microbial growth yet allow the combined use of mass spectrometry and spectroscopic imaging methods to characterize processes in a realistic soil-like environment.

TerraForms are manufactured from an ultraviolet curable resin with a conductive indium tin oxide–coated glass backing using a combination of Bosch etching and soft lithography techniques. The soil habitats can be custom built to reproduce the heterogeneous pore size distribution of soil properties, including soil pores and soil microaggregates. Moreover, research can amend the soil habitats with soil minerals such as potassium (K) feldspar, hematite, and kaolinite to study effects of mineralogy on microbial and plant growth. The study has demonstrated that these polymer glass-based platforms are compatible with mass spectrometry imaging techniques, X-ray photoelectron spectroscopy, scanning electron microscopy/energy dispersive X-ray analysis, and synchrotron-based techniques such as X-ray fluorescence (XRF) and X-ray absorption near edge structure spectroscopy (XANES).

This presentation will highlight results from collaborations built over the past years that helped create the TerraForms capability and user program. These collaborations include early adopters and advisors who contributed to the creation of several of the TerraForms platforms and helped design rigorous tests that ensured



TerraForms. TerraForms are tunable environments for exploring soil processes. [Courtesy Pacific Northwest National Laboratory]

workflow development and deployment to the user community, including Pacific Northwest National Laboratory's (PNNL) Soil Microbiome Science Focus Area (SFA) and the Trail Ecosystems for the Advancement of Microbiome Science (TEAMS) group. Deployment to the user community has included several outreach activities, such as Environmental Molecular Sciences Laboratory webinars, workshops, and TerraForms publications, to train users how to use these platforms and downstream sample preparation for multimodal analysis. This talk will demonstrate specific examples of user science that used TerraForms for their work as well as collaborations with Stanford Synchrotron Radiation Lightsource, Lawrence Livermore National Laboratory's carbon negative Earthshot (TerraForming Soil Energy Earthshot Research Center), and PNNL's Soil Microbiome SFA and TEAMS. Specifically, this talk will demonstrate the use of mineral micromodels for visualizing mineral-derived K uptake and transformation by fungi under drought. This presentation will also discuss using mineral RhizoChips to study mineral transformations and inorganic nutrient uptake by arbuscular mycorrhizal fungi. Finally, some of the newer developments within the TerraForms capability will be discussed, such as creating workflows for developing TerraForms representative of soil cores collected from field sites and upscaling pore scale biogeochemistry.

*Presenter

From Roots to Atmosphere: Utilizing DOE Facilities to Track Drought's Carbon Impact on Ecosystem Processes

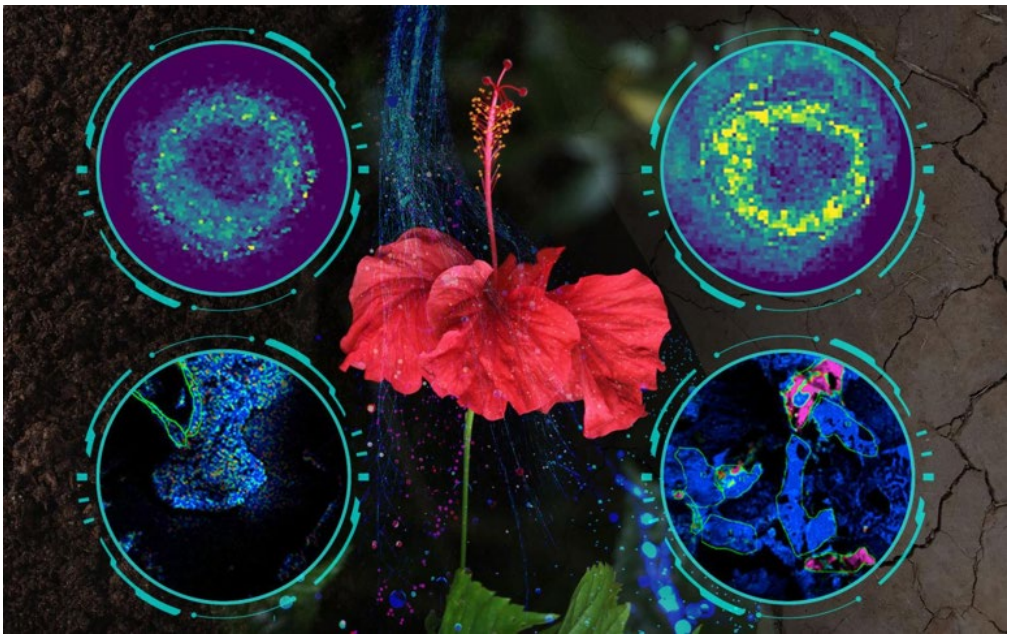
Malak M. Tfaily[1]* (tfaily@arizona.edu, PI), Laura Meredith[1], Linnea Honeker[1], Gina Hildebrand[1], Dusan Velickovic[2], David Hoyt[2], Rosalie Chu[2], Jason Toyoda[2], Christopher Anderton[2], Ben Yang[1], Mekayla Crawford[1]

[1]University of Arizona; [2]Pacific Northwest National Laboratory

Utilizing DOE user facilities, this team investigated plant-soil-microbe interactions that drive rhizosphere processes contributing to metabolite turnover and nutrient cycling. As climate warming increases water scarcity frequency and severity, understanding how plant-mediated processes like root exudation influence rhizosphere soil organic matter turnover is critical. Using 16S rRNA gene amplicon sequencing, rhizosphere metabolomics, and position-specific carbon-13 pyruvate labeling, researchers examined three tropical plant species (*Piper auritum*, *Hibiscus rosa sinensis*, and *Clitoria fairchildiana*) and their associated microbial communities' effects on rhizosphere soil organic carbon turnover.

The findings indicate the rhizosphere metabolome is primarily shaped by the roots' drought response rather than direct rhizosphere bacterial community composition shifts. Specifically, reduced root exudation notably affected the *P. auritum* rhizosphere metabolome, with less reliance on neighboring microbes. Contrary to *P. auritum*, *H. rosa sinensis* and *C. fairchildiana* experienced exudate composition changes during drought, altering bacterial communities and collectively impacting the rhizosphere metabolome. Furthermore, excluding phylogenetically distant microbes shifted the rhizosphere metabolome. Under drought, *C. fairchildiana* associated with only a subset of symbiotic bacteria.

These results indicate plant species-specific microbial interactions systematically change with the root metabolome. As roots respond to drought, associated microbial communities adapt, potentially reinforcing plant roots' drought tolerance strategies. These findings have significant implications for maintaining plant health and performance during drought stress and improving plant performance under climate change.



Hibiscus rosa sinensis. Scientists are studying this tropical beauty to learn more about the ways it and its neighbors respond to drought. [Courtesy Pacific Northwest National Laboratory]

Utilizing Metaomic and Chemical Imaging Approaches to Explore the Dynamics of Nutrient Exchange and Environmental Adaptation Mediated by Plant–Fungal Symbiosis

Hui-Ling Liao[1]* (sunny.liao@ufl.edu), Haihua Wang[1], Kaile Zhang[1], Ryan Tappero[3], Tiffany Victor[3], Rytas Vilgalys[4], Kerrie Barry[5], Keykhosrow Keymanesh[5], Sravanthi Tejomurthula[5], Igor V. Grigoriev[5], William R. Kew[6], Elizabeth K. Eder[6], Carrie D. Nicora[6], John Cliff[6], Jeremy Bougoure[6], Dehong Hu[6], Sarah Leichty[6], Edward Brzostek[7], **Jennifer Bhatnagar[2]** (PI)

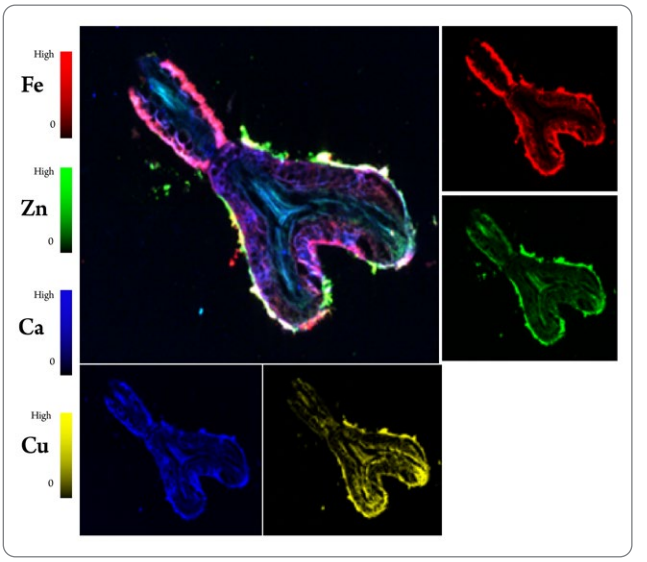
[1]University of Florida; [2]Boston University; [3]Brookhaven National Laboratory;[4]Duke University; [5]DOE Joint Genome Institute; [6]Pacific Northwest National Laboratory; [7]West Virginia University

soils.ifas.ufl.edu/microbiome

Interactions among terrestrial free-living microorganisms, root symbiotic microbes, and plants play a crucial role in modulating biogeochemical cycling within ecosystems. However, the dynamic mechanisms of these interactions and their responses to changing environmental conditions, such as soil nutrient status and stresses are poorly understood. Consequently, terrestrial nutrient fluxes in ecosystem models often overlook plant–symbiont–microbe interactions and cannot accurately reflect the impacts of environmental changes on their functions. Using pine trees as a model, researchers employ advanced metaomic, NanoSIMS and X-ray fluorescence (XRF) tools to uncover the mechanisms by which plants, ectomycorrhizal fungi, and free-living soil microorganisms interact to shape nutrient exchange and decomposition processes.

This experiment was established in a phytotron-based synthetic ecosystem for pine and its soil microbiomes. Treatments include the combination of soil sources harboring diverse soil biomes, elevated carbon dioxide (CO_2) levels, soil organic matter (SOM), inorganic nitrogen (N), iron (Fe), and pine-associated ectomycorrhizal fungi (EMF).

Findings from the subset of these treatment combinations indicated that an enriched abundance of EMF facilitated soil molecular and extracellular enzyme activities, involved in soil carbon (C), N, and Fe processes. NanoSIMS images revealed the ability of EMF-extended hyphae to acquire ^{15}N from ^{15}N-labeled SOM. Metaomic studies demonstrated that



X-Ray Fluorescence. Images indicate the uptake of iron (Fe), zinc (Zn), calcium (Ca), and copper (Cu) in a mycorrhizal root tip of *Pinus taeda* inoculated by *Suillus cothurnatus* under conditions of relatively low inorganic nitrogen and high soil organic matter. [Courtesy University of Florida and Brookhaven National Laboratory]

elevated CO_2 levels enhanced EMF pathways involved in symbiosis-mediated SOM decomposition and N transfer from the soil to the host plants.

However, the treatments with inorganic N or Fe compromised symbiotic C/N processes; such negative impacts can be mitigated by enhancing the diversity of the EMF community. The ongoing study integrating data from metaomic analyses and XRF will discuss the interactive effects of SOM, N, and Fe on free-living microorganisms and EMF-mediated nutrient fluxes.

The team's next goal is to integrate these data into an ecosystem model to represent plant–microbe interactions under global change scenarios and their feedback to ecosystem-level nutrient fluxes. The team anticipates that this refined model will better capture the impacts of plant–microbe interactions on biogeochemical cycling in changing environments and inform forest management practices.

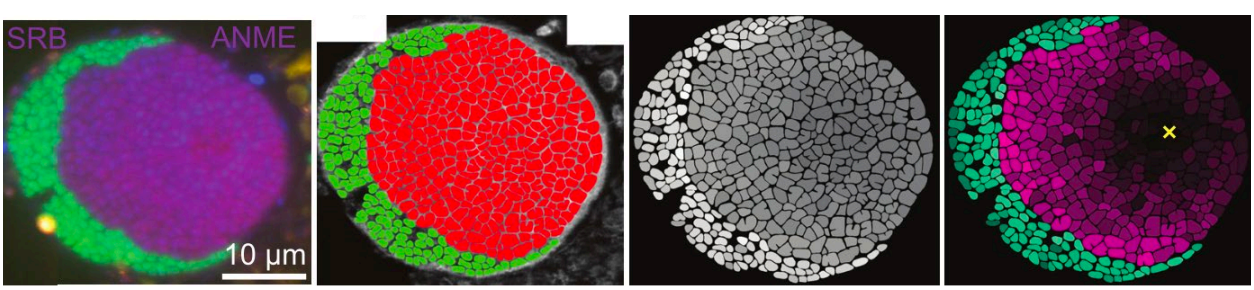
*Presenter

A Single Consortium View of Syntrophic Methane Oxidation by Environmental Archaea and Bacteria

Victoria Orphan[1]* (vorphan@caltech.edu, PI), Dan Utter[1], Yongzhao Guo[1], John Magyar[1], Ranjani Murali[4], Roland Hatzenpichler[2], Gray Chadwick[3], Hang Yu[10], Shawn McGlynn[6], Rex Malmstrom[5], Danielle Goudeau[5], Mark Ellisman[7], Guy Perkins[7], Tom Deerinck[7], Jocelyn Richardson[8], Sam Webb[9]

[1]California Institute of Technology; [2]Montana State University; [3]University of California–Berkeley; [4]University of Nevada–Las Vegas; [5]DOE Joint Genome Institute; [6]Tokyo Institute of Technology; [7]University of California–San Diego; [8]SLAC National Accelerator Laboratory; [9]Stanford Synchrotron Radiation Lightsource; [10]Peking University

The syntrophic anaerobic oxidation of methane (AOM) between methanotrophic archaea and sulfate-reducing bacteria is a critical microbially controlled process in the global methane cycle. These metabolic partnerships are diverse, representing several genera and species-level clades that co-occur in the same environment, segregated into individual two-member consortia believed to consist of a single archaeal and bacterial lineage. The high strain level diversity in sediments harboring AOM consortia frequently complicates the bioinformatic assembly of high-quality genomes from metagenomic datasets and the field currently lacks pure culture representatives for detailed physiological study. Working with the Joint Genome Institute (JGI), Environmental Molecular Sciences Laboratory (EMSL), and the Stanford Synchrotron Radiation Lightsource (SSRL), this research team has been developing culture-independent strategies to investigate AOM consortia. Through support from the Facilities Integrating Collaboration for User Science program with JGI, the team developed flow cytometry sorting and sequencing protocols for individual AOM consortia from sediment in combination with click-chemistry enabled fluorescent tagging of translationally active microbes using a technique called BONCAT (BioOrthogonal Noncanonical Amino Acid Tagging). Application of the BONCAT-FACS method to these environmental methane-fueled microbial communities has enabled novel genomic insights into specific syntrophic partner pairings and hidden variation in metabolic potential (e.g., nitrogen fixation) among co-existing strains of methanotrophic archaea. To complement these genome-guided analyses, researchers are working with SSRL and collaborators at National Center for Microscopy and Imaging Research to develop multimodal imaging for environmental AOM consortia that enables various combinations of fluorescence, electron, X-ray (μ-XRF, XANES), and secondary ion (NanoSIMS) imaging. As part of this effort, researchers have been testing SSRL's upgraded 14-3 beamline for high resolution analysis of sulfur distribution and speciation in methane and sulfate-respiring consortia, enabling testing of hypotheses regarding the underlying syntrophic mechanism(s) of AOM. Collectively, these molecular, microscopy, and chemical/isotopic imaging approaches are providing key insights into the physiology, ecology, and evolution of this globally important syntrophic partnership and importantly, offer a methodological roadmap that can be used with diverse microbial ecosystems.



Overview of AOM Consortium Structure, NanoSIMS Data Acquisition and Single Cell Analysis of $^{15}NH_4$ Assimilation Activity. Fractional abundance of ^{15}N is calculated as $^{15}N^{12}C-/(^{15}N^{12}C-+^{14}N^{12}C-)$. **Left to Right:** Fluorescence *in situ* hybridization (FISH) image indicating phylogenetic identity of cells (green, bacterial probe; red, anaerobic methanotrophic archaea (ANME)-2 archaea-specific probe); **2nd Panel:** Segmentation image showing sulfate-reducing bacteria (SRB) and ANME cells manually segmented based on observation of FISH and NanoSIMS data; **3rd Panel:** Individual segmented cells shaded by their total ^{15}N fractional abundance; **4th Panel:** SRB and ANME cells scaled by minimum and maximum ^{15}N enrichment values within the consortium. The yellow X marks the approximate minimum ANME cell activity several cell lengths away from their nearest syntrophic SRB partner. [Modified from He, X., et al., 2021. "Controls on Interspecies Electron Transport and Size Limitation of Anaerobically Methane-Oxidizing Microbial Consortia," *MBio* **12**(3), 10-1128. Republished under Creative Commons license (CC BY 4.0)]

CHAPTER 2

Quantum Imaging and Sensing

Fundamental quantum science–enabled research can overcome current challenges related to suboptimal stability and photobleaching, enabling prolonged imaging studies. For example, quantum-entangled pairs of single photon–emitting probes can potentially offer an advantage over conventional methods by enhancing spatial and temporal resolution, measurement speed, long-term sample stability, or bioimaging technology sensitivity. This chapter describes research on imaging probes, detectors, and sensors related to BER science.

Biological Imaging Using Entangled Photons

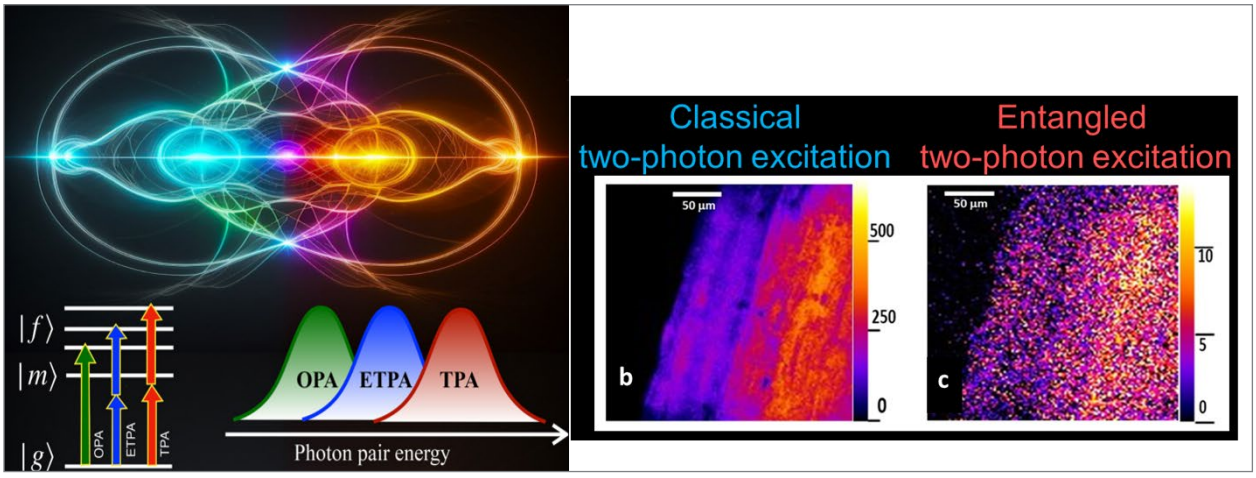
Theodore Goodson III[1]* (tgoodson@umich.edu, PI), George C. Schatz[2], Jennifer Morrell-Falvey[3], Ying-Zhong Ma[3]

[1]University of Michigan; [2]Northwestern University; [3]Oak Ridge National Laboratory

sites.lsa.umich.edu/goodgrp

The organization and function of microbial communities in the plant rhizosphere are influenced by interactions among organisms within a complex and dynamic physical and chemical environment. This environment, which includes growing plants, microbial communities, complex metabolites, and nutrient gradients, cannot be characterized in a noninvasive and nondestructive manner using current bioimaging technologies. Imaging with quantum light offers the opportunity to realize long-term nondestructive imaging investigations of new species in the plant-microbial system.

In this talk, the reports of experiments with quantum light based on entangled two-photon absorption for specific probes will be discussed. In addition, group members describe new work concerning theoretical predictions of important cross-sections for the probes as well as the construction of a pump pulse-shaping setup for entangled photon generation and dispersion compensation.



Two-Photon Absorption and Excitation. Left: Schematic representation of the entangled two-photon absorption peak with respect to the absorption peak position for classical two-photon absorption (TPA) and for one-photon resonant absorption (OPA). **Right:** The first entangled two-photon emission images of a plant root system. [Left: Republished with permission from Varnavski, O., et al., 2023. "Colors of Entagled Two-Photon Absorption," *Proceedings of the National Academy of Sciences* **120**(35), e2307719120 under Creative Commons license (CC BY-NC-ND); Right: Courtesy University of Michigan]

Nondestructive, Three-Dimensional Imaging of Processes in the Rhizosphere Utilizing High-Energy Photons

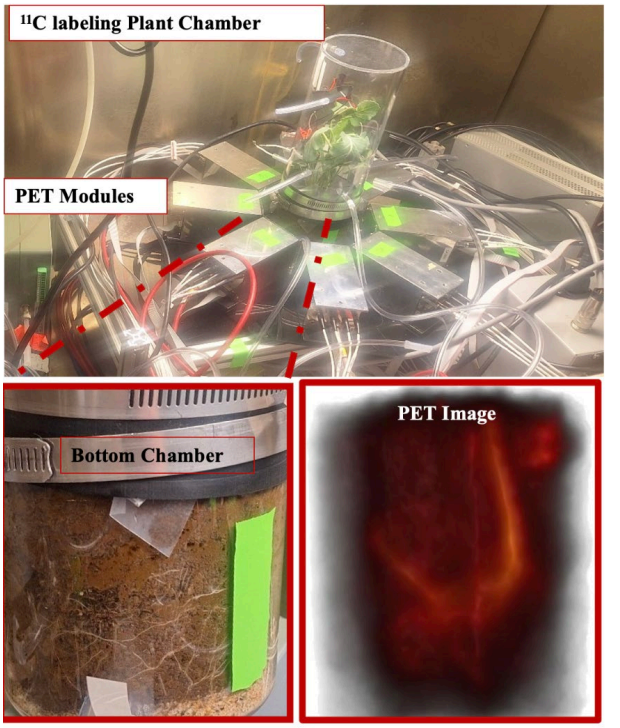
Shiva Abbaszadeh[1]* (sabbasza@ucsc.edu, PI), Weixin Cheng[1], Craig S. Levin[2], Adam S. Wang[2]

[1]University of California–Santa Cruz; [2]Stanford University

Root–soil interactions play pivotal roles in biogeochemical processes from local scales to the global scale. The movement and transformation of nutrients, water, and soil organic matter in the root-soil interface or the rhizosphere are commonly referred to as rhizosphere processes, which include nutrient cycling, water fluxes, and carbon sequestration. Because of these crucial roles of the rhizosphere, better understanding of rhizosphere dynamics is highly needed. To augment scientists' capabilities to study the rhizosphere, this research group is exploring the potential of utilizing positron emission tomography (PET) and computed tomography (CT) for dynamic 3D imaging of intact rhizospheres. The CT system can provide high-resolution structural information, such as soil aggregates and pore networks, and the PET system can provide functional information at a super high temporal resolution by using a carbon-11 labeled carbon dioxide tracer ($^{11}CO_2$).

The group has designed and developed a ^{11}C labeling and tracing chamber inside the fume hood (130 × 60 × 80 cm^3) located at the Cyclotron and Radiochemistry Facility at Stanford University. In collaboration with Jefferson Laboratory, a PET system consisting of eight PhytoPET modules was built and tested using the ^{11}C labeling and tracing system (see figure, this page). Currently, this research team is performing image normalization and attenuation correction to understand the quantitative accuracy of the PET system. Once the PET system is ready for further experimenting, researchers will demonstrate its capability using a plant-soil system with a different soil matrix to analyze the 3D dynamic coordination between rhizospheric hotspots from PET imaging and the 3D soil matrix from CT imaging. These rhizospheric hotspots often represent a very small fraction of the soil volume but can be responsible for a substantial portion of the overall biogeochemical functions. Current methods do not have the capability to visualize these crucial hotspots inside intact soil matrix.

An amorphous selenium (a-Se) direct conversion detector on a scalable Complementary Metal-Oxide-Semiconductor (CMOS) readout for a large capability is under development at University of California–Santa Cruz for high spatial resolution CT imaging (<20 micrometer resolution). The potential capability of the proposed hybrid detector for imaging very fine soil structures at less than 20 μm resolution has been demonstrated based on a previously developed prototype [1k × 1k Readout Integrated Circuit (ROIC) by KA Imaging]. Successful development of accessible PET/CT systems for rhizosphere imaging and direct observation of belowground ecosystems can reveal the puzzling complexity and crucial Earth system functions of root–soil systems. This new capability also has important implications for other disciplines in Earth system science.



^{11}C Labeling and Tracing Chamber. Shown inside the fume hood with eight PhytoPET detector modules surrounding the bottom portion of the chamber, which is sealed/separated from the top chamber. The carbon-11 labeled carbon dioxide tracer gas is release to the top chamber containing plant canopy. The root soil is contained in the bottom chamber; the ^{11}C activity position emission tomography image of the bottom chamber is shown, which represents active root systems and the soil in the rhizosphere network receiving root exudates. The bright curved picture on the right and the bottom portions of the image shows the tap root and bundles of fine roots. [Courtesy University of California–Santa Cruz]

*Presenter

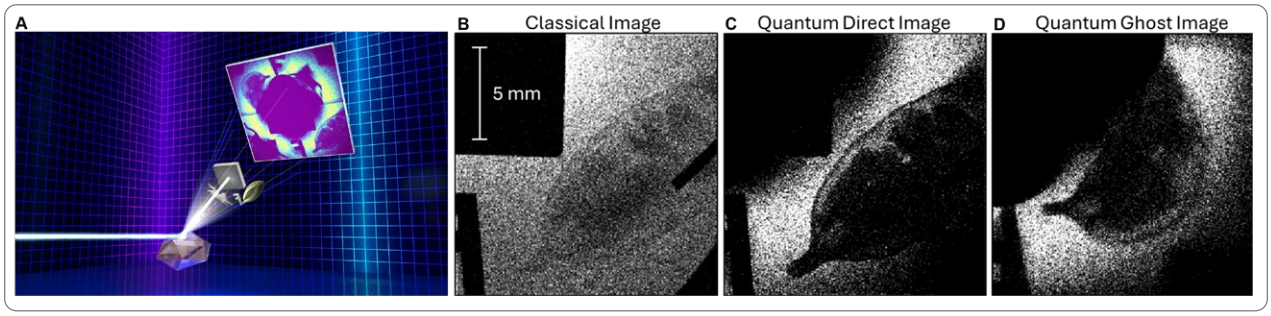
Progress Toward a Quantum Enhanced X-Ray Microscope

Justin C. Goodrich[1]* (jgoodrich@bnl.gov), Ryan Mahon[1], Joseph Hanrahan[1], Lonny Berman[1], Andrei Fluerasu[1], Monika Dziubelski[1], Raphael A. Abrahao[1], Sanjit Karmakar[2], Cinzia DaVià[2], Andrei Nomerotski[3], **Sean McSweeney[1]** (PI)

[1]Brookhaven National Laboratory; [2]Stony Brook University; [5]Florida International University

The use of the quantum properties of light offers a new opportunity for imaging, in that the use of quantum correlations of the (two-photon) system allows for image construction with minimal dose requirements. The method primarily consists of generating two photon beams; of these, only one beam interacts with the object being measured. The resulting image and other observable effects are then discerned by examining the spatial correlations between these two beams. This team's most recent investigations have explored the use of spontaneous parametric down-conversion spontaneous parametric down-conversion of hard X-rays as a source of a biphoton X-ray state strongly correlated in position and energy.

In this work, the group shares the world record rates of generating such correlated X-ray pairs using the high brightness of the National Synchrotron Light Source II and cutting-edge Timepix detector technology. Furthermore, researchers demonstrate a proof-of-concept quantum correlation imaging technique to image several objects, including a biological object—an *Elettaria cardamomum* seed. This work is a significant milestone in the field of X-ray quantum ghost imaging and provides a pathway to demonstrating the "quantum advantage" in this energy regime.



Quantum X-Ray Correlation Imaging. (A) Nonlinear X-ray diffraction spontaneous parametric down conversion imaging setup: 15 kiloelectron volt (keV) pump X-rays from a synchrotron light source generate spontaneous parametric down-conversion biphotons from a diamond crystal. Tungsten objects shaped like a cat and the letter "F", along with an *Elettaria cardamomum* seed, are positioned inside the detector ring to block the lower chips; coincidence measurements produce a quantum direct correlation image and a deformed quantum ghost image on the upper chips. **(B)** Classical reference image of the cardamom seed using scattered 15 keV X-rays. **(C)** Quantum direct correlation image of the seed on the signal detector, with photon energies from 5 to 9 keV. **(D)** Quantum ghost correlation image of the seed on the idler detector, adjusted for alignment, showing distortion from non-degeneracy of the X-ray photons. [Courtesy Brookhaven National Laboratory]

Quantum-Entangled Hyperpolarized Spin States for Noninvasive Imaging of Nitrogen Assimilation

Thomas Theis* (ttheis@ncstate.edu, PI)

North Carolina State University

theislab.wordpress.ncsu.edu

Nitrogen (N) fertilizer synthesis for agriculture sustains about half of the human population. Recent studies show that N input from N-fertilizer synthesis and river runoff poses a serious and growing problem with intensifying climate change. To address these major societal challenges, improvements to today's agricultural strategies are necessary; however, the scientific community's knowledge of plant–soil–microbe interactions in unperturbed soil remains extremely limited because of lacking noninvasive technology to probe metabolism. Here, this team develops new, noninvasive quantum sensing to directly observe metabolic transformation in the rhizosphere to acquire currently inaccessible knowledge.

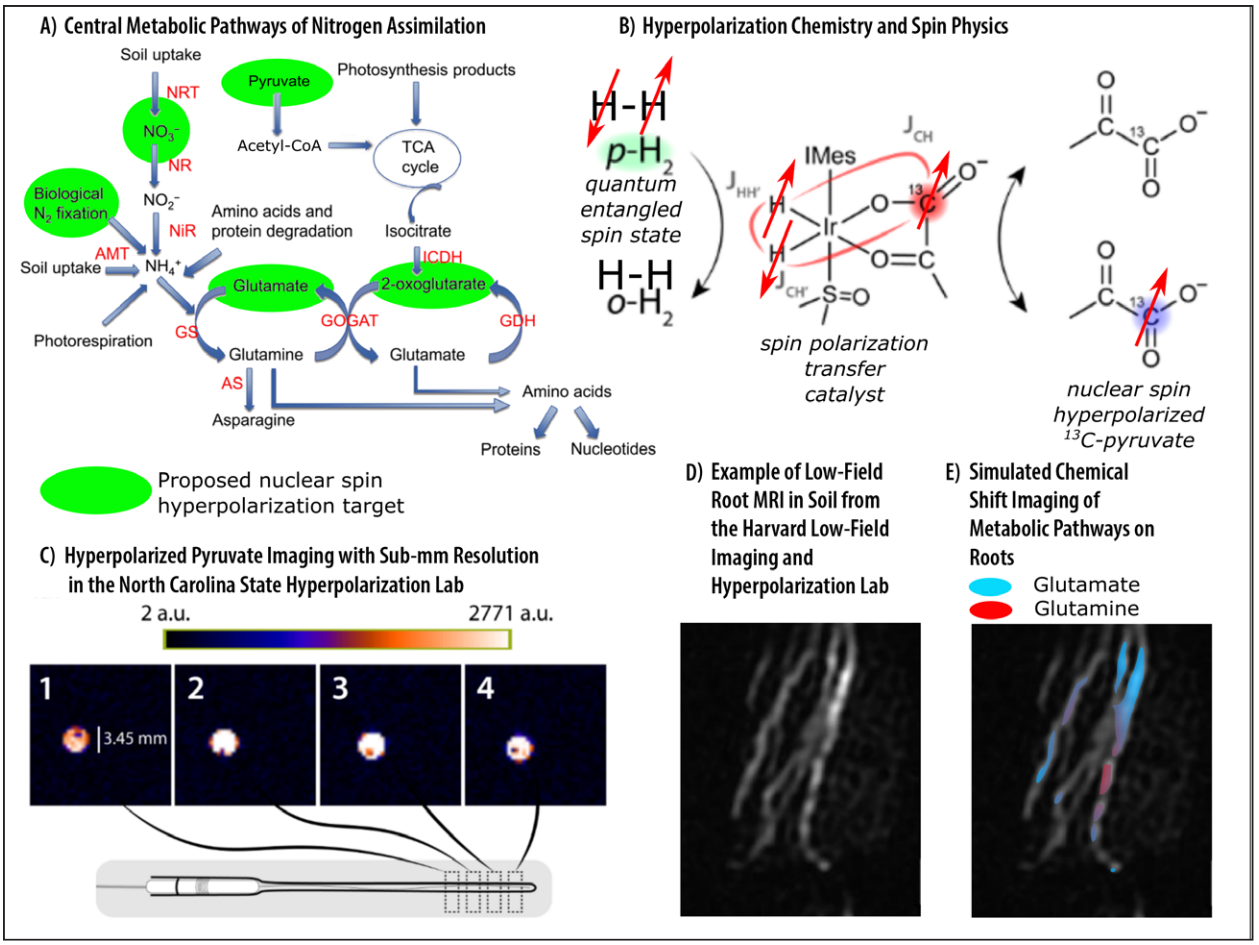
This group's approach takes advantage of the quantum-entangled nuclear spin state in hydrogen gas (i.e., parahydrogen), which can readily be enriched to ~99.5% to enhance magnetic resonance imaging (MRI) signals by up to seven orders of magnitude, such that metabolites at low, physiological concentrations become detectable by MRI even at low magnetic fields. The figure (see p. 14; A) shows the target metabolites central to N assimilation and their metabolic pathways; (B) illustrates hyperpolarization chemistry, which transfers the nuclear spin hyperpolarization to the metabolites of interest; (C) shows images of the hyperpolarized metabolite at low concentrations and low field in a test tube; (D) shows an example of noninvasive low-field MRI; and (E) illustrates the goal of obtaining molecular imaging of metabolic transformations in the rhizosphere with portable devices that could be employed directly in the field.

This poster shows significant progress towards several goals:

- The team has significantly expanded the substrate scope of its parahydrogen-based hyperpolarization technique.
- The team has detailed the fundamental quantum chemistry and spin dynamics governing the polarization transfer processes.
- The team has demonstrated MRI of the hyperpolarized molecular sensors at low magnetic fields and low concentrations.
- The team has enabled the detection of long-lived carbon-13 hyperpolarized metabolites on existing low-field hydrogen MRI systems that are already deployed in crop fields.
- The team has demonstrated the very first detection of metabolic conversions with this technique using cryogen-free MRI systems.

In summary, exciting progress toward noninvasive imaging of nitrogen assimilation using quantum-entangled hyperpolarized spin states will be presented.

*Presenter



Using Quantum-Entangled Nuclear Spin States for Noninvasive Magnetic Resonance Imaging of Nitrogen Assimilation in the Rhizosphere. **(A)** Metabolic conversions of relevance in nitrogen assimilation. **(B)** Hyperpolarization chemistry: the entangled nuclear spin state of hydrogen is used to induce spin alignment on a metabolite (pyruvate) mediated by a polarization transfer catalyst. **(C)** Recent results demonstrating pyruvate imaging at a low mm concentration with sub-mm resolution in the Theis Lab. **(D)** Low-field root magnetic resonance imaging (MRI) images in soil by Rosen Lab. **(E)** Simulated chemical shift images that report on metabolism and molecular turnover from glutamate to glutamine. [(A) Adapted and reused under a Creative Commons license (CC by 4.0) from Lu, J., et. al. 2016. "Expression of a Constitutively Active Nitrate Reductase Variant in Tobacco Reduces Tobacco-Specific Nitrosamine Accumulation in Cured Leaves and Cigarette Smoke," *Plant Biotechnology Journal* **14**, 1500–10. (B) Courtesy North Carolina State University. (C) Adapted with permission from TomHon, P., et al. 2022. "Temperature Cycling Enables Efficient ^{13}C SABRE-SHEATH Hyperpolarization and Imaging of [1-^{13}C]-Pyruvate," *Journal of the American Chemical Society* **144**(1), 282–87. DOI:10.1021/jacs.1c09581. (D, E) Bagnall, C., et al. 2020. "Low-Field Magnetic Resonance Imaging of Roots in Intact Clayey and Silty Soils," *Geoderma* **370**, 114356. DOI:10.1016/j.geoderma.2020.114356. Reused under a Creative Commons license (CC BY-NC-ND 4.0)]

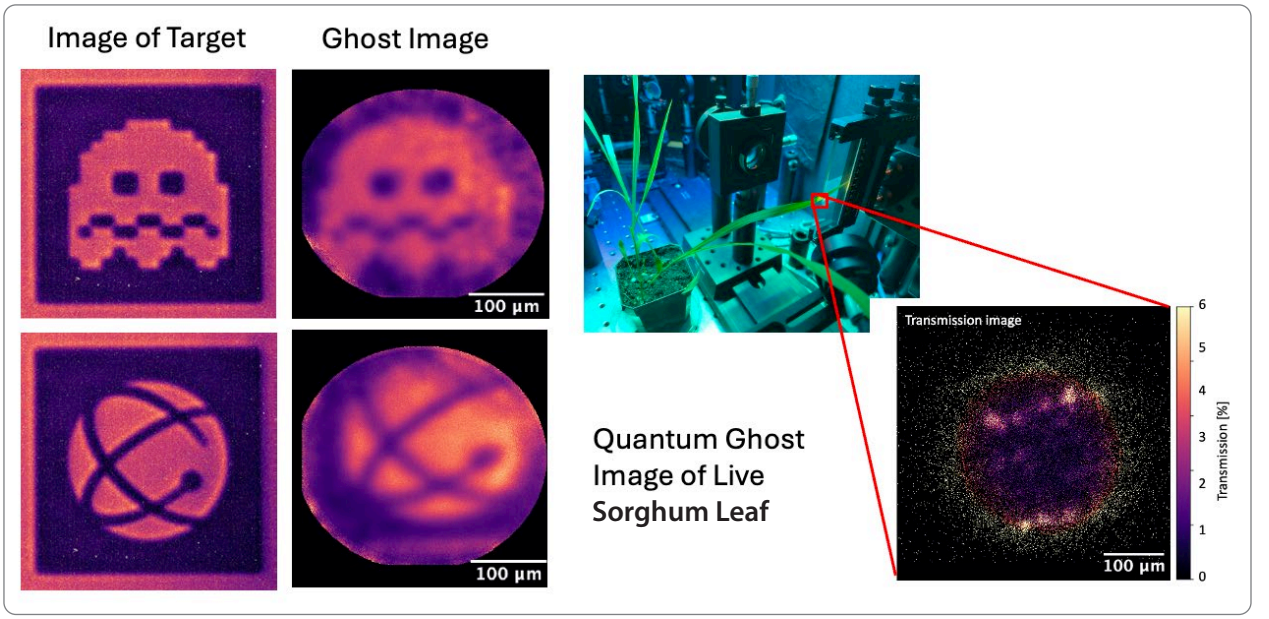
Quantum Ghost Imaging of Water Content and Plant Health with Entangled Photon Pairs

James Werner[1]* (jwerner@lanl.gov, PI), Duncan Ryan[1], Kristina Meier[1], Kati Seitz[1], David Thompson[1], Rebecca Holmes[1], Demosthenes Morales[1], Buck Hanson[1], David Hanson[2], Peter Goodwin[1], Raymon Newell[1], David Thompson[1]

[1]Los Alamos National Laboratory; [2]University of New Mexico

The near infrared (NIR) and mid-infrared (MIR) portions of the electromagnetic spectrum are sensitive to absorption features of specific molecular bonds and chemical species in a sample. For example, lignan and proteins in plants have specific absorption signatures in the NIR. However, because detectors are inefficient in the NIR and MIR, infrared spectroscopy requires high light levels to overcome detector limitations. Cameras in particular do not perform well in this spectral range, and microscopy methods such as Fourier transform infrared spectroscopy typically rely on scanning confocal arrangements with single-element detectors to spatially map chemical information.

To overcome these limitations, the team developed and exploited a new quantum ghost imaging microscope for obtaining absorption measurements in the NIR without the need of scanning or high light intensities. The team reports on the use of a novel detector—NCam—in quantum ghost imaging using non-degenerate photon pairs generated by spontaneous parametric down conversion (SPDC). NCam records single-photon arrival events with ~100 picosecond resolution, enhancing the correlation window of SPDC pairs over previous wide-field ghost imaging by 30-fold. This permits ghost imaging of living and intact plant samples at light levels lower than what the plants would experience from starlight. For photosynthesizing organisms, this low-light imaging method enables the study of plants without disturbing or eliciting responses from the plant due to the measurement itself. Following the development of quantum ghost imaging in the near IR to visualize binary test targets, the team has demonstrated imaging of water content in live sorghum plants (see figure, this page).



Quantum Ghost Imaging. Left: A near infrared image of selected binary targets (a Pac-Man ghost and the Los Alamos National Laboratory logo) and their corresponding quantum ghost images. Ghost images were acquired by probing the target at ~1450 nm with one entangled photon, with image formation performed at ~540 nm with the corresponding visible entangled photon pair. **Right:** A picture of a live sorghum plant in our quantum ghost imaging microscope and the corresponding ghost image. Similar to the binary targets on left, image formation is at ~540 nm, while sample is probed with ~1450 nm light. Rows of highly transmissible regions in the sorghum leaf are most likely stomata. Images acquired with an entangled photon flux orders of magnitude less than ambient starlight. [Courtesy Los Alamos National Laboratory]

*Presenter

The 3DQ Microscope: A Novel System Using Entangled Photons to Generate Volumetric Fluorescence and Scattering Images for Bioenergy Applications

Audrey Eshun[1]* (eshun2@llnl.gov), Dominique Davenport[1], Brandon Demory[1], Paul Mos[2], Yang Lin[2], Sam Jeppson[1], Ashleigh Wilson[1], Shervin Kiannejad[1], Tiziana Bond[1], Mike Rushford[1], Chuck Boley[1], Edoardo Charbon[2], Claudio Bruschini[2], **Ted Laurence[1]** (PI)

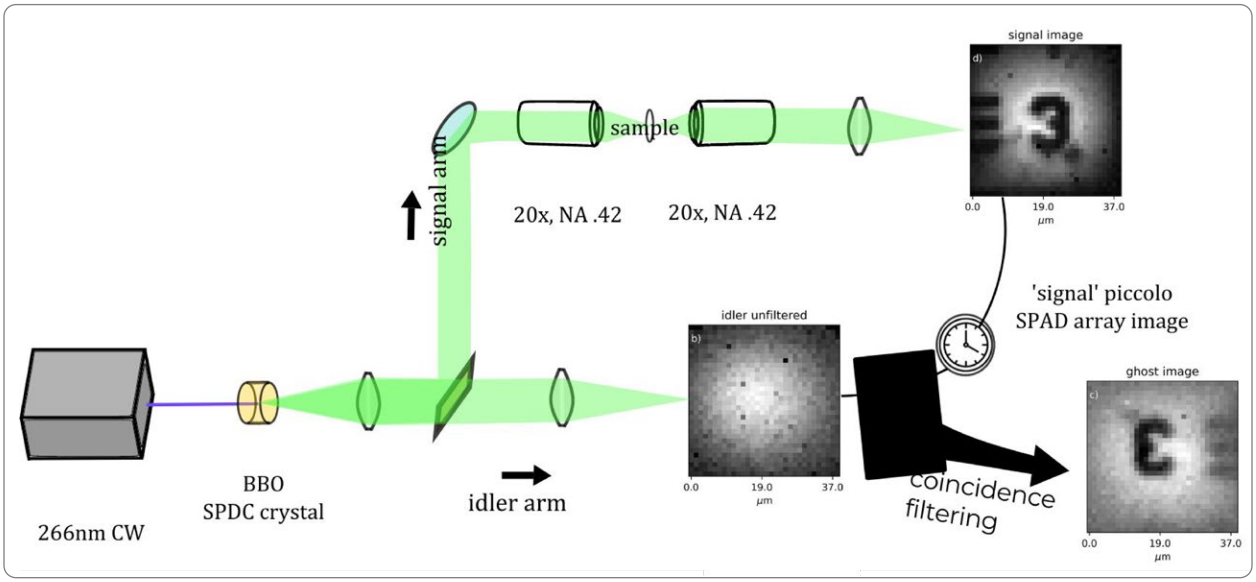
[1]Lawrence Livermore National Laboratory; [2]École Polytechnique Fédérale de Lausanne

In the study of biological systems, real-time 3D microscopy is an important tool in understanding how live cells move and interact with other cells, microbes, and other external elements. Although these dynamics can currently be studied with confocal and light sheet microscopy, for example, these approaches require scanning either the beam or the sample, exposing the sample to higher excitation energies and limiting time resolution of the imaging process.

An alternative approach that is ideal for dynamic information is to simultaneously capture the scene from two perspectives. This limits the time resolution only by the acquisition rate of each sensor. However, recreating the 3D scene from two views requires correlating features in both views, posing challenges at higher densities.

This study proposes that quantum-entangled light can provide the 3D information using a new detection architecture while keeping peak excitation intensity low to preserve the integrity of the sample and avoiding biases that can be caused by scanning. Quantum-entangled light can be spatially separated while preserving the momenta and temporal relations of entangled photons.

Utilizing quantum-entangled light generated by a beta barium borate crystal and the concepts of quantum ghost imaging, the team presents a microscope that views microscopic specimens in 3D from a single snapshot. This is achieved by utilizing two event-based 2D sensors in which information is relayed, through correlations, to generate two perspectives from a single scene. The quantum-entangled light source allows researchers to correlate signal and idler, both spatially and temporally. Group members characterize the



Simplified Ghost Image. Setup showing the direct signal image and the ghost image obtained after coincidence filtering. [Courtesy Lawrence Livermore National Laboratory]

microscope by imaging resolution targets at various depths, using models to guide them in the optimization of the crystal orientation, and attempt to understand the achievable depth of field with the specific light source.

After characterization, researchers image gold nanoclusters with the ghost imaging microscope. Additionally, the team presents its first use of utilizing scattering correlations with the 3DQ microscope for the imaging of nanoparticles. This group foresees a large potential for the 3DQ microscope in various areas, including measuring the dynamics of microbial symbiosis in bioenergy algal ponds and plants. More broadly, the 3DQ concept can be applied to many biological systems and extended to longer wavelength and spectroscopy applications requiring more dimensions of information while retaining high resolution and sensitivity.

Funding Statement: This work was performed under the auspices of DOE by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Probing Photoreception with New Quantum-Enabled Imaging

James E. Evans[1]* (james.evans@pnnl.gov, PI), Kevin Crampton[1], Samantha Powell[1], Jory Brookreson[1], Nick Black[2], Saleem Iqbal[2], Patrick El-Khoury[1], Robert Boyd[2]

[1]Pacific Northwest National Laboratory; [2]University of Rochester

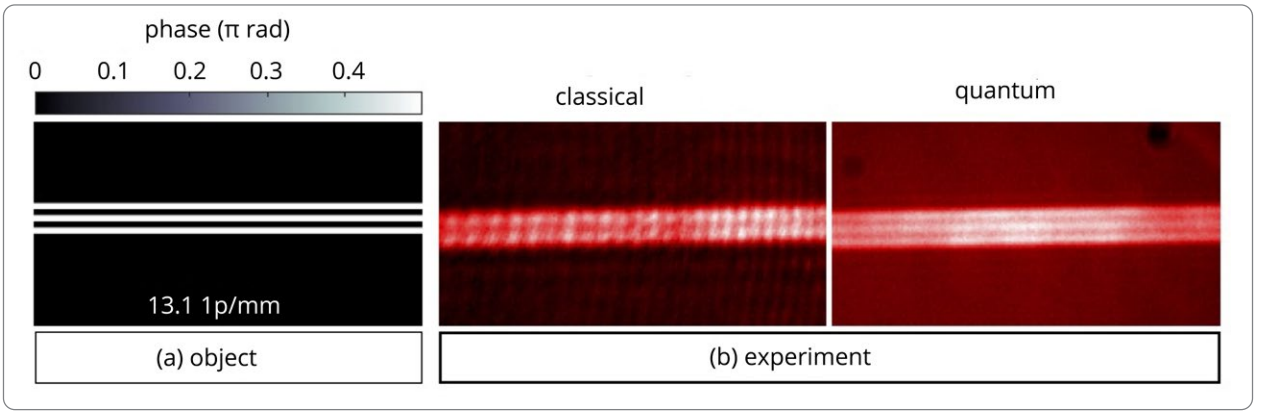
This project is developing new hybrid quantum-enabled imaging platforms that combine advances in adaptive optics, quantum entanglement, coincidence detection, ghost imaging, quantum phase-contrast microscopy, and multidimensional nonlinear coherent spectromicroscopy. The approach has three parallel aims. The first two aims focus on developing new quantum imaging approaches in which entangled photons are employed to investigate samples with lower flux or lower-energy photons but with increased spatial resolution (Aim 1) and detection sensitivity (Aim 2). Aim 3 focuses on using coherent (nonentangled) photons and four-wave mixing along with structured illumination for super-resolution nonlinear imaging.

During the current project period, the team has been developing ghost imaging, quantum Differential Interference Contras and super-resolution second harmonic generation capabilities. The group also published a paper, "Quantum-Enhanced Phase Imaging Without Coincidence Counting", in *Optica* (Black et al. 2023), which demonstrated a 1.7-fold increase in resolution. The team has since expanded that capability from a proof-of-principle low magnification setup to a setup more in line with plant, algal, and fungal bioimaging specifications. The study has acquired preliminary images off each new instrument in all three aims and begun initial experiments with test biological samples to evaluate real-world resolution and sensitivity. This poster will highlight the overall goals of the project and showcase recent results from each imaging modality.

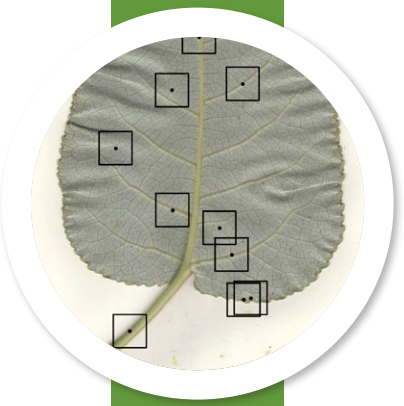
Funding Statement: Pacific Northwest National Laboratory is operated by Battelle for DOE under contract DE-AC05-76RL01830. This program is supported by the DOE Office of Science through the Genomic Science program within BER under FWP 76295. The work was performed at the Environmental Molecular Sciences Laboratory (grid.436923.9), a DOE Office of Science user facility sponsored by BER.

Reference

Black, A. N., et al. 2023. "Quantum-Enhanced Phase Imaging Without Coincidence Counting," *Optica* **10**(7), 9522–8. DOI:10.1364/OPTICA.482926.



A Comparison of Resolution Between Classical Phase-Shifting Holography and Quantum Phase-Shifting Holography. **(a)** A series of three horizontal bars with a maximum phase shift of π/2 was used to measure resolution with a spatial frequency of the bars at 13.3 line pairs/mm. **(b)** Experimental results (interferograms) indicate that only the quantum phase–shifting holography scheme can resolve the bars at this spatial frequency. [Adapted with permission from Black, A. N., et al. 2023. "Quantum-Enhanced Phase Imaging Without Coincidence Counting." *Optica* **10**(7), 952–58. DOI:10.1364/OPTICA.482926]



CHAPTER 3
Phenomics and Function

Understanding how organisms and their communities respond to their changing environment requires a detailed knowledge of both the genetic makeup of these organisms and the phenotypes these organisms display in their environments.

Phenotyping organisms, either as individuals or populations, involves measurable traits, which can be physical, chemical, or biological. This branch of study, known as phenomics, involves parallel, detailed measurements of system stimuli, mechanisms, and responses promote an integrated understanding of biosystem function. Stimuli include environmental and genetic changes; mechanisms include transcripts, signaling pathways, and biomolecular interactions; and responses include physical and chemical characteristics. Evaluating input parameters and the resulting phenotypic states provides a stimulus-response framework for an integrated understanding of biosystem function. This chapter describes researchers' work measuring traits of interest at different scales, from single cells to a whole plant's interaction with its environment.

Building a Genome-Wide Atlas of Cell Morphology

James Neal* (jtneal@broadinstitute.org, PI)

Broad Institute of Massachusetts Institute of Technology and Harvard University

A key challenge of the modern genomics era is developing empirical data-driven representations of gene function. This study presents the first unbiased morphology-based genome-wide perturbation atlas in human cells, containing three genome-wide genotype-phenotype maps comprising CRISPR/Cas9-based knockouts of >20,000 genes in >30 million cells. The optical pooled cell profiling platform (PERISCOPE) combines a de-stainable high-dimensional phenotyping panel (based on cell painting) with optical sequencing of molecular barcodes and a scalable open-source analysis pipeline to facilitate massively parallel screening of pooled perturbation libraries. This perturbation atlas comprises high-dimensional phenotypic profiles of individual cells with sufficient resolution to cluster thousands of human genes, reconstruct known pathways and protein-protein interaction networks, interrogate subcellular processes, and identify culture media-specific responses. Using this atlas, researchers identified the poorly characterized disease-associated TMEM251/LYSET as a Golgi-resident transmembrane protein essential for mannose-6-phosphate-dependent trafficking of lysosomal enzymes. In summary, this perturbation atlas and screening platform represents a rich and accessible resource for connecting genes to cellular functions at scale.

Measuring the Molecules of Life: Why It Is So Important, and Why We're Not Very Good at It

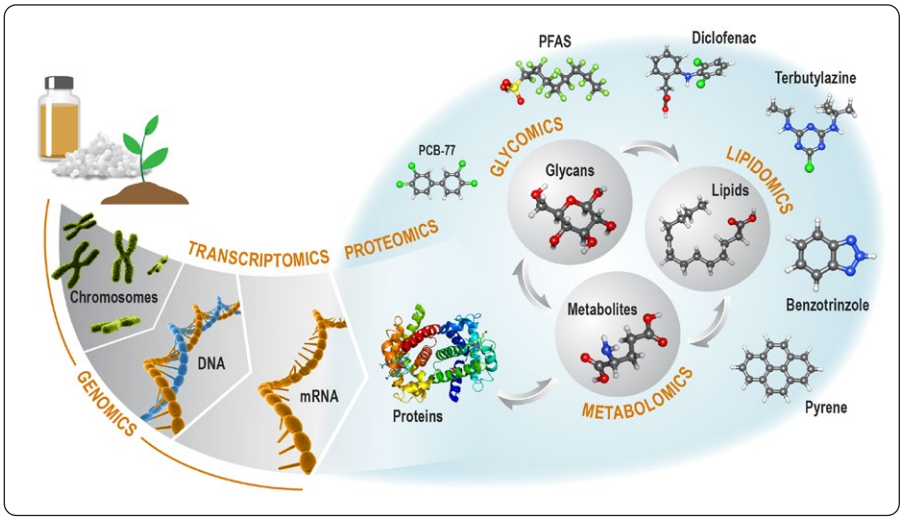
Thomas O. Metz* (thomas.metz@pnnl.gov, PI)

Pacific Northwest National Laboratory

Genome-enabled interrogation and manipulation of biological systems is converging with technologies for measuring the phenome and with advanced computational methods for integrating related and disparate data. The scientific community is poised to capitalize on this to explain the bases for health and disease and to harness biological processes for the betterment of mankind and the environment through the toggling of system function.

Molecular components of the phenome include genes, transcripts, proteins, and small molecules, and their measurement is made possible by a spectrum of technologies with broad diversity in efficiency, robustness, and impact. Genomes and transcriptomes are now routinely determined and measured with near completeness using high-throughput sequencing technologies. The measurement of proteomes and metabolomes lags that of their genetic counterparts, although technologies for broadly measuring proteins in high throughput (i.e., proteomics) are much more mature, in large part due to the high correlation of the proteome to the genome. For example, proteins are direct readouts of the genetic code, and if the genome is known, then the proteome can be predicted. While proteins are chemically more complex than genes and transcripts, they are chemically less complex than metabolites. Molecular structures of the latter, having nearly the same chemical composition as proteins, are limited in their complexity only by thermodynamics, and the chemical space is estimated to be 1,060 molecules.

Technologies for the measurement of metabolites, other biomolecules, and anthropogenic compounds have not significantly evolved since the late 1960s, when Linus Pauling implemented the concept of "orthomolecular medicine" to identify biosignatures that were correlated to phenotype. New, paradigm-disrupting technologies and concepts are needed to drive the next revolution in multiple fields of strategic importance to the nation: biological sciences, medical sciences, national security, chemistry, and the bioeconomy. This presentation will critically review the state of the art in measuring molecules integral to the phenome in the context of the respective chemical space and will propose a multidimensional evolution of the analytical approach.



The Molecules of Life. The genome and transcriptome are comprised of DNA and RNA, respectively, which are polymers of four defined nucleotides. Similarly, proteins are polymers composed of 20 defined amino acids. In contrast, the metabolome and related small molecule omes are comprised of molecules with much greater chemical diversity. [Adapted from Metz, T., et al. 2017. "Integrating Ion Mobility Spectrometry Into Mass Spectrometry-Based Exposome Measurements: What Can it Add and How Far Can it Go?" *Bioanalysis* **9**(1), 81–98. DOI:10.4155/bio-2016-0244. © 2017 Future Science Ltd and republished under Creative Commons license (CC BY-NC-ND 4.0)]

*Presenter

Understanding Plant–Environmental Interactions Using Single-Cell Approaches

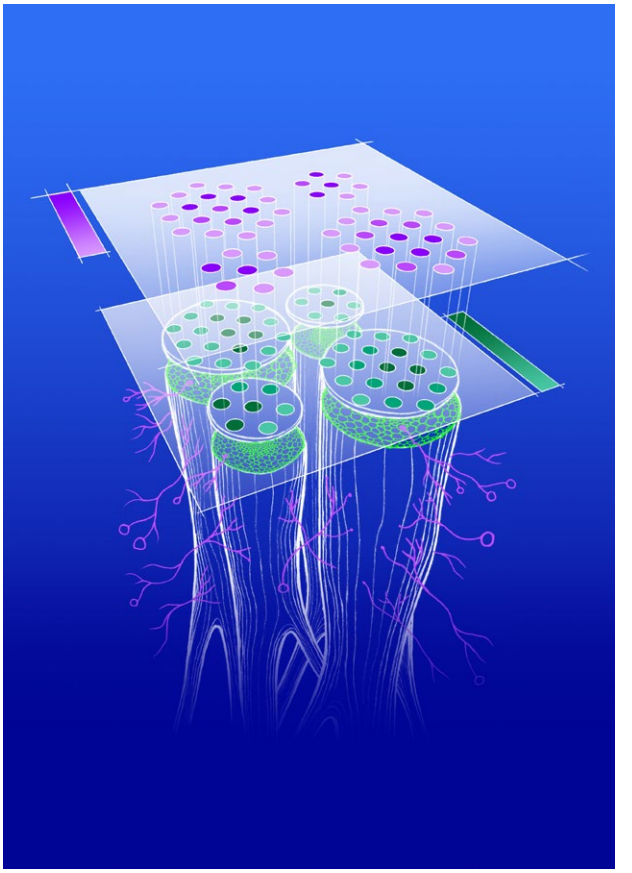
Benjamin Cole[2]* (bjcole@lbl.gov, PI), Karen Serrano[1], Margot Bezrutczyk[2], Danielle Goudeau[2], Thai Dao[1,2], Rex Malmstrom[2], Henrik Scheller[1]

[1]Joint BioEnergy Institute; [2]DOE Joint Genome Institute

Biomass derived from plant feedstocks is a renewable and sustainable energy resource, but these resources are vulnerable to environmental stress such as water and nutrient limitations. Understanding how cells work independently and in concert to regulate plant responses to their environment, including their surrounding microbial community, as well as abiotic stress will be crucial to improving their performance.

This group has applied several cutting-edge, single-cell, and spatially resolved transcriptome sequencing approaches to several plant species and is constructing a comprehensive single-cell resource for plants to better understand the complexity behind environmental responses among diverse cell types. In particular, the group has leveraged both single-nuclei and spatial transcriptomics to profile interactions between plants and arbuscular mycorrhizal fungi (AMF) using the *Medicago truncatula* and *Rhizophagus irregularis* system. The team developed a cross-kingdom transcriptome map for this crucial symbiosis, profiling both plant and fungal expression patterns. Team members are also working toward establishing more precise spatial omics tools to profile tissues from bioenergy grasses (including sorghum, switchgrass, and *Brachypodium*) reacting to AMF colonization, as well as physical stresses including nutrient deprivation and osmotic stress. The team hopes to build a multi-species model of cell type–specific environmental responses.

Funding Statement: This work was supported by the U. S. Department of Energy, Office of Science, BER program, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and DOE. This work was also supported by an Early Career Research program to B.C., the Laboratory Directed Research and Development program at Lawrence Berkeley National Laboratory, and by the DOE Joint BioEnergy Institute.



Spatial Co-Transcriptomics Reveals Discrete Stages of the Arbuscular Mycorrhizal Symbiosis. This 2024 publication shows stylized *Medicago* root colonized by arbuscular mycorrhizal fungi and profiled using spatial transcriptomics technology. [2024. "Symbiosis in Time and Space," *Nature Plants* **10**(4). Reused under Creative Commons license (CC BY 4.0)]

Artificial Intelligence for Image-Based Plant and Microbial Phenotyping

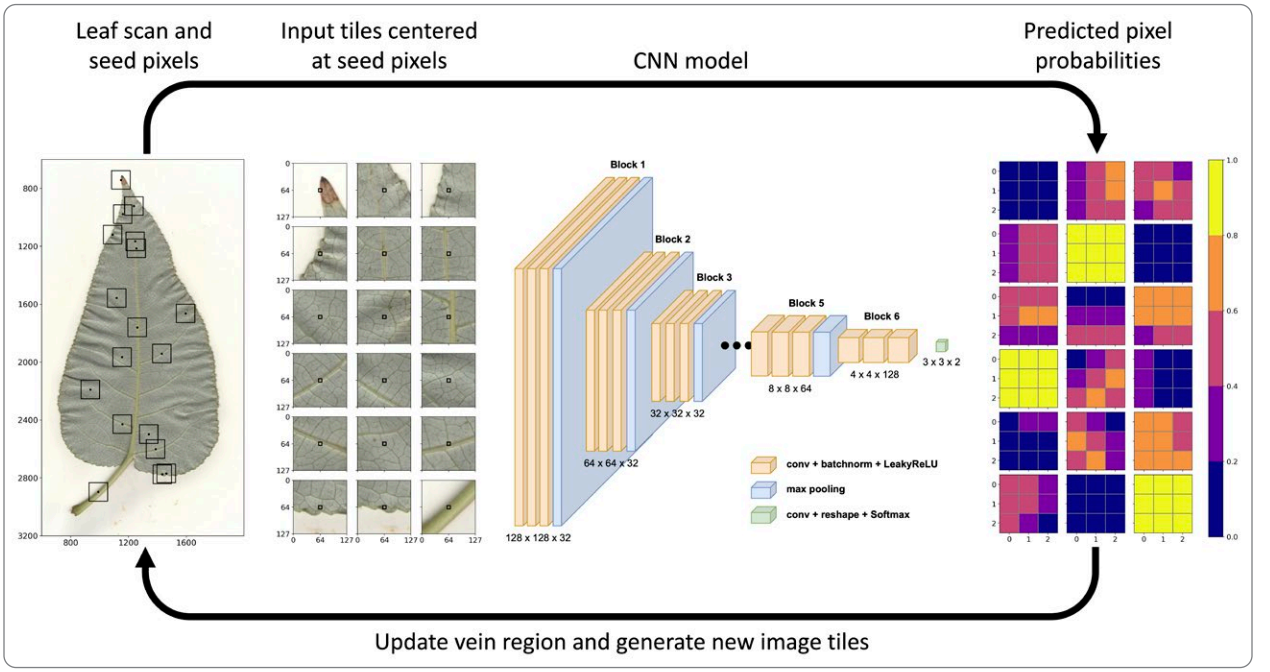
John Lagergren[1,2]* (lagergrenjh@ornl.gov), Larry York[1,2], Mirko Pavicic[1,2], Hari Chhetri[1,2], Daniel Jacobson[1,2], Tomas Rush[1], Wellington Muchero[1,2], David Weston[1], Dale Pelletier[1], Mitchel Doktycz[1], **Gerald Tuskan[1,2]** (PI)

[1]Oak Ridge National Laboratory; [2]Center for Bioenergy Innovation

Project Goals: The Center for Bioenergy Innovation (CBI) vision is to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations along the bioenergy and bioproduct supply chain while focusing on sustainable aviation fuels (SAF). CBI has four overarching innovation targets: (1) develop sustainable, process-advantaged biomass feedstocks; (2) refine consolidated bioprocessing with cotreatment to create fermentation intermediates; (3) advance lignin valorization for biobased products and aviation fuel feedstocks; and (4) improve catalytic upgrading for SAF blendstocks certification.

There is a growing need to develop sustainable perennial crops and beneficial microbes that thrive in suboptimal environments, are resilient to biotic and abiotic stresses, and aid conversion to advanced bioproducts (e.g., sustainable aviation fuel). These goals are facilitated by connecting gene functions to observable traits through high-throughput measurement of key plant and microbe characteristics. However, crucial bottlenecks include the availability of large, high-quality datasets and the burden of time-, cost-, and labor-intensive phenotyping. Deep learning has enabled fast and accurate image-based phenotyping but is similarly challenged by the lack of annotated image data.

This presentation discusses ongoing efforts to address these challenges through the development of deep learning methods (Lagergren et al. 2023) that (1) maximize predictive accuracy while minimizing the amount of annotated training data; (2) ensure biologically realistic morphological features (e.g., contiguity); (3) scale to high-throughput, population-scale datasets for genomic analysis; and (4) extend across kingdoms and species (e.g., *Populus* spp., *Bacillus* spp.) and image modalities (e.g., RGB, hyperspectral,



Leaf Tracing Algorithm. A neural network algorithm developed by John Lagergren that accelerates phenotyping by analyzing scanned images to identify key biologically relevant traits. [Reprinted from Lagergren, J., et al. 2023. "Few-Shot Learning Enables Population-Scale Analysis of Leaf Traits in *Populus trichocarpa*," *Plant Phenomics* **5**. DOI:10.34133/plantphenomics.0072. Republished under Creative Commons license (CC BY 4.0)]

*Presenter

confocal microscopy). This team also shows how the model outputs (e.g., segmentations) are used to extract multiple traits (e.g., area, perimeter) using open-source tools (Seethepalli et al. 2021) that are validated with real-world physical measurements and used in downstream scientific analyses like genome-wide association studies.

Finally, the team introduces preliminary results that demonstrate how typical narrow artificial intelligence (e.g., models trained on specific datasets for a specific purpose as described above) struggle to generalize outside of their training distribution (e.g., adapting to new species, plant characteristics), and how foundation models for computer vision have recently emerged as powerful tools for image-based phenotyping with the ability to perform multiple tasks without the need to fine-tune or train new models from scratch. These new research directions are given in the context of the Advanced Plant Phenotyping Laboratory, a state-of-the-art plant phenotyping facility at Oak Ridge National Laboratory, which produces multimodal image data for DOE mission–relevant bioenergy crops at a rate of up to one petabyte per year. Harnessing the power of artificial intelligence through intentional development for biological sciences will be imperative to achieve sustainable energy independence for the nation.

Funding Statement: Funding was provided by the Center for Bioenergy Innovation (CBI) led by Oak Ridge National Laboratory. CBI is funded as a U.S. DOE Bioenergy Research Center supported by the BER program in the DOE Office of Science under FWP ERKP886. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for DOE under contract DE-AC05-00OR22725.

References

Lagergren, L., et al. 2023. "Few-Shot Learning Enables Population-Scale Analysis of Leaf Traits in *Populus trichocarpa*," *Plant Phenomics* **5**(75). DOI:10.34133/plantphenomics.0072.

Seethepalli, A., et al. 2021. "RhizoVision Explorer: Open-Source Software for Root Image Analysis and Measurement Standardization," *AoB Plants* **13**(6). DOI:10.1093/aobpla/plab056.

CHAPTER 4

BER Data Science and Infrastructure

BER-supported national scientific user facilities and experimental and observational facilities create high-quality, large-scale scientific data collections. These data structures require interoperability to enable BER scientists to leverage integrated analyses of these various data types for deeper understanding of system processes and knowledge discovery. This chapter highlights efforts within BER to integrate computational and data science platforms and to develop interoperability of data, tools, and supporting information that span biological and environmental sciences. Presentations will include efforts to develop data compatibility, metadata standards, new software workflows, and tools for data transfer services to accelerate scientific discovery through the convergence of experimental and simulation data.

A Unified Data Infrastructure for Biological and Environmental Research

Kerstin Kleese van Dam* (kleese@bnl.gov, PI)

Brookhaven National Laboratory; Chair, BERAC Subcommittee and Working Group on Unified Data Infrastructure

In October 2022, the BER Advisory Committee (BERAC) received a charge letter from the DOE Office of Science director requesting a review of existing capabilities in data management and infrastructure relevant to BER science. The charge also requested a recommended strategy for next-generation data management and analysis within a unified framework. Further goals included identifying new science opportunities that could be enabled by increased integration of BER's facilities while considering advances in artificial intelligence and machine learning (AI/ML). The charge asked BERAC to examine synergistic investments within DOE and at other agencies and the impact of a more unified data infrastructure on the scientific workforce. To address these goals, the appointed subcommittee established five working groups focusing on (1) environmental science; (2) biological science; (3) BER data infrastructure services; (4) workforce development, inclusion, and diversity; and (5) data infrastructure technologies. The subcommittee organized a two-day virtual community workshop that included discussions on new unified data infrastructure–enabled science opportunities, barriers to broader inclusion of minorities, support for early career scientists, and potential unified data infrastructure solutions for BER. The results of the workshop and a public request for information were evaluated by the subcommittee and its working groups and its findings summarized in a report. The talk will outline the initial science opportunities identified that could be enabled by a new unified data infrastructure for BER sciences. The talk will also review how a unified data infrastructure may increase the accessibility of BER science and support early career and minority scientists. Furthermore, it will discuss existing data infrastructure capabilities available at BER and elsewhere, and finally, outline the subcommittee recommendations.

A Unified Data Infrastructure for Biological and Environmental Research. This 2024 report details a review of existing capabilities in data management and infrastructure relevant to BER science and provides a recommended strategy for next-generation data management and analysis within a unified framework. Full report found here: ess.science.energy.gov/berac-data-infrastructure-report

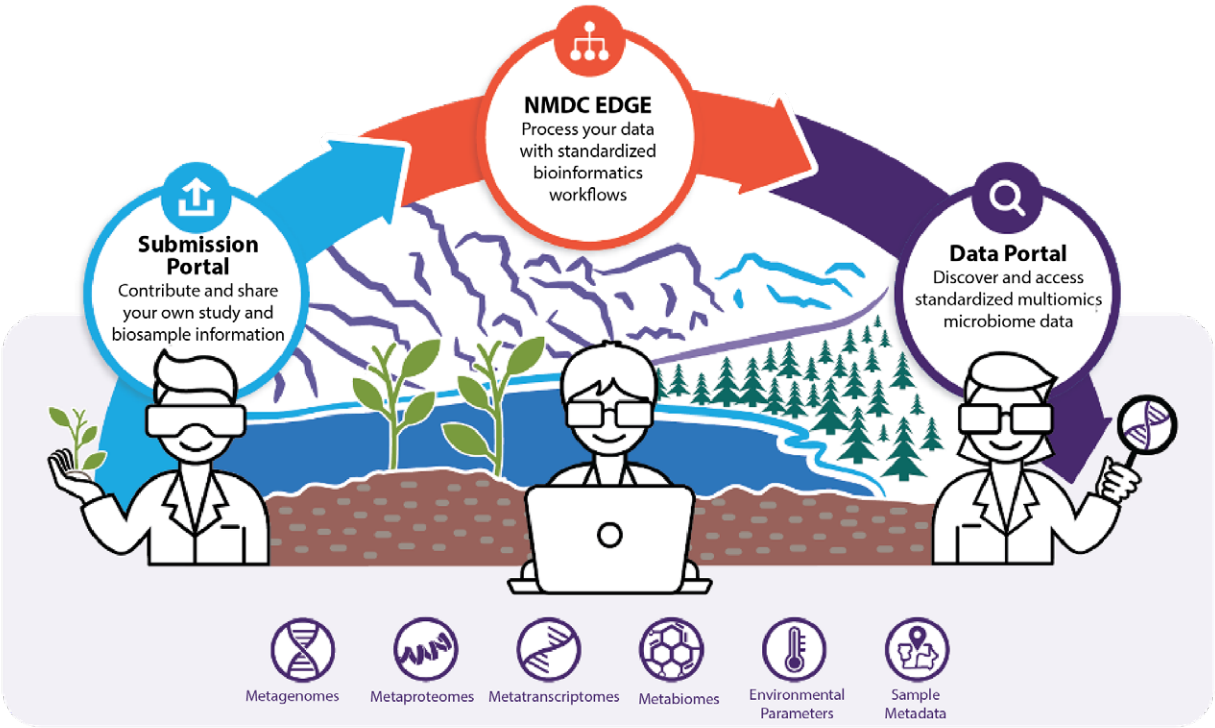
Advancing Microbiome Research with the National Microbiome Data Collaborative

Lee Ann McCue[1]* (leeann.mccue@pnnl.gov),
Emiley Eloe-Fadrosh[2] (PI)

[1]Pacific Northwest National Laboratory;
[2]Lawrence Berkeley National Laboratory

microbiomedata.org

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. The three core infrastructure elements of the NMDC framework are: (1) the Submission Portal to support collection of standardized study and biosample information; (2) NMDC Empowering the Development of Genomics Expertise (EDGE), an intuitive user interface to access standardized bioinformatics workflows; and (3) the Data Portal, a resource for consistently processed and integrated multiomics data enabling search, access, and download. 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. This open-access framework lowers the barrier for capture of contextual information about data and facilitates the effective use of microbiome data for applications in energy, environment, and agriculture.



National Microbiome Data Collaborative (NMDC) Products. These three NMDC products provide microbiome data standards, a robust data sharing infrastructure, and community building opportunities. [Courtesy Pacific Northwest National Laboratory]

*Presenter

Partnerships to Improve FAIRness

Kjiersten Fagnan[1,2*] (kmfagnan@lbl.gov, PI), Chris Beecroft[1], Neil Byers[1], Chuck Parker[1], Emiley Eloe-Fadrosh[1,2], Elisha Wood-Charlson[3], A. J. Ireland[3], Jeffrey Johnson[4], Nigel Mouncey[1]

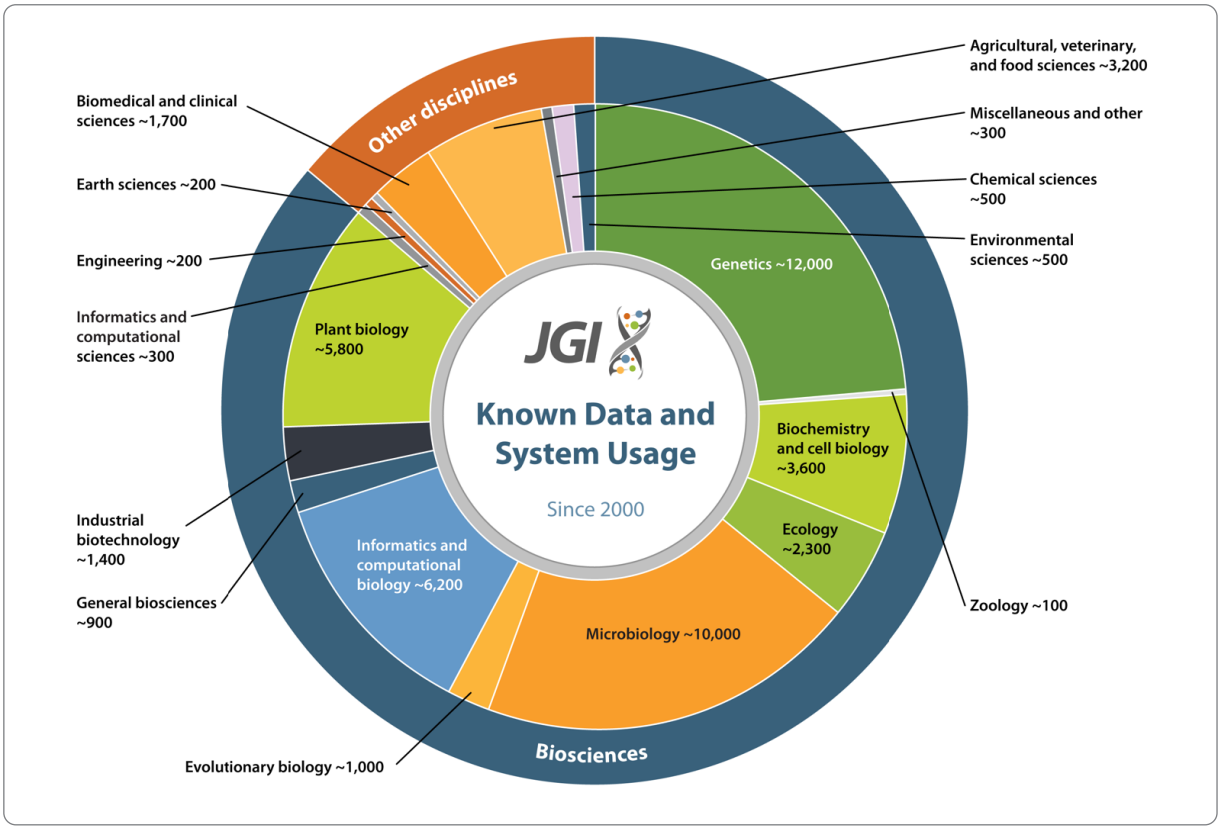
[1]DOE Joint Genome Institute; [2]National Microbiome Data Collaborative; [3]KBase, Lawrence Berkeley National Laboratory; [4]Cohere Consulting, LLC

data.jgi.doe.gov, data.microbiomedata.org, kbase.us

Project Goals: DOE Joint Genome Institute, National Microbiome Data Collaborative, and KBase are collaborating to make data discovery and transfers easier.

Data discoverability is a challenge. BER supports the generation of petabytes of high-quality environmental data leading to exciting discoveries. Once the data are archived in national repositories the return on the original investment is amplified through reuse. In order to create more opportunities for data to be identified and utilized by the scientific community, organizations have pursued implementation of the FAIR (findable, accessible, interoperable, reusable) principles. Data and metadata quality are improving, however biological data requires more structure to ensure interoperability across studies. DOE JGI, NMDC, and KBase are collaborating on software infrastructure including common data models, application programming interfaces, and transfer protocols to improve data flow between resources. The group's shared vision is a consistent data discovery experience across BER platforms. This talk will share details of the team's efforts to work toward this vision, including JGI's planned reuse of the NMDC Submission Portal, KBase and JGI's co-development of the Data Transfer Service (DTS), and exploration of data reuse through JGI's Data Citation Explorer.

Funding Statement: The work conducted by the DOE Joint Genome Institute (ror.org/04xm1d337), a DOE Office of Science user facility, is supported by the DOE Office of Science operated under contract number DE-AC02-05CH11231.



JGI Known Data and Storage Usage. Genetics and microbiology comprise the largest share of DOE Joint Genome Institute's data and storage usage. Altogether, the majority of data and storage usage is attributed to the biosciences. [Courtesy DOE Joint Genome Institute]

Developing a Community Approach to Data Integration and Data Science in KBase

Adam P. Arkin[1]* (aparkin@berkeley.edu, PI), Christopher S. Henry[2], Robert Cottingham[3], Shane Canon[1], Paramvir S. Dehal[1], Elisha Wood-Charlson[1], Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], Joseph Bezouska[2], David Dakota Blair[4], John-Marc Chandonia[1], Zachary Crockett[3], Ellen G. Dow[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Andrew Freiburger[2], Tianhao Gu[2], Prachi Gupta[1], A. J. Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Dileep Kishore[3], Dan Klos[2], Filipe Lui[2], Mikaela McDevitt[1], Christopher Neely[1], Cody O'Donnell[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Pamela Weisenhorn[2], Sijie Xiang[1], Ziming Yang[4], Shinjae Yoo[4], Qizhi Zhang[2]

[1]Lawrence Berkeley National Laboratory; [2]Argonne National Laboratory; [3]Oak Ridge National Laboratory; [4]Brookhaven National Laboratory

kbase.us

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into a user-friendly platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform, enabling scientists to analyze their own data alongside public and collaborator data, then share their findings across the system and ultimately publish reproducible analyses.

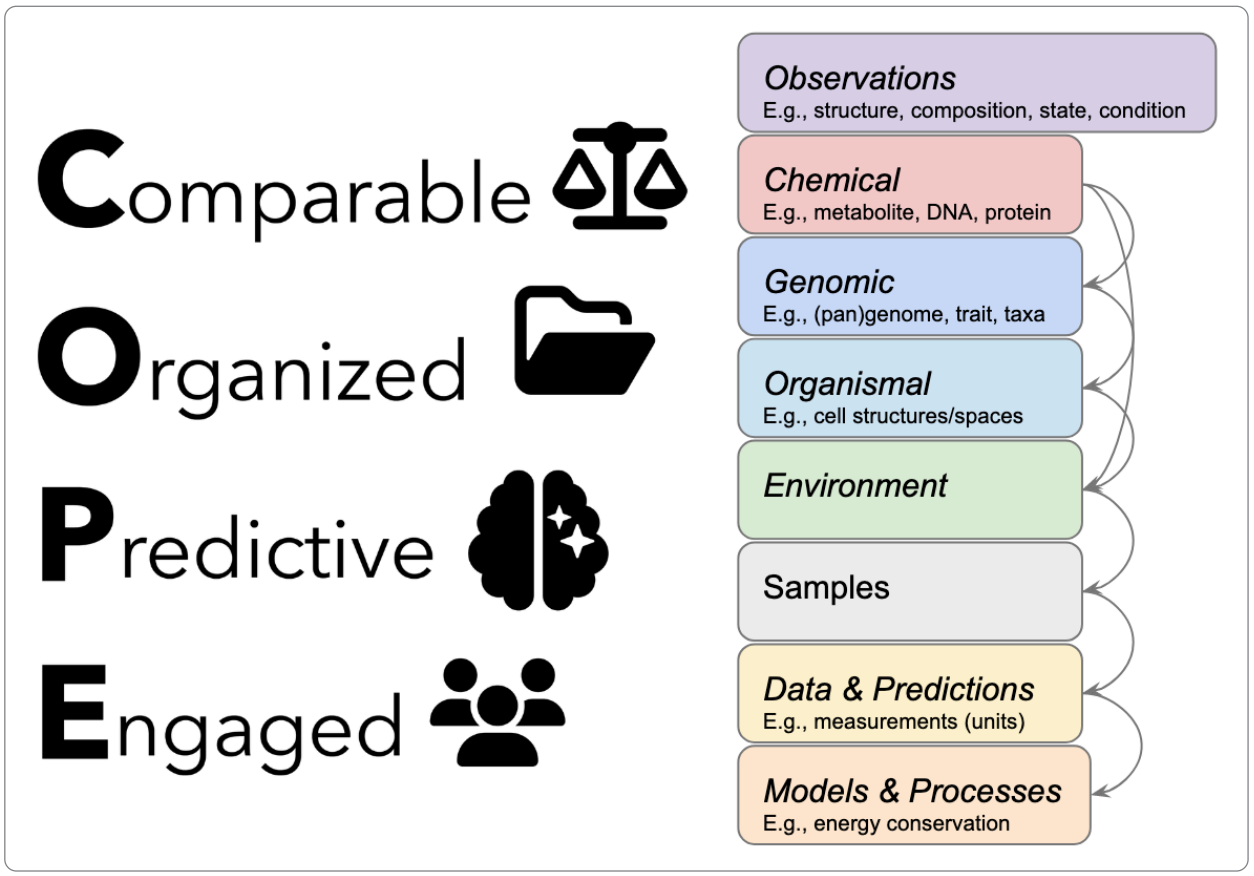
KBase aims to empower its users to predict, control, and design the behavior of biological systems from subcellular to ecosystem processes. A critical capability for such research is the ability to find and integrate relevant data from the larger scientific community that can be used to strengthen and test the generality of user analyses, and to help identify gaps in both personal and collective knowledge that reduce the effectiveness of such analyses. To address the integration problem, this research group is leading two central efforts.

First, the group is working with partners at the DOE Joint Genome Institute (JGI), National Microbiome Data Collaborative (NMDC), and Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE), among others, to develop a Data Transfer Service that streamlines finding and transporting data easily among, initially, BER platforms, while ensuring that provenance and ownership are tracked and credited. Researchers are developing an integrated system for scalable inference generation from user data, comprised of a central data model (CDM) containing the knowledge representation and data organization schema for the team's system; a relation engine (RE) that powers the population of the CDM with public reference data; and a knowledge engine (KE) that interfaces with the RE to create a wide range of inferences for data entities within the CDM and for user data. Within KBase, these three elements (CDM, RE, and KE) work together to ensure data from diverse sources are linked by common concepts and thereby become comparable for analytical purposes.

The CDM is iteratively designed to represent biological, physical, and experimental relationships among data that are brought together from various resources and instantiated during intake into KBase. It will also enable queries supported by artificial intelligence (AI) to find and organize data relevant to a user's question suitable for downstream analysis. This group intends for the CDM to serve projects beyond KBase and are assembling community members to aid in its design, testing, and iterative revision. The RE maps user data to the CDM and enables the creation, maintenance, and query of relationships within the CDM. The KE provides predicted and inferred relationships among data referenced by the CDM and creates facilities for data-driven search that enhance relevant data retrieval. The relationships include calculated similarity of genomes or genes, and predictions, such as phenotype and environmental distribution. The data-driven searches include sequenced or functional abundance profile-based queries that might return similar genomes or metagenomes to the user. The KE will also exploit new innovations in large language models and their interface to systems like the CDM to create AI-based assistants that enable users to employ natural language to state the problems they are trying to solve, and then navigate retrieving relevant data, organizing it alongside their own for analysis, and designing and executing the analyses with KBase tools.

This talk outlines the rationale and principles driving the development of the CDM and emphasize

the importance of iterative community engagement throughout the process. In particular, this presentation outlines how it supports integration across resources and advancing biological data science within KBase. The goal is to ensure the CDM, and the tools it enables, will help lower the bar to data integration across BER, expand the types of science questions researchers can ask, and advance the field of data science to better handle the complexity in and among biological datasets this team's scientists and the broader community are creating. This group will motivate this vision with examples drawn from causal microbial ecology that interfaces with a number of the goals of collaborating DOE programs.



COPE. Going beyond findable, accessible, interoperable, reusable data requires a community effort to make data comparable and organized, and to increase its predictive potential by engaging in feedback and validation. [Courtesy DOE Systems Biology Knowledgebase]

Microbiome Data Science: from the Earth Microbiome to the Global Virome

Nikos C. Kyrpides* (nckyrpides@lbl.gov, PI)

DOE Joint Genome Institute

The field of microbiome research is experiencing a transformative shift towards data science, propelled by the massive influx of microbiome data. This burgeoning volume of data presents both formidable challenges in terms of establishing standards and management frameworks, and simultaneously unlocks unprecedented opportunities for groundbreaking discoveries. Current exploration into computational analysis of microbiome samples, including those from previously uncultured organisms, is significantly enriching scientists' understanding of microbial community structures and functions. This, in turn, is broadening scientists' grasp of the genetic and functional diversity within individual microorganisms. This talk will elucidate cutting-edge computational methodologies, underscoring the pivotal role of big data processing and integration in mining metagenomic datasets. Such approaches are instrumental in unveiling novel insights and fostering discoveries. This talk details the latest strategies for data analysis and share illustrative science vignettes that highlight the exploration of microbial, viral, and functional diversities. This talk aims to showcase the transformative potential of integrating big data with microbiome research, paving the way for scientific breakthroughs in understanding the complexity and dynamism of microbial ecosystems.

*Presenter

The Landscape of Data Infrastructure from the National Virtual Biosecurity for Bioenergy Crops Center Perspective

Shanetenu Jha* (sjha@pppl.gov, PI), Martin Schoonen

Brookhaven National Laboratory

Brookhaven National Laboratory was awarded a pilot project in FY22 under the DOE Office of Science Biopreparedness Research Virtual Environment (BRaVE) initiative to define research priorities, needs, and requirements for a national virtual center devoted to the biosecurity of bioenergy crops. The proposed center's mission, the National Virtual Biosecurity for Bioenergy Crop Center (NVBBCC), would be to provide the scientific basis and tools to detect, characterize, model, and mitigate biothreats to bioenergy crops. This function will ensure increased U.S. reliance on essential plant-based energy products, e.g., biojet fuel, over the next few decades The NVBBCC is envisioned as a distributed, virtual center with multiple national laboratories at its core to maximize the use of existing unique facilities and expertise across the DOE complex. To underpin this collaborative and distributed effort, a flexible computational platform that supports high-performance computing workflows and data management and allows for efficiently conducting modeling simulations is needed.

This presentation outlines the computing infrastructure capabilities toward these goals, derived from a community requirements workshop:

- Develop an integrated research infrastructure that enables meaningful integration of data, computing, instrumentation, and related resources to allow researchers access to needed computational/data resources from anywhere.
- Employ robust data management systems that can manage diverse data types to ensure quality while adhering to FAIR (findability, accessibility, interoperability, and reusability) principles and supporting better metadata.
- Explore and adopt advanced technologies, such as 5G/6G wireless communication for high-throughput transmission with low latency in poorly connected areas.
- Develop scalable and generalizable models that link model design to downstream decision-making while using tools and techniques to create a unified, integrated modeling approach.

CHAPTER 5

KBase Science: Data Integration to Support (or Refute) Predictions

Much of biological science knowledge is based on prediction–algorithmic assessment of taxonomy and function determined by parameters based on current knowledge. But how certain is the scientific community that those parameters are accurate across ecosystems, time, and the tree of life? By exploring and integrating multiple lines of evidence (e.g., multiomics data, experimental validation, mechanistic modeling, and machine learning), researchers can build a story that supports or refutes their predictions. This chapter features examples of research questions and tools that integrate data to build trust in scientists' assertion of the world.

Measuring Microbial Phenotypes for Improving Genome-Based Predictions

Dale A. Pelletier[1]* (pelletierda@ornl.gov), Amber N. Bible[1], Leah H. Burdick[1], Elisha Wood-Charlson[2], Paramvir S. Dehal[2], Janaka N. Edirisinghe[3], Jennifer Morrell-Falvey[1], Andrew P. Freiburger[3], Christopher S. Henry[3], Dileep Kishore[1], Jennifer Kuehl[2], Hira Lesea[2], Christopher Neely[2], Trent R. Northen[2], Markus de Raad[2], Priya Ranjan[1], Adam P. Arkin[2], **Mitchel J. Doktycz[1]** (PI)

[1]Oak Ridge National Laboratory; [2]Lawrence Berkeley National Laboratory; [3]Argonne National Laboratory

Microbes perform numerous essential roles in terrestrial ecosystems including biogeochemical cycling of nutrients, soil structuring, and plant productivity. Despite the tremendous gains in knowledge of the metabolism of microbes, predicting microbial phenotypes from genomic information across the vast diversity of microbes remains a challenge. This presentation will discuss ongoing collaborative efforts between the Plant-Microbe Interfaces and the Ecosystems and Networks Integrated with Genes and Molecular Assemblies Science Focus Areas (SFAs), to establish a cross-SFA characterized strain collection with measured microbial phenotypes for improving genome-based predictions. These collections represent isolates of plant microbiomes and subsurface microbiomes, respectively. Standardized methodologies and phenotyping datasets are being developed for expanding this dataset and for serving as a data standard for genotype-to-phenotype prediction tools and methodologies within Systems Biology Knowledgebase.
This dataset and related tools lay the groundwork for broader community engagement and will be useful for developing more robust phenotype prediction classifiers that cover a broader array of carbon sources and a phylogenetically diverse set of microbes. With this integration and standardized approaches, the research team strives to enable users to predict microbial phenotypes from a wide array of genomes, thereby significantly advancing microbial research.

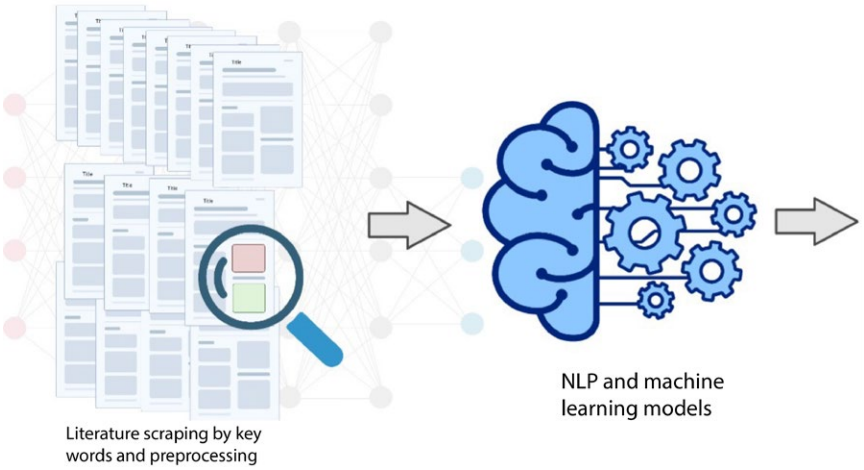
Knowledge Extraction from Literature

Shinjae Yoo[1]* (sjyoo@bnl.gov), Carlos X. Soto[1], Gilchan Park[1], Christopher Neely[2], Vivek K. Mutalik[2], **Paramvir S. Dehal[2]** (PI)

[1]Brookhaven National Laboratory;
[2]Lawrence Berkeley National Laboratory

Project Goals: This presentation details a proof-of-concept demonstration that applies state-of-the-art natural language processing (NLP) techniques to automatically extract biological entities and genetic tools from the literature for synthetic biology research. This work seeks to address important knowledge gaps in this field, while simultaneously providing a meaningful staging ground to expose new NLP tools to the KBase community and to gather feedback on their efficacy and use.

Genetic tool engineering in non-model organisms remains a major challenge in the field of synthetic biology and is typically throttled by the large literature



Project Overview. The group proposes to use best natural language processing and machine learning models to process and learn from literature data about growth characteristics, conditions, traits (e.g., antibiotic and stress tolerance, fitness traits) and available *in silico* models and genetic tools to engineer DOE-BER centric microorganisms. [Courtesy Lawrence Berkeley National Laboratory]

*Presenter

searches that invariably accompany the project. Indeed, with decades of publications containing a vast corpus of non-standardized data formats and methods, synthesizing an adequate protocol to guide the study can be a daunting task. But as global issues like climate change, degradation of ecosystems, and increasing food scarcity continue to emerge and grow, the need to engineer these organisms and their relevant toolsets becomes ever greater. Clearly, there is a pressing need for fast and comprehensive searches that help inform and guide laboratory research, as incomplete, cursory searches may increase the time it takes to complete the project or may even preempt its success.

Advances in the field of natural language processing have resulted in the development of powerful large language models (LLMs) to provide solutions to such problems. Fine-tuning these models to identify genetic engineerability terms and to perform biological entity extraction can direct researchers towards useful and informative answers that are driven by existing literature. This team presents a workflow for automating this extraction and subsequently incorporating the data into model file-tuning. Working with a large number of publications extracted from *bioRxiv*, the team uses text-mining techniques guided by human experts to extract biologically relevant entities from a large publication corpus: organism names and genetic tools, including plasmids, promoters, reporters, and other entities of interest. Researchers test various publicly available LLMs, including Falcon (Almazrouei et al. 2023), LLaMA-2 (Touvron et al. 2023), MPT-Chat (MosaicML 2023), and others, to identify the best-performing model and augment it using state-of-the-art techniques to mitigate model hallucination.

As a proof of principle, this group presents a web application that interfaces a chatbot driven by this model with a visualization tool. User queries are highlighted on the National Center for Biotechnology Information (Schoch et al. 2020) taxonomy tree at the genus-level, highlighting the organisms of interest and displaying their nearest relatives. Datasets collected from various isolate reference databases, including BacDive (Reimer et al. 2022), as well as specific genetic tool databases like the Phage-Host Database (Albrycht et al. 2022), the Plasmid Database (Schmartz et al. 2022), and others, are searched for relevant matches to display to the user. In addition, matches from the literature, mined by the LLM, are also provided to the user. Integrating this tool into KBase infrastructure provides another bridge for DOE BER researchers to access this information, linking it not only with biologically relevant organisms for laboratory experiments, but also with KBase's own ecosystem to allow subsequent analyses and publication of results.

Funding Statement: This material is based upon work supported by the DOE Office of Science, BER program under award number DE-AC02-05CH11231 (Lawrence Berkeley National Laboratory) and DE-SC-0012704 (Brookhaven National Laboratory).

References

Albrycht, K., et al. 2022. "Daily Reports on Phage-Host Interactions," *Frontiers in Microbiology* **13.** DOI:10.3389/fmicb.2022.946070.

Almazrouei, E., et al. 2023. "The Falcon Series of Open Models," *arXiv*. DOI:10.48550/arXiv.2311.16867.

MosaicML NLP Team. 2023. "Introducing MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs." Databricks. www.mosaicml.com/blog/mpt-7b.

Reimer, L. C., et al. 2022. "BacDive in 2022: The Knowledge Base for Standardized Bacterial and Archaeal Data," *Nucleic Acids Research* **50**. DOI:10.1093/nar/gkab961.

Schmartz, G. P., et al. 2022. "PLSDB: Advancing a Comprehensive Database of Bacterial Plasmids," *Nucleic Acids Research* **50.** DOI:10.1093/nar/gkab1111.

Schoch, C. L., et al. 2020. "NCBI Taxonomy: A Comprehensive Update on Curation, Resources and Tools," Database (Oxford). DOI:10.1093/database/baaa062.

Touvron, H., et al. 2023. "LLaMA 2: Open Foundation and Fine-Tuned Chat Models," *arXiv*. DOI:10.48550/arXiv.2307.09288.

Predicting Protein Function Using Structure and Sequence Similarity in KBase

Christopher S. Henry[1]* (chenry@anl.gov, PI), Nidhi Gupta[1], Claudia Lerma-Ortiz[1], Janaka N. Edirisinghe[1], Nils Oberg[2], John Gerlt[2], KBase Team[1,3,4,5], Robert Cottingham[4], Adam P. Arkin[3]

[1]Argonne National Laboratory; [2]University of Illinois Urbana-Champaign; [3]Lawrence Berkeley National Laboratory; [4]Oak Ridge National Laboratory; [5]Brookhaven National Laboratory

kbase.us

Project Goals: Protein families of unknown function are a significant challenge facing the DOE BER research community. While many tools in KBase and elsewhere today permit the discovery of entirely new protein families, very few tools exist to study the function of these families. The Enzyme Function Initiative (EFI; enzymefunction.org) offers tools to address this critical problem. This project aims to integrate the EFI toolset into KBase fully, with complete ties to DOE BER sequencing sources, including all sequence data in KBase and the DOE Joint Genome Institute Integrated Microbial Genomes database. Further, the team will ensure the interoperability of these tools with other functional genomics tools in KBase, particularly tools to integrate structural data from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB).

One of the most significant challenges currently inhibiting understanding of complex biological systems from genomic and multiomic data is the staggering number of proteins with unknown functions. Tools are needed to integrate multiple sources of evidence to decode the functions of uncharacterized protein families and understand the limits of annotation propagation. EFI toolkit supports protein function discovery through Sequence Similarity Networks (SSNs) (Zallot et al. 2019; Oberg et al. 2023). Here, researchers will demonstrate how the EFI toolkit (now partially in KBase) is combined with other tools, particularly tools for integrating structural insights from RCSB-PDB, to study the propagation of function through the members of a close protein family.

The team's first demonstration of the protein function discovery pipeline in KBase focuses on the aconitase superfamily. In the quest to enable automated rapid reconstruction of high-quality fungal metabolic models, researchers detected essential functions often misannotated in fungal genomes. The team focused on three examples: aconitases AcnN and AcnD and 2-methylisocitrate lyase. First, researchers constructed a protein family around correctly annotated instances of each function; then, for each family, researchers constructed phylogenetic protein trees and SSNs (generated by the EFI tools). The team correctly annotated the problematic protein families across diverse fungal genomes using these constructs, improving annotation consistency and the corresponding metabolic models. Furthermore, group members investigated the structure of the aconitase proteins using KBase-RCSB tools, observing that yeast aconitase D's structure is similar to that of bacterial proteins. In contrast, the bacterial aconitase A has only 6% identity with its fungal mitochondrial equivalent (AcnN).

The second demonstration explores enzymes involved in microbial degradation of pyridine, specifically recently discovered Group C mono-oxygenases pdbA and pyrA and an alternative route involving Vanillate O-demethylase oxygenase VanA. Protein families were constructed in KBase around these enzymes; protein members were then organized into a tree and compiled into an SSN. Researchers used these approaches to study the evolutionary patterns of variation within these protein families to explore the potential phylogenetic breadth of the pyridine degradation activity discovered or proposed for these genes. The team also applied AlphaFold to produce structures for representative genes, which were compared with related structures in PDB and applied to perform docking simulations in KBase. Together, these tools reveal insights into accurately propagating these relatively new annotations to new genomes, improving the representation of the new pyridine degradation pathway in metabolic models produced by KBase.

This group is working to validate the new monooxygenase annotations in the pyridine degradation pathway using the self-driven laboratory system at Argonne National Laboratory and the strain *Acinetobacter* sp. ADP1. Using these capabilities, researchers can track the capacity of ADP1 to grow on pyridine with complementation and knockout of the candidate proteins. This team is exploring using this platform as an automated testbed for validating new protein function discoveries emerging from KBase.

*Presenter

Funding Statement: This work is supported as part of the BER Genomic Science program. The DOE Systems Biology Knowledgebase (KBase) is funded by the DOE Office of Science BER program under DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886. Plans for Integration of Enzyme Function Initiative (EFI) Tools into the KBase Platform is funded by the DOE Office of Science BER program under PRJ1010515. Self-driven laboratory work was funded by Laboratory Directed Research Development at Argonne National Laboratory.

References

Oberg, N., et al. 2023. "EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools," *Journal of Molecular Biology* **435**(14). DOI:10.1016/j.jmb.2023.168018.

Zallot, R., et al. 2019. "The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways," *Biochemistry* **58**(41), 4169–82. DOI: 10.1021/acs.biochem.9b00735.

Integrating Data to Predict Functions for Gaps in Metabolic Models

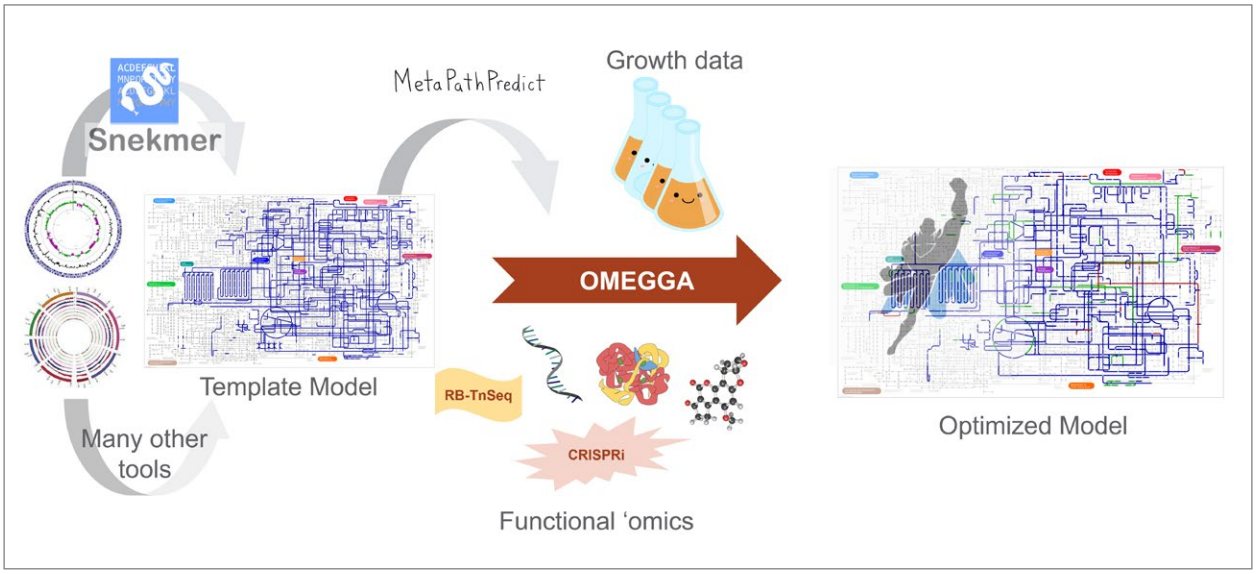
William C. Nelson[1]* (william.nelson@pnnl.gov), Jason McDermott[1,2], David Geller-McGrath[3], Christine Chang[1], Jeremy Jacobsen[1], Tara Nitka[1], Travis Wheeler[4], **Robert Egbert[1]** (PI)

[1]Pacific Northwest National Laboratory; [2]Oregon Health & Science University; [3]Woods Hole Oceanographic Institute; [4]University of Arizona

Project Goals: The Persistence Control Science Focus Area (PerCon SFA) at Pacific Northwest National Laboratory seeks to understand plant-microbiome interactions in bioenergy crops to establish plant growth–promoting microbiomes that are contained to the rhizosphere of a target plant. This vision requires the discovery of exudate catabolism pathways from plant roots, the elimination of genes that support fitness in bulk soil environments without decreasing rhizosphere fitness, and the engineering of rhizosphere niche occupation traits in phylogenetically distant bacteria. The team anticipates the impacts of these efforts will be to increase understanding of plant-microbe interactions and to extend high-throughput systems and synthetic biology tools to non-model microbes.

Metabolic modeling in bacterial genomes or metagenomes depends on several critical steps. Assigning functions to predicted genes and populating known metabolic pathways and modules with annotated genes precede the assembly a draft model. Typically, the limitations of functional annotation reveal critical reaction steps, or gaps, in a draft model for which no gene was identified. Further, comparison of the draft model against experimentally derived phenotypic data, such as cell growth experiments, can identify additional gaps or inconsistencies in the model. Gaps are typically filled either by simply assuming the that the function exists with no specific gene assignment to the step or by relaxing thresholds and computationally or manually searching original annotations for lower confidence or less specific annotations. There are multiple significant impediments to the metabolic modeling and gap-filling process: (1) incomplete genome information, including for metagenome-assembled genomes (MAGs); (2) low-confidence annotations from bacteria that are evolutionarily distant from well-studied organisms; and (3) the large pool of protein families that have either no functional assignment or a non-specific function.

This team has developed two tools to overcome these deficiencies. MetaPathPredict (Geller-McGrath et al. 2022) is a deep learning framework for prediction of complete metabolic modules in genome-scale models from incomplete genome data. Employing a gradient boosted trees [XGBoost (Chen and Guestrin 2016)]



Metabolic Model. Machine learning is the basis of many new tools for processing a wide and diverse range of experimental data for the purpose of building (Snekmer), improving (MetaPathPredict), and parameterizing (OMics-Enabled Global Gapfilling app) metabolic network models to optimize their accuracy and biological relevancy. [Courtesy Pacific Northwest National Laboratory]

*Presenter

and neural network stacked ensemble classification framework, MetaPathPredict can accurately predict the presence of metabolic modules with as much as 60 to 70% of the genome missing by learning from complete genomes. The team also developed Snekmer (Chang et al. 2023), a generalized computational framework to model protein function families using sequence recoding into reduced-complexity amino acid alphabets and a k-mer based approach. Snekmer can be used to rapidly develop models for novel protein families and to screen genomes or metagenomes to evaluate the distribution of the family across organisms or environments. Using these tools to fill gaps will improve researchers' ability to develop metabolic models from limited genome data, including MAGs. Team members have successfully integrated Snekmer as a KBase app, and next plan to integrate MetaPathPredict. Both tools will function with the new OMics-Enabled Global Gapfilling (OMEGGA) app in KBase to allow iterative model building, gap filling, and model refinement.

Funding Statement: This work was supported by the DOE BER program and is a contribution of the Science Focus Area Persistence Control of Engineered Functions in Complex Soil Microbiomes, the DOE/BER "Improved Protein Annotation in KBase Using Machine Learning, Multiomics Data Integration, and Structural Prediction" project, the DOE/BER "Machine-Learning Approaches for Integrating Multiomics Data to Expand Microbiome Annotation" project, National Science Foundation (1856556), the Defense Threat Reduction Agency, under Interagency Agreement (DTRA13081–37739), and the Proxy Applications for Converged Workloads (PACER) project, a component of the Laboratory Directed Research and Development program at Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

References

Chang, C. H., et al. 2023. "Snekmer: A Scalable Pipeline for Protein Sequence Fingerprinting Based on Amino Acid Recoding," *Bioinformatics Advances* **3**(1), vbad005. DOI:10.1093/bioadv/vbad005.

Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–95. DOI: 10.1145/2939672.2939785.

Geller-McGrath, D., et al. 2022. Preprint. "MetaPathPredict: A Machine Learning-Based Tool for Predicting Metabolic Modules in Incomplete Bacterial Genomes," *bioRxiv*. DOI:10.1101/2022.12.21.521254.

Leveraging Large Language Models to Synthesize and Develop New Questions

Paramvir S. Dehal[1]* (psdehal@lbl.gov), Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], Joseph Bezouska[2], David Dakota Blair[4], Mikaela Cashman[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Ellen G. Dow[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Andrew Freiburger[2], Tianhao Gu[2], Prachi Gupta[1], A. J. Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Dileep Kishore[3], Dan Klos[2], Filipe Lui[2], David Lyon[1], Cody O'Donnell[1], Christopher Neely[1], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Pamela Weisenhorn[2], Ziming Yang[4], Shinjae Yoo[4], Sijie Xiang[1], Qizhi Zhang[2], Shane Canon[1], Elisha Wood-Charlson[1], Robert Cottingham[3], Christopher S. Henry[2], **Adam P. Arkin[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Argonne National Laboratory; [3]Oak Ridge National Laboratory; [4]Brookhaven National Laboratory

kbase.us

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.

In the rapidly evolving landscape of biological data analysis, KBase is uniquely positioned with respect to its capabilities that combine analytic tools, large-scale data compendia, user data, and publishing and sharing. To leverage these capabilities, this team is starting two new initiatives that are driven by custom trained large language models (LLMs). The first initiative is the use of LLM-powered artificial intelligence (AI) agents that assist users in their analysis and the interpretation of those results. And second, LLMs can assist users with search and discovery of data within KBase to help formulate and evaluate hypotheses. This presentation shows the progress made in these areas. This includes the creation of an AI agent with a natural language interface that guides the user through the analysis of a microbial genome from the reads through to a genome paper. This agent is capable of invoking all the necessary tools and assisting the user in interpreting the output of those tools. In addition, this presentation will discuss the development of an LLM infrastructure to enhance the search and discovery of data within KBase. This includes creating a natural language query interface, personalizing the search experience, and employing intelligent data retrieval and reasoning to answer complex scientific questions. By implementing these AI-enhanced capabilities, KBase aims to offer a more intuitive and effective platform for scientific exploration through a collaborative environment.

Getting Credit for Contributions in a Big Data World

Elisha Wood-Charlson[1]* (elishawc@lbl.gov), Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], Joseph Bezouska[2], David Dakota Blair[4], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Ellen G. Dow[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Andrew Freiburger[2], Tianhao Gu[2], Prachi Gupta[1], A. J. Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Dileep Kishore[3], Dan Klos[2], Filipe Lui[2], David Lyon[1], Cody O'Donnell[1], Mikaela McDevitt[1], Christopher Neely[1], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Pamela Weisenhorn[2], Ziming Yang[4], Shinjae Yoo[4], Sijie Xiang[1], Qizhi Zhang[2], Shane Canon[1], Paramvir S. Dehal[1], Robert Cottingham[3], Christopher S. Henry[2], **Adam P. Arkin[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Argonne National Laboratory; [3]Oak Ridge National Laboratory; [4]Brookhaven National Laboratory

kbase.us

Project Goals: The DOE Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.

Many of the concerns researchers have around sharing data include knowledge barriers, reuse concerns, and disincentives (Gomes et al. 2022). KBase already addresses components of knowledge barriers through its outreach strategy (e.g., in-person and virtual training sessions, documentation, and robust hands-on activities) and reuse concerns through the basic functionality of the platform (e.g., provenance, interoperability, and reproducibility) (Arkin et al. 2018). KBase is now working to address disincentives: concerns around getting scooped, the time it takes to curate and share data, and lack of clarity around the rewards that come from embracing open science and good data management.

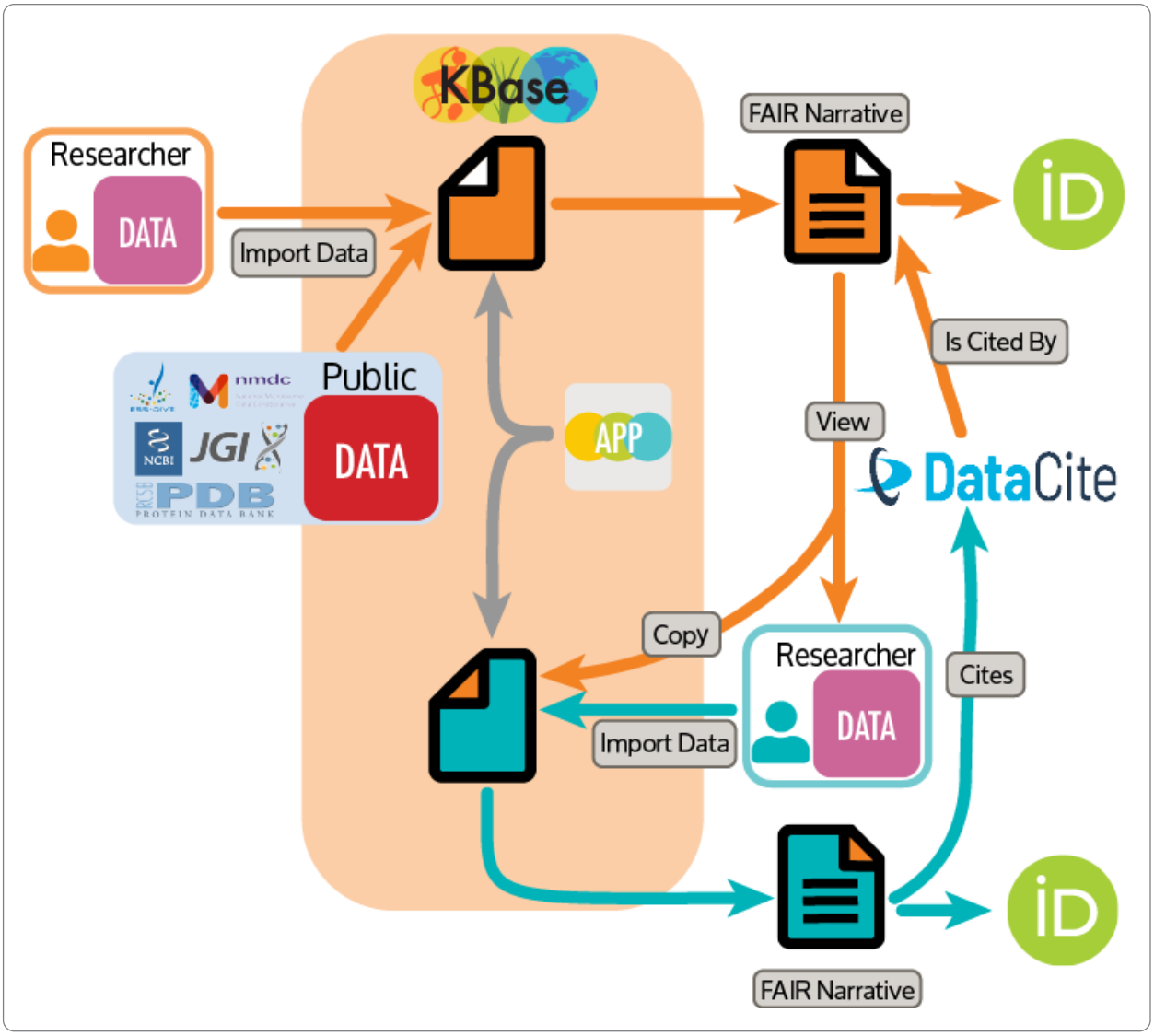
Behind much of this is the fundamental tenet that neither the current culture of science nor the publishing infrastructure value data outside of a publication. In partnership with several efforts across BER and the publishing world, KBase is establishing its platform as a change agent focused on "getting and giving credit to data" (Wood-Charlson et al. 2022). Leveraging Persistent Identifiers (PIDs), group members are developing linkages to/from and within KBase that support data management best practices and ensure credit is retained. An example user (PID: ORCID) research workflow:

- Collect samples and assign them International General Sample Numbers (PID: IGSN),
- Submit sample metadata (Standard: MIxS) to the National Microbiome Data Collaborative (NMDC) Sample Submission Portal,
- Send sample material to the DOE Joint Genome Institute (JGI) for sequencing,
- Send sample material to the Environmental Molecular Sciences Laboratory (EMSL) for Fourier-transform ion cyclotron resonance (FTICR) mass spectrometry analysis, and
- Submit geochemistry measurements on those samples to Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE).

Leveraging the (in development) Data Transfer Service (DTS), the user could request sample metadata from NMDC, sequence data from JGI, and FTICR data from EMSL to do a global community analysis and build community models. When the user is ready to publish their data and reproducible analyses, the KBase "Credit Engine" provides the user's workflow ("Narrative") with a dataset DOI that captures (Ireland and Wood-Charlson 2023):

1. Important credit metadata [e.g., authors/contributors (PID: ORCID), funder information (PID: Research Organization Registry, ROR)]
2. Citations for the funded proposal/data management plan (PID: DOI), samples (PID: IGSN), public data (PID: DOI, when available), and tools (DOI, typically a publication) that contributed to the analysis.

The KBase DOI is registered by Office of Scientific and Technical Information and submitted to DataCite, which links the now FAIR (findable, accessible, interoperable, and reusable) Narrative to the broader publishing infrastructure, and shared back to the



DOE Biology Knowledgebase (KBase). Researchers using KBase can automatically get credit for their research by requesting a DOI for their KBase Narrative associated with a publication. [Courtesy KBase Credit Engine Team]

ESS-DIVE project. These connections enable KBase to start tracking and reporting the reuse of shared data.

Funding Statement: This work is supported as part of the Genomic Science program of BER. The DOE Systems Biology Knowledgebase (KBase) is funded by the DOE Office of Science, BER program under DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

References

Arkin, A. P., et al. 2018. "KBase: The United States Department of Energy Systems Biology Knowledgebase," *Nature Biotechnology* **36**(7), 566–9. DOI:10.1038/nbt.4163.

Gomes, D. G. E., et al. 2022. "Why Don't We Share Data and Code? Perceived Barriers and Benefits to Public Archiving Practices," *Proceedings of the Royal Society B: Biological Sciences* **289**, 20221113. DOI:10.1098/rspb.2022.1113.

Ireland, A. J., and E. M. Wood-Charlson. 2023. "KBase Credit Metadata Schema." Accessed April 2024. KBase. DOI:10.25982/1984203.

Wood-Charlson, E. M., et al. 2022. "Ten Simple Rules for Getting and Giving Credit for Data," *PLoS Computational Biology* **18**(9), e1010476. DOI:10.1371/journal.pcbi.1010476.

CHAPTER 6
Bioimaging

Biological Systems Science Division supports research that focuses on creating new bioimaging instrumentation capabilities, enabling selective and deeper penetration to non-destructively resolve cellular structures. These tools are intended for imaging biological processes within and among cells in living plants and microorganisms. Multimodal imaging capabilities are featured prominently.

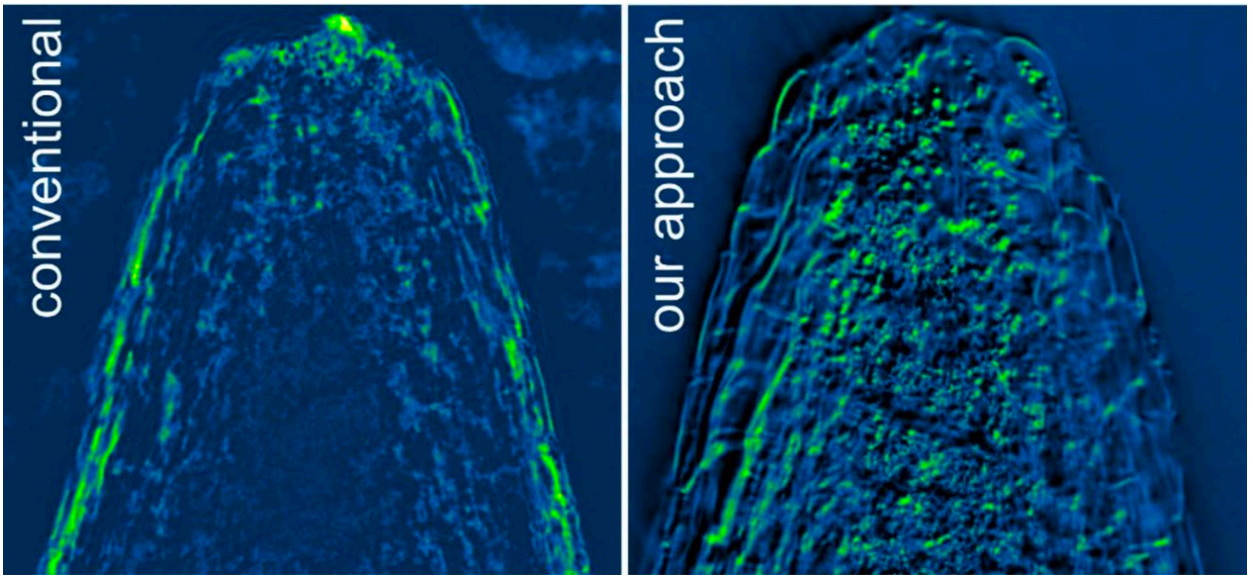
Integrative Imaging of Plant Roots During Symbiosis with Mycorrhizal Fungi

Andreas E. Vasdekis[1]* (andreasv@uidaho.edu, PI), Scott E. Baker[2], Demetrios Christodoulidis[3,4], Maria Harrison[5], Luke Sheneman[1], Ram Kasu[1], Jinming Zhang[1], Haokun Luo[3,4]

[1]University of Idaho; [2]Pacific Northwest National Laboratory; [3]University of Central Florida; [4]University of Southern California; [5]Boyce Thompson Institute

Research Plan: The goal of this project is to create an integrative optical imaging platform capable of independently quantifying the growth of plant roots and their metabolic interactions with symbiotic mycorrhizal fungi. To accomplish this, this research team is integrating interferometric (or quantitative phase) imaging with light-sheet fluorescence and Raman microscopy. This team's approach to enable deep imaging within root tissue also involves the strategic combination of specially designed optical beams with ultralow light detection. Researchers deploy this strategy in order to overcome the challenges of the spatiotemporal degradation experienced by light as it propagates through tissue. To support these methodological advancements, group members deploy a mix of theoretical analyses, deep learning techniques for image reconstruction, and the development of dedicated biomarkers.

Current and Anticipated Accomplishments: Since its beginning (2021), the project has concentrated on four areas. First, the team combined photon-sparse imaging with Airy light-sheet microscopy to demonstrate video Raman imaging rates at more than 1,000-fold lower irradiance than current approaches (Dunn et al. 2023; Dunn et al. 2024). This technique allowed researchers to quantify fungal metabolic using a deuterated biomarker, while the group is presently concluding its investigations on further improving these gains via alternative image reconstruction strategies (Sheneman et al. in review 2024). Second, the team developed a quantitative-phase microscope capable of measuring the dry density and mass within root tissue. This method integrates asymmetric illumination interferometry with standard differential interference contrast microscopes and enables the visualization of distinct features in the meristem of roots up to ~500 μm in diameter (see figure, this page; Zhang et al. 2023 in review; Zhang and Vasdekis 2023). Third, the group constructed a light-sheet fluorescence microscope that improves imaging efficiency in tissue via specially designed optical beams that can self-heal after scattering, and demonstrated the benefits of this technique in root tissue imaging. Fourth, the group is expanding its



Quantitative-Phase Image of a *Medicago truncatula* Root-Tip. Left: Using conventional methods and **(Right)** this team's approach. [Adapted from Zhang, J., et al. 2024. "Quantitative Phase Imaging by Gradient Retardance Optical Microscopy," *Scientific Reports* **14**, 9754. DOI:10.1038/s41598-024-60057-y. Republished under Creative Commons license (CC BY 4.0)]

*Presenter

palette of biomarkers to track cellular changes underlying host cell accommodation of symbionts. Next steps involve: (1) enhancing the imaging depth in root tissue by refining optical fields temporally and spatially; and (2) investigate plant growth in microfluidics prior to transitioning to imaging the fungal-root symbiosis.

Benefits and Applications: This optical imaging system provides quantitative insights into root-fungi interactions, supporting the DOE's energy prosperity goals with innovative tools, while it relies on commercial hardware and open-source software, improving its availability to the wider scientific community.

References

Dunn, L. et al. 2023. "Video-Rate Raman-Based Metabolic Imaging by Airy Light-Sheet Illumination and Photon-Sparse Detection," *Proceedings of the National Academies of Sciences* **120**(9), e2210037120 (2023). DOI:10.1073/pnas.2210037120.

Dunn, L., et al. 2024. "Video-Rate Spontaneous Raman Imaging and Method for Using." Patent Application 18/408,198.

Sheneman, L., et al. 2024. In review. "Imaging with No Photons."

Zhang, J., and A. E. Vasdekis. 2023. "Cost-Effective Interferometric Imaging Module Compatible with Standard Commercial Microscopes." Provisional Patent Application 63/601582.

Zhang, J., et al. In Review. "Gradient Retardance Optical Microscopy."

Ultra-Sensitive High-Resolution Label-Free Nonlinear Optical Microscopy for Imaging Plant-Microbe Interactions *In Vivo*

Na Ji[1]* (jina@berkeley.edu, PI), Trent Northen[2], John Vogel[2]

[1]University of California–Berkeley; [2]Lawrence Berkeley National Laboratory

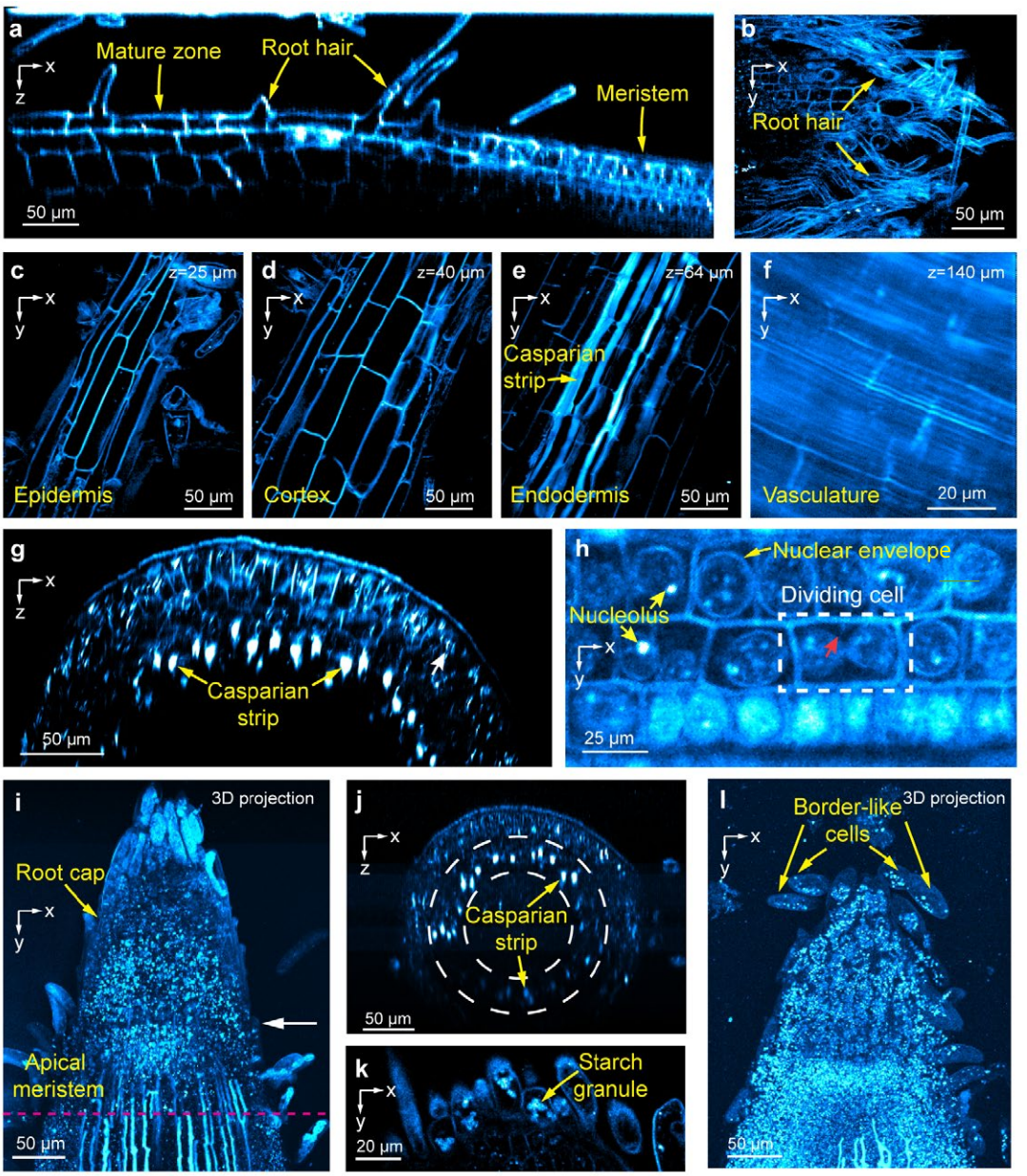
Project Goals: Root biology is pivotal in addressing global challenges including sustainable agriculture and climate change. However, roots have been relatively understudied among plant organs, partly due to the difficulties in imaging root structures in their natural environment. This team aims to develop advanced optical microscopy techniques to reduce photodamage and improve imaging resolution and depth for live plant roots and microbes grown in microfabricated ecosystems (EcoFABs).

Researchers have succeeded in using microfabricated ecosystems (EcoFABs) to establish growing environments with optical access and employing nonlinear multimodal microscopy of third-harmonic generation (THG) and three-photon fluorescence (3PF) to achieve label-free, *in situ* imaging of live roots and microbes at high spatiotemporal resolution.

THG has enabled researchers to observe key plant root features in mature and meristem roots including laminar structures down to the vasculature, Casparian strips, dividing meristematic cells, root cap and border-like cells, as well as resolve subcellular features including nuclear envelopes, nucleoli, starch granules, and putative stress granules (see figure, p. 48). THG from the cell walls of bacteria and fungi also provides label-free contrast for visualizing these microbes in the root rhizosphere. With simultaneously recorded 3PF fluorescence signal, the team has achieved single-bacterium tracking and subcellular imaging of fungal spores and hyphae in the rhizosphere, indicating this method's potential for studying plant-microbe interactions.

To improve image resolution, the team combined adaptive optics with THG microscopy. By measuring and correcting sample-induced optical aberrations on the excitation light, adaptive optics has led to a substantial increase in root THG signal.

The group is now working on optimizing homodyne-mixing setups for both THG and second harmonic generation (another coherent light scattering process that provides structural contrast without external labels) microscopy. By interfering a larger reference harmonic signal with the harmonic signal generated by the sample, researchers can enhance the detectability of the sample signal thus reduce the amount of light needed for harmonic generation and its associated photodamage.



Label-Free Structural Imaging of Live *Brachypodium distachyon* Roots Using Third-Harmonic Generation Microscopy. (a) Axial (xz) image of a root from mature zone to meristem. **(b)** Lateral (xy) images of structural features in the mature root including **(c)** root hairs enveloping root surface, **(d)** epidermis, **(e)** cortex, **(f)** endodermis, and **(g)** vasculature of mature root. **(h)** Axial image of meristem root and **(i)** lateral view of subcellular structures. **Red arrow:** Emerging cell wall. Images of root tips, including **(j)** brightest-spot projection with depth cueing (100% to 50%) of a 230-micrometer (µm)-thick image stack through a root tip, **(k)** axial image of apical meristem along the pink line in i, **(l)** root cap cells, and brightest-spot projection with depth cueing (100% to 50%) of an 88-µm-thick image stack of a root tip. Post-objective power of 1300 nanometer excitation light: **(a)** 3 milliwatts; **(b)** 4 milliwatts; **(c-e)** 2.6 milliwatts; **(f)** 14 milliwatts; **(g)** 5.3 milliwatts; **(h)** 7 milliwatts; **(i-j)** 3 to 15 milliwatts; **(k)** 4 milliwatts; **(l)** 2.4 to 2.8 milliwatts. [Modified from Pan, D., et al. 2024. "Label-Free Structural Imaging of Plant Roots and Microbes Using Third-Harmonic Generation Microscopy," *bioRxiv*. Preprint. DOI:10.1101/2024.04.13.589377. Republished under Creative Commons license (CC BY-NC-ND 4.0)]

Deep Chemical Imaging of the Rhizosphere

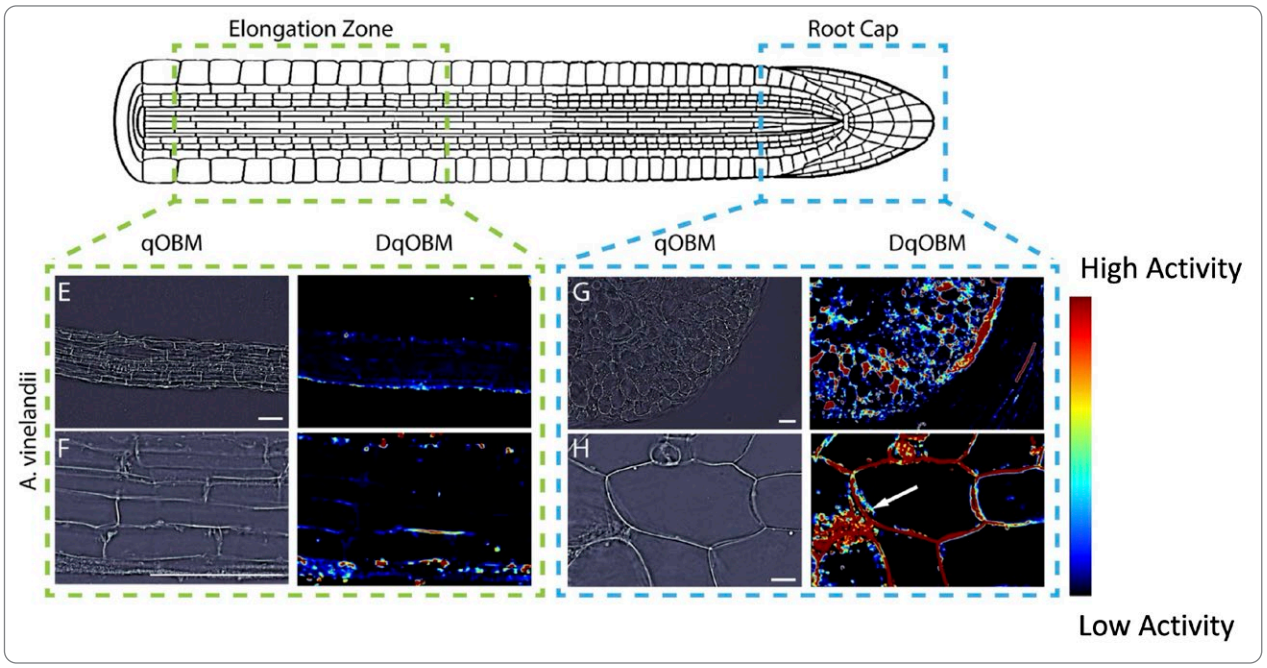
Marcus T. Cicerone[1]* (cicerone@gatech.edu, PI), Abigail Diering[1], Madison Green[1], Rajas Poorna[1], Caroline Filan[1], Lily S. Cheun[1], Joel E. Kostka[1], Francisco E. Robles[1,2]

[1]Georgia Institute of Technology; [2]Emory University

Beneficial diazotrophic microbes promote plant growth and productivity by consuming sugars and other compounds exuded by roots and, in turn, provide fixed nitrogen to the plant. Although a large fraction of the carbon fixed by plants during photosynthesis is secreted through roots to sustain the root microbiome, this metabolic exchange is not well understood. Understanding the nitrogen-carbon nexus will help to develop transformative biofertilization technologies that require a smaller carbon commitment from plants to the nitrogen-fixing microbes. This research team is building a label-free microscope to image metabolic activity and chemical exchange between plants and bacteria deep within thick living plant roots and their associated rhizosphere microbial communities.

In year three, the team continued to develop quantitative phase and broadband coherent Raman imaging modalities for quantifying dynamic and chemical aspects of the carbon-nitrogen nexus in the rhizosphere. The team used quantitative oblique back-illumination microscopy (DqOBM) to image bacterial culture of three different diazotrophs (*Sinorhizobium meliloti*, *Azotobacter vinelandii*, and *Rahnella aquatilis*) cultured on media. Results from these imaging experiments were compared with DqOBM images from the root cap and elongation zone of *Arabidopsis thaliana* primary roots that had been inoculated with either *A. vinelandii* or *R. aquatilis* the primary root system. This study's findings demonstrate that quantitative dynamic phase imaging can effectively characterize microbial dynamics and provide insights into plant-microbe interactions *in situ.*

Trends in dynamics observed with DqOBM were consistent with a dependence of energy level on carbon and nitrogen, and the team demonstrated strong linear correlations between the nitrogen fixation and the



***Arabidopsis thaliana* Roots Inoculated with *Azotobacter vinelandii*. (E,F)** The elongation zone with minimal dynamics. **(G,H)** Increased activity in the root cap. [Adapted from Filan, C., et al. 2024. "Label-Free Functional Analysis of Root-Associated Microbes with Dynamic Quantitative Oblique Back-Illumination Microscopy,' *Scientific Reports* **14**, 5812. DOI:10.1038/s41598-024-56443-1. Republished under Creative Commons license (CC BY 4.0)]

*Presenter

DqOBM signal energy. Since the signal energy level measured by DqOBM is a sum of all microbial activities, it is necessary to have orthogonal information to directly link dynamics observed by imaging with rates of growth or nitrogen fixation.

Researchers also used broadband coherent anti-Stokes Raman imaging to identify spectral changes in bacteria under the same conditions where DqOBM was used. Raman spectra of *S. meliloti* are obtained during lag, exponential, and stationary growth phases under varying media conditions designed to induce nitrogen fixation. Known Raman signatures of general metabolic activity and nitrogen fixation were correlated to these culture conditions.

Additionally, broadband coherent anti-Stokes Raman scattering (BCARS) was used to image *Medicago truncatula* nodules that host *S. meliloti* in the infection thread and nodule zones where bacteria exhibit characteristics of these different growth and media conditions. Researchers used SampleMap—a variant of uniform manifold approximation and projection the team adapted for Raman imaging—to map the Raman signatures obtained from the Medicago nodules to phenotypes expressed in the culture experiments. These SampleMaps will ultimately be compared to high-resolution spatial transcriptomic maps of sectioned root nodules whose sister serial slices have been imaged with BCARS.

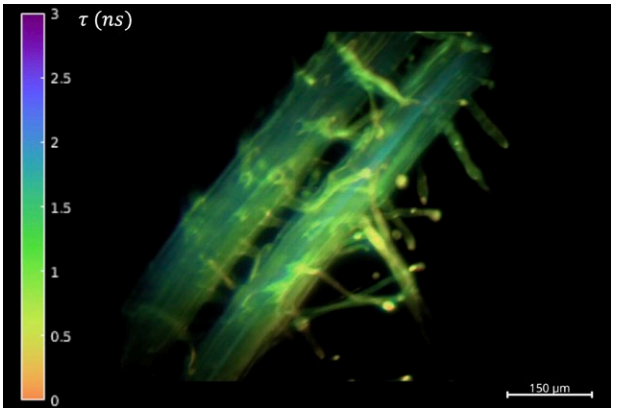
Development of High-Throughput Light-Sheet Fluorescence Lifetime Microscopy for 3D Functional Imaging of Metabolic Pathways in Plants and Microorganisms

Adam Bowman[1,3*] (abowman@salk.edu), Dara Dowlatshahi[1,2], Rose Knight[1,2], Nils Bode[1], Soichi Wakatsuki[1,2], **Mark Kasevich[1]** (PI)

[1]Stanford University; [2]SLAC National Accelerator Laboratory; [3]Salk Institute for Biological Studies

The project goal is to realize a high-speed lifetime imaging platform for light-sheet microscopy of metabolic pathways and interactions between plants and soil bacteria using electro-optic fluorescence lifetime microscopy (EO-FLIM) (Bowman et al. 2019; Bowman and Kasevich 2021). Wide-field optical modulators allow efficient lifetime capture combined with low noise readout on standard scientific cameras. This group presents results from its fluorescence lifetime light-sheet microscopy platform. Images acquired in a selective plane illumination microscope are gated using a Pockels cell driven at 80 megahertz, enabling light-sheet FLIM with up to 800 micrometer (μm) field of view. Volumetric lifetime acquisitions are demonstrated on live *Arabidopsis thaliana* root samples using both genetically encoded fluorescent proteins and endogenous autofluorescence. The group also presents application of EO-FLIM to record neuron activity *in vivo* at kilohertz frame rates using a genetically encoded voltage indicator (Bowman et al. 2023).

Funding Statement: This research was supported by the DOE Office of Science, through the Biomolecular Characterization and Imaging Sciences program, BER program, grant DE-SC0021976.



Fluorescence Lifetime Light-Sheet Microscopy. *Arabidopsis* root labeled with green fluorescent protein (maximum intensity projection). [Courtesy Stanford University]

References

Bowman, A. J., and M. A. Kasevich. 2021. "Resonant Electro-Optic Imaging for Microscopy at Nanosecond Resolution," *ACS Nano* **15**(10), 16043–54. DOI:10.1021/acsnano.1c04470.

Bowman, A. J., et al. 2019. "Electro-Optic Imaging Enables Efficient Wide-Field Fluorescence Lifetime Microscopy," *Nature Communications* **10**(1) DOI:10.1038/s41467-019-12535-5.

Bowman, A. J., et al. 2023. "Wide-Field Fluorescence Lifetime Imaging of Neuron Spiking and Subthreshold Activity *In Vivo*," *Science* **380**(6651), 1270–5. DOI:10.1126/science.adf9725.

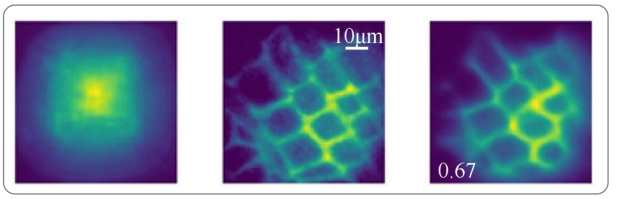
*Presenter

Novel *In Vivo* Visualization of Bioenergy Metabolic and Cellular Phenotypes in Living Woody Tissues

Andrew Groover[2]* (andrew.t.groover@usda.gov), Al Ingold[1], Ruipeng Guo[1], Rajesh Mennon[1], Reed S. Sorenson[1], Gayatri Mishra[1,2], Dasmeet Kaur[3], Chang-Jun Liu[3], **Leslie E. Sieburth[1]** (PI)

[1]University of Utah; [2]U.S. Forest Service, Northern Research Station; [3]Brookhaven National Laboratory

This research team is developing new approaches for live cell imaging within the wood-forming tissues of trees enabling deeper understanding of developmental and physiological processes underlying wood formation and function directly in woody bioenergy feedstocks. A primary challenge is that the dividing and differentiating cells of interest are embedded under thick layers of light-scattering bark tissues and are well beyond the working distances of traditional light microscopes. The group is developing two types of microendoscopic, implantable imaging probes to access and visualize wood-forming tissues: miniscopes and fiber optic cannulas. These microscopy approaches are being tested by applying them to the following problems relevant to bioenergy feedstock development in poplar: (1) analysis of lignification and impacts of the altered cell wall polysaccharides on lignin formation; (2) vessel element differentiation and the impact of abscisic acid levels on vessel cell properties; and (3) fiber development in both tension-wood-inducing and normal growth conditions. Previously, the group reported progress in fabricating and using embedded optical probes to carry out live cell imaging in poplar stems. This poster reports improvements to epifluorescence miniscopes fitted with implantable Gradient Index (GRIN) lenses inserted to reach internal tissues with minimal tissue damage. This approach has the advantage of directly rendering images and video in real time. It can be used in combination with fluorescent probes and dyes compatible with a range of different excitation and emission filters. However, miniscope resolution is challenged by light scattering in woody stems with complex autofluorescence. The group is using ring deconvolution microscopy (Kohli et al. 2023), a technique for computational aberration with a calibration image of randomly distributed fluorescent microspheres. The second approach uses computational cannula microscopy with an individual cannula or an optrode array (Guo et al. 2023), which can be modified to enable hyperspectral imaging. This approach has the advantages of smaller diameter probes (0.22 mm) and larger fields of view (0.20 mm). However, in contrast to the miniscope, machine learning–based image processing algorithms are necessary to convert spatially scrambled fluorescent signals into images. The group implemented a generative adversarial convolutional neural network that surpasses the capabilities of the prior U-Net architecture (Isola et al. 2018). This advancement enables reconstructed images of living poplar branches, which display low contrast and dense structures and capture of dynamics at a cellular level. Finally, researchers will discuss opportunities to extend live cell imaging capabilities for woody plant bioenergy applications.



Imaging Sections of Stems. **Left:** A cannula directs excitation light into the plant and collects emitted fluorescence, which is then recorded by a camera. **Middle:** Ground-truth image obtained using a conventional reference microscope. **Right:** Reconstructed image after computational post-processing of the recorded image. [Adapted with permission from Guo, R., et al. 2023. "Overcoming the Field-of-View to Diameter Trade-Off in Microendoscopy via Computational Optrode-Array Microscopy," *Optica Express* **31**(5). DOI:10.1364/OE.478314. © 2024 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement]

Funding Statement: This research was supported by the DOE Office of Science BER program, under award number DE-SC0021996, and Interagency Agreement Number 89243021SSC000074 to AG.

References

Guo, R., et al. 2023. "Overcoming the Field-of-View to Diameter Trade-Off in Microendoscopy via Computational Optrode-Array Microscopy," *Optics Express* **31**(5), 7505–14. DOI:10.1364/OE.478314.

Isola, P., et al. 2018. "Image-to-Image Translation with Conditional Adversarial Networks," *arXiv.* DOI:10.48550/arXiv.1611.07004.

Kohli, A., et al. 2023. "Ring Deconvolution Microscopy: An Exact Solution for Spatially-Varying Aberration Correction," *arXiv*. DOI:10.48550.arXiv.2206.08928.

Novel Multimodal Chemical Nano-Imaging Technology to Visualize and Identify Small Biomolecules Exchanged in Microbial Communities

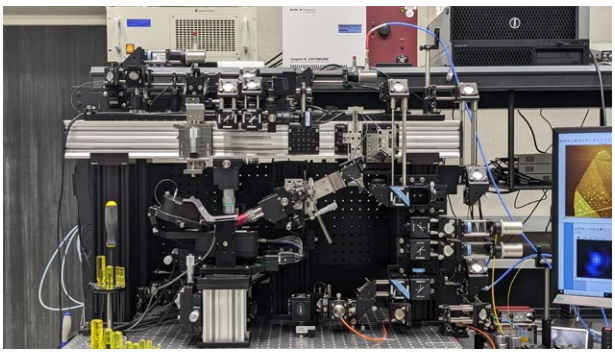
Scott Lea[1]* (scott.lea@pnnl.gov, PI), Patrick El-Khoury[1], Sarah Akers[1], Edo Aprà[1], Victoria Orphan[2]

[1]Pacific Northwest National Laboratory;
[2]California Institute of Technology

The goal of this project is to develop bioimaging technologies that will significantly advance understanding of microbial metabolism and communication in real space. This is an outstanding challenge because existing approaches do not have the required spatial resolution. This project will combine nano-optics tools with multimodal (non)linear optical spectroscopy to identify and image small biomolecules with nanometer spatial resolution under physiological conditions. Specifically, the team plans to enhance the spatial resolution in optical extinction, Raman/fluorescence, and coherent Raman/two-photon fluorescence/second-harmonic generation spectroscopy down to ~1 to 2 nanometers under ambient conditions to visualize metabolites involved in a wide range of microbial and plant processes.

Though linear nano-optical measurements have been demonstrated, nonlinear nano-optical measurements comprise a novel high risk, high reward aspect of this project. Nonlinear nano-optical measurements, while providing improved signal-to-noise ratios, are challenging when chemical imaging and identification of biomolecules is the goal as this requires spectrally resolved detection schemes, e.g., in coherent Raman-based vibrational nanoimaging and nanospectroscopy. These coherent nano-optical measurements, however, require long collection times that restrict their usefulness in a point scanning hyperspectral nanoimaging scheme where full-time traces must be recorded at every position. The team plans to overcome this difficulty by decreasing collection times by orders of magnitude using time-series analysis in combination with machine learning.

With these measurements, an entirely novel set of nanoscopic selection rules is expected. Team members are developing a theoretical framework to assign experimental observables—primarily the linear and nonlinear optical signatures of biomolecules—by coupling *ab initio* molecular dynamics computed optical spectra to classical finite difference time-domain simulations to reproduce experimental plasmon-enhanced spectral nanoimages. As these simulations are time-consuming, the team will apply machine learning and time-series analysis to dramatically accelerate these simulations.



BIGTUNA. Instrumentation showing the AFM (lower left) and associated light sources and optics. The instrumentation provides top-, side- and bottom-access to the sample using numerous linear and nonlinear near-field imaging modalities (as described in the text) for nanoscale biochemical imaging. [Courtesy Pacific Northwest National Laboratory]

To optimize and validate the performance of this technology for bioimaging applications, this research team will benchmark the system using environmental consortia of anaerobic methane-oxidizing archaea and syntrophic partner sulfate-reducing bacteria. Using these consortia, this technology will allow researchers to spatially resolve the electronic and vibrational signatures of large multiheme cytochromes embedded in the extracellular matrix, thereby providing the first direct evidence that these proteins predicted in the genomes are exported into the extracellular space.

While this current effort is devoted to optimization of this multimodal nanoimaging technology capable of in-liquid operation, the team is also developing novel nano-optical methods, including (1) ultralow frequency tip-enhanced Raman scattering (Wang et al. 2023a) and (2) broadband extinction nanoimaging and nanospectroscopy (Wang et al. 2023b). The first approach not only allows tracking the chemical identities of bioanalytes, but also enables tracking crystallinity on the nanoscale. Further developments would allow this method to mature into a nano-optical

analogue of X-ray diffraction. The second approach is a novel measurement that boasts subnanometer spatial resolution under ambient laboratory conditions. Initial measurements tracked spatially varying plasmon resonances throughout the formation of a junction plasmon. This effort sheds light on the fundamental mechanisms behind optical nanospectroscopy and nanoimaging, which is important for multimodal spectral nanoimaging of biomolecules. More generally, the use of the probe as a nanoscopic broadband light source will allow measuring the nano-extinction spectra of yoctomolar concentrations of biomolecules.

References

Wang, C.-F., et al. 2023a. "Probing Local Optical Fields via Ultralow Frequency Raman Scattering from a Corrugated Probe," *Journal of Physical Chemistry Letters* **14**, 8334–8. DOI:10.1021/acs.jpclett.3c02122.

Wang, C.-F., et al. 2023b. "Subnanometer Visualization of Spatially Varying Local Field Resonances that Drive Tip-Enhanced Optical Spectroscopy," *Nano Letters* **23**(19), 9114–8. DOI:10.1021/acs.nanolett.3c03028.

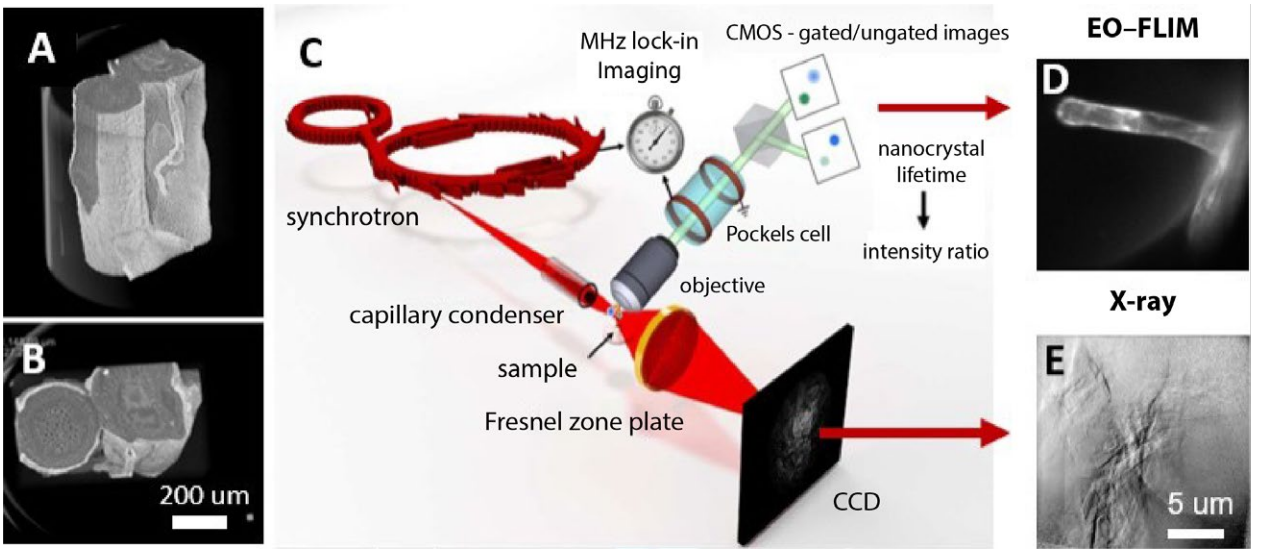
Optical and X-Ray Multimodal-Hybrid Microscope Systems for Imaging of Plant-Pathogen Interactions

Soichi Wakatsuki[1,2]* (soichi.wakatsuki@stanford.edu, PI), Rose Knight[1,2], Dara Dowlashahi[1,2], Adam Bowman[2,4], David Nana Agyeman-Budu[1,2], James Safranek[1], Johanna Nelson Weker[1], Jung-Gun Kim[2], Christopher Takacs[1], Jacob Summers[2], Mary Beth Mudgett[2], Federica Brandizzi[3], Mark Kasevich[2]

[1]SLAC National Accelerator Laboratory; [2]Stanford University; [3]Michigan State University; [4]Salk Institute for Biological Studies

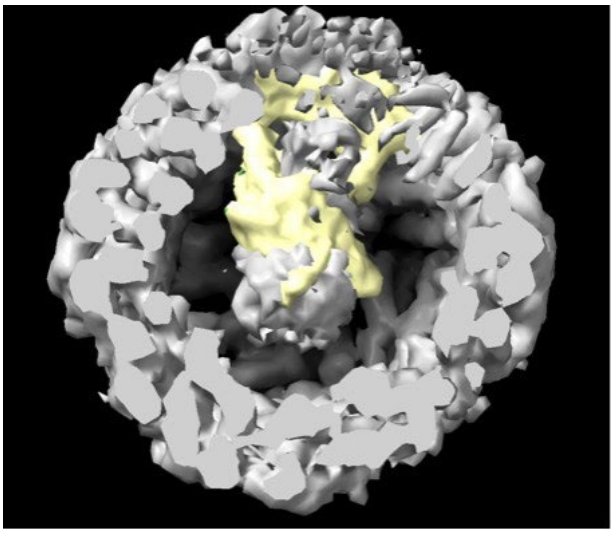
Plant-pathogen interactions are complex and dynamic phenomena, relevant to fundamental, environmental and bioenergy biology. Imaging plays important roles in understanding the interactions at the atomic, molecular, subcellular, cellular, to tissue and whole plant scales. Each bioimaging method has its intrinsic limitations in spatial and/or temporal resolution, field of view and depth, and sensitivity. Fluorescence-based optical microscopes have huge advantages in monitoring dynamics of cellular and subcellular events using wide spectral ranges with super-resolution, light-sheet 3D, and wide-field imaging, but cannot go beyond 10 or 20 nanometers spatial resolution. X-ray imaging and tomography can penetrate deeper than light, and with high brilliance synchrotron sources one can reach 10 nm spatial resolution, but with a high risk of sample damage from radiation dose. Cryogenic electron tomography (cryo-ET) can offer tremendous insight into the subcellular organization of organelles and macromolecules down to several nm resolution, but information gained is largely static. Large efforts have gone into correlative imaging but, to date, there is no easy way to correlate more than two modalities directly. Correlating these different modalities in an unequivocal manner will substantially advance understanding of multiscale complex biological phenomena.

This talk presents updates on the development of the next generation correlative X-ray, light, and electron tomography systems (see figure, this page). In particular, the team has modularized electro-optic fluorescence lifetime imaging microscopy (EO-FLIM) (Bowman et al. 2019; Bowman and Kasevich 2021) for use in correlative microscopies, in the deep ultraviolet, or at synchrotron beamlines, preparing for a first implementation alongside transmission X-ray microscopy (TXM), beamline 6-2c, with correlative X-ray/vis optics at SLAC and Stanford Synchrotron Radiation Lightsource. The team will utilize both the micro- and nano-computed tomography (CT) imaging capability of beamline 6-2 to gain hierarchical insights into the dynamics of fungal-pathogen interactions. Towards this



Electro-Optic Fluorescence Lifetime Imaging Microscopy. (A, B) Micro-computed tomography of *Arabidopsis thaliana* root sample at the cellular length scale. **(C)** Proposed lock-in synchronization between X-ray photoluminescence, X-ray microscopy, and FLIM to give both **(D)** EO-FLIM and **(E)** X-ray images. [Courtesy SLAC National Accelerator Laboratory]

*Presenter



Molecular Cage for Correlative Multimodal Imaging. Cryogenic electron microscopy structure of a protein cage containing one fluorescence protein. [Courtesy SLAC National Accelerator Laboratory]

end, the team has successfully collected preliminary tomography data using a laboratory-based micro-CT X-ray source on plant root and leaf samples. The team has also established a standing EO-FLIM microscope at the synchrotron for further uncorrelated data collection and to explore extremes of resonant drive frequencies.

Finally, the team presents advances towards fluorescent protein- and nanocrystal-containing cages (see figure, this page) as molecular tracers for X-ray imaging/microscopy and EO-FLIM and fiducial markers for cryo-ET, including proof-of-concept synchrotron TXM images of leaf samples bombarded with 400 nm nanogold particles. The presentation discusses further plans for the initial application examining fungal-plant pathogen interactions via chitin binding domains at the plant cell surface and subsequent applications in an ongoing study of extracellular vesicles.

Funding Statements: This research is supported by the DOE Office of Science, through the Biomolecular Characterization and Imaging Sciences program, BER program, under BCIS FWP 100878.

References

Bowman, A. J., and M. A. Kasevich. 2021. "Resonant Electro-Optic Imaging for Microscopy at Nanosecond Resolution," *ACS Nano* **15**(10), 16043–54. DOI:10.1021/acsnano.1c04470.

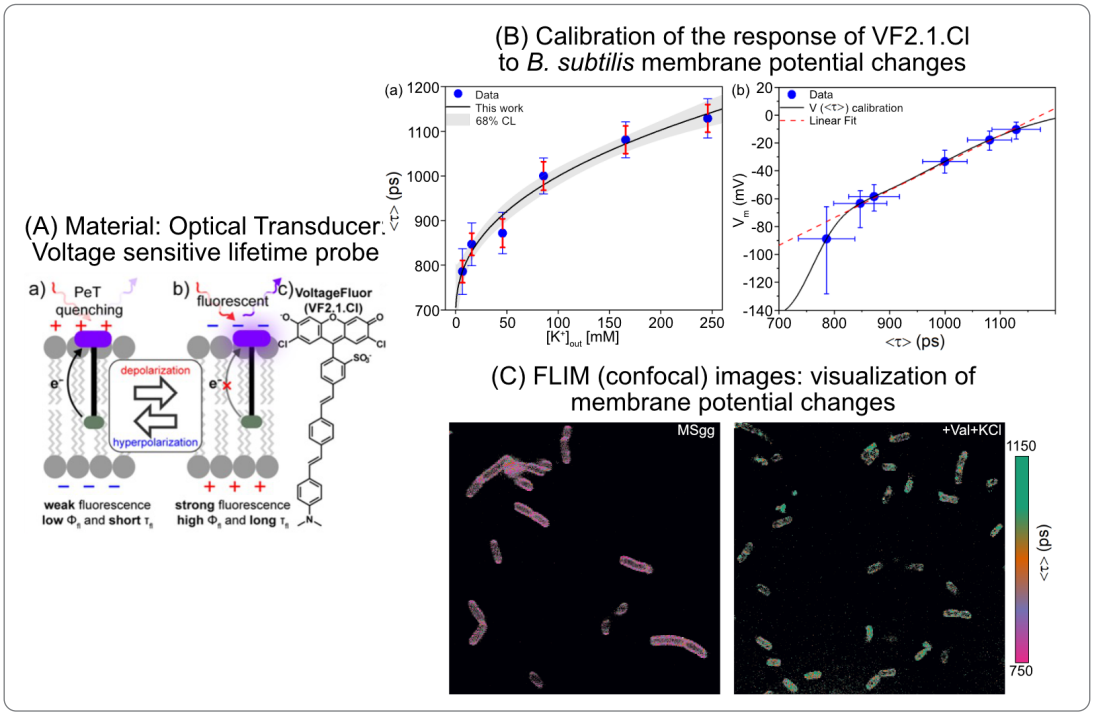
Bowman, A. J., et al. 2019. "Electro-Optic Imaging Enables Efficient Wide-Field Fluorescence Lifetime Microscopy," *Nature Communications* **10**(1) DOI:10.1038/s41467-019-12535-5.

Fluorescence Lifetime-Based Imaging of *Bacillus subtilis* Membrane Potential

Debjit Roy[1]*(deb9804@g.ucla.edu), Xavier Michalet[2], Kiran Bharadwaj[1], Evan W. Miller[3], **Shimon Weiss[1,4]** (PI)

[1]UCLA-DOE Institute for Genomics and Proteomics; [2]University of California–Los Angeles; [3]University of California–Berkeley; [4]Bar-Ilan University

Membrane potential (MP) changes can provide a simple readout of bacterial functional and metabolic state or stress levels. While several optical methods exist for measuring fast changes in MP in excitable cells, there is a dearth of such methods for precise (and calibrated) measurements of steady-state MPs in bacterial cells. Conventional electrode-based methods for the measurement of MP are not suitable for small bacterial cells. Existing optical electrophysiological techniques based on fluorescent Nernstian probes have been successfully used in many studies, but they do not provide precision or absolute quantification of MP or their changes. This team presents a novel, calibrated MP recording approach to address this gap. This group's method uses (1) a unique optical transducer (a chromophore-wire-donor construct), that utilizes intrinsic photoinduced electron transfer (PeT) mechanism to measure MP via its fluorescence lifetime and (2) a quantitative fluorescence lifetime imaging microscopy (FLIM) data analysis based on phasor analysis. In order to visualize individual bacterial cells' MPs under different extracellular conditions, amplitude-averaged lifetime maps were computed from pixel-wise phasor fractions. This allows group members to accurately measure even small MP changes in single bacterial cells. Calibration of membrane potential estimation via phasor-FLIM measurements has been achieved by modulating MP artificially through changing ionic (potassium +) concentration gradients across the membrane utilizing ionophores. Applying this technique to *Bacillus subtilis*, researchers estimated their normal MP at -86 millivolts and a chemically modulated depolarized state at +1 mV. This breakthrough work paves the way for deeper insights into bacterial electrophysiology and bioelectricity research.



Fluorescence Lifetime-Based Imaging of *Bacillus subtilis* Membrane Potential. [Courtesy University of California–Los Angeles and Bar-Ilan University]

*Presenter

CHAPTER 7

BER Biopreparedness Research Virtual Environment (BRaVE)

The DOE Office of Science launched several initiatives in fiscal years 2022 and 2023. These collaborative efforts across different Office of Science research programs include Energy Earthshots Research Centers (EERCs); Science Foundations for Energy Earthshots; Reaching a New Energy Sciences Workforce (RENEW); Funding for Accelerated, Inclusive Research (FAIR); Accelerate Innovations in Emerging Technologies (ACCELERATE); and Biopreparedness Research Virtual Environment (BRaVE). BSSD-supported BRaVE projects are highlighted in this section because they aim to support national biopreparedness and response capabilities that can be advanced through engagement with DOE's unique faciliites and resources.

Developing A National Virtual Biosecurity for Bioenergy Corps Center

Martin Schoonen* (mschoonen@bnl.gov, PI), Kerstin Kleese van Dam, John Shanklin, Alistair Rogers, Robert McGraw, Qun Liu, Shantenu Jha, Paul Freimuth

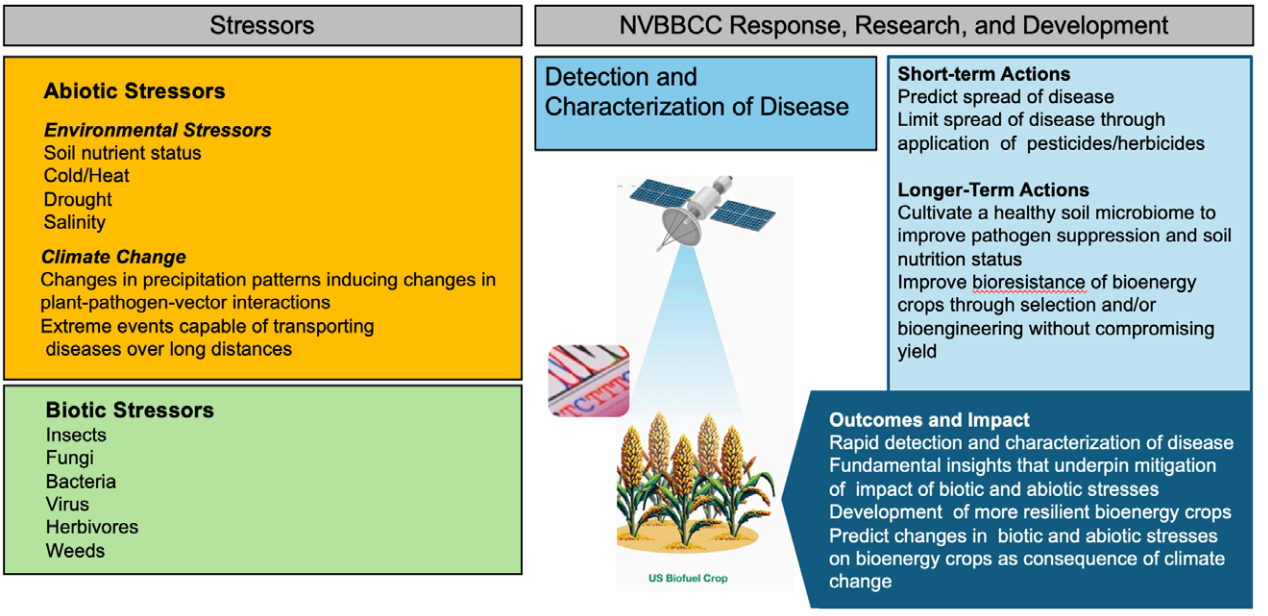
Brookhaven National Laboratory

Background. The development of resilient and sustainable bioenergy crops is an important part of the U.S. strategy to transition to a net-zero economy. An important consideration in developing the U.S. bioeconomy is the biosecurity of crops grown for bioenergy production. The most likely biosecurity threats to bioenergy crops are either known pests or pathogens that emerge in new areas, possibly driven by climate change; or new pests or pathogens that are genetically related to known ones. Here, researchers report on an effort to develop a roadmap for a National Virtual Biosecurity for Bioenergy Crops Center (NVBBCC). In FY23, community input was gathered through workshops. This talk summarizes the results from these workshops.

Approach and Activities. Six community workshops with participants drawn from the DOE laboratory complex, U. S. Department of Agriculture, Department of Human Services, academia, and the private sector were held. Four workshops focused on the following science topics: (1) detection of diseases using remote sensing based on unmanned aerial vehicles; (2) aerial dispersal of disease vectors; (3) biomolecular characterization of plant-pathogen-vector interactions; and (4) disease mitigation strategies. In each of these four workshops, researchers asked: (1) what are the key knowledge gaps; (2) which of these gaps is DOE uniquely positioned to address; and (3) what investments in research infrastructure are needed. A fifth crosscutting workshop focused on computing needs for analytics, modeling, and data distribution as well as workforce development. A sixth focused on a framework for preparedness to respond to the emergence of biothreats in bioenergy crops.

In parallel to the workshops, the team engaged in an experimental study on a known disease in sorghum. Furthermore, a focused ion beam for cryo-tomography, a computer platform to support data-intensive collaborative research, and drone-based sensors were acquired and commissioned.

The input from these workshops as well as insights stemming from research activities conducted during this pilot study are now being summarized in an NVBBCC roadmap document to be delivered to BER.



Schematic Diagram of Stressors on Bioenergy Crops, Disease Detection, and Possible Actions. [Courtesy Brookhaven National Laboratory]

*Presenter

Recommendations. The overarching recommendation is for BER to stand up a long-term research program that aims to safeguard bioenergy crops and establish a response capability to detect and identify emerging diseases in bioenergy crops. Specific recommendations are:

- Apply a systems-level approach to bioenergy crop health that would augment and extend the smart farming approaches that the USDA has adopted to safeguard the nation's agricultural crops. In particular, DOE and BER resources could be applied to advance basic understanding of pathogen-host plant interactions at the molecular level and then apply this knowledge to guide genetic engineering of bioenergy crop plants for increased resilience against disease.
- Develop a capability to predict the impact of climate change on bioenergy (and agricultural) crop security. Building on DOE's long expertise in Earth system modeling, a capability to predict climate-induced changes in host-pathogen ecosystems could guide efforts to avoid or limit the spreading of crop diseases.
- DOE's expertise in stand-off detection and biosensing also could be harnessed to augment the USDA's efforts in crop surveillance.

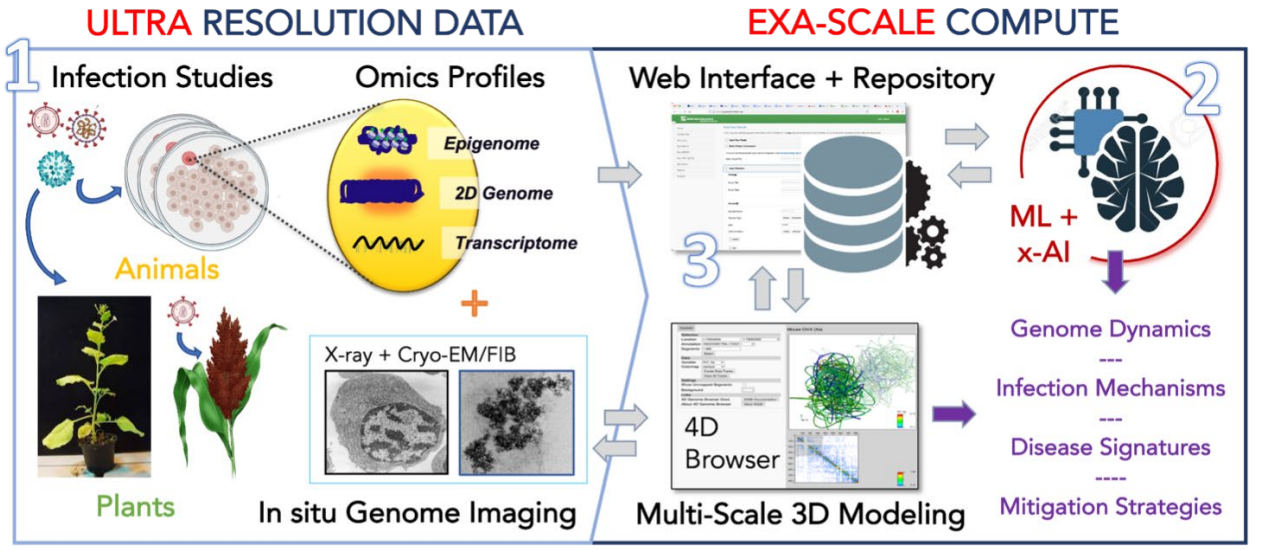
Finally, the envisioned center (NVBBCC) would allow scientists from DOE, USDA, and academia to collaborate and share dedicated research and computing facilities.

Decoding Host-Pathogen Interactions with 4D (Epi)Genomics

Shawn R. Starkenburg[1]* (shawns@lanl.gov, PI), Samantha Adikari[1], Shounak Banerjee[1], Judy Brown[2], Paul Freimuth[3], Daniel Jacobsen[4], Jessica Kubicek-Sutherland[1], Anna Lappala[5], Qun Liu[3], David Rogers[1], Cullen Roth[1], Karissa Sanbonmatsu[1], Ankush Singhal[1], Eric Small[1], Vrinda Venu[1], John Watt[1], Christina R. Steadman[1]

[1]Los Alamos National Laboratory; [2]University of Arizona; [3]Brookhaven National Laboratory; [4]Oak Ridge National Laboratory; [5]Harvard University and Massachusetts General Hospital

The ability to counter biological threats is limited given the enormous lack of knowledge of host resilience mechanisms in the face of pervasive pathogens. The resilience of multicellular hosts (e.g., plants, animals, and humans) is predicated on the susceptibility of individual cells and effective defense mechanisms to halt the spread of infection for pathogen elimination. Recent work in the field of epigenomics suggests that epigenetics play a key role in host defense. Epigenetic mechanisms collectively function to open or close regions of chromosomes to control gene expression. The resulting dynamic structural changes in the genome underpin most biological functions, including responses to infection. Conversely, genome structure can be altered by pathogens to reorient host cellular function to enhance pathogen replication or to establish latent or persistent infections. In response, hosts employ epigenetic modifications to counter infection, thereby altering the expression and 3D spatial configuration of their own genomes. These researchers hypothesize that epigenetic modifications vary between resilient vs. susceptible host cells and that underlying changes to the epigenome and genome are characteristic of pathogen classes. Although recent progress has been made, scientists and decision-makers currently lack methods to quickly compare and identify these pathogen-induced changes to host genomes to understand susceptibility and resiliency. This team proposes to address this gap by taking a holistic, 'one health' approach to (1) develop an experimental workflow to characterize and survey early onset molecular signatures of infection found in the genome and epigenome; (2) leverage Advanced Scientific Computing Research user facilities to create an exascale computational and explainable artificial intelligence (XAI) workflow that integrates this data (and data from the user community) to enable interactive, comparative 4D (3D + time) (epi)genome exploration and predictive dynamic



A Novel Experimental and Computational 4D (Epi)Genomics Platform and Data Repository to Decipher Host-Pathogen Interactions and Infection Mechanisms. (Task 1) In this project, researchers generate single cell (epi)genomic datasets from viral infection panels in plants and mammalian systems **(Task 2)** that will enable exascale artificial intelligence and machine learning approaches. **(Task 3)** Visualization and modeling of 3D genomes will be validated with *in situ* genome imaging. The web interface and data repository will enable the user community to expand the dataset to refine and improve infection signatures. [Courtesy Los Alamos National Laboratory]

*Presenter

modeling; and (3) integrate 3D genomic structural maps, epigenetic modifications, and ultra-resolution physical images (using cryo-electron microscopy) to validate genome structure–function relationships. This platform (see figure, p. 61) is agnostic by design and can be adapted for any pathogen (e.g., bacteria, fungi, viruses). The research group will first apply this platform to examine viral infection(s) in mammalian and plant systems to determine (for the first time) realistic 3D spatial architecture and dynamic reconfigurations of key host genomes induced by viral pathogens. Broad knowledge of epigenetic regulation of host-pathogen interactions would greatly advance the ability to predict pathogens that have high potential to cause the next global scale catastrophe or pandemic and will directly advance genomics capabilities in biopreparedness to transform the nation's ability to prepare for, and respond to, future biological threats.

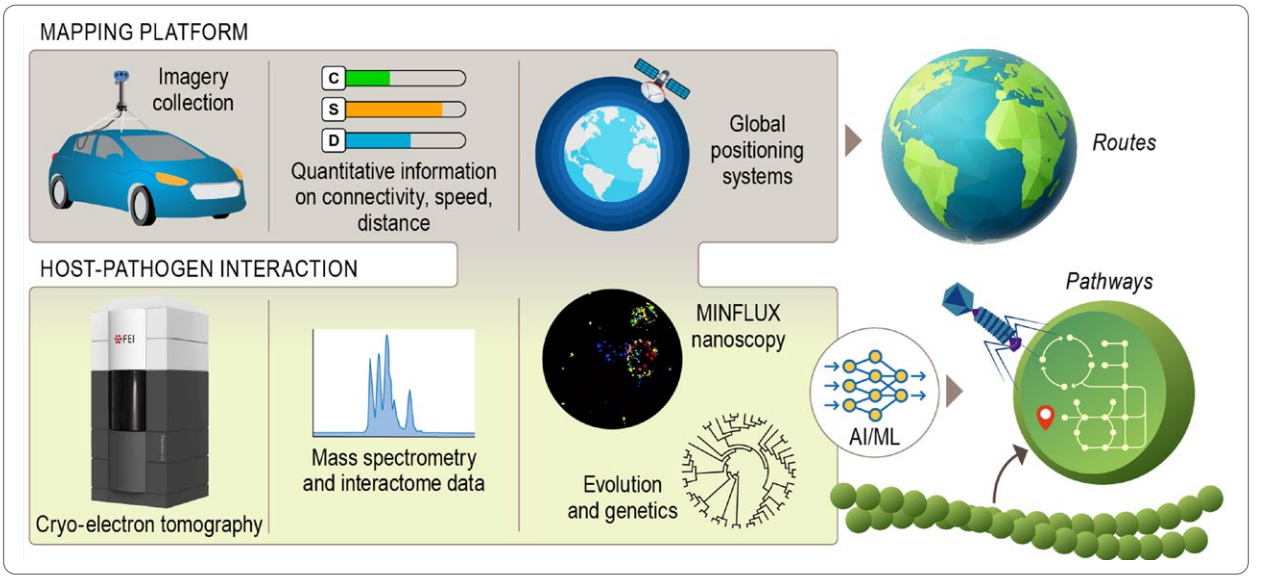
Enhancing Biopreparedness Through a Model System to Understand the Molecular Mechanisms that Lead to Pathogenesis and Disease Transmission

Margaret S. Cheung[1,2*] (margaret.cheung@pnnl.gov, PI), David Pollock[2,3], James Evans[1,2], Weijun Qian[1], Marat Valiev[1], Olga Kuchar[4], Arvind Ramanathan[5], Melissa Peacock[6], Omonike Olaleye[7], Greg Morrison[8], Zaida A. Luthey-Schulten[9], Shannon Matzinger[10]

[1]Pacific Northwest National Laboratory; [2]Environmental Molecular Sciences Laboratory; [3]University of Colorado–Anschutz; [4]Oak Ridge National Laboratory; [5]Argonne National Laboratory; [6]Northwest Indian College; [7]Texas Southern University; [8]University of Houston; [9]University of Illinois Urbana-Champaign; [10]Colorado Department of Public Health and Environment

The science of biopreparedness to counter biological threats hinges on understanding the fundamental principles and molecular mechanisms that lead to pathogenesis and disease transmission. One vision to address this challenge is to create a powerful and user-friendly platform to elucidate the fundamental principles of how molecular interactions drive pathogen-host relationships and host shifts. Researchers will enable groundbreaking discoveries by integrating a wide range of structural, genomics, proteomics, and other advanced omics measurements, along with evolutionary and artificial intelligence predictions. To make sure the system is applicable to real-world problems, it will be developed in the context of a tractable model system, the small, abundant, and accessible photosynthetic cyanobacteria and their constantly co-adapting viral pathogens, cyanophages. This model will maintain the system's applicability to real-world problems and techniques, but the overall focus will be on elucidating general principles of detecting, assessing, and surveilling molecular interaction, adaptation, and coevolution that are system agnostic and therefore extensible to any other viral-host interaction.

The objectives are to (1) identify the molecular complexes that comprise the cyanobacteria redox macromolecular subsystem and how they dynamically change with bacteriophage infection *in situ*, using cryo-electron tomography; (2) profile regulatory changes during infection using proteomics, multi-omics, and experimental validation, and integrate the data with *in situ* structures; (3) use genomics and metagenomics to determine environmental and population factors across time scales that impact the interactions between marine cyanobacteria and their cyanophage parasites, predicting the evolutionary origins of *in situ* structural and functional interactions, convergence and coevolution; and (4) develop a data



Platform for Mapping Host-Pathogen Interactions. A platform for integrating multimodal datasets and modeling approaches toward targeting the host-pathogen interactions with equitable access to resources. [Courtesy Pacific Northwest National Laboratory]

*Presenter

integration and transformation platform that facilitates the integration of *in situ*, proteomic, and evolutionary measurements of molecular interactions to surveil diverse hosts and parasites in various environmental contexts.

This powerful and user-friendly platform will enhance connections between the often-siloed fields of structure, molecular phenotype, and evolutionary genomics that are key to biopreparedness, but in need of integration. The team will do this by building a navigation tool to facilitate the effective use of globally distributed experimental data for integrated analysis and predictive modeling. The impact of the project will be to develop, implement, and test a platform to assess host-pathogen molecular interactions, adaptation to hosts and host shifts, and coevolution between hosts and pathogens. A successful project outcome will transform researchers' ability to study any host-pathogen interaction, encourage diverse community contributions, and gain fundamental insights into how proteins adapt to new contexts. This ability will be critical for designing early interventions to address future threats. Researchers will build surveillance training capability, aiming for a fair and equitable response to future pandemics and biothreats.

Phage Foundry: A High-Throughput Platform for Rapid Design and Development of Countermeasures to Combat Emerging Drug-Resistant Pathogens

Brady Cress[2]* (cress@berkeley.edu), Mohamad Alayouni[1], Archana Anand[4], Adam P. Arkin[2], Hans Carlson[1], Adam Deutschbauer[1], Darian Doakes[2],Michael Hajkowski[4], Jamie Inman[1], Alexey Kazakov[1], Britt Koskella[2], Petr Leiman[5], Catherine Mageeney[6], Ryan Melnyk[1], Mark Mimee[3], Harshini Mukundan[1], Avery Noonan[1], Denish Piya[1], Ella Rotman[3], Hemaa Selvakumar[1], Antoine Snijders[1], Jessica Trinh[6], Simon Roux[1], **Vivek K. Mutalik[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]University of California–Berkeley; [3]University of Chicago; [4]San Francisco State University; [5]University of Texas Medical Branch; [6]Sandia National Laboratory

Project Goals: There is an urgent need to develop effective antimicrobials to address the public health crisis arising from antimicrobial-resistant (AMR) bacteria. Phages represent a promising alternative to antibiotics as they tend to specifically target a few bacterial hosts and can therefore be applied as precise antimicrobials without collateral disruption of the microbiome. However, isolation of phages against a bacterial strain of interest currently relies on a tedious workflow and may not be achieved in a timely manner. In this project, this team is building the largest collection of phages ("phage banks") targeting a panel of ESKAPE pathogens AMR pathogens (*Klebsiella pneumoniae, Acinetobacter baumannii,* and *Pseudomonas aeruginosa*) and a plant pathogen (*Pseudomonas syringae*). The phage bank consisting of natural phages may not necessarily represent the optimal genotype for therapeutic use. The group will overcome this limitation by directed evolution of natural phages towards broadening the host-range and by pre-adapting phages to common bacterial resistance mechanisms. In addition, the group will also develop technologies to rapidly engineer effective phages in the absence of optimal phages in the phage bank. Specifically, researchers aim to learn phage engineering design principles by applying CRISPR-based genome-scale phage functional genomics, determining which phage genes are dispensable and which are useful for host-range engineering. The group will also use CRISPR-Cas tools to build synthetic phages and will apply "rebooting" technologies to facilitate *in vitro* phage engineering. Taken together, the new resources and capabilities developed as part of the Biopreparedness Research Virtual Environment (BRaVE) Phage Foundry should enable a broader and more efficient use of phages as therapeutics.

Funding Statement: This material by Biopreparedness Research Virtual Environment (BRaVE) Phage Foundry at Lawrence Berkeley National Laboratory is based upon work supported by the DOE Office of Science BER program under contract number DE-AC02-05CH11231.

Unlocking the Molecular Basis of Plant-Pathogen Interactions to Create Resilient Bioenergy Crops

Qun Liu[1]* (qunliu@bnl.gov, PI), Sean McSweeney[1], Shinjae Yoo[1], Ljiljana Paša-Tolić[2], Yasuo Yoshikuni[3], Christopher S. Henry[4], Huimin Zhao[5], Clint W. Magill[6], Jeffery Dangl[7]

[1]Brookhaven National Laboratory; [2]Pacific Northwest National Laboratory; [3]Lawrence Berkeley National Laboratory; [4]Argonne National Laboratory; [5]University of Illinois Urbana-Champaign; [6]Texas A&M University; [7]University of North Carolina–Chapel Hill

The development of resilient and sustainable bioenergy crops such as sorghum, poplar, and switchgrass is a focal point within BER. Bioenergy crops, like all crops, are susceptible to diseases that can vastly impact yield and quality. With the large-scale deployment of bioenergy crops, pathogen outbreaks will inevitably occur. With climate change and growth in marginal conditions without competition with food crops, bioenergy crops are facing biothreats and diseases. Plant pathogens (fungi, bacteria, and viruses) produce a stunning array of virulence effector proteins and other molecules that interact and hijack plant defense systems resulting in infection and disease. Conversely, all plants encode intracellular innate immune receptors called nucleotide-binding leucine-rich repeat proteins (NLRs) that recognize effectors to elicit successful immune responses. The co-evolution of plants and pathogens drives cycles of infection and immunity. Researchers, therefore, are integrating systems biology, biomolecular characterization, and synthetic biology with computation and artificial intelligence/machine learning to provide foundational insights into the dynamic plant–pathogen interactions. The output of this project will contribute to the development of a resilient U.S. bioeconomy, which includes the bioengineering and breeding of broad pathogen-resistant bioenergy crops and biocontrol of disease through mutualistic plant-bacteria interactions. The technologies and resources developed in this proposal may be rapidly deployable for combating emerging biotreats. Sorghum is the second most common biofuel crop in the United States and is the primary source of biodiesel production worldwide. However, a devastating anthracnose disease, caused by a fungal pathogen *Colletotrichum sublineola* can lead to yield losses of up to 67%. The co-evolution and genetic diversity of both sorghum and *C. sublineola* make this a highly relevant model system for studying plant-pathogen interactions. This team's primary objective is to advance a fundamental understanding of plant-pathosystem interactions by investigating the molecular interactions between sorghum, its anthracnose-disease causative fungal pathogen *C. sublineola*, and antifungal biocontrol bacteria to create disease-resilient bioenergy crops. The proposed project is organized into four linked aims. Aim (1) Identify molecular interactions underlying the pathogenicity of *C. sublineola* and its inhibition by bacteria. Aim (2) Characterize the molecular basis of key interactions determining *C. sublineola* pathogenicity, anthracnose resistance, and its susceptibility to biocontrol. Aim (3) Create synthetic pathogen infections to study pathogenicity, resilience, and disease biocontrol. Aim (4) Develop innovative computational resources to study plant-pathogen interactions across biological scales.

CHAPTER 8
ECR Discussion: Research Obstacles and Opportunities

Moderators: Na Ji, University of California–Berkeley;
Christopher S. Henry, Argonne National Laboratory;
Hugh O'Neill, Oak Ridge National Laboratory

In a roundtable discussion about BER enabling technologies and computational research, attendees considered scientific opportunities, key biological questions, novel methods, and promising new ways to build collaboration. These included:

- Artificial intelligence (AI): Concerns and opportunities
- Improving experimental and computational collaboration
- Gaps in experimental biosystem characterization for understanding the response to genomic and environmental changes
- Connecting measurement and prediction of biosystem function to advance higher level BER bioeconomy objectives (i.e., bioenergy, carbon sequestration, sustainable circular bioeconomy for bio-derived products)
- Computational and experimental coordination among GSP and ECR research communities including facilities, biomolecular characterization, bioimaging, and quantum measurement
- Need for data generators to incorporate current omics with new data types including imaging, spatial and temporal changes, chemical mapping, and environmental conditions
- Computational needs to address data-integration challenges

Artificial Intelligence: Concerns and Opportunities

Concerns:

- Increased reliance on computational technologies will reduce the depth of scientific thought underlying research. AI has the potential to amplify this. Researchers need to better integrate AI into science without losing focus on the underlying biology questions.
- AI training of many large language models is focused completely on text in papers, while ignoring underlying data. This is a concern because many papers have incorrect interpretations of the underlying data. AI efforts like AuroraGPT (ScienceGPT; lifearchitect.ai/auroragpt) are working to translate biological databases into training text for AI to mitigate this concern.

- Computers training computers are likely to be misdirected. AI must be grounded with validated biological data and mechanistic models. Researchers must be cautious about distinguishing theoretical computational predictions from AI.
- AI tools will be used by people who don't completely understand how they work and thus will be unaware of AI's many caveats.

Opportunities:

- AI could be useful in building/reviewing/refining metadata by drawing inspiration from text-based computational fields. The scientific community should carefully consider whether metadata is needed for everything. What about data that cannot be regimentally characterized? What can be learned from data already collected? Transfer learning can help here.
- There is potential for AI to provide more context and instruction on accurate use of tools. Researchers can leverage AI to understand tools and interpret output from tools and pipelines. Researchers should invest in AI based on domain- and physics-informed models, not just text. Some teams are already converting data, including plots, into text for training AI.

Improving Experimental and Computational Collaboration

- Collaboration between experimental and computational efforts is becoming increasingly vital. It's important to ensure that both sides understand the collaboration's objectives. Otherwise, it's easy to misapply data or methods. Researchers need more integrative approaches. Standards are essential to make this happen.
- The 2014 EERE Incorporating Bioenergy on Sustainable Landscape Designs workshop (energy.gov/eere/bioenergy/articles/incorporating-bioenergy-sustainable-landscape-designs-workshop-two-summary) was an example of a workshop that featured experimental-computational collaboration. More sophisticated technologies are emerging to better organize experimental and computation efforts. For example, building magic goggles to see what happens in a tree was discussed. However, older methods are still being used by experimentalists, and this is something that needs to be reconciled.
- Genotype-to-phenotype mapping using real-time, nondestructive imaging and sensing techniques of synthetic subsystems of plants, microbes, and their interfaces was discussed. Phenotypes can be generated through natural variants, gene editing, or environmental stresses (including abiotic and biotic).
- Studying complex systems requires coordinated computational and experimental efforts. Large projects often have computational efforts, but they are typically pushed to the side. Large projects should do both theoretical and experimental work—bridging theory and data. Theory-driven research is needed. Computational analysis is science.
- BER facilities play an important role in bridging experiments and computation. For example, data-producing facilities like the Joint Genome Institute (JGI) and the Environmental Molecular Sciences Laboratory (EMSL) collaborate closely with computational resources like DOE Systems Biology Knowledgebase (KBase) and DOE National Microbiome Data Collaborative (NMDC). Together, they endeavor to use cutting-edge methods to ensure computational methods make the best possible use of emerging data. All facilities actively coordinate with the community. JGI and EMSL help users develop and utilize new experimental approaches, while KBase and NMDC do the same computationally. In this way, facilities serve as research hubs—learning from their users and incorporating these lessons into future interactions.
- Good communication among biologists, computational scientists, and experimental tool builders is increasingly important for the overall mission of BER. Biologists identify the phenotypes and structures of biological importance for research on bioenergy and the bioeconomy. Tool builders develop imaging, sensing, and measurement methods to quantify experimental design parameters and biological, chemical, and physical results. Computational scientists extract information from experimental data to identify patterns and build models identifying environmental and biological variables that predict optimal biosystem phenotypes for useful bioenergy and bioproducts.

Gaps in Experimental Biosystem Characterization for Understanding Biosystem Response to Genomic and Environmental Changes

- Available technology limits some desired measurements. There is opportunity to employ physics-based machine learning (ML) approaches to expand the information obtainable from biological systems.
- The ability to predict phenotypes is limited, even with the availability of large datasets using genomics and transcriptomics. There is an opportunity to develop computational approaches (i.e., AI and network analysis) to help address inherent variability in datasets as well as identify missing components and connections. The ability to correlate experimental observations in the time domain and to connect information from different imaging modalities will be critical.
- There is a gap in the ability to measure chemical activity in plant cells. This is crucial for understanding cellular responses to perturbations such as environmental changes. There may be an opportunity to leverage approaches used by the medical community to develop sensors and reporters that can be utilized by plant-imaging methods.
- Current understanding of systems biology has greatly benefited from development of a broad array of omics technologies. However, investment in structural characterization of the biomacromolecules responsible for cellular processes is lagging. Connecting back to the molecular scale is a critical component of biosystem characterization.
- Ground-truth cell models are needed to address current data-fidelity issues. With them, it will be possible to correlate information from different measurement modalities and to understand causation in cellular models. These models will also be important for developing reliable AI models.

Connecting Measurement and Prediction of Biosystem Function to Advance Higher Level BER Bioeconomy Objectives in Bioenergy, Carbon Sequestration, and the Sustainable Circular Bioeconomy for Bio-Derived Products

- Gaps in protein structure prediction, such as flexibility and disorder, need to be addressed. The ability to model protein-ligand interactions *in silico* is critical for understanding cellular metabolic processes in bioconversion and for discovering new ones.
- As AI/ML approaches become more widely implemented, it will be necessary to develop standards and quantify uncertainty in predictions. This is currently lacking in the field.
- Exploiting living systems for development of a sustainable bioeconomy and to address present and future environmental challenges will only be possible if the underlying mechanisms of how proteins function are uncovered. One example is transporter proteins such as G protein-coupled receptors (GPCRs), which are poorly understood in nonhuman systems.
- As the focus of bioenergy research, carbon sequestration, and climate change research evolves, it is important to have closer interactions with stakeholders to ensure that fundamental research is relevant to the needs of the applied community.
- New tools need to be developed to combat fungal infections in bioenergy crops to increase resilience.

Computational and Experimental Coordination Among GSP and ECR Research Communities Including Facilities, Biomolecular Characterization, Bioimaging, and Quantum Measurement

- Community coordination could be at scientific, technical, management, and communication levels. For example, there could be incentives that encourage the use of specific experimental facilities and computational resources or to pilot and promote experiment-computation coordination, which could produce preliminary data for full proposals and collaborations.
- Cross-community communication tools such as newsletters, mail groups, or webinars to foster scientific and technical collaborations would be helpful. It would be very useful if GSP and ECR could have small discretionary funds for topic-driven workshops and sustained communication among investigators.

Need for Data Generators to Incorporate Current Omics with New Data Types Including Imaging, Spatial and Temporal Changes, Chemical Mapping, and Environmental Conditions

- Data incorporation and integration will help researchers understand the biological complexity of bioenergy research. Imaging and sensing tools developed by BER should be integrated with current omics for systems understanding and modeling.
- Data integration is best carried out across the DOE system to facilitate sharing among the supported facilities and centers and prevent duplication of effort. Data integration should be a continuous and ongoing process. This is especially important for BRaVE projects with different data types and purposes.
- There is pressure for computational resources, such as KBase and NMDC, to incorporate different datatypes quickly to facilitate sharing. Researchers need to improve approaches to doing this.
- A grand challenge for the future is how to manage multidimensional datasets from samples acquired across scales (i.e., from molecules, to fungi, to forest), which would require broader involvement within DOE than just BER.

Computational Needs to Address Data-Integration Challenges

- Data integration challenges include: (1) scarcity of data, particularly high-quality data; (2) differences of simulated data, laboratory data, and field data; (3) lack of standardized data generation, annotation, and data/metadata format; (4) lack of consensus on data and credit sharing; (5) integration of multimodal, multiscale, and heterogenous data; (6) data annotation and validation under the context of biological systems.
- Provision of storage resources is needed for investigators who lack the resources to maintain large, shared datasets such as high-throughput imaging datasets.
- Because funding is often limited to cutting-edge techniques (e.g., traditional AI is no longer funded; only foundational AI is currently fundable), there is widespread need for computational engineers who will take on the necessary work of implementing tools already developed by computation communities for biologists and interfacing/translating different data storage standards. These tasks may be of interest to small companies or small computational groups.
- An index of tools and resources available to the community would benefit their implementation and adoption.
- DOE-sponsored sessions at biological conferences and webinars for biologists by technology developers could be effective mechanisms for encouraging collaboration and technology adoption.

PART 2

Poster Presentations

CHAPTER 9 | POSTERS

Bioimaging

Developing a High-Throughput Functional Bioimaging Capability for Rhizosphere Interactions Utilizing Sensor Cells, Microfluidics, Automation, and AI-guided Analyses

Gyorgy Babnigg[1]* (gbabnigg@anl.gov, PI), Jessica Johnson[1], Chase Akins[1], Si Chen[1], Tomoya Honda[2], Shanshan Qi[4], Neil Getty[1], Sora Yu[2], Rosemarie Wilton[1], Fangfang Xia[1], Kenneth Kemner[1], Jean T. Greenberg[3], Yasuo Yoshikun[2]

[1]Argonne National Laboratory; [2]DOE Joint Genome Institute; [3]University of Chicago

wordpress.cels.anl.gov/bioimaging

Project Goals: The complex dynamics of root-microbe interactions in the rhizosphere drives recognizable spatial structures. However, knowledge of the specific factors that lead to their development and sustain them for plant health and productivity is sparse.

This project aims to develop a unique functional imaging technique that exploits native sense-and respond circuits of plant growth–promoting rhizobacteria (PGPR) to monitor chemical exchange between the plant root and microbe during the different phases of colonization. Several native PGPRs are turned into biosensor cells, and root colonization is evaluated with *Arabidopsis*, camelina, and poplar. Genetic variants of *Arabidopsis* with gain or loss of function provide drastically altered local environments, resulting in colonization patterns that differ from those observed previously. An orthogonal X-ray imaging approach provides high resolution elemental analysis of the local environment, and imaging throughput in general is accelerated by automation and analysis driven by artificial intelligence (AI). In addition, the team aims to advance the throughput of current bioimaging capabilities that leverage imaging chips developed with BER funding with automation, and an AI-guided image analysis strategy.

Updates from this project will be presented with specific focus on promoter library development, biosensor design and testing with *Arabidopsis*, phenotyping and genotyping of new PGPRs, AI-based image analysis, and automation.

Funding Statement: Argonne National Laboratory is managed by UChicago Argonne, LLC for DOE under contract number DE-AC02-06CH11357. This program is supported by the U. S. DOE Office of Science, through the Biomolecular Characterization and Imaging Sciences program, BER program, under FWP 39156.

Development of High-Throughput Light-Sheet Fluorescence Lifetime Microscopy for 3D Functional Imaging of Metabolic Pathways in Plants and Microorganisms

Adam Bowman[1,3*] (abowman@salk.edu), Dara Dowlatshahi[1,2], Rose Knight[1,2], Nils Bode[1], Soichi Wakatsuki[1,2], **Mark Kasevich[1,3]** (PI)

[1]Stanford University; [2]SLAC National Accelerator Laboratory; [3]Salk Institute for Biological Studies

See abstract, p. 51.

Spatial Omics Using Coherent Raman Microscopy

Marcus T. Cicerone[1]* (cicerone@gatech.edu, PI), Abigail Diering[1], Madison Green[1], Rajas Poorna[1], Caroline Filan[1], Lily S. Cheung[1], Joel E. Kostka[1], Francisco E. Robles[1,2]

[1]Georgia Institute of Technology; [2]Emory University

Spectroscopic coherent Raman microscopy can provide transcriptome-correlated metabolic information at unprecedented speed and spatial resolution in live organisms. Spectra from spontaneous Raman scattering provide semi-quantitative information about the relative abundance of major structural and metabolic cell components. It is now well established that even this semi-quantitative information can be used to

uniquely identify the activation of signaling pathways in eukaryotic and prokaryotic cells. Unfortunately, the spontaneous Raman scattering cross-sections of most biological molecules are small enough that spectral acquisition requires seconds or minutes even at cytotoxic light flux levels, so it is not conducive to imaging. Broadband coherent anti-Stokes Raman scattering (BCARS) is >1,000 times more efficient than spontaneous Raman scattering and it is fully quantitative. The group recently used BCARS microscopy on a living model organism to obtain Raman spectra at 300 nm spatial resolution that recapitulated the information content of a 6,000-gene transcriptomic study in the same organism. In the present work, researchers use BCARS to map out phenotypic heterogeneity in plant and bacterial components of rhizosphere systems with the aim of better understanding the complex signaling and metabolite exchange between nitrogen fixation symbionts. Raman spectra of *Sinorhizobium meliloti* are obtained during lag, exponential, and stationary growth phases under varying media conditions designed to induce nitrogen fixation. Bacteria single-cell transcriptomics under each condition will be correlated with the BCARS phenotypes. Using SampleMap—a variant of Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) adapted for Raman imaging—group members map these bacteria spectra to the infection thread and nodule zones where bacteria exhibit characteristics of these different growth and media conditions. These SampleMaps will ultimately be compared to high-resolution spatial transcriptomic maps of sectioned root nodules whose sister serial slices have been imaged with BCARS.

Visualizing Spatial and Temporal Responses of Plant Cells to the Environment

Peter Dalhberg[1]* (pdahlb@stanford.edu, PI), Magda Zaoralová[1], Christopher A. Azaldegui[2], Yue Rui[3], Anthony V. Sica[2], Jose Dinneny[3], Lydia-Marie Joubert[1], Wah Chiu[1]

[1]SLAC National Accelerator Laboratory; [2]University of Michigan; [3]Stanford University

Cryogenic electron tomography (cryo-ET) is a powerful approach to observe subcellular architecture and can even achieve near atomic resolution when specific complexes can be computationally identified, aligned, and averaged. Advances in this area have led to a situation where biological insight is often not limited by resolution, but instead by a lack of contextual information with which to interpret observed structures and by an inability to work with non-model systems, such as plant roots. This work aims to tackle both these issues through the development of biosensor cryogenic correlative light and electron microscopy (BioCryoCLEM) and advanced sample preparation techniques, including custom electron microscopy grids.

BioCryoCLEM correlates fluorescent biosensor data with electron tomography, providing essential physiological context alongside high-resolution structural information. This research group has calibrated biosensors for calcium, pH, and molecular crowding, demonstrating the workflow using the molecular crowding sensor. While broadly applicable, this project's focus is on investigating the plant plasma membrane-cell wall interface and its response to biotic (microbes, pathogens) and abiotic (salinity, drought, nutrients) effectors.

Unfortunately, achieving high-quality cryogenic electron microscopy is challenging as soon as one deviates from model systems. Thick plant tissues pose specific difficulties due to their size and the presence of large vacuoles which both serve to slow freezing and makes sample preparation prone to crystalline ice formation. This presentation discusses the team's work in employing the latest of sample preparation techniques, including high pressure freezing, cryogenic-lift-out, and custom grids to hold the roots and minimize sample volume as researchers work towards the goal of obtaining cryo-ET of the plasma membrane-cell wall interface.

Next-Generation Stimulated Raman Scattering (SRS) Microscopy Using Squeezed Light

Bryon Donohoe[1]* (bryon.donohoe@nrel.gov, PI), Yining Zeng[1], Christopher Anderton[2], Zhenhuan Yi[3], Girish Agarwal[3], Alexei Sokilov[3], Davinia Salvachua[1], Eric Knoshaug[1], Jonathan Jacobs[4], Marlan Scully[3]

[1]National Renewable Energy Laboratory; [2]Pacific Northwest National Laboratory; [3]Texas A&M University; [4]The Ohio State University

It is challenging to visualize the dynamic metabolic processes of living plants, algae, and fungi as they are exposed to environmental stressors. This is especially accurate for tracking biomolecules that are difficult to label with fluorescent probes, such as lipids and carbohydrates. A prototype of an advanced stimulated Raman scattering (SRS) microscope for conducting innovative studies in this field is currently being

designed and assembled. This microscope utilizes squeezed light and structured illumination to enable prolonged examination and direct chemical analysis of biological processes without compromising the system's structural integrity or dynamics. The term squeezed refers to the quantum uncertainty of the electromagnetic field strength of the light. Light in a squeezed state has an uncertainty of the field strength that is smaller than that of a coherent state. The squeezed light source will increase the signal-to-noise ratio of SRS by up to ten times. Consequently, the range of chemical imaging studies that are possible will increase, and the likelihood of photodamage will decrease to allow for the examination of extensive regions of interest and prolonged image capture.

Molecular Tracer Systems for Visualizing Plant-Pathogen Interactions Compatible with Fluorescence Imaging and Cryo-Electron and X-Ray Tomography

Dara Dowlatshahi[1,2*] (darapd@stanford.edu), Jacob Summers[2], David Nana Agyeman-Budu[1,2], Deepak Bhandari[3], Jung-Gun Kim[2], Rose Knight[1,2], Adam Bowman[2,4], James Safranek[1], Johanna Nelson Weker[1], Christopher Takacs[1], Mary Beth Mudgett[2], Federica Brandizzi[3], Mark Kasevich[2], **Soichi Wakatsuki[1,2]** (PI)

[1]SLAC National Accelerator Laboratory; [2]Stanford University; [3]Michigan State University; [4]Salk Institute for Biological Studies

Plant-pathogen interactions are complex and dynamic phenomena, relevant to fundamental, environmental and bioenergy biology. Imaging plays important roles in understanding the interactions at the atomic, molecular, subcellular, cellular, to tissue and whole plant scales. Each bioimaging method has its intrinsic limitations in spatial and/or temporal resolution, field of view and depth, and sensitivity. Fluorescence-based optical microscopes have huge advantages in monitoring dynamics of cellular and subcellular events using wide spectral ranges with super-resolution, light-sheet 3D, and wide-field imaging, but cannot go beyond 10 or 20 nanometer spatial resolution. X-ray imaging and tomography can penetrate deeper than light, and with high brilliance synchrotron sources one can reach 10 nm spatial resolution, but with a high risk of sample damage from radiation dose. Cryogenic electron tomography (cryo-ET) can offer tremendous insight into the subcellular organization of organelles and macromolecules down to several nm resolution, but information gained is largely static. Experimental validation of the spatial position and size of molecules observed in these methods, typically requires some sort of reference probe or fiducial marker in the case of tomography. These markers are typically exogenously added for correlation, and this project arm aims to improve upon this by developing molecular tracers with protein nanocages containing metal nanocrystals that also serve as intrinsic fiducials. The group presents advances towards fluorescent protein- and nanocrystal-containing cages as molecular tracers for X-ray imaging/microscopy and EO-FLIM and fiducial markers for cryo-ET, including proof-of-concept synchrotron TXM images of leaf samples bombarded with 400 nm nanogold particles. The group shares ongoing development and discuss further plans for the initial application examining fungal-plant pathogen interactions via chitin binding domains at the plant cell surface using split fluorescence complementation probes.

Funding Statement: This research is supported by the U. S. DOE, Office of Science, through the Biomolecular Characterization and Imaging Sciences program, BER program, under BCIS FWP 100878.

References

Bowman, A. J., and M. A. Kasevich. 2021. "Resonant Electro-Optic Imaging for Microscopy at Nanosecond Resolution," *ACS Nano* **15**(10), 16043–54. DOI:10.1021/acsnano.1c04470.

Bowman, A. J., et al. 2019. "Electro-Optic Imaging Enables Efficient Wide-Field Fluorescence Lifetime Microscopy," *Nature Communications* **10**(1) DOI:10.1038/s41467-019-12535-5.

Stefano, G., et al., et al. (2018). "Plant Endocytosis Requires the ER Membrane-Anchored Proteins VAP27-1 and VAP27-3," *Cell Reports* **23**(8), 2299–307. DOI:10.1016/j.celrep.2018.04.091.

Zhang, K., et al. 2022. "Cryo-EM, Protein Engineering, and Simulation Enable the Development of Peptide Therapeutics Against Acute Myeloid Leukemia," *ACS Central Science* **8**(2), 214–22. DOI:10.1021/acscentsci.1c01090.

Novel *In Vivo* Visualization of Bioenergy Metabolic and Cellular Phenotypes in Living Woody Tissues

Andrew Groover[2*] (andrew.t.groover@usda.gov) Al Ingold[1], Ruipeng Guo[1], Rajesh Mennon[1], Reed S. Sorenson[1], Gayatri Mishra[1,2,] Dasmeet Kaur[3], Chang-Jun Liu[3], **Leslie E. Sieburth[1]** (PI)

[1]University of Utah; [2]U.S. Forest Service, Northern Research Station; [3]Brookhaven National Laboratory

See abstract, p. 52.

Label-Free Structural Imaging of Plant Roots and Microbes Using Third-Harmonic Generation Microscopy

Na Ji[1,2,4*] (jina@berkeley.edu, PI), Daisong Pan[1], Jose A. Rivera[1], Peter Kim[1,2], Max Miao[3], Tomas Tyml[3], Cristina Rodriguez[1], Nathalie H. Elisabeth[2], John Vogel[3], Trent Northen[2]

[1]Department of Physics, University of California–Berkeley; [2]Lawrence Berkeley National Laboratory; [3]DOE Joint Genome Institute; [4]Department of Molecular and Cell Biology, University of California–Berkeley

Root biology is pivotal in addressing global challenges, including sustainable agriculture and climate change. However, roots have been relatively understudied among plant organs, partly due to difficulties in imaging root structures in their natural environment. Here, researchers used microfabricated ecosystems (EcoFABs) to establish growing environments with optical access and employed nonlinear multimodal microscopy of third-harmonic generation (THG) and three-photon fluorescence (3PF) to achieve label-free, *in situ* imaging of live roots and microbes at high spatiotemporal resolution. THG enabled us to observe key plant root features in mature and meristem roots including laminar structures down to the vasculature, Casparian strips, dividing meristematic cells, and root cap cells, as well as resolving subcellular features including nuclear envelopes, nucleoli, starch granules, and putative stress granules. THG from the cell walls of bacteria and fungi also provide label-free contrast for visualizing these microbes in the root rhizosphere. With simultaneously recorded 3PF fluorescence signal, the team demonstrated its ability to investigate root-microbe interactions by achieving single-bacterium tracking and subcellular imaging of fungal spores and hyphae in the rhizosphere.

Modular Platform for Correlative Electro-Optic Fluorescence Lifetime Microscopy and X-Ray Tomography

Rose Knight[1*] (vrknight@stanford.edu), Dara Dowlatshahi[1,2], Adam Bowman[1], David Agyeman-Budu[1,2], James Safranek[1,2], Johanna Weker[1,2], Christopher Takacs[1,2], Mark Kasevich[1], **Soichi Wakatsuki[1,2]** (PI)

[1]Stanford University; [2]SLAC National Accelerator Laboratory

The ability to view a single biological sample or time-resolved cross-kingdom interaction in three dimensions with each of visible light, X-rays, and electrons would yield amazing insight. Fluorescence-based optical microscopes have advantages in monitoring dynamics of cellular and subcellular events using wide yet distinguishable spectral ranges, where lifetime measurement and modulation offer a yet further dimension of rich microenvironment information but were traditionally limited for quick dynamics. X-ray imaging and tomography can penetrate deeper than light, and high brilliance synchrotron sources can enable 10 nm spatial resolution, but available modes come with high sample damage. Cryogenic electron tomography (cryo-ET) offers tremendous insight into the subcellular organization of organelles and macromolecules down to <2 nm resolution, but samples are frozen and yield static information. Correlating these different modalities in an unequivocal manner will substantially advance understanding of multiscale complex biological phenomena.

This team presents updates on the development of the next generation correlative X-ray, light and electron tomography systems. The team has modularized electro-optic fluorescence lifetime imaging microscopy (EO-FLIM) (Bowman et al. 2019; Bowman and Kasevich 2021), which produces a widefield lifetime image from single camera exposures via electro-optic gating, for use at synchrotron beamlines, allowing simultaneous X-ray and visible imaging. In addition, the team will be able to utilize newly developed bunch modes at SLAC National Accelerator Laboratory to lock the synchrotron X-ray repetition rate with the visible excitation and optical gating at 76 megahertz and 158 MHz, allowing simultaneous studies of time-resolved X-ray excited optical luminescence. The team has also developed multiexponential analysis methods and established a standing EO-FLIM microscope at the synchrotron for further uncorrelated data collection and to explore extremes of resonant drive frequencies. In concert with time-resolved fluorescence studies, the team will utilize both the micro- and

nano-computed tomography imaging capability of beamline 6-2 for Transmission X-ray Microscopy, Stanford Synchrotron Radiation Lightsource, SLAC to gain hierarchical insights into the dynamics of fungal–pathogen interactions.

References

Bowman, A. J., and M. A. Kasevich. 2021. "Resonant Electro-Optic Imaging for Microscopy at Nanosecond Resolution," *ACS Nano* **15**(10), 16043–54. DOI:10.1021/acsnano.1c04470.

Bowman, A. J., et al. 2019. "Electro-Optic Imaging Enables Efficient Wide-Field Fluorescence Lifetime Microscopy," *Nature Communications* **10**(1) DOI:10.1038/s41467-019-12535-5.

Metabolic Imaging at Video Rates Using Raman with Airy Light-Sheet Illumination and Sparse Photon Detection

Haokun Luo[1]* (haokunlu@usc.edu), Ramachandran Kasu[2], Demetrios N. Christodoulides[1], **Andreas E. Vasdekis[2]** (PI)

[1]University of Southern California; [2]University of Idaho

Nowadays, Raman imaging represents only a modest fraction of all research and clinical microscopy to date even though it exhibits great potential. This limited adoption is primarily attributed to the ultralow Raman scattering cross-sections of most biomolecules, resulting in low-light or photon-sparse conditions. Imaging biological samples under such conditions is suboptimal, leading to either extremely low frame rates or the need for higher levels of irradiance. In this study, researchers address this tradeoff by introducing Raman imaging capable of operating at both video rates and with irradiance levels 1,000 times lower than existing methods. To achieve this, the team utilized a carefully designed Airy light-sheet microscope, which efficiently images large specimen areas (Dunn et al. 2023). The Airy beam, known for its unique diffraction-free properties such as self-healing and refocusing, has been employed in light-sheet microscopy (LSI) (Subedi et al. 2020; Subedi et al. 2021) and selective plane illumination imaging schemes. The team investigated the diffraction-free behavior of Airy beams as a function of cubic phase modulation 'α', both theoretically and experimentally. Additionally, group members implemented subphoton per pixel image acquisition and reconstruction techniques to address challenges arising from sparse photon availability during short integration times. This project demonstrates the versatility of this approach through successful imaging of various samples, including the three-dimensional metabolic activity of individual microbial cells and their associated cell-to-cell variability. Moreover, to visualize small-scale targets, the group leveraged photon sparsity and photon superlocalization to increase magnification without sacrificing field-of-view, thereby overcoming another significant limitation in modern light-sheet microscopy.

References

Dunn, L. et al. 2023. "Video-Rate Raman-Based Metabolic Imaging by Airy Light-Sheet Illumination and Photon-Sparse Detection," *Proceedings of the National Academies of Sciences* **120**(9), e2210037120. DOI:10.1073/pnas.2210037120.

Subedi, N. R., et al. 2020. "Integrative Quantitative-Phase and Airy Light-Sheet Imaging," *Scientific Reports* **10**(1), 20150. DOI:10.1038/s41598-020-76730-x.

Subedi, N. R., et al. 2021. "Airy Light-sheet Raman Imaging," *Optics Express* **29**(20), 31941–51. DOI:10.1364/OE.435293.

Multimodal Optical Nanoscopy for In-Liquid Bioimaging with Few Nanometer Spatial Resolution

Brian T. O'Callahan* (brian.ocallahan@pnnl.gov), Chih-Feng Wang, Patrick Z. El-Khoury, **Scott Lea** (PI)

Pacific Northwest National Laboratory

The ability to perform chemical nanoimaging of biosystems in liquid remains challenging despite recent experimental advances. Fluorescence-based super-resolution techniques such as stimulated emission depletion microscopy and stochastic optical reconstruction microscopy allow tracking of tagged analytes with nanoscale spatial resolution. However, the development of generic chemical nanoimaging techniques is needed to study systems or analytes for which fluorescent tags are unavailable or infeasible. Towards this goal, this team designed and built a multimodal hyperspectral micro/nanoscope (MHNano), which combines optical spectroscopy with scanning probe microscopy to enable (non)linear optical measurements with scales from hundreds of micrometers to a few nanometers in a single platform. Compatible with nonlinear imaging, perform micro- to nanoscale resolution the linear and nonlinear optics, [e.g., tip-enhanced (two-photon) photoluminescence (TEPL/TE2PPL)]. In this work, researchers demonstrate the spatial resolution of TEPL and TE2PPL with sub 5 nm from cadmium selenide and zinc sulfide semiconductor quantum dots (QDs) using sputtered plasmonic gold probe under ambient conditions. A custom liquid

cell allows high numerical aperture split excitation and collection for *in situ* Raman and nonlinear measurements. The capability of TEPL/TE2PPL paves the way for (non)linear photoluminescence-based or Raman spectral nanoimaging of biosystems in their native environment.

Nanometrology of Lignin Deposition on Cellulose Nanofibrils: Paving the Way for Advanced Bioenergy and Quantum Bioimaging Studies

Ali Passian[1]* (passianan@ornl.gov), Sam Bhagia[1], Rubye Farahi[1], Patrick Snyder[2], **Xu Yi[3]** (PI)

[1]Oak Ridge National Laboratory; [2]University of Illinois Urbana-Champaign; [3]University of Virginia

Cellulose and lignin, the primary constituents of plant biomass, are essential to the development of sustainable bioenergy solutions and the advancement of the bioeconomy. Their abundant availability and renewable nature make them ideal candidates for biofuel production, biocomposite materials, and as models in cutting-edge characterization. This team investigates the synthesis and deposition of guaiacyl lignin on cellulose nanofibrils, emulating the process of secondary cell wall formation in plants. Such a focus is crucial for enhancing understanding of plant biomass's resilience and efficiency in bioenergy conversion processes. Utilizing coniferyl alcohol and employing biocatalysis with horseradish peroxidase and hydrogen peroxide, the team mimics the natural polymerization of lignin, offering a controlled environment to study its interaction with cellulose at the highest achievable nanoscale classical resolution. By combining Micro-Infrared Spectroscopy, Confocal Raman Spectroscopy, Atomic Force Microscopy, and Nano-IR Spectroscopy, the team aims to provide a broad understanding of the bulk and nanoscale properties of these biopolymer composites. Such polymer-scale bioimaging and chemical characterization are indispensable for revealing the molecular-level interactions and structural arrangements, enabling bio-based material science. Moreover, this work serves as a foundational study for exploring the properties of tension wood, which exhibits unique characteristics in its mechanical, cell-level structure, and compositional behavior. Using these cellulose and lignin samples as control systems, the team can achieve tension and compression in a controlled manner, establishing baseline data crucial for understanding the material response in grown tension wood measurements. This approach not only aids in studying the complex polymers within tension wood but also sets the stage for comparing these natural systems with these bioengineered samples, enhancing understanding of plant biomass mechanics. The group's motivation extends beyond traditional studies, aiming to bridge the gap between classical and quantum bioimaging techniques. By establishing a solid understanding of the classical Raman spectroscopy limits and characteristics of cellulose-lignin interactions, this study paves the way for research in quantum bioimaging. This biosystem will allow researchers to explore the quantum features of Raman measurements, providing a biomolecular framework for addressing the limitations faced by current imaging methodologies. Furthermore, in preparation for quantum microscopy studies of enzyme-plant cell interactions, the group will introduce enzymes to these cellulose and lignin films. Such preliminary characterizations are vital for understanding how enzymatic actions modify the plant cell walls at the nanoscale, ultimately informing the quantum bioimaging of real enzyme-plant sample interactions. This sequential approach, from classical imaging to quantum measurements, offers a comprehensive strategy for dissecting the complex dynamics of plant biomass at the forefront of bioenergy research and quantum science. This investigation contributes to the fundamental understanding of plant biomass structure and introduces a methodological approach towards utilizing quantum bioimaging for bioenergy applications. By elucidating the interactions between lignin and cellulose, this study unlocks potential avenues for optimizing biomass conversion into biofuels and developing sustainable materials, aligning with the goals of a circular economy and pushing the boundaries of material science into the quantum realm.

Integrative Imaging of Plant Roots During Symbiosis with Mycorrhizal Fungi

Andreas E. Vasdekis[1]* (andreasv@uidaho.edu, PI), Scott E. Baker[2], Demetrios Christodoulidis[3,4], Maria Harrison[5], Luke Sheneman[1], Ram Kasu[1], Jinming Zhang[1], Haokun Luo[3,4]

[1]University of Idaho; [2]Pacific Northwest National Laboratory; [3]University of Central Florida; [4]University of Southern California; [5]Boyce Thompson Institute

See abstract, p. 45.

Quantitative Phase Imaging of Live Roots by Gradient Retardance Optical Microscopy

Jinming Zhang[1*] (jinmingz@uidaho.edu), Maria J. Harrison[2], **Andreas E. Vasdekis[1]** (PI)

[1]University of Idaho; [2]Boyce Thompson Institute

Quantitative phase imaging (QPI) has recently emerged as a widespread optical imaging method for measuring the dry mass and density of individual cells, two key metabolic parameters in biological systems. Despite its potential, applying QPI techniques to specimens that are thicker than 500 wavelengths faces significant challenges. In such cases, optical scattering from thick specimens compromises image quality by increasing background noise and reducing contrast. To overcome these challenges, various strategies have been explored, including laser-based tomographic methods and asymmetric illumination/detection interferometry that uses incoherent light to avoid speckle-driven image degradation. However, these methods require expensive optical elements, such as spatial light modulators or polarization-sensitive cameras, that additionally are known to reduce imaging efficiency due to energy losses. To address these shortcomings, this research team developed Gradient Retardance Optical Microscopy (GROM), a QPI technique that is compatible with 3D imaging and requires no computational image reconstruction. GROM operates by transforming asymmetrically illuminated intensity images into phase gradient images and enables fully automated 3D acquisition of interferometric images using custom-made routines in open access platforms. Further, GROM can transform any standard microscope into a QPI platform by placing only a liquid crystal retarder between the illumination condenser and the sample. Through this method, the group has successfully reconstructed a variety of imaging targets, including conducting 3D volume viewing of individual bacteria and fungi, as well as a 500 micrometer–diameter plant roots tissue of the model system *Medicago truncatula*, showcasing the depth and versatility of GROM's capabilities.

Funding Statement: The research team gratefully acknowledges funding from the U.S. DOE, project number DE-SC0022282.

Reference

Zhang, J., et al. In review. "Quantitative Phase Imaging by Gradient Retardance Optical Microscopy (GROM)."

Nondestructive, Three-Dimensional Imaging of Processes in the Rhizosphere Utilizing High-Energy Photons

Shiva Abbaszadeh[1*] (sabbasza@ucsc.edu, PI), Weixin Cheng[1], Craig S. Levin[2], Adam S. Wang[2]

[1]University of California–Santa Cruz; [2]Stanford University

See abstract, p. 11.

CHAPTER 10 | POSTERS

Biopreparedness

Phage Foundry: A High-Throughput Platform for Rapid Design and Development of Countermeasures to Combat Emerging Drug-Resistant Pathogens

Brady Cress[2*] (cress@berkeley.edu), Mohamad Alayouni[1], Archana Anand[4], Adam P. Arkin[2], Hans Carlson[1], Adam Deutschbauer[1], Darian Doakes[2], Michael Hajkowski[4], Jamie Inman[1], Alexey Kazakov[1], Britt Koskella[2], Petr Leiman[5], Catherine Mageeney[6], Ryan Melnyk[1], Mark Mimee[3], Harshini Mukundan[1], Avery Noonan[1], Denish Piya[1], Ella Rotman[3], Hemaa Selvakumar[1], Antoine Snijders[1], Jessica Trinh[6], Simon Roux[1], **Vivek K. Mutalik[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]University of California–Berkeley; [3]University of Chicago; [4]San Francisco State University; [5]University of Texas Medical Branch; [6]Sandia National Laboratory

See abstract, p. 65.

Genome-Scale Metabolic Modeling to Study Interactions and Coevolution Between Cyanobacteria and Cyanophages

Song Feng[1*] (song.feng@pnnl.gov), Ruonan Wu[1], Pavlo Bohutskyi[1], Tong Zhang[1], John Melchior[1], Young-Mo Kim[1], Xiaolu Li[1], August George[1], Youngki You[1], Aaron Chan[2], Zan Luthey-Schulten[2], David D. Pollock[1,3], Wei-Jun Qian[1], **Margaret S. Cheung[1]** (PI)

[1]Pacific Northwest National Laboratory; [2]University of Colorado School of Medicine; [3]University of Illinois Urbana-Champaign

Marine cyanobacteria are well-known for their role in fixing nearly 30% of organic matter on Earth. Up to 60% of the cyanobacterial cells, however, are infected by phages. During phage infection, cyanobacterial metabolisms are reprogrammed towards phage replication, and the carbon dioxide (CO_2) fixation functional module is inhibited via expressing phage auxiliary metabolic genes (AMGs). Such phenotypic change in primary producers is especially concerning given the impact of climate change. Moreover, different phages induce different host responses and life cycle changes that are likely driven by metabolic and molecular interaction network reprogramming. Here, group members describe a genome-scale metabolic model of cyanobacteria combined with two additional biomass objective functions driving replication of two cyanophage strains, P-HM2 and P-SSP7. Using the dynamic flux balance analysis, the group will compare host-optimal solutions and phage-optimal solutions in the diurnal cycles with changing light intensities. Then the group will use this functional information of AMGs to compare the metabolic fluxes infected by different phage strains by enforcing AMG-associated reactions. These analyses will reveal interactions between host and phages as well as metabolic reprogramming by cyanophages. Furthermore, this model will provide a basis for integrating multiomics data with a whole-cell systems model of cyanophage infection to better understand host-virus interactions. This research team will use the metabolic network as functional coordinates for enzymes. The abundance and state changes (e.g., post-translational modifications, conformational and interactional changes) can be mapped to the metabolic network model and whole-cell model to study their subsequent effects on phenotypes of cyanobacteria and/or cyanophages.

BREAD: Bioenergy-Crops Resilience and Evolution Dynamics

Qun Liu[1*] (qunliu@bnl.gov, PI), Sean McSweeney[1], Shinjae Yoo[1], Ljiljana Paša-Tolić[2], Yasuo Yoshikuni[3], Christopher S. Henry[4], Huimin Zhao[5], Clint W. Magill[6], Jeffery Dangl[7]

[1]Brookhaven National Laboratory; [2]Pacific Northwest National Laboratory; [3]Lawrence Berkeley National Laboratory; [4]Argonne National Laboratory; [5]University of Illinois Urbana-Champaign; [6]Texas A&M University; [7]University of North Carolina–Chapel Hill

Sorghum is the second most cultivated U.S. biofuel crop and is the primary source of biodiesel production worldwide. Due to its drought resistance and fast growth of biomass as a source for ethanol production, sorghum is one of DOE's flagship bioenergy crops. However, sorghum diseases such as anthracnose, stalk rot, downy mildew, grain mold, and leaf blight reduce the yield of sorghum biomass production.

Anthracnose alone can lead to yield losses of up to 67% in susceptible sorghum cultivars. Therefore, improving anthracnose resistance and biocontrol in sorghum directly improves its biomass production and bioeconomy.

Sorghum anthracnose disease is caused by the hemibiotrophic fungal pathogen *Colletotrichum sublineola*. *C. sublineola* produces a stunning array of virulence effector proteins and other molecules that interact and hijack plant defense systems resulting in infection and disease in sorghum. Conversely, some sorghum cultivars encode intracellular innate immune receptors called nucleotide-binding leucine-rich repeat proteins (NLRs) that recognize effectors to elicit successful immune responses. The co-evolution of sorghum and *C. sublineola* drives cycles of infection and immunity. This project hypothesizes that the dynamic coevolution of effectors and NLR receptors forms the basis of sorghum immunity in response to *C. sublineola* and other pathogen infections. A key knowledge gap is that none of *C. sublineola* effectors are biochemically or structurally characterized. Also, there is a lack of mechanistic data regarding the interactions between specific NLR proteins and specific effectors. This lack of molecular understanding of sorghum *C. sublineola* interactions is the biggest impediment to the rational design of resistance to anthracnose in sorghum crops and is thus the focus of this project. Leveraging the available reference genome sequences for both sorghum and *C. sublineola*, group members are employing comparative genomics, transcriptomics, proteomics, and metabolomics to identify *C. sublineola* effectors and their respective NLRs. The group is using structural biology and bioimaging to characterize the temporal and spatial NLR-effector interactions across scales from atoms to cells and plants. Subsequently, engineering and biocontrol strategies targeting the characterized NLRs or *C. sublineola* are used to develop resilient and sustainable bioenergy crops. This team is developing computational resources to analyze molecular interactions and genetic co-evolution of the plant–pathosystem, and foundational large natural language models to integrate and train multimodal text and imaging data for predictive understanding of pathogenicity and biocontrol.

While this project focuses on the sorghum *C. sublineola* system, the results will lay down a groundwork for studying plant-pathogen interactions more broadly. The research strategies and techniques developed under this project will advance scientists' ability to rapidly respond to emerging biothreats impacting bioenergy crops and plants in unmanaged ecosystems.

Phage Foundry: Establishing Capabilities for High-Throughput Phage-Host Interaction Characterization and Prediction

Denish Piya[1]* (denish.piya@lbl.gov), Mohamad Alayouni[1], Archana Anand[4], Adam P. Arkin[2], Hans Carlson[1], Brady Cress[2], Adam Deutschbauer[1], Darian Doakes[2], Michael Hajkowski[4], Jamie Inman[1], Alexey Kazakov[1], Britt Koskella[2], Petr Leiman[5], Catherine Mageeney[6], Ryan Melnyk[1], Mark Mimee[3], Harshini Mukundan[1], Avery Noonan[1], Ella Rotman[3], Hemaa Selvakumar[1], Antoine Snijders[1], Jessica Trinh[6], Simon Roux[1], **Vivek K. Mutalik[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]University of California–Berkeley; [3]University of Chicago; [4]San Francisco State University; [5]University of Texas Medical Branch; [6]Sandia National Laboratory

Increase in the incidence of antimicrobial-resistant (AMR) bacterial pathogens currently poses an immense threat to normal world order. In addition to the tragic impact on human health, AMR is estimated to have a worldwide economic cost running into trillions of U.S. dollars by severely debilitating agriculture, dairy, aquaculture, livestock, and poultry industries. Bacteriophages (phages) have been proposed as an alternative to antibiotics due to a dearth of new antimicrobial molecules in the discovery pipeline. There have already been some successes in treating AMR pathogenesis by using phages under compassionate use protocols; however, biological tools for broad-scale mechanistic characterization of phage-host interactions in clinically and agriculturally relevant bacteria are still limited, hampering development and application of phages as robust antimicrobial agents. In this project, researchers are developing tools for high-throughput characterization of phage-host interactions on highest priority ESKAPE human pathogens: *Klebsiella pneumoniae*, *Acinetobacter baumannii* and *Pseudomonas aeruginosa*, as well as important crop pathogen *Pseudomonas syringae*. By using panels of clinically relevant and genomically diverse strains, the team is building a highly diverse collection of phages ("phage banks") and conducting large-scale phage-susceptibility assays for genome-wide association studies (GWAS)-like analyses to identify bacterial and phage genes that drive host range and specificity. To guide phage isolation efforts and build a baseline understanding of the population diversity, mobile genetic elements, and phages associated with pathogens, the team is using a multiomics approach on a large wastewater system. To generate systematic genotype-phenotype mapping, powerful high-throughput genetic technologies, such as genome-wide loss-of-function randomly

barcoded transposon sequencing, CRISPR interference and overexpression dual barcoded shotgun expression library sequencing, are being applied to a set of select strains. This systematic effort in phenotyping and high-throughput genetics will provide fitness landscapes in presence of phages, antibiotics, metals, and other stressors, as well as enable researchers to map the cross-resistance and collateral sensitivities between phages and antibiotics. The characterization workflows, resources and datasets generated at the Biopreparedness Research Virtual Environment Phage Foundry will provide crucial foundational knowledge necessary to develop machine-learning models and facilitate quick and effective prediction of therapeutic formulations for countering any emerging recalcitrant infections.

Funding Statement: This material by Biopreparedness Research Virtual Environment (BRaVE) Phage Foundry at Lawrence Berkeley National Laboratory is based upon work supported by the DOE, Office of Science, BER program under contract number DE-AC02-05CH11231.

Generation of High-Resolution Chromatin Configuration and Epigenomics Datasets to Decipher Host-Pathogen Interactions

Christina R. Steadman[1]* (crsteadman@lanl.gov), V. Venu[1], S. Adikari[1], E. Small[1], Cullen Roth[1], J. Brown[2], J. Kubicek-Sutherland[1], P. Freimuth[3], **Shawn R. Starkenburg[1]** (PI)

[1]Los Alamos National Laboratory; [2]University of Arizona; [3]Brookhaven National Laboratory

The ability to counter biological threats is limited given the lack of knowledge of host resilience mechanisms in the face of pervasive pathogens. Research suggests that epigenetic mechanisms and associated chromatin structure regulate the functionality of the genome and play profound roles in host-pathogen interactions. As such, this team hypothesizes that these processes vary between resilient versus susceptible hosts, potentially providing specific signatures (patterns) of infectivity. Further, these signatures may be attributed to classes of pathogens allowing for early detection and mitigation. Yet, the paucity of epigenomics and chromatin structural datasets from systematically tested pathogen exposures precludes identifying these proposed signatures for surveillance and diagnostics in novel species. As such, this project's goal is to develop an experimental workflow to generate large omics datasets to characterize and survey early onset molecular signatures of infection, with particular focus on viruses. The workflow is designed to utilize representative molecules for various classes of viruses in the same primary cell culture system (mammalian and plant) to generate single cell sequencing assessments for deep learning and exascale computing analysis. The team's initial assessments demonstrate the relationship between specific local (epigenetic) and global (genomic) structures and their variability in response to infection, providing novel signatures.

CHAPTER 11 | POSTERS

Computational Biology

Cyanobacteria-Cyanophage Environmental Sampling in the Salish Sea

Noelani Boise[1]* (noelani.boise@pnnl.gov), Scott Edmundson[1], Owen Leiser[1], David Pollock[1, 2], **Margaret Cheung[1]** (PI)

[1]Pacific Northwest National Laboratory; [2]University of Colorado–Anschutz Medical Campus

Marine picocyanobacteria are a ubiquitous component of phytoplankton and play an outsized role in global primary production. As part of a larger project to use cyanobacteria and cyanophage as a model system to understand molecular mechanisms of pathogenesis and disease transmission, this team is performing sampling, metadata collection, and sequencing to understand their complex natural host-parasite relationships in relation to environmental factors. To achieve this, the group sampling coastal Salish Sea waters along the northern Olympic Peninsula at various times of day, times of year, depths, and tidal flow conditions. Here, the poster describes preliminary sampling, sequencing results, and metadata analysis results. Sampling involved collecting surface-level serial filtration samples from the Pacific Northwest National Laboratory-Sequim floating dock, in-line filtration of a raw seawater line, and cultured specimens. Filtering was performed to collect whole plankton, picocyanobacteria, and viral fractions, the last two to enrich sequencing efforts to focus on the target organisms. Preprocessed sequence reads were aligned to known cyanophage as well as the cyanobacteria *Prochlorococcus marinus* and *Synechococcus elongatus* genomes, demonstrating the presence of organismal targets but at low frequency, as expected during the cold, low light winter months. These initial survey results provide a promising foundation to establish standard sampling and metadata collection protocols to sample over the next three years to better understand cyanobacterial-cyanophage coevolution. Further steps will integrate genomic, multimodal structural and proteomic data with evolutionary models to better understand cyanobacteria-cyanophage ecotypes and coevolutionary interactions.

Development of Computational Tools for Integrated, Exascale Analysis of Chromatin Configurations and Epigenomics Datasets for Profiling Host-Pathogen Interactions

Cullen Roth[1]* (croth@lanl.gov), Daniel Jacobsen[2], Karissa Sanbonmatsu[1], Ankush Singhal[1], Anna Lappala[3,4], David Rogers[1], Shounak Banerjee[1], **Christina R. Steadman[1]** (PI), **Shawn R. Starkenburg[1]** (PI)

[1]Los Alamos National Laboratory; [2]Oak Ridge National Laboratory; [3]Massachusetts General Hospital; [4]Harvard University

Epigenetic mechanisms and associated chromatin structure regulate the functionality of the genome and play profound roles in host-pathogen interactions. However, the field currently lacks systems-level algorithms, methods, and visualization tools to quickly compare and identify pathogen-induced changes to host epigenomes and chromatin structure to understand susceptibility and resiliency. In this project, group members are building a singular platform that includes tools, pipelines, and software for analysis and visualization of chromatin and epigenomic datasets. Here, the team presents two novel components of this platform: (1) SLUR(M)-py; a Python-based pipeline that leverages the Simple Linux Utility for Resource Management tool (SLURM) to processes paired-end sequencing data from several different sequencing strategies, including whole genome, assay for transposase-accessible chromatin with sequencing, chromatin immunoprecipitation assays with sequencing, and Hi-C sequencing, and generates outputs for further analysis and publication; and (2) 4D Genome Browse (4DGB), a visualization and analysis tool that transforms chromatin conformation data into 3D genome models. The 4DGB workflow takes user Hi-C files, produces 3D chromosome reconstructions, and paints the chromosome structures with epigenetic information. The first implementation of the workflow encapsulates data preparation, a molecular dynamics simulation, and a web server for the 4DGB in a single executable. Furthermore, to analyze complex omics datasets and determine the nonlinear relationships

among them, the team is conducting additional research to utilize explainable artificial intelligence from the SLUR(M)-py output to enable interactive, comparative exploration and predictive dynamic modeling across time. Collectively, the team envisions that these novel analysis tools will propel deeper understanding of host-pathogen interactions in mammalian and plant systems.

Outreach and User Development for the KBase Science Community

Benjamin Allen[3*] (allenbh@ornl.gov), Ellen G. Dow[1] (egdow@lbl.gov), Zachary Crockett[3], Elisha Wood-Charlson[1]; Jason Baumohl[1], Kathleen Beilsmith[2], Joseph Bezouska[2], David Dakota Blair[4], John-Marc Chandonia[1], Dylan Chivian[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Andrew Freiburger[2], Tianhao Gu[2], Prachi Gupta[1], A. J. Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Dileep Kishore[3], Dan Klos[2], Filipe Lui[2], David Lyon[1], Cody O'Donnell[1], Mikaela McDevitt[1], Chris Neely[1], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Pamela Weisenhorn[2], Ziming Yang[4], Shinjae Yoo[4], Sijie Xiang[1], Qizhi Zhang[2]; Shane Canon[1], Paramvir S. Dehal[1]; Robert Cottingham[3], Christopher S. Henry[2], **Adam P. Arkin[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Argonne National Laboratory; [3]Oak Ridge National Laboratory; [4]Brookhaven National Laboratory

kbase.us

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into a user-friendly platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform, enabling scientists to analyze their own data alongside public and collaborator data, then share their findings across the system and ultimately publish reproducible analyses.

The KBase user development and outreach team works to accelerate user research, develop skills and knowledge of computational biology across experience levels, and promote advances in research and software development for the betterment of the scientific community. User development activities promote applications of many analytical pathways available, including protocol and tutorial development for the KBase platform, as well as conducting workshops and webinars. Outreach and user engagement helps identify, grow, and support the scientific communities using KBase for research. The KBase Educators program supports professors and teachers integrating hands-on computational biology exercises and analyses through KBase. The variety of programming targets different cross-sections of the BER research community, with the goal to empower the next generation of researchers using KBase.

Outreach and User Development: The KBase team hosts training events to support institutions, research groups, collaborators, and educators advance their research. Training events include workshops and webinars to demonstrate use of the platform and showcase features of interest. Workshops are designed to support specific institutional research needs and facilitate collaboration between research groups utilizing KBase. A recent example is the Long-Reads Isolate Sequencing and Assembly workshop hosted by the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) Science Focus Area and KBase to demonstrate new tools for microbial genomics. Webinars are broadcast to a global audience and showcase powerful workflows, new tools and features, and speakers from community developers to subject matter experts. Recorded webinars are shared via the KBase YouTube channel (youtube.com/DOEKBase) as a persistent resource for training materials.

Help Desk: KBase is committed to helping users utilize KBase for their research. The KBase Help Board allows users to report bugs or ask questions to be addressed by KBase staff in a forum that is public so that questions are findable by future users. Users may also submit new feature requests to inform KBase development to better serve their needs.

KBase Educators: The KBase Educators Community (kbase.us/engage/educators/) includes biology and data science instructors teaching students at community colleges, primarily undergraduate, and doctoral research institutions that represent diverse populations, including minority serving and emerging research institutions. Instructors find KBase enables their students to conduct hands-on data science research and analysis without programming skills or additional computational resources. The KBase Educators Org provides free, open-access to instructional workflows—teaching resources, data, analysis tools, and mark-down utility to customize instructions and learning goals, including Spanish versions. Working groups seeking

new opportunities for their students have designed and are piloting a modular curriculum based on the open, reproducible scientific process: (1) research question and hypothesis development, (2) experimental design and sample metadata, (3) sample collection and processing, which build upon the existing resources for (4) data analysis, and (5) conclusions and publishing.

Through each of these approaches, KBase empowers researchers and inspires the next generation of biologists and data scientists by providing a platform to investigate systems biology problems with sophisticated tools and share knowledge across the scientific community.

Funding Statement: This work is supported as part of the Genomic Science program of BER. The DOE Systems Biology Knowledgebase (KBase) is funded by the DOE Office of Science, BER program under award numbers DE-AC02- 05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886. The Educator Working Groups described is funded by a National Science Foundation RCN-UBE Incubator under award number 2316244.

Deep-Plant: A Deep Learning Platform for Plant Genomics

Asa Ben-Hur* (asa@colostate.edu, PI)

Colorado State University

Gene regulation is governed by a multitude of proteins and RNAs, and especially transcription factors. Transcription factors control gene expression by binding proximal to genes in their promoter regions or at distal enhancers. The binding of transcription factors is modulated by the state of the DNA molecule, namely whether it is accessible or wrapped around histones, and by DNA and histones modifications. In recent years, several databases that provide vast amounts of plant genomics data from various types of assays have been curated from thousands of published studies. These include genome-wide expression, transcription factor binding, histone modifications, and DNA accessibility. Deep learning has demonstrated its value in modeling large and complex genomics compendia in mammals, providing insights into gene regulation in those systems. However, very little such work has been carried out in plants. This initiative proposes to leverage the wealth of data available in plants to create a deep learning framework called DEEP-PLANT to model plant chromatin state and its consequences for gene regulation. More specifically, the DEEP-PLANT model will predict transcription factor binding and chromatin state directly from sequence. These models will provide a detailed picture of gene regulation and will support downstream applications including the prediction of gene expression, enhancers, and the effects of genetic variation. This work will be carried out in *Arabidopsis* and rice and shed light on conserved aspects of gene regulation across dicots and monocots and provide plant biologists with tools to form hypotheses on the factors that drive gene expression. The poster will provide a summary of preliminary work towards these goals.

Understanding Bacterial and Fungal Protein Function Using Structure and Sequence Similarity Networks in KBase

Christopher S. Henry[1]* (PI), Nidhi Gupta[1] (ngupta@anl.gov), Claudia Lerma-Ortiz[1], Janaka N. Edirisinghe[1], Nils Oberg[2], John Gerlt[2], KBase Team[1,3,4,5], Robert Cottingham[4], Adam P. Arkin[3]

[1]Argonne National Laboratory; [2]University of Illinois Urbana-Champaign; [3]Lawrence Berkeley National Laboratory; [4]Oak Ridge National Laboratory; [5]Brookhaven National Laboratory

See abstract titled "Predicting Protein Function Using Structure and Sequence Similarity in KBase," p. 37.

Principles of Fungal Metabolism, Growth and Bacterial Interactions

William Cannon[1,4]* (william.cannon@pnnl.gov, PI), Austin Hansen[1], Yi-Syuan Guo[2,3], Tomas Rush[2], Dale Pelletier[2], Scott Retterer[2], Mark Alber[1], Bruce Palmer[3], Connah Johnson[3], Andrew Myers[4], Ann Almgren[4]

[1]University of California–Riverside; [2]Oak Ridge National Laboratory; [3]Pacific Northwest National Laboratory; [4]Lawrence Berkeley National Laboratory

Project Goals: The goals of this project are to elucidate fundamental principles of species interactions using hybrid machine learning/simulation models of fungal-bacterial interactions and dynamics. These hybrid data analytic/simulation models are used to carry out virtual experiments and develop fundamental understanding of the interactions between fungi and bacteria, specifically the mycorrhizal fungus *Laccaria bicolor* and *Pseudomonas sp.* helper. At the same time, researchers carry out experiments aimed at developing and testing quantitative assays to measure the same

interactions, and whose data inform views of biology. Researchers are:

- Evaluating the fundamental physical operating principles of cells and using these principles to develop physics-informed machine learning models of metabolism.
- Evaluating how fungal lifestyles impact growth in sparse nutrient environments.
- Computationally and experimentally investigating pseudomonas chemotaxis in response to *L. bicolor* produced metabolites.

The exchange of metabolites between microbes is an emergent property that evolves because the exchanged metabolites allow for increased growth of both species by reducing the thermodynamic cost of growth. Instead of each species producing every metabolite needed, metabolite exchange allows each microbe to specialize and efficiently produce a metabolite, such as trehalose, in exchange for one that it cannot produce as cheaply, such as thiamine. In order to evaluate the benefits of such microbial trade, physics-based models are needed that are capable of modeling the thermodynamic costs and benefits. The long-term goal of this project is to understand fundamental principles of fungal-bacterial interactions through physics-based, simulation and machine learning models of metabolism, protein expression and gene expression, and to couple these models to the mycelial growth and bacterial chemotaxis.

In this regard, the maximization of entropy, such as is done through metabolic exchange discussed above, has been alluded to historically or directly stated as the goal or an emergent property of biological systems by both physicists and ecologists, e.g., Lotka 1922a; Lotka 1922b; Odum and Pinkerton 1955; Prigogine 1978; Vallino 2009. Yet, the concept has been underdeveloped regarding application to systems such as metabolism, and because of the abstract nature of the concept, it has gained insufficient recognition as an operational principle among microbiologists and cell biologists. There are several issues regarding the application of the physical principles to biological systems the team addresses in this project.

This project first demonstrates how one can determine a species phenotype based on knowledge of the genome and the environmental conditions by using physical principals. Specifically, using statistical thermodynamics, this study shows how a cell's most probable state (phenotype) can be determined (Cannon et al. 2023), and how physicochemical constraints can be used to predict the internal regulation of the cell (Britton et al. 2020). The team demonstrates these concepts using a sophisticated model of metabolism (King et al. 2023).

To scale these concepts beyond the cell to interactions between fungi and bacteria, this group has developed realistic 2D and 3D models of fungal and bacterial growth in which each cell can contain sophisticated metabolisms and exchange nutrients with other species. These models require sophisticated, large-scale computing. Researchers have teamed with the DOE Exascale Computing project to implement Adaptive Mesh Refinement (AMR) using high performance computing in these models. The AMReX (AMR for the Exascale) project at Lawrence Berkeley National Laboratory supports the development of block-structured AMR algorithms for solving systems of partial differential equations on exascale architectures. The team has now developed large-scale, hybrid central processing unit-graphics processing unit simulations for growth of bacterial colonies (Palmer et al. 2023), growth of fungal mycelia, and the integrated growth and metabolic exchange between fungi and bacteria. By the end of the year, these methods will enable sophisticated computer experiments that can be used to complement experimental field work and omics measurements.

Funding Statement: This project is supported by the DOE's BER program under contract 78460. The AMReX developers (Almgren, Myers) were supported by the DOE Exascale project.

References

Britton, S., et al. 2020. "Enzyme Activities Predicted by Metabolite Concentrations and Solvent Capacity in the Cell," *Journal of the Royal Society, Interface* **17**(171), 20200656. DOI:10.1098/rsif.2020.0656.

Cannon, W. R., et al. 2023. "Probabilistic and Entropy Production Modeling of Chemical Reaction Systems: Characteristics and Comparisons to Mass Action Kinetic Models," *arXiv* 2310.06135. DOI:10.48550/arXiv.2310.06135.

King, E., et al. 2023. "An Approach to Learn Regulation to Maximize Growth and Entropy Production Rates in Metabolism," *Frontiers in Systems Biology* 3. DOI:10.3389/fsysb.2023.981866.

Lotka, A. J. 1922a. "Natural Selection as a Physical Principle," *Proceedings of the National Academies of Sciences of the United States America* **8**(6), 151–4. DOI:10.1073/pnas.8.6.151.

Lotka, A. J. 1922b. "Contribution to the Energetics of Evolution," *Proceedings of the National Academies of Sciences of the United States of America* **8**(6), 147–51. DOI:10.1073/pnas.8.6.147.

Odum, H.T., and R. T. Pinkerton. 1955. "Time's Speed Regulator: The Optimum Efficiency for Maximum Power Output in Physical and Biological Systems," *American Scientist* **43**(2).

Palmer, B. J., et al. 2023. "BMX: Biological Modelling and Interface Exchange," *Scientific Reports* **13**(1), 12235. DOI:10.1038/s41598-023-39150-1.

Prigogine, I. 1978. "Time, Structure, and Fluctuations," *Science* **201**(4358), 777–85. DOI:10.1126/science.201.4358.77.

Vallino, J. J. 2009. "Ecosystem Biogeochemistry Considered as a Distributed Metabolic Network Ordered by Maximum Entropy Production," *Philosophical Transactions of the Royal Society of London Biological Sciences* **365**, pp. 1417–27. DOI:10.1098/rstb.2009.0272.

The Data Transfer Service: FAIR Data Delivery Made Easy

A. J. Ireland[1]* (aireland@lbl.gov), Jeffrey Johnson[2], Kjiersten Fagnan[1], **Elisha M. Wood-Charlson[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Cohere Consulting LLC

Project Goals: Enable integrated scientific research across BER and beyond by facilitating the transfer of data and accompanying metadata between sample, data, and analysis platforms.

Researchers working in the DOE BER program have access to a wide range of data and analysis resources across scientific domains. Each platform provides unique capabilities, ranging from user facilities generating data and preliminary analyses (e.g., DOE Joint Genome Institute and Environmental Molecular Sciences Laboratory), linking sample metadata to standardized data analyses (National Microbiome Data Collaborative), enabling complex reproducible analyses for publication [DOE Systems Biology Knowledgebase (KBase)], and serving as a BER project data repository (Environmental System Science Data Infrastructure for a Virtual Ecosystem). However, researchers typically need to use more than one platform for data integration, processing, publication, and/or collaboration with others, and so need to move data from one platform to another. The lack of cross-platform coordination poses significant challenges to researchers: data transfer is usually performed manually, by downloading from one platform and uploading at another, which tends to be time consuming, difficult to automate or perform at scale, and error prone; it also removes useful metadata, including citation information and data provenance. As established components of the FAIR (findable, accessible, interoperable, and reusable) data principles, it is important that credit information, provenance, and file metadata are preserved; keeping this metadata also enables the tracking and reporting of the impact of samples and data generated by BER researchers and BER-funded data platforms.

This team is embarking on a new effort to build a Data Transfer Service (DTS) (Wood-Charlson et al. 2023) designed to streamline cross-platform research by providing a simple way to search, access, and transfer data between platforms. The DTS will leverage persistent identifiers (e.g., ORCID, DOI) and community-defined standards (e.g., Frictionless; PROV-O) for capturing file information and provenance. The data package delivered by the DTS will include the data file alongside file-level metadata, data citation, and funder information. These will be captured using the KBase Citation Metadata Schema (Ireland and Wood-Charlson 2023), which seamlessly integrates with OSTI and DataCite publishing schemas. The aim is to streamline and incentivize the practice of citing datasets in the same way that one might cite a publication (Wood-Charlson et al. 2022). In the future, the DTS will support the movement not only of public data but also private datasets with secure authentication, and the design is extensible enough that it could be used to connect resources outside BER (e.g., National Center for Biotechnology Information, National Aeronautics and Space Administration), making it a versatile tool for a wide range of scientific investigations.

The ability to easily move data between BER platforms, without the hassle of manual transfers or the risk of losing valuable metadata and provenance information, will be a significant benefit to the many researchers who use more than one platform for data analysis, management, and publication. The DTS not only promises to improve the efficiency of data transfers and facilitate cross-platform collaboration, but also to enhance the integrity and usability of the data itself, paving the way for new insights and facilitating advances in scientific research.

Funding Statement: This work is supported as part of the Genomic Science program of BER. The Data Transfer Service project is funded by the DOE, Office of Science, BER program under award number DE-AC02-05CH11231.

References

Ireland, A. J., and E. M. Wood-Charlson. 2023. "KBase Credit Metadata Schema." Accessed April 2024. KBase. DOI:10.25982/1984203.

Wood-Charlson, E. M., et al. 2022. "Ten Simple Rules for Getting and Giving Credit for Data," *PLoS Computational Biology* **18**(9), e1010476. DOI:10.1371/journal.pcbi.1010476.

Wood-Charlson, E. M., et al. 2023. "Data Transfer Service," DMPTool. DOI:10.48321/D1W96D.

Collaboratively Assembling a Toolkit in KBase to Leverage Probabilistic Annotation and Multiomics Data to Improve Mechanistic Modeling of Metabolic Phenotypes

José P. Faria[1] (jplfaria@anl.gov), Filipe Liu[1], Patrik D'haeseleer[2], Jeremy Jacobson[3], Bill Nelson[3], Jason McDermott[3], Aimee K. Kessell[4], Hugh C. McCullough[4], Hyun-Seob Song4, Janaka N. Edirisinghe[1], Nidhi Gupta[1], Samuel Seaver[1], Andrew P. Freiburger[1], Qizhi Zhang[1], Pamela Weisenhorn[1], Neal Conrad[1], Raphy Zarecki[5], Matthew DeJongh[6], Aaron A. Best[6], KBase Team[1,7,8,9], Robert W. Cottingham[7], Rhona Stuart[2], Kirsten Hofmockel[3], Christopher S. Henry[1], **Adam P. Arkin[8]** (PI)

[1]Argonne National Laboratory; [2]Lawrence Livermore National Laboratory; [3]Pacific Northwest National Laboratory; [4]University of Nebraska–Lincoln; [5]Newe Ya'ar Research Center, Agricultural Research Organization; [6]Hope College; [7]Oak Ridge National Laboratory; [8]Lawrence Berkeley National Laboratory; [9]Brookhaven National Laboratory

kbase.us

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for biologists and bioinformaticians. KBase integrates many data and analysis tools from DOE and other public services into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer-extensible platform that enables scientists to analyze their data within the context of public data and share their findings across the system. This presentation describes a new modeling pipeline developed by a collaborative project between KBase and the μBiosphere Science Focus Area (SFA) and Pacific Northwest National Laboratory (PNNL) Soil SFA.

Mechanistic understanding of a biological system begins with accurate functional annotation of the system's proteins. Unfortunately, in most cases, protein annotations are uncertain and error-prone, while most analytical pipelines treat annotations as either present or absent. Genome-scale metabolic models (GEMs) permit the evaluation of metabolic annotations within the broader context of the living machines they characterize and, thus, are ideal tools for considering and resolving uncertainty to arrive at the optimal set of annotations that best describe all experimental observations about an organism.

Here, researchers describe an ecosystem of metabolic modeling tools collaboratively developed within KBase to accomplish this goal. The system begins by annotating protein sequences using Rapid Annotation using Subsystems Technology (RAST), Protein Data Bank (PDB), Distilled and Refined Annotation of Metabolism (DRAM), Prokka, and GLM4EC. The system also supports the upload of annotations produced outside of KBase. These tools provide a pool of probabilistic protein annotations that this modeling framework will draw upon to mechanistically explain organism phenotypes.

Next, the newly developed ModelSEED2 tool is used to build a draft GEM. This tool offers significant enhancements over the previous reconstruction apps in KBase, including (1) dramatically improved representation of energy metabolism; (2) improved archaea and cyanobacteria reconstruction; and (3) curation of all metabolic pathways with mappings to RAST subsystems annotations (Faria et al. 2023). The ModelSEED2 generates larger models with more reactions and genes and fewer gaps. Applying ModelSEED2 to thousands of diverse species, group members see conserved patterns in the adenosine triphosphate biosynthesis mechanism across phylogeny and identify clades where understanding of energy biosynthesis is still poor. Gaps in the draft GEMs also offer a metric to evaluate annotation quality at the systems level.

New ensemble modeling tools then sample from the probabilistic pool of hypothesized protein annotations to produce an ensemble of potential draft GEMs. GEM ensembles are evaluated based on: (1) adenosine triphosphate biosynthesis mechanisms, (2) gap filling needed to replicate observed phenotypes, and (3) agreement of simulated flux with multiomics data. A subset of best-performing models can then be extracted and retained for further analysis.

Gap filling is essential to this ecosystem as it selects the most probable annotations that best fit experimental observations (e.g., observed growth phenotypes or

multiomics data). The new OMEGGA gap-filling algorithm globally fits a GEM to available phenotype data using reactions associated with the highest probability annotations and genes with expression in multiomics datasets.

With KBase building the ModelSEED2, μBiosphere SFA building the probabilistic annotation system, and PNNL Soil SFA building OMEGGA, this has been a collaborative endeavor. This project demonstrates the efficacy of the tools by applying them to study isolates and omics-datasets from the μBiosphere and PNNL's Soil Microbiome SFA (McClure et al. 2022). The group annotates each isolate, constructs and optimizes GEMs, and fits the GEMs to phenotype and expression data generated for the isolates. As a result, researchers greatly reduce gaps in GEM pathways and improve isolate annotations.

Funding Statement: This work is supported as part of the Genomic Science program of BER. The DOE Systems Biology Knowledgebase (KBase) and the Science Focus Area-KBase supplements are funded by the DOE, Office of Science, BER program under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, DE-AC02-98CH10886, DEAC0576RL01830, and DE-AC52-07NA27344.

References

Faria, J. P., et al. 2023. Preprint. "ModelSEED v2: High-Throughput Genome-Scale Metabolic Model Reconstruction with Enhanced Energy Biosynthesis Pathway Prediction," *bioRxiv.* DOI:10.1101/2023.10.04.556561.

Henry, C. S., et al. 2010. "High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models," *Nature Biotechnology* **28**(9), 977–82. DOI:10.1038/nbt.1672.

McClure, R., et al. 2022. "Interaction Networks are Driven by Community Responsive Phenotypes in a Chitin Degrading Consortium of Soil Microbes," *mSystems* **7**(5), e00372–22. DOI:10.1128/msystems.00372-22.

KBase Research Assistant and Genome Annotation Agent

Prachi Gupta[1]* (prachigupta@lbl.gov), William J. Riehl[1], Meghan Drake[1,2], Sean P. Jungbluth[1], Christopher Neely[1], Ziming Yang[3], Marcin P. Joachimiak[1], Mikaela Cashman[1], R. Shane Canon[1], Paramvir S. Dehal[1], **Adam P. Arkin[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Oak Ridge National Laboratory; [3]Brookhaven National Laboratory

kbase.us

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. It is a powerful platform for biological research and analysis, providing a breadth of reference data, analysis applications, and resources to the scientific community. Usage of these apps and narrative descriptions of their results can be easily presented in the KBase Narrative Interface. This poster presents a novel set of artificial intelligence–driven tools to assist researchers in using KBase to analyze their data and publish their findings.

Rapid advances are being made in artificial intelligence using large language models (LLMs) as a natural language processing tool for various applications. This project takes advantage of the growth of this technology to produce a KBase Research Assistant that will serve as a guide, facilitating navigation and analysis within the KBase platform. The team envisions the Assistant as a tool that can converse with a KBase user, understand their data and its relationship with public data on the system, and leverage this to reach the user's analysis goals. This Assistant will also aid in communicating results via static narratives and academic publications.

The initial target for the KBase Assistant will help a user start with a set of sequenced reads from a microbial isolate to produce an annotated genome that can be used in further analyses by the community. The general workflow here uses KBase apps to assemble and annotate the reads, accompanied by quality assurance and control at each step. The Assistant will help with interpretation of the output of each app and craft each subsequent step in the workflow, with variation where needed.

To build this first genome builder assistant, several LLM-related tools are being created to interact with each other and the user's data in a user-driven workflow. After uploading their reads to a KBase Narrative,

a user will be able to activate the Assistant, which will orchestrate different groups of artificial intelligence agents. The first agent will be a modular reasoning, knowledge, and language agent that will make use of retrieval augmented generation (RAG) tools to perform app recommendation. These RAG tools provide knowledge from KBase documentation and tutorials to ensure proper use of KBase apps. The second agents will manage running the apps and provide results to the assistant for analysis and interpretation, followed by suggestions of the next steps. A third set of agents will ensure that the narrative gets populated properly with the apps and summaries of results. Once analysis is complete and findings are gathered, an agent will assist with developing a publication.

The KBase Research Assistant will be a powerful step forward in enabling KBase users to take advantage of the full breadth of computational tools and public data that KBase provides. Although its initial focus will be on genome annotation, the Assistant will grow to provide insight and utility for other biological analyses.

Funding Statement: This work is supported as part of the BER Genomic Science program. The DOE Systems Biology Knowledgebase (KBase) is funded by the DOE, Office of Science, BER program under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886. The Educator Working Groups described is funded by a National Science Foundation RCN-UBE Incubator under award number 2316244.

Uncovering Characteristic Traits of Earth's Microbiomes

Marcin P. Joachimiak[1]*(mjoachimiak@lbl.gov), Ziming Yang[1], William J. Riehl[1], Christopher Neely[1], Prachi Gupta[1], Sean P. Jungbluth[1], Paramvir S. Dehal[1], **Adam P. Arkin[1,2]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]University of California–Berkeley

Project Goals: The goal of this project is to use a vast, standardized collection of metagenome-sequenced samples from diverse ecosystems, to uncover microbial traits predictive of different environmental niches. Through advanced machine learning techniques, this research team aims to identify signature metagenome features—spanning sequence domains, functions, and taxa—that correspond to traits characteristic of specific ecosystems and indicate underlying ecosystem relationships at the microbial biogeographic scale.

Earth's biosphere is an interconnected, dynamically changing network of ecosystems, with microbes playing a significant environmental role. Advances in microbial metagenomics have recently provided extensive data on microbial communities across ecosystems, including biological sequences, taxonomy, and functional annotations. Previous research, focusing on 16S taxonomic data, has shown promising results and indicated vast uncharacterized biological diversity, but 16S barcoding-based methods are constrained by the weak connections to functional annotation data. This project hypothesized that using diverse metagenome features, including sequence domains, functions, and taxa, could reveal traits essential for survival in different ecosystems. Using the largest standardized metagenome sample collection across varied ecosystems, the group trained machine learning models to predict the source ecosystem for metagenome samples. Group members identified optimal metagenome feature types and model parameters, resulting in models that performed well in cross-validation, and training at different ecosystem classification levels improved performance for ecosystems with sparse training data. Model interpretation methods identified signature metagenome features for distinguishing 41 ecosystems, leading to insights about traits that are characteristic of specific ecosystems. This collection of traits, which may have adaptive significance, reveals examples of direct linkages between microbial functions and environmental properties, highlights important unknown functions, and implies ecosystem relationships that align well with established classifications but with ecosystems being more interlinked than is currently appreciated.

Funding Statement: This work is supported as part of the Genomic Science program of BER. The DOE Systems Biology Knowledgebase (KBase) is funded by the DOE, Office of Science, BER program under DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

Reference

Park, H., et al. 2024. "A Bacterial Sensor Taxonomy Across Earth Ecosystems for Machine Learning Applications," *mSystems* **9**(1), e0002623. DOI:10.1128/msystems.00026-23.

Leveraging Machine Learning for Enhanced Prediction of Microbial Carbon Utilization Phenotypes

Priya Ranjan[1]* (ranjanp@ornl.gov), Dileep Kishore[1], Christopher Neely[2], William Riehl[2], Marcin P. Joachimiak[2], Paramvir S. Dehal[2], Janaka N. Edirisinghe[3], José P. Faria[3], Dale A. Pelletier[1], Mitchel J. Doktycz[1], **Robert W. Cottingham[1]** (PI), **Christopher S. Henry[3]** (PI), **Adam P. Arkin[2]** (PI)

[1]Oak Ridge National Laboratory; [2]Lawrence Berkeley National Laboratory; [3]Argonne National Laboratory

kbase.us

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for biologists and bioinformaticians. KBase integrates many data and analysis tools from DOE and other public services into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer-extensible platform that enables scientists to analyze their data within the context of public data and share their findings across the system. Here, the team describes a microbial carbon-utilization phenotype prediction pipeline developed as a collaboration between KBase, the Oak Ridge National Laboratory's Plant-Microbe Interfaces (PMI) Science Focus Area (SFA), and the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) SFA.

Deciphering the mechanisms of microbial nutrient utilization phenotypes from genomic data is necessary for understanding microbial niches in ecosystems. However, incomplete gene annotations and limited, inconsistent training data hamper the accuracy of current predictive models. This study tackles these obstacles by integrating diverse annotation sources, including metabolic [e.g., Rapid Annotation using Subsystem Technology (RAST), KOfam), protein functional (e.g., UniProt)], and *de novo* protein clustering, to enrich feature representations. The group employed a comprehensive dataset comprising 626 diverse microbial genomes and their individual growth outcomes across 98 different carbon sources, facilitating the development of 98 phenotype-specific classifiers.

This study employed a range of feature preprocessing and selection strategies alongside a standardized evaluation framework, to facilitate the comparison of classifier performance and enable the effective integration of models that use different feature representations. Group members evaluated the accuracy and robustness of these classifiers utilizing various feature representations and observed that metabolic features, specifically RAST, exhibit the highest average accuracy across the different phenotypes. However, specific phenotype classifiers exhibit improved performance when utilizing protein function annotations or *de novo* protein clusters, suggesting that these genomes may possess incomplete metabolic annotations in pathways relevant to the given phenotype. These findings highlight the potential limitations of current genome annotation methods and the need for continued research to enhance understanding of metabolic pathways and their associated phenotypes. Moreover, the group observed that while feature selection enhances classifier accuracy, methods like non-negative matrix factorization, which reduce feature dimensionality, detrimentally impact performance. This loss in accuracy indicates the critical role of smaller sets of specific enzymes or proteins in phenotype expression. Overall, the merger of feature sets notably boosts prediction accuracy for challenging phenotypes, underscoring this method's effectiveness in addressing annotation inaccuracies.

Furthermore, as a part of the PMI and the ENIGMA SFAs, group members are currently curating gold-standard phenotypic datasets under the same experimental protocols. These standardized datasets will form the foundation for further developing more robust phenotype prediction classifiers that cover a broader array of carbon sources and a phylogenetically diverse set of microbes. Ultimately, the group aims to integrate these classifiers into the KBase platform, making them available as applications and as part of the relation engine-driven pipelines. With this integration, the group strives to enable users to predict microbial phenotypes from a wide array of genomes, including all imported Reference Sequence genomes, thereby significantly advancing microbial research.

This study aims to improve microbial phenotype prediction by utilizing multifaceted feature representations, advanced machine-learning techniques, and standardized datasets to create accurate classifiers for specific phenotypes. These classifiers have the potential to advance scientists' understanding of microbial growth phenotypes and serve as essential resources for improving annotations of metabolic pathways and understanding of microbial ecology.

Funding Statement: This work is supported as part of the Genomic Science program (GSP) of BER. The DOE Systems Biology Knowledgebase (KBase) is

funded by the DOE, Office of Science, BER program under DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

This work is also supported by DOE, Office of Science, BER program, GSP as part of the Plant–Microbe Interfaces Science Focus Area at Oak Ridge National Laboratory (pmiweb.ornl.gov). Oak Ridge National Laboratory is managed by UT-Battelle, LLC., for DOE under contract DE-AC05-00OR22725.

Natural Language Processing for Synthetic Biology: Providing Generalizable Literature Mining Through KBase

Christopher Neely[2]* (cjneely@lbl.gov), Shinjae Yoo[1], Carlos X. Soto[1], Gilchan Park[1], Vivek K. Mutalik[2], **Paramvir S. Dehal[2]** (PI)

[1]Brookhaven National Laboratory; [2]Lawrence Berkeley National Laboratory

See abstract titled "Knowledge Extraction from Literature," p. 35.

Enhancing KBase Security: Strengthening Platform Integrity for Biological Research

Boris Sadkhin[2]* (bsadkhin@anl.gov), Christopher S. Henry[2], Gavin Price[1], A. J. Ireland[1], **Shane Canon[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Argonne National Laboratory

kbase.us

For over a decade, the DOE Systems Biology Knowledgebase (KBase) platform has been an essential resource for DOE Biological researchers and the broader biological research community. It provides a comprehensive web-based system for analysis, collaboration, and publishing results. However, as the platform has evolved, it faces increased complexity and security challenges, including end-of-life dependencies and potential vulnerabilities. This project aims to enhance the platform's security and maintenance capabilities to mitigate high-impact outages, data integrity risks, and security breaches.

This poster presents an overview of the KBase platform architecture, highlighting its wide array of services and underlying databases. This poster outlines the project's objectives, emphasizing the team's proactive approach to prioritize updates based on a thorough audit, focusing on critical services and dependencies. Leveraging automated scanning tools like Dependabot and Trivy, the team can efficiently identify and address vulnerabilities, ranging from straightforward dependency updates to complex migrations ensuring both security and operational integrity.

Additionally, the group delves into the specifics of the platform services, emphasizing their importance and role within the KBase ecosystem. This presents progress to date, showcasing milestones achieved in addressing critical security and end-of-life issues for the platform's core services. This poster concludes with an outline of the next steps, including ongoing maintenance processes and automation, aimed at sustaining these improvements.

Integration of Computational Tools to Explore the Diversity of Temporal Regulation in Plant Specialized Metabolism

Samuel Seaver[2]* (seaver@anl.gov), Angela Ricono[1], Sara El Alaoui[2], **Kathleen Greenham[1]** (PI)

[1]University of Minnesota; [2]Argonne National Laboratory

Plants produce an amazing diversity of specialized metabolites (SM) that offer many benefits to human society. SMs are essential for pharmaceutical products and non-medicinal applications in the chemical industry, food additives, dyes, perfumes, cosmetics, and nutraceuticals. These products offer the potential to increase the return on investment of current biofuel crops by providing high-value co-products. While many specialized metabolic enzymes have been characterized, their spatial and temporal regulation is less understood, creating a challenge for engineering and optimizing metabolite levels. Understanding how diverse plants differentially regulate the production of the products arising from the same SM pathway will enable researchers to engineer such plants with greater reliability. The goal of this project is to build a computational tool in DOE Systems Biology Knowledgebase (KBase) that would enable researchers to integrate transcriptome data with metabolic networks of general and specialized metabolism for different plant species. This tool will enable researchers to explore different combinations of SM precursors and identify key enzyme targets for engineering. The research team aims to build a set of classifiers through the application of machine learning techniques that would enable this prediction. To do so, the team will focus on the glucosinolate (GSLs) class of SMs within the plant order Brassicales. This project's

objectives are to (1) experimentally design and benchmark the biosynthesis of multiple GSLs in eight phylogenetically distinct species from diverse families within the Brassicales using high resolution time series datasets; (2) reconstruct the general and specialized metabolic networks for GSL biosynthesis, enabling the integration of omics data; (3) train and test the model to predict GSL biosynthesis; and (4) use the KBase platform to encode this approach in a series of apps that will enable other researchers to apply this approach to their pathway of interest. To disseminate the utility of the tool for target identification for SM production, the team will host virtual and onsite training workshops. This will help to spur research into engineering plants as platforms for co-production of biofuel and co-products and also increase the plant user community on KBase.

Develop Software Tools to Discover Genotype-Specific RNA-Splicing Variants and Microexon Alternative Splicing in Plant Populations

Chi Zhang* (czhang5@unl.edu, PI), Yu Shi

University of Nebraska–Lincoln

Pre-mRNA splicing is an essential step in the regulation of gene expression. Although extensive research has been conducted, there remain open questions, significant knowledge gaps, and requirements for bioinformatic tools in plants, especially at the population level and for microexon (exons ≤ 15 nucleotides) splicing. What is lacking are efficient and accurate methods to analyze population transcriptome datasets to define and identify the full spectrum of RNA-splicing variants and link these RNA-splicing variants to phenotype in plant populations. This project's goals are to create broadly applicable software tools to quantify genotype-specific RNA-splicing (GSS) variants based on multiomics data in a population to facilitate gene function discovery and develop frameworks to associate RNA-splicing variants to identify candidate genes that influence plant adaptability to the environment. Once a bioinformatics framework capable of analyzing multiomics data to identify high-quality RNA-splicing variations in plant populations is established, it will reveal the link from genotype-specific RNA splicing, to protein functions, and finally to the phenotypic response. This will enable the researchers to comprehend how RNA-splicing regulation translate into organismal phenotypes and underlying molecular mechanisms that determine phenotypic outcomes in response to perturbations of complex biological systems. To achieve these goals and construct integrative computational tools, the team is currently (1) developing an online database of microexons for 132 plant genomes; (2) developing the second version of microexon modeling software tools (ME modeler version 2) to model microexons in plants more efficiently and more accurately; and (3) developing VaSP2 for alternative splicing analysis in plant populations.

Expanding Python Library Scikit-Bio for Efficient Multiomic Data Integration and Complex Community Modeling

Qiyun Zhu[1]* (qiyun.zhu@asu.edu, PI), James Morton[2], Daniel McDonald[3], Matthew Aton[1], Lars Hunger[2], J. Gregory Caporaso[4], Rob Knight[3]

[1]Arizona State University; [2]Gutz Analytics; [3]University of California–San Diego; [4]Northern Arizona University

scikit.bio

Project Goals: This team is expanding scikit-bio, a popular and versatile bioinformatics Python library. The team is implementing functionality for large-scale multiomic data analysis to examine complex relationships between plants, microbes, and the environment.

Scikit-bio is an open source Python library offering an extensive range of bioinformatics functions to support microbiome research and beyond. The team's continuous efforts to optimize fundamental algorithms have enabled the analysis of extremely large communities. The GSP award has empowered the team to accelerate scikit-bio development since September 2023.

Team Building: A competitive developer team has been successfully reassembled. Software engineer Matthew Aton and bioinformatician Dr. Lars Hunger have been recruited and are effectively working on the project with the senior members. Three undergraduate students from Arizona State University and University of California–San Diego have been engaged in the project. The revived scikit-bio has also attracted multiple community contributors. The team has been meeting monthly to ensure a cohesive overview of progress and plans.

Overall Advancements: In the first six months of this project, a total of 40 pull requests have been merged into the codebase. A redesigned website (scikit.bio) is online, featuring reorganized documentation for users

and contributors. The codebase has been rigorously refactored to match modern standards. For example, Ruff was adopted to standardize code style across the project. Support for the latest Python ecosystem, including Python 3.12 and SciPy 1.12, has been unblocked. A new release of scikit-bio is anticipated by the meeting time.

Sparse Matrix: The Biological Observation Matrix (BIOM) library has been integrated into scikit-bio, marking a significant enhancement in its ability to represent and manipulate sparse data matrices, which are characteristic of various omic data types. This integration not only streamlines the handling of large-scale omic data but also optimizes computational resources by focusing on the non-zero values. Efforts are ongoing to further optimize algorithms to take advantage of sparse matrices.

Metadata Object: The team adapted and augmented the metadata module from the popular QIIME 2 package. This improved module supports a wider range of metadata types, extending from numeric and categorical to also include Boolean, ordinal, temporal, and free text, among others. Additionally, it introduces standardization for sample identifiers, like specimen ID and host subject ID. Efforts are underway to develop a data dictionary object, which will provide essential context for metadata values, facilitating harmonization of data across omic layers and studies.

Diversity Metrics: Multiple phylogeny-aware diversity metrics such as balance weighted phylogenetic diversity (BWPD) have been implemented to facilitate modeling of complex communities in light of the evolutionary relationships among microbes. Meanwhile, the team refined the implementation and documentation of existing metrics.

Biological Sequences: The team has expanded the functionality of biological sequences. The sequence alignment function is being redesigned to improve efficiency and usability. Sequences can be converted into tokens to facilitate feature annotation using machine learning frameworks.

Workshop: The team's proposal for hosting a full-day tutorial of scikit-bio at the Intelligent Systems for Molecular Biology 2024 conference in July has been accepted. The team anticipates enrolling up to 40 mentees, including researchers, educators, and developers. These efforts aim to make scikit-bio increasingly useful to the science community.

Funding Statement: This project is currently supported by DOE's Office of Science under DE-SC0024320, awarded to Dr. Qiyun Zhu (lead principal investigator), Dr. James Morton, and Dr. Rob Knight.

Multiscale Computational Digital Twins for Whole-Body to Subcellular Radiation Effects

Greeshma Agasthya[1]* (agasthyaga@ornl.gov), Paul Inman[1], Matthew Andriotty[1,2], Zachary Fox[1], Arpan Sircar[1], Xiao Wang[1], Zakaria Aboulbanine[1], Jayasai Rajagopal[1], Debsindhu Bhowmik[1], Christopher Stanley[1], Miguel Toro Gonzalez[1], Sandra Davern[1], Anuj Kapadia[1], **Rick Stevens[3]** (PI)

[1]Oak Ridge National Laboratory; [2]Georgia Institute of Technology; [3]Argonne National Laboratory

This research group has developed a high-performance computing (HPC)–based multiscale computational digital twin framework to assess radiation exposure effects on humans across subcellular, multicellular, and whole-body scales. This includes a multiscale perfusion model for radioisotope therapy, estimating radioactivity delivery and uptake at each scale and predicting biological effects. Here is how each scale integrates into the framework: at the subcellular level, the team has validated a pipeline quantifying DNA chromosome aberrations post-radiation. Using experimental Hi-C data (Sanders et al. 2020), researchers create 3D chromosome models imported into TOPAS n-bio (a Monte-Carlo simulation platform) to estimate DNA double-strand breaks (DSB; Chatzipapas et al. 2020). Mechanistic repair models from MEDRAS predict post-radiation aberrations over time (McMahon and Prise 2021). Validation against *in vitro* studies and cell experiments with a mouse breast cancer cell line confirms accuracy. Moving to multicellular and whole-body scales, researchers have automated dose calculations for organs, individual cells in tumors, and cell survival probabilities based on DNA damage. The team's pipeline integrates a tumor growth model in Compucell3D with eXtended CArdiac Torso (XCAT) phantoms for whole-body simulations (Segars et al. 2013). GEANT4, a Monte Carlo simulation software, calculates absorbed dose in radiation therapy protocols, feeding back into Compucell3D to assess cell survivability (Allison et al. 2006, 2016; Swat et al. 2012).Validation against recent multicellular models shows promising results, with ongoing efforts to validate against spheroid-based experiments. For multiscale perfusion modeling, the team has implemented a physiologically based

pharmacokinetic (PBPK) model simulating radioisotope distribution at the organ level, feeding into a computational fluid dynamics model for tissue-level perfusion. Work is underway to demonstrate subcellular and multicellular scale radiopharmaceutical perfusion. Preliminary results involve growing spherical tumors with multiple cell types in a controlled nutrient environment, exposing them to Actinium 225 in GEANT4 to estimate cell radiation dose and survivability. This fully automated pipeline is optimized for graphics processing unit computation. In conclusion, this framework offers a robust tool for modeling radiation effects across various applications. With potential applications in radiation treatment protocol comparisons, as well as future clinical use pending further validation, the group envisions a multiscale *in silico* digital twin framework for precision dosimetry in radiation therapy. This framework would predict cell survival and treatment outcomes across different cancer types, integrating absorbed dose, biodistribution, cell toxicity, and repair mechanisms.

References

Allison, J., et al. 2006. "Geant4 Developments and Applications," *IEEE Transactions on Nuclear Science* **53**, 270–78. DOI:10.1109/TNS.2006.869826.

Allison, J., et al. 2016. "Recent Developments in Geant4," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835**, 186–225. DOI:10.1016/j.nima.2016.06.125.

Chatzipapas, K. P., et al. 2020. "Ionizing Radiation and Complex DNA Damage: Quantifying the Radiobiological Damage Using Monte Carlo Simulations," *Cancers* **12**(4), 799. DOI:10.3390/cancers12040799.

McMahon, S. J., and K. M. Prise. 2021. "A Mechanistic DNA Repair and Survival Model (Medras): Applications to Intrinsic Radiosensitivity, Relative Biological Effectiveness and Dose-Rate," *Frontiers in Oncology* **11**. DOI: 10.3389/fonc.2021.689112.

Sanders, J. T., et al. 2020. "Radiation-Induced DNA Damage and Repair Effects on 3D Genome Organization," *Nature Communications* **11**, 6178. DOI:10.1038/s41467-020-20047-w.

Segars, W. P., et al. 2013. "Population of Anatomically Variable 4D XCAT Adult Phantoms for Imaging Research and Optimization," *Medical Physics* **40**. DOI:10.1118/1.4794178.

Swat, M. H., et al. 2012. "Chapter 13: Multi-Scale Modeling of Tissues Using CompuCell3D," *Methods in Cell Biology* **110**, 325–66. Eds. Asthagiri, A. R. and A. P. Arkin. Academic Press. DOI:10.1016/B978-0-12-388403-9.00013-8.

Unveiling Molecular Interactions and Metabolic Contributions in Sorghum Anthracnose Defense: Towards the Integration of Fungal Pathogen and Host Sorghum Models

Janaka N. Edirisinghe[1]* (janakae@anl.gov), Claudia Lerma-Ortiz[1], Clint Magill[2], Sam Seaver[1], Sara El Alaoui[1], Filipe Liu[1], José P. Faria[1], Christopher S. Henry[1], **Qun Liu[3]** (PI)

[1]Argonne National Laboratory; [2]Texas A&M University–College Station; [3]Brookhaven National Laboratory

Project Goals: To gain deeper insights into the virulence mechanisms of the fungus *Colletotrichum sublineola* in Anthracnose, it is essential to comprehend the pivotal metabolic interactions occurring between *C. sublineola* and its host organism, sorghum. Through the development of mechanistic models, researchers can delve into these crucial metabolic interactions, shedding light on the roles of both the fungus and the host sorghum in the process.

Anthracnose, a devastating disease in sorghum, is caused by the hemibiotrophic fungal pathogen *Colletotrichum sublineola* (Crouch and Tomaso-Peterson 2012). Sorghum's nucleotide-binding leucine-rich repeat proteins (NLRs) play a pivotal role in recognizing virulence effectors of pathogens like *C. sublineola* (Cs), triggering effective immune responses vital for plant defense (Rausher 2001). Nonetheless, the intricate molecular interactions governing *C. sublineola* pathogenicity, anthracnose resistance, and susceptibility in sorghum remain poorly elucidated. Further, *C. sublineola* metabolism during its appressoria and haustoria invasion stages is not well understood. To address these knowledge gaps, the overall goal within the wider context of the BRaVE project is to build a metabolic model of the interactions between sorghum and *C. sublineola*. To this end, the team sequenced, assembled, and annotated a newly isolated *C. sublineola* pathogen, followed by the reconstruction of a genome-scale metabolic model (GEM) around a well-curated core of central carbon metabolism, fermentation, electron transport chains, and energy biosynthesis. A comparative study was also performed between the new *C. sublineola* genome and many existing nearby fungal strains.

This GEM reconstruction builds upon extensive curation efforts to develop an improved fungal modeling template for KBase by consolidating and reconciling

data from thirteen diverse published fungal models (Edirisinghe et al. 2023). This integration involved harmonizing the biochemistry of each published model with the ModelSEED biochemistry database, effectively minimizing redundancy and inconsistency in biochemical pathways and the underlying protein annotations.

This approach is further supported by protein family data computed across the thirteen fungal species and a set of additional well-sequenced fungal strains mapped to pertinent biochemistry data. This system systematically captures the unique biochemistry of each model, ensuring consistency in annotation mapping with the relevant biochemical context.

The curated and reconciled pathways refined and consistently annotated protein families and enhanced fungal template model are bundled together within the readily accessible user-friendly Build Fungal Model application on the DOE Systems Biology Knowledgebase (KBase) platform (narrative.kbase.us). The work also resulted in significant improvements to all the published fungal models that were reconciled in the efforts to build the fungal modeling tool. Additionally, researchers present a KBase narrative workflow illustrating the construction of the central carbon model for *C. sublineola*, utilizing data from the newly sequenced *C. sublineola* genome and other closely related *C. sublineola* genomes.

Concurrently, for the plant side of the eventual host-pathogen mechanistic model, researchers applied the PlantSEED (Seaver et al. 2018) approach to generate a high-quality metabolic model of sorghum using the latest set of gene models. Previously, the conservative approach of predicting orthologs using OrthoFinder meant researchers had a high rate of false negatives. Researchers adopted a new approach that takes into account the distribution of pairwise sequence identity for each protein family, allowing both orthologs and in-paralogs to be predicted as carrying the functions of plant primary metabolism, enabling researchers to assign sorghum enzymes to an additional 10% of plant primary metabolism.

This team is actively applying the fungal and plant models independently to integrate and mechanistically interpret multiomics datasets produced by other project team members, while the ultimate goal is to merge these models into a predictive dynamic host-pathogen interaction model.

Funding Statement: This work is supported as part of the BER Biopreparedness Research Virtual Environment (BRaVE) project and by KBase. KBase and BRaVE are funded by the DOE Office of Science, BER program under DE-AC02-06CH11357, and DE-AC02-98CH10886.

References

Crouch, J. A., and Tomaso-Peterson, M. 2012. "Anthracnose Disease of Centipede Grass Turf Caused by *Colletotrichum eremochloae*, a New Fungal Species Closely Related to *Colletotrichum sublineola*," *Mycologia* **104**(5), 1085–96. DOI:10.3852/11-317.

Edirisinghe, J. N., et al. 2023. "Consolidating Published Fungal Models: Strategies and Challenges in the Integration of Diverse Fungal Biochemistry for Multi-Omics Analyses," Metabolic Pathway Analysis (MPA) International Conference. Seoul, Republic of Korea.

Rausher, M. D. 2001. "Co-Evolution and Plant Resistance to Natural Enemies," *Nature* **411**(6839), 857–64. DOI:10.1038/35081193.

Seaver, S. M. D., et al. 2018. "PlantSEED Enables Automated Annotation and Reconstruction of Plant Primary Metabolism with Improved Compartmentalization and Comparative Consistency," *Plant Journal* **95**(6), 1102–13. DOI:10.1111/tpj.14003.

Analyzing Biotic and Abiotic Stress Responses in Sorghum Using Comprehensive Field Phenomics Data

Emmanuel M. Gonzalez[1]* (emmanuelgonzalez@arizona.edu), Ariyan Zarei[1], Brenda Huppenthal[1], Jeffrey Demieville[1], Travis Simmons[1], Sebastian Calleja[1], Bruno Rozzi[1], Clay Christenson[1], Yuguo Xiao[2], Indrajit Kumar[2], Maxwell Braud[2], Brian P. Dilkes[3], Eric Lyons[1], Duke Pauli[1,4], **Andrea L. Eveland[2]** (PI)

[1]University of Arizona; [2]Donald Danforth Plant Science Center; [3]Purdue University; [4]Center for Agroecosystem Research in the Desert (ARID)

PhytoOracle GitHub Organization:
github.com/phytooracle

CyVerse Data Commons:
datacommons.cyverse.org/browse/iplant/home/shared/phytooracle

Charcoal Rot of Sorghum App:
charcoal-dryrot-quantification.streamlit.app

Project Goals:

- Use the Field Scanner at the University of Arizona to gather field phenomics data from ethyl methanesulfonate–mutagenized sorghum populations and a custom diversity panel, under both well-watered and water-limited conditions.
- Develop software and machine learning (ML) models to analyze field phenomics data, quantifying individual plant traits to study plant responses to biotic and abiotic stresses.
- Leverage trait data to investigate genotype-phenotype associations and elucidate gene functions.

Sorghum [*Sorghum bicolor* (L.) Moench], the fifth most cultivated cereal crop, is increasingly grown in the United States due to its adaptability to marginal lands and diverse uses as a food, feed, and biofuel crop (Ndlovu et al. 2021; Hossain et al. 2022; Yang et al. 2022). Expanding sorghum cultivation requires understanding its natural and induced resistance to biotic and abiotic stress. Recent technological advances have resulted in small, low-cost, and high-resolution sensors that can be used to rapidly collect phenotypic trait data at regular time intervals in field or greenhouse settings (Li et al. 2020; Sooriyapathirana et al. 2021). Today, high spatial and temporal resolution field phenomics data are being collected to extract information on dynamic plant responses to abiotic and biotic stress under real world field conditions.

The University of Arizona houses the world's largest outdoor plant phenotyping system, the Field Scanner. It uses various sensors to collect plant trait data, including red-green-blue (RGB), photosystem II chlorophyll fluorescence, thermal imagery, and 3D point clouds. This raw data is processed using PhytoOracle, a collection of scalable, modular pipelines for phenomic data (Gonzalez et al. 2023). The PhytoOracle (PO) pipelines facilitate extraction of phenotypic trait data at multiple levels, from whole plants to individual organs. The machine learning (ML) models segment plant point clouds to gather detailed morphological data. This includes traditional shape descriptors like height, volume, and angle. Additionally, topological data analysis (TDA) is used to study subtle shape nuances. Common TDA methods like persistence diagrams and Euler characteristic curves capture topological signatures for a more nuanced shape study (Amézquita et al. 2020; Chazal and Michel 2021; Amézquita et al 2022).

Additionally, ML models are being utilized to identify particular stress factors, including biotic stress. Sorghum, while drought-resistant, is vulnerable to various pathogens, including the destructive soil-borne fungus *Macrophomina phaseolina* (Tassi) Goid. This fungus causes charcoal rot of sorghum (CRS), disrupting the plant's water and nutrient transport, leading to symptoms often confused with other conditions like drought stress and frost damage. An automated method distinguishing CRS from other stresses could improve selection accuracy, heritability, and genetic gain, ultimately facilitating the development of more resilient crop cultivars. Various ML models trained to identify and quantify CRS in RGB images are available for use on a web-based application where users can easily analyze their own images: charcoal-dryrot-quantification.streamlit.app.

In studying both abiotic and biotic stress factors, this research seeks to enhance crop productivity by pinpointing variations in stress resilience. Through the application of fine-scale phenotyping, this research contributes to the development of improved, climate-resilient crop varieties.

Funding Statement: This material is based upon work supported by the DOE BER award numbers DE-SC0020401 and DE-SC0023305, the DOE Advanced Research Projects Agency-Energy OPEN Award Number DE-AR0001101, and the National

Science Foundation CyVerse Award Number DBI-1743442.

References

Amézquita, E. J., et al. 2020. "The Shape of Things to Come: Topological Data Analysis and Biology, from Molecules to Organisms," *Developmental Dynamics* **249**, 816–33. DOI:10.1002/dvdy.175.

Amézquita, E. J. et al. 2022. "Measuring Hidden Phenotype: Quantifying the Shape of Barley Seeds Using the Euler Characteristic Transform," *Silico Plants* **4**, diab033. DOI:10.1093/insilicoplants/diab033.

Chazal, F. and Michel, B. 2021. "An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists," *Frontiers in Artificial. Intelligence* **4**. DOI:10.3389/frai.2021.667963.

Gonzalez, E. M. et al. 2023. "PhytoOracle: Scalable, Modular Phenomics Data Processing Pipelines," *Frontiers in Plant Science* **14**. DOI:10.3389/fpls.2023.1112973.

Hossain, Md. S., et al. 2022. "Sorghum: A Prospective Crop for Climatic Vulnerability, Food and Nutritional Security," *Journal of Agriculture and Food Research* **8**, 100300. DOI:10.1016/j.jafr.2022.100300.

Li, B. et al. 2020. "Phenomics-Based GWAS Analysis Reveals the Genetic Architecture for Drought Resistance in Cotton," *Plant Biotechnology Journal* **18**, 2533–44. DOI:10.1111/pbi.13431.

Ndlovu, E., et al. 2021. "Morpho-Physiological Effects of Moisture, Heat and Combined Stresses on *Sorghum bicolor* [Moench (L.)] and its Acclimation Mechanisms," *Plant Stress* **2**, 100018. DOI:10.1016/j.stress.2021.100018.

Sooriyapathirana, S. D. S. S., et al. 2021. "Photosynthetic Phenomics of Field- and Greenhouse-Grown Amaranths vs. Sensory and Species Delimits," *Plant Phenomics* 2021, 1–13 DOI:10.34133/2021/2539380.

Yang, Q., et al. 2022. "Genetic Analysis of Seed Traits in *Sorghum bicolor* that Affect the Human Gut Microbiome," *Nature Communications* **13**, 5641. DOI:10.1038/s41467-022-33419-1.

Machine Learning-Assisted Genome-Wide Association Study Uncovers Copy-Number Variations of Tandem Paralogs Driving Stress Tolerance Evolution in *Issatchenkia orientalis*

Ping-Hung Hsieh[1,2*] (phhsieh@lbl.gov), Yusuke Sasaki[1,2], Chu-I Yang[3], Zong-Yen Wu[1,2], Andrei S. Steindorff[4], Sajeet Haridas[4], Jing Ke[4], Zia Fatma[1,5], Zhiying Zhao[4], Dana A. Opulente[6], Siwen Deng[7], Chris Todd Hittinger[6], Igor V. Grigoriev[2,4,7], Bruce Dien[1,8], Huimin Zhao[1,5], Yi-Pei Li[3], Yasuo Yoshikuni[1,2,4,9,10], **Andrew Leakey[1]** (PI)

[1]Center for Advanced Bioenergy and Bioproducts Innovation; [2]Lawrence Berkeley National Laboratory; [3]National Taiwan University; [4]DOE Joint Genome Institute; [5]University of Illinois Urbana-Champaign; [6]Great Lakes Bioenergy Research Center; [7]University of California–Berkeley; [8]USDA Agricultural Research Service; [9]Hokkaido University; [10]Tokyo University of Agriculture and Technology

cabbi.bio

Project Goals: This project aims to exploit the genetic and phenotypic diversity of *Issatchenkia orientalis* through population genomics and machine learning approaches to (1) explore the potential molecular mechanisms behind its multi-stress tolerance, (2) develop predictive models for various stress tolerances, and (3) identify key genes associated with tolerance to industrial stresses and resistance to different fungicides.

The environmental yeast *I. orientalis* plays a dual role in human society, owing to its multi-stress tolerance. Its ability to withstand common industrial stressors, such as low pH and high temperatures, makes it an ideal candidate for engineered biosynthesis of bioproducts. However, it also poses significant health risks as a multidrug-resistant pathogen capable of causing invasive fungal diseases and is recognized by the World Health Organization as a priority fungal pathogen. Deciphering the molecular mechanisms and evolutionary paths of its stress tolerance is key for managing risks and exploiting the benefits of this species. Here, researchers report the potential mechanisms driving the adaptive evolution of *I. orientalis* to various stressors using population genomics and machine learning approaches.

This research group conducted whole-genome sequencing of 170 *I. orientalis* isolates and assessed the growth of 161 isolates under 57 different stressors, including heat, low pH, organic acids, lignocellulosic inhibitors, and fungicides from different families (e.g.,

Azoles, Polyenes, and Echinocandins). The group also developed a machine-learning-assisted analysis pipeline, Machine-Learning-Assisted Engineering of Stress-Tolerance Rational Optimization (MAESTRO), to streamline the analysis. MAESTRO revealed that pleiotropic effects of copy-number variations (CNVs) among a small set of genes (less than 3.5%) play a significant role in *I. orientalis'* stress tolerance. Using CNVs as features, the team successfully correlated genetic variation with phenotypic variation of stress tolerance, demonstrating a median R_2 of 0.67 and a median Pearson's correlation of 0.92 across 57 stress conditions when comparing actual fitness to predicted fitness values. Notably, many of these genes (17 to 23%) were tandem repeat paralogs (TRPs), a genomic configuration known as hot spots for gene amplification, reduction, or even shuffling to invent new activities. Additionally, TRPs were significantly enriched with transporters (52%, compared to the genome-wide average of 3%), likely composing the resistome network. Finally, as a proof of concept, the team engineered a strain with enhanced tolerance to the lignocellulosic inhibitor 5-hydroxymethyl furfural but lower resistance to the fungicide fluconazole by deleting four TRPs. This work demonstrates the potential of leveraging fungal genetic variation to predict their potential risks in society and designing more robust industrial strains to develop a sustainable bioeconomy.

Funding Statement: This work was supported by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (DOE, Office of Science contract DE-SC0018420 and DE-AC02-05CH1123) and Biosystems Design program (DOE, Office of Science contract DE-SC0018260 and DE-AC02-05CH1123). The work (proposal: 10.46936/10.25585/60001019) conducted by the DOE Joint Genome Institute (ror.org/04xm1d337), a DOE Office of Science user facility, is supported by the DOE Office of Science operated under contract DE-AC02-05CH11231. Y.-P.L. is supported by the Taiwan NSTC Young Scholar Fellowship Einstein Program (111-2636-E-002-025). Research in the Hittinger Lab is supported by National Science Foundation under grant DEB-2110403, the USDA National Institute of Food and Agriculture (Hatch Project 1020204), in part by the Great Lakes Bioenergy Research Center (DOE BER Office of Science DE–SC0018409), and an H. I. Romnes Faculty Fellowship, supported by the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. These researchers are grateful that the U.S. Department of Agriculture, Agricultural Research Service Culture Collection (Northern Regional Research Laboratory) Database provided them with nearly 60 strains free of charge.

Artificial Intelligence for Image-Based Plant and Microbial Phenotyping

John Lagergren[1,2*] (lagergrenjh@ornl.gov), Larry York[1,2], Mirko Pavicic[1,2], Hari Chhetri[1,2], Daniel Jacobson[1,2], Tomas Rush[1], Wellington Muchero[1,2], David Weston[1], Dale Pelletier[1], Mitchel Doktycz[1], **Gerald Tuskan[1,2]** (PI)

[1]Oak Ridge National Laboratory; [2]Center for Bioenergy Innovation

See abstract, p. 23.

A Prompt Engineering Approach for Root Confocal Image Segmentation Using Segment-Anything-Model

Song Li[1*] (songli@vt.edu), Upasana Sivaramakrishnan[1], Andrea Ramirez[2], Sanchari Kundu[1], **Jose Dinneny[2]** (PI)

[1]Virginia Tech; [2]Stanford University

Project Goals: Establishing a digital anatomical atlas for roots of 11 members of the Brassicaceae family to inform the understanding of gene function and connection between genotype and phenotype. The long-term goal is to develop stress tolerant oilseed crops to advance a sustainable production of biofuel.

Comparative anatomical studies of diverse plant species are vital for the understanding of changes in gene functions such as those involved in solute transport and hormone signaling in plant roots. Through the extraction of quantitative phenotypic data of root cells, researchers can further characterize their response to environmental stimuli, facilitating an in-depth characterization of how genes control root cell development. As the first step for comparative anatomical analysis of root cells, accurate segmentation of individual cells is essential to the analysis of whole root traits. Existing software such as PlantSeg and MorphographX utilized neural networks called U-Net for cell wall segmentation. U-Net was a last generation neural network model that requires training with large amount of manually labeled confocal images. It is time consuming to retrain the model in order to adapt to new images. Foundational models like the Segment Anything Model (SAM) hold promise across various domains due to its zero-shot learning capability alongside prompt engineering can reduce the effort and time traditionally consumed in dataset annotation,

facilitating a semi-automated training process. This research evaluates SAM's segmentation capabilities against PlantSeg, a state-of-the-art model for plant cell segmentation. The team found that PlantSeg was able to segment 2,332 plant cells from 20 confocal images of arabidopsis roots. However, 792 such segmentations (34.0% of total segmented cells) were incorrect based on manual inspection. In contrast, SAM model without finetuning (Vanilla SAM, or V-SAM) was able to segment 1,052 cells, with only 7.8% of cells incorrectly segmented. Although V-SAM can only find 68.3% of correct cells found by PlantSeg, this is a surprisingly good performance because V-SAM was never trained on root confocal images. The team further fine-tuned V-SAM with human prompt of ~1,000 cells by drawing rectangle bounding boxes around cells that were not segmented by V-SAM. Note this is a substantially simpler annotation as compared to the required labeling by U-Net, which is to label every pixel of the cell wall from each training image. With the fine-tuned SAM, the team was able to segment 2,885 cells correctly from the 20 confocal images, which is 187% of that obtained by PlantSeg. These findings demonstrate the efficiency of SAM in confocal image segmentation, showcasing its adaptability and performance compared to existing tools. By addressing challenges specific to confocal images, this team's approach offers a robust solution for studying plant structure and dynamics. Overall, this research highlights the potential of foundational models like SAM in specialized domains and underscores the importance of tailored approaches for achieving accurate semantic segmentation in confocal imaging.

Funding Statement: This project is supported by DOE BER, DE-SC0022985.

Evolutionary Flexibility and Rigidity in the Bacterial Methylerythritol Phosphate (MEP) Pathway

Bailey Marshall[1,2*] (bamarshall2@wisc.edu), Kaustubh Amritkar[1], Michael Wolfe[1,2], Betül Kaçar[1], Robert Landick[1,2], **Timothy J. Donohue[1,2]** (PI)

[1]University of Wisconsin–Madison; [2]Great Lakes Bioenergy Research Center

Project Goals: Identify potential alternative routes in bacterial methylerythritol phosphate (MEP) metabolic pathway which is the pathway used to generate high value terpenoid products.

Terpenoids are a diverse class of compounds with wide-ranging uses such as industrial solvents, fragrances, and more. Industrial production of most terpenoids relies on nonrenewable feedstocks making alternative production methods desirable. Fermentation of engineered microbes using renewable feedstocks like lignocellulose is an attractive strategy for large-scale production of key terpenoids because it has the potential to be sustainable and relatively inexpensive. To achieve large-scale production of terpenoids, there are widespread efforts to engineer the metabolic pathway that generates terpenoids. All terpenoids are made from the final products of the methylerythritol phosphate (MEP) pathway, which is composed of seven enzymatic steps. Efforts in engineering the MEP pathway have identified some of these enzymes as having unfavorable characteristics, so this research group is interested in identifying alternative enzymatic routes which may have evolved that are functionally redundant to the canonical MEP pathway. The group used comparative genomics to search for alternative enzymes to the canonical MEP enzymes and found that enzymes early in the pathway likely evolved alternatives as supported by literature. In contrast, researchers found enzymes late in the pathway appear to have no alternatives in the database of 4,400 genomes in this study. Early pathway flexibility suggests that researchers may be able to identify the genes responsible for an incomplete canonical pathway and implement these alternatives enzymes in metabolic engineering should they have more favorable qualities. For the late pathway steps, if alternative enzymes have evolved at all, they are rare or their host organisms have not been sequenced. The ever-growing repository of sequenced bacterial genomes has great potential to provide metabolic engineers with alternative metabolic pathway solutions. This team's finding that late MEP pathway enzymes are evolutionarily

indispensable informs both metabolic engineering efforts and understanding of the evolution of terpenoid biosynthesis pathways.

Funding Statement: This material is based upon work supported by the Great Lakes Bioenergy Research Center of DOE's Office of Science BER program under award number DE-SC0018409.

Extracting Switchgrass Features Through Minirhizotron and Hyperspectral Image Processing

Mira Saldanha[1]* (mirasaldanha@ufl.edu), Jeffrey Drew[1]* (jeffreydrew@ufl.edu), Justin Rossiter[1], Jason Bonnette[2], Felix Fritschi[3], Alina Zare[1], **Thomas Juenger[2]** (PI)

[1]University of Florida; [2]University of Texas–Austin; [3]University of Missouri

sites.cns.utexas.edu/juenger_lab/switchgrass

Project Goals: This research team's goal is to develop computer vision software pipelines for efficient analysis of minirhizotron and hyperspectral images of switchgrass.

Both minirhizotrons and unmanned aerial vehicles (UAVs) can provide a massive amount of image data on plants like switchgrass (*Panicum virgatum*), a potential source of biofuel. However, manual analysis of these images is time-consuming. This team focuses on developing computer vision software pipelines to segment the images, or classify pixels based on their respective imagery, and accurately quantify data for these segmentations and raw images. Minirhizotrons allows researchers to track the growth of the same plant roots over time. For any automated analysis, it will be necessary to align the images so that calculations from pixel differences are accurate. The team uses the Binary Robust Invariant Scalable Keypoints (Leutenegger et al. 2011) algorithm for feature detection, and then RANSAC to calculate a homography between matched points, to align two images from different dates. Minirhizotrons provide eight images, in color, of the roots at different depths. Team members experiment with aligning raw minirhizotron images versus segmented images, where roots have been identified, and with aligning each level separately versus all at once, when images have been stitched together. Additionally, researchers demonstrate an analysis pipeline for UAV hyperspectral data of individual switchgrass plants. This pipeline produces both individual segmentations of switchgrass plants, and extractions of vegetation indices for the respective plants using the hyperspectral data and segmentations. The team first obtains segmentations of individual plants by using a combination of Sparsity Promoting Iterated Constrained Endmembers (SPICE) (Zare and Gader 2007), and manually inputting key-points for each plant. Team members then apply the watershed algorithm to assign boundary labels for each plant from the binary SPICE output. This minirhizotron analysis pipeline is able to identify a change in biomass over time, and alignment results are promising. The team's hyperspectral analysis pipeline calculates all vegetation indices after performing radiance, reflectance, and orthorectification processing and stitching together all data cubes. Team members' combined aim with this research is to expedite data collection and analysis for biologists.

Funding Statement: This research was supported by the Office of Science (BER program), grant DE-SC0021126.

References

Leutenegger, S., et al. 2011. "BRISK: Binary Robust Invariant Scalable Keypoints," 2011 International Conference on Computer Vision, 2548–55. DOI:10.1109/ICCV.2011.6126542.

Zare, A., and P. Gader. 2007. "SPICE: A Sparsity Promoting Iterated Constrained Endmember Extraction Algorithm with Applications to Landmine Detection from Hyperspectral Imagery," Proc. SPIE 6553, *Detection and Remediation Technologies for Mines and Minelike Targets XII.* DOI:10.1117/12.722595.

Designing Novel Enzymes for Complete Degradation of Recalcitrant Polyamides

Amandeep Sangha* (amandeep.sangha@arzeda.com), **Alexandre Zanghellini** (PI)

Arzeda

As of 2015, a total of 6.3 billion tons of plastic waste had been generated globally. It is estimated that only 9% of this total has been recycled, while 12% has been incinerated to recover energy values, and the remainder has entered landfills. New technologies are needed to address this ever-growing problem. An alternative approach, harnessing the power of biology to not just depolymerize plastics back to their monomer precursors but convert them into higher-value products, offers stronger economic incentives and in turn would be expected to drive more rapid and widespread adoption. Toward that end, this team's work focuses

on combining cutting-edge computational protein design and synthetic biology to address the challenge of complete biodegradation and upcycling of the recalcitrant polymers nylon 6 and nylon 66. Although natural enzymes have been shown to be able to degrade amorphous portions of polyamides such as nylon 6 and nylon 66, complete enzymatic degradation has not been demonstrated. These researchers hypothesize this to be due in large part to a lack of natural enzymes able to efficiently catalyze degradation of the crystalline portion of the polymer. To alleviate this limitation, these researchers are using a combination of physics-based and generative deep learning-based protein design methods to engineering new and improved enzymes, with optimized active sites for binding and hydrolysis of polyamides. In conjunction, the researchers are screening and engineering bacterial strains able to metabolize nylon 6 and nylon 66 degradation byproducts directly into central metabolism. Such platform strains can be used to produce a wide variety of fermentation products from central metabolites. Integration of nylon 6 and nylon 66 depolymerizing enzymes into these engineered hosts will provide a novel, elegant, and cost-effective consolidated fermentation process for nylon upcycling to higher-value sustainable materials.

Artificial Intelligence Foundation Models for Understanding Cellular Responses to Radiation Exposure

Abraham Stroka* (astroka@anl.gov), Rebecca Weinberg, Sara Forrester, Casey Stone, Rafael Vescovi, Dan Schabacker, Thomas S. Brettin, Arvind Ramanathan, **Rick Stevens** (PI)

Argonne National Laboratory

The use of image classification machine learning models has the potential for great impact on the speed and accuracy of medical diagnoses. The ability to accurately identify genetic perturbations based on cellular morphology would be crucial to the medical field. A large amount of research has been conducted on the effects of acute, high-dose radiation on the morphological profile of human cells. However, the effects of low-dose radiation on cellular morphology have yet to be investigated. For the purposes of this study, researchers focused on the analysis of human umbilical vein endothelial cells (HUVEC) that have undergone low-dose radiation exposure. If phenotypic features of the HUVEC cell's morphology can be identified using this model, it could lead to advanced and more efficient screening for low-dose radiation exposure. The team implemented a vision transformer pipeline as the image classification model for this study, specifically the mura vision transformer, which has shown reliable validation accuracy. In order to train this model, the team implemented a pipeline that utilizes the CellProfiler cell image analysis software to perform cell segmentation on HUVEC cell painting images. The CellProfiler pipeline the team developed allows researchers to stack the several channels of cell painting images and export the images of each individual cell into the vision transformer.

Molecular and Cellular Responses of Human Endothelial Cells to Low-Dose Radiation

Rebecca Weinberg* (rweinberg@anl.gov), Sara Forrester, Justin Podowski, Abraham Stroka, Thomas S. Brettin, Dan Schabacker, **Rick Stevens** (PI)

Argonne National Laboratory

Radiation exposure has a wide spectrum of impacts on human health, notably in carcinogenesis, but also in neurological and cardiovascular disorders. While acute toxicity from high doses of radiation is well characterized, understanding the range of outcomes following exposure to low-dose radiation is more challenging. This research team is establishing new experimental workflows that will enable high throughput experiments across molecular and cellular scales to facilitate more comprehensive modeling. In the pilot study, a monolayer of human umbilical vessel endothelial cells (HUVECs) was exposed to a point source of ^{137}Cs at a low-dose rate (6 milligrays/hour). Cells were exposed for one week in culture (a total dose of 1,008 milligrays), and then harvested for RNA or replated for Cell Painting staining. Cell Painting is a streamlined multiparameter approach to fluorescence microscopy that provides rich feature data of cell structure and function.

A major advantage of Cell Painting is a robust, publicly available dataset spanning thousands of small molecular and genomic perturbations produced by a collaborative team (the JUMP Consortium). The scale of characterized phenotypes has facilitated development of predictive models that incorporate chemical structural information, biological mechanism of action, and gene expression, which this team will expand into the realm of radiation exposure. With Cell Painting, features can be extracted based on staining of nuclear, endoplasmic reticulum, plasma membrane and Golgi, actin, nucleoli, and mitochondria. Principle component

analysis of control and irradiated cells provided a proof-of-principle demonstration that Cell Painting enables detection of features impacted by irradiation. The team's transcriptome analysis revealed that in endothelial cells, radiation robustly induced cell response pathways integral to cytokine and chemokine pathways, such as the Tumor Necrosis Factor pathway. Underscoring the relevancy of HUVECs to cardiovascular disease, pathways associated with "lipid" and "atherosclerosis" were also activated. Two Kyoto Encyclopedia of Genes and Genomes terms shed light on the molecular mechanisms of these processes, namely the HIF-1 and NF-kappa B signaling pathways.

To compare these results to previously obtained studies of low-dose radiation exposure, the team compared its data with gene expression datasets from the RadBioBase, a publicly available comprehensive transcriptome repository of irradiated mammalian samples. Researchers selected datasets that used human cells and doses below 0.5 grays to identify 235 genes impacted by radiation across four published datasets. Of these, 35 genes were also seen in this study's data, notably the inflammatory cytokines IL6 and IL1B, as well as the genes PTGS2 (COX2) and CXCL12, which are involved in inflammatory processes underlying cardiovascular disease.

To overcome the limitations (variable dose field, high activity) of the point radiation source in the pilot study, a major goal of the next phase of this project is to prototype and deploy new source geometries (96-well plate format) for high-throughput experimental exposures. New source geometries will require minimal activity, provide uniform dose fields, and allow for multiple dose rate exposures in parallel. Researchers will then assess the impact of low-dose radiation harnessing molecular (multiomic) and cellular (Cell Painting) assays that can be used to develop advanced multiscale models of the impacts of low-dose radiation.

CHAPTER 12 | POSTERS

Computational Infrastructure and Resources for Omics Technologies

Supporting Continental Scale Research in the National Microbiome Data Collaborative Data Portal

Alicia Clum[1]* (aclum@lbl.gov), Anastasiya Prymolenna[2], Antonio Camargo[1], Bea Meluch[2], Bin Hu[3], Brynn Zalmanek[2], Cameron Giberson[2], Camilo Posso[2], Chien-Chi Lo[3], Chris Mungall[1], Donny Winston[5], Eric Cavanna[1], Francie Rodriguez[3], Grant Fujimoto[2], James Tessmer[2], Jeff Baumes[4], Jing Cao[5], Julia Kelliher[3], Kaitlyn Li[3], Katherine Heal[2], Kaylee Kudish[3], Kevin Fox[2], Kjiersten Fagnan[1], Leah Johnson[3], Lee Ann McCue[2], Mark Flynn[3], Mark Miller[1], Mary Salvi[4],Michael Thornton[1], Michal Babinski[3], Migun Shakya[3], Mike Nagler[4], Montana Smith[2], Patrick Chain[3], Patrick Kalita[1], Paul Piehowski[2], Po-E (Paul) Li[3], Samuel Purvine[2], Set Sarrafan[1], Shalki Shrivastava[1], Shane Canon[1], Shreyas Cholia[1], Sierra Moxon[1], Simon Roux[1], Sujay Patil[1], Wendi Lynch[1], Yan Xu[3], Yuri Corilo[2], **Emiley Eloe-Fadrosh[1]** (PI)

[1]Lawrence Berkeley National Laboratory; [2]Pacific Northwest National Laboratory; [3]Los Alamos National Laboratory; [4]Kitware Inc; [5]Polyneme LLC

Continental scale research is important to understand global processes such as climate change and ecosystem dynamics and can be used to identify patterns and trends including spatial variations and temporal trends. The National Microbiome Data Collaborative (NMDC) Data Portal (Eloe-Fadrosh et al. 2022) has sample and processing information as well as standardized workflow results for several continental scale datasets of high value to the research community. To support a continental-scale understanding of terrestrial ecosystems, the NMDC hosts soil data from National Ecological Observatory Network (NEON) sites, Environmental Molecular Sciences Laboratory's (EMSL) 1,000 Soils Research Campaign, and the Earth Microbiome Project 500 (EMP500). To enable continental-scale research of aquatic ecosystems, the NMDC hosts freshwater and benthic data from NEON sites and freshwater samples used to generate the Genome Resolved Open Watershed (GROW) database. These efforts all leverage standardized and documented sampling protocols, enabling comparisons of datasets across sites.

The Data Portal is focused on making multiomics datasets more findable to enable data reuse. It provides search tools to find information by principal investigator or study name, by sample information like geographic location, collection date, and depth, or by information about how the omics data was generated. Additionally, data can be searched by Kyoto Encyclopedia of Genes and Genomes (KEGG) terms to identify samples by molecular function. Once datasets have been identified, standardized workflow results can be downloaded in bulk via the website.

Reference

Eloe-Fadrosh, E. A., et al. 2022. "The National Microbiome Data Collaborative Data Portal: An Integrated Multi-Omics Microbiome Data Resource," *Nucleic Acids Research* **50**(D1), D828–36. DOI:10.1093/nar/gkab990.

Community Engagement and User-Centered Design Underpin the Product Development of the National Microbiome Data Collaborative (NMDC)

Julia Kelliher[3*] (jkelliher@lanl.gov), Alicia Clum[1], Anastasiya Prymolenna[2], Antonio Camargo[1], Bea Meluch[2], Bin Hu[3], Brynn Zalmanek[2], Cameron Giberson[2], Camilo Posso[2], Chien-Chi Lo[3], Chris Mungall[1], Donny Winston[5], Eric Cavanna[1], Francie Rodriguez[3], Grant Fujimoto[2], James Tessmer[2], Jeff Baumes[4], Jing Cao[5], Kaitlyn Li[3], Katherine Heal[2], Kaylee Kudish[3], Kevin Fox[2], Kjiersten Fagnan[1], Leah Johnson[3], Lee Ann McCue[2], Mark Flynn[3], Mark Miller[1], Mary Salvi[4], Michael Thornton[1], Michal Babinski[3], Migun Shakya[3], Mike Nagler[4], Montana Smith[2], Patrick Chain[3], Patrick Kalita[1], Paul Piehowski[2], Po-E (Paul) Li[3], Samuel Purvine[2], Set Sarrafan[1], Shalki Shrivastava[1], Shane Canon[1], Shreyas Cholia[1], Sierra Moxon[1], Simon Roux[1], Sujay Patil[1], Wendi Lynch[1], Yan Xu[3], Yuri Corilo[2],
Emiley Eloe-Fadrosh[1] (PI)

[1]Lawrence Berkeley National Laboratory; [2]Pacific Northwest National Laboratory; [3]Los Alamos National Laboratory; [4]Kitware Inc; [5]Polyneme LLC

microbiomedata.org

The National Microbiome Data Collaborative (NMDC) is a multinational laboratory initiative focused on advancing innovation and discovery in the field of microbiome science through the project's development of products and tools for the environmental microbiome research community (Wood-Charlson et al. 2020). The NMDC provides the community with three products: (1) The Submission Portal, (2) NMDC Empowering the Development of Genomics Expertise (EDGE), and (3) The Data Portal (Eloe-Fadrosh et al. 2022), each aimed at making multiomics microbiome data findable, accessible, interoperable, and reusable (FAIR). The NMDC team utilizes a user-centered design approach through the implementation of insights gleaned from user research, usability testing, and community feedback to continuously improve its products. This team routinely engages with microbiome researchers to discuss how they want the NMDC products to look and operate, as well as understand what new functionality would benefit future research.

The NMDC communicates and engages with many types of stakeholders, including funding agencies, publishers, institutions, programs, projects, and individual scientists. As part of these collaborative efforts, the NMDC hosts and co-hosts workshops (Vangay et al. 2021), webinars, presentations, panel discussions, and other events aimed at spreading awareness of and lowering barriers to adoption of FAIR principles in microbiome research and data generation. The NMDC Ambassador Program (microbiomedata.org/ambassadors) allows early career researchers to host some of these events, thus expanding the overall reach of the content and training materials, while providing the Ambassadors with valuable experiences and career opportunities. The NMDC Champions Program (microbiomedata.org/community/championsprogram) brings together microbiome researchers from diverse backgrounds to contribute to the NMDC (e.g., by beta-testing the NMDC products, co-authoring publications with the NMDC team, providing feedback). The NMDC will continue to prioritize community engagement as its products and network grow. The NMDC engagement strategy focuses on promoting a collaborative ecosystem for diverse microbiome researchers and implementing community feedback in all of the NMDC efforts and products.

References

Eloe-Fadrosh, E. A., et al. 2022. "The National Microbiome Data Collaborative Data Portal: An Integrated Multi-Omics Microbiome Data Resource," *Nucleic Acids Research* **50**(D1), D828–36. DOI:10.1093/nar/gkab990.

Vangay, P., et al. 2021. "Microbiome Metadata Standards: Report of the National Microbiome Data Collaborative's Workshop and Follow-On Activities," *mSystems* **6**, e01194–20. DOI:10.1128/mSystems.01194-20.

Wood-Charlson, E. M., et al. 2020. "The National Microbiome Data Collaborative: Enabling Microbiome Science," *Nature Reviews Microbiology* **18**(6), 313–4. DOI:10.1038/s41579-020-0377-0.

Determination of Metabolic Fluxes by Multi-Isotope Tracing and Machine Learning

Richard C. Law[1*] (richardlaw@g.ucla.edu), Glenn Nurwono[1], Samantha O'Keeffe[1], Rachel Ki[1], Aliya Lakhani[1], Pin-Kuang Lai[2], **Junyoung O. Park[1]** (PI)

[1]University of California–Los Angeles; [2]Stevens Institute of Technology

Metabolic fluxes are a fundamental descriptor of cellular state, representing the rates at which organisms operate metabolic pathways. Mass spectrometry and isotope tracing have been instrumental in quantifying fluxes, as metabolic pathways imprint unique isotope labeling patterns on metabolites corresponding to their

fluxes. Metabolic flux analysis (MFA) is a commonly used computational framework that identifies the set of fluxes that best simulate observed isotope labeling patterns. However, quantitative flux analysis remains an expert method, and the relationships between isotopic labeling patterns and fluxes remain elusive in complex metabolic environments.

Here, researchers aimed to quantify fluxes in dynamic and complex biological systems including microbial communities. Using multiple isotope tracers, the group elucidated the evolutionary benefit of the Entner-Doudoroff (ED) pathway, which is parallel to textbook (EMP) glycolysis. Tracing from two asymmetrically labeled glucose on a minutes timescale revealed that the ED pathway flux accelerates faster than the textbook glycolysis in response to nutrient upshift. The rapid utilization of the ED pathway endows *Escherichia coli* cells with rapid adaptability and evolutionary benefits in a microbial community during intermittent nutrient supply. Additionally, to make flux quantitation tools more scalable and accessible, the group innovated a two-stage machine learning (ML) framework termed ML-Flux. ML-Flux is trained using data from five universal models of central carbon metabolism and 26 different carbon-13 (^{13}C) and dihydrogen (^{2}H) glucose and glutamine tracers to convert isotope labeling patterns into metabolic fluxes. Using ML-Flux with multi-isotope tracing, the group determined fluxes and free energies through central carbon metabolism at orders-of-magnitude faster speeds than traditional MFA. Taken together, dynamic multi-isotope tracing identified the role of parallel pathways in balancing metabolic stability and adaptability as a key design principle. ML-assisted multi-isotope tracing is a promising step toward making flux quantitation in complex biological systems increasingly accessible and expanding understanding and control of metabolism.

Illuminating Novel Terpenoid Biosynthesis Pathways in *Yarrowia lipolytica* by Metabolomics

Glenn Nurwono* (nurwono@ucla.edu), Zuodong Sun, Edward Hu, Yi Tang, **Junyoung O. Park** (PI)

University of California–Los Angeles

Yarrowia lipolytica is an emerging microbial host for the bioconversion of low-value carbon into natural products, but its endogenous terpenoid metabolism has yet to be fully mapped. Here, this research group aimed to illuminate novel terpenoid biosynthetic pathways and quantify metabolic flux and free energy therein by employing metabolomics, isotope tracing, and genetic engineering. The group engineered a strain to push increased carbon flux through the mevalonate pathway and to farnesyl pyrophosphate (FPP) by overexpression of 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) and FPP synthase (FPPS). Overexpression of HMGR and FPPS led to a 150-fold increase in mevalonate production and a 1.5-fold increase in isopentenyl diphosphate (IPP) production after one day of growth, indicative of increased metabolic activity through the mevalonate pathway and terpenoid metabolism. However, the group observed a lower amount of IPP during the second and third days, suggesting the activation of secondary metabolism and prompting an investigation how the isoprenoid backbone was being utilized. Upon untargeted metabolomic analysis using liquid chromatography-mass spectrometry (LC-MS), researchers discovered several new metabolites being produced in the engineered strain but absent in the wild-type strain. Based on measured monoisotopic mass-to-charge ratios and proposed molecular formulas, the group hypothesized that these molecules were oxygenated terpenoids. After compound purification and nuclear magnetic resonance (NMR) spectroscopy, the group confirmed that these compounds were terpenoids, with a farnesyl scaffold and bifunctionalized with carboxylic acids. To the group's knowledge, this is the first observed biosynthesis of such diacid compounds. To map the novel terpenoid biosynthetic pathway, the group reconstituted the putative enzymatic steps in *Saccharomyces cerevisiae* and successfully conferred full biosynthetic capabilities. Furthermore, isotope tracing and direct farnesol feeding were utilized to elucidate biosynthetic intermediates. Notably, a P450 enzyme previously shown to be involved in alkane assimilation was responsible for the hydroxylation of the allylic carbon-hydrogen bond, demonstrating the substrate promiscuity and multifunctionality of involved enzymes. This work demonstrates the utility of increasing precursor availability to activate untapped metabolic pathways for the discovery of new natural products. Furthermore, the new compounds and their biosynthetic intermediates represent an exciting pool of organic building blocks that can be accessed for renewable fuel, polymer, and natural product synthesis.

CHAPTER 13 | POSTERS
Quantum Imaging and Sensing

Novel Quantum Sensing Tools for the Rhizosphere

Ashok Ajoy[1,2*] (ashokaj@berkeley.edu, PI) , Brooke Newell[1,2], Zhao Hao[2], Benjamin Gilbert[2]

[1]University of California–Berkeley; [2]Lawrence Berkeley National Laboratory

Project Goals: Development of quantum sensing tools for chemical analytes of relevance to the rhizosphere.

This poster outlines the collaborative efforts between University of California–Berkeley (UCB) and Lawrence Berkeley National Laboratory (LBNL) towards innovating quantum sensing technologies for the rhizosphere. Work at UCB is bifurcated into two main streams: firstly, the development of novel quantum sensor probes utilizing hyperpolarized carbon-13 (^{13}C) nuclei in nanodiamonds. These probes, capitalizing on the unique properties of ^{13}C nuclear spins, act as sensitive magnetometers for time-varying fields under strong bias magnetic fields. Group members report on protocols for their application as nuclear magnetic resonance (NMR) sensors, providing unprecedented chemical and spatial resolution. Notably, these sensors exhibit exceptionally long coherence times, surpassing T2'=800s at 100 K, while inherently filtering out common-mode instrumental noise. Secondly, the group is advancing a new technology platform for detecting rhizosphere-specific analytes using nanodiamonds embedded in monodisperse, picoliter-volume microdroplets. This approach aims to encapsulate and sense chemical analytes from the rhizosphere efficiently within a rapidly flowing system. This research shows chemical sensing for model target paramagnetic analytes with an excellent limit of detection (~100 nm).

At LBNL, group members discovered that nitrogen vacancy relaxometry can provide a new contrast mechanism for plant tissue imaging. Researchers observed gradients in the relaxation variables at subcellular length-scales and are working to understand the underlying processes or species that couple to the quantum sensing centers. Currently, the team is working on enhancing the signal-to-noise ratio through surface preparation and various instrumentation improvements. Simultaneously, the team is developing complementary approaches to enable biologically relevant characterization under the existing quantum sensing microscope. Additionally, researchers are constructing novel imaging and data analysis capabilities, e.g., using hydrogen relaxometry, to study water movements within biological systems.

Funding Statement: The research team gratefully acknowledges funding from DOE BER DE-SC0023065.

Quantum Optical Microscopy of Biomolecules near Interfaces and Surfaces (QuOMBIS)

Mikael Backlund* (mikaelb@illinois.edu, PI), Simeon Bogdanov, Paul Kwiat, Virginia Lorenz

University of Illinois Urbana-Champaign

Project Goals: Group members are working to develop three complementary microscopy techniques that exploit quantum correlations in light: Hong-Ou-Mandel interferometric tomography, $g^{(2)}$ correlation function imaging, and passive and active transverse mode sorting. Researchers will subsequently incorporate these methods into a single platform for tracking and imaging individual and few fluorescently labeled biomolecules, including cellulases, in the context of nearby biological interfaces and surfaces in order to unravel the fundamental processes involved in the conversion of lignocellulosic biomass into renewable fuels.

Since the publication of Hooke's *Micrographia* in 1665, the scientific disciplines of light microscopy and (sub) cellular biology have progressed in lockstep with one another. Advances in the spatial and temporal resolution, specificity, and sensitivity of optical methods have continually led to new capabilities and insights in biological imaging. The pace of this evolution has quickened in the past century, as a mastery of the physics of light according to Maxwell's equations has been wielded to more fully exploit classical effects like interference and diffraction. As the classical limits of light microscopy near saturation, however, sustained improvement in bioimaging technology is ultimately untenable without a more fundamental shift in research

direction. Just as the field of quantum computing has gained prominence in anticipation of the inevitable breakdown of Moore's Law, quantum-enabled light microscopy will likely provide the path forward for (sub)cellular biological imaging.

This team aims to help lead this effort by developing three complementary quantum microscopy modalities that each address a different challenge inherent to (sub) cellular microscopy:

1. Hong-Ou-Mandel Interference Microscopy to enable loss- and noise-tolerant depth imaging with exquisite resolution;
2. $g^{(2)}$ Microscopy to facilitate orders-of-magnitude sensitivity improvement in focusing and tracking single quantum emitters atop oppressive classical backgrounds at reduced excitation powers; and
3. Transverse Mode Sorting Microscopy to enable super-resolution microscopy at low excitation powers and high temporal resolution.

This poster presented results demonstrating progress in developing these constituent techniques. The team will ultimately incorporate them into a common imaging platform that can provide access to the many scales of interest in energy-relevant plant and microbial biology. The combined technique, QuOMBIS (Quantum Optical Microscopy of Biomolecules near Interfaces and Surfaces), will be especially powerful for tracking and imaging individual and few fluorescently labeled biomolecules in the context of nearby biological interfaces and surfaces. Upon development of the methods, the team will apply the platform to unravel and harness the enzymatic conversion of biomass into renewable fuels.

Funding Statement: This research was supported by the DOE Office of Science, BER program grant DE-SC0023167.

Multi-Pass Stimulated Raman Scattering Microscopy

Shaun C. Burd[1]* (scburd@stanford.edu), Joshua Reynolds[1], Tzu-Chieh Yen[1], Dara Dowlatshahi[1,2], Samsuzzoha Mondal[1,2], Soichi Wakatsuki[1,2], **Mark Kasevich[1]** (PI)

[1]Stanford University; [2]SLAC National Accelerator Laboratory

web.stanford.edu/group/kasevich/cgi-bin

These researchers present results from two multi-pass microscopes that provide metrological advantages in stimulated Raman scattering (SRS) microscopy. SRS is a nonlinear optical process that is quantitative, bond-specific, and label-free. However, for fragile specimens such as live cells, the required optical intensities result in specimen damage or death, and photon shot noise limits the absolute achievable sensitivity for these dynamic and dose-sensitive samples. A multi-pass microscope interrogates a sample sequentially with the same probe field in a programmable and deterministic fashion, increasing the sensitivity of measurements of weak scatterers (Juffmann et al. 2016; Israel et al. 2023). In such cases, multi-pass measurements are competitive with or can outperform measurements using squeezed or other quantum states using purely classical resources (Giovannetti et al. 2006). This research demonstrates multi-passed SRS microscopy using both infrared and visible light modalities. The team quantified the metrological advantage in comparison with a conventional measurement scheme. These quantum-optimal imaging protocols will advance microscopy and flow cytometry for studying the life cycles and interactions of soil microbes and plants and can be shared with the BER science community.

References

Giovannetti, V., et al. 2006. "Quantum Metrology," *Physical Review Letters* **96**(1), 010401. DOI:10.1103/PhysRevLett.96.010401.

Israel, Y., et al. 2023. "Continuous Wave Multi-Pass Imaging Flow Cytometry," *Optica* **10**(4), 491–6. DOI:10.1364/OPTICA.482316.

Juffmann, T., et al. 2016. "Multi-Pass Microscopy," *Nature Communications* **7**, 12858. DOI:10.1038/ncomms12858.

Probing Photoreception with New Quantum-Enabled Imaging

James E. Evans[1]* (james.evans@pnnl.gov, PI), Kevin Crampton[1], Samantha Powell[1], Jory Brookreson[1], Nick Black[2], Saleem Iqbal[2], Patrick El-Khoury[1], Robert Boyd[2]

[1]Pacific Northwest National Laboratory; [2]University of Rochester

See abstract, p. 18.

The 3DQ Microscope: A Novel System Using Entangled Photons to Generate Volumetric Fluorescence and Scattering Images for Bioenergy Applications

Audrey Eshun[1]* (eshun1@llnl.gov), Dominique Davenport[1], Brandon Demory[1], Paul Mos[2], Yang Lin[2], Sam Jeppson[1], Ashleigh Wilson[1], Shervin Kiannejad[1], Tiziana Bond[1], Mike Rushford[1], Chuck Boley[1], Edoardo Charbon[2], Claudio Bruschini[2], **Ted Laurence[1]** (PI)

[1]Lawrence Livermore National Lab; [2]École Polytechnique Fédérale de Lausanne

See abstract, p. 16.

Progress Towards a Quantum-Enhanced X-Ray Microscope

Justin C. Goodrich[1]* (jgoodrich@bnl.gov), Ryan Mahon[1], Joseph Hanrahan[1], Lonny Berman[1], Andrei Fluerasu[1], Monika Dziubelski[1], Raphael A. Abrahao[1], Sanjit Karmakar[2], Cinzia DaVià[2], Andrei Nomerotski[3], **Sean McSweeney[1]** (PI)

[1]Brookhaven National Laboratory; [2]Stony Brook University; [3]Florida International University

See abstract, p. 12.

Biological Imaging Using Entangled Photons

T. Goodson III[1]* (tgoodson@umich.edu, PI), George C. Schatz[2], Jennifer Morrell-Falvey[3], Ying-Zhong Ma[3]

[1]University of Michigan; [2]Northwestern University; [3]Oak Ridge National Laboratory

See abstract, p. 10.

Squeezed-Light Multimodal Nonlinear Optical Imaging of Microbes

Ralph Jimenez[1,2,3]* (rjimenez@jila.colorado.edu, PI), Miles SanSoucie[1,3], Ying-Zhong Ma[4], Jennifer L. Morrell-Falvey[4]

[1]University of Colorado; [2]National Institute of Standards and Technologies; [3]University of Colorado; [4]Oak Ridge National Laboratory

jila.colorado.edu/jimenez

Project Goals: The overarching goal of this project is to develop multimodal quantum nonlinear optical imaging based on a squeezed light source for co-registered, steady-state two-photon excited fluorescence, two-photon-excited fluorescence lifetime imaging, and second harmonic generation microscopies. The capabilities and advantages of these quantum light modalities will be validated through imaging the growth and dynamics of bacterial strains with intrinsic fluorescence or strains expressing fluorescent proteins in a synthetic microbial community or during plant colonization.

The team's initial efforts focused on whether phenazines produced by *Pseudomonas* sp. could be used as a biomarker for live cell imaging. To this end, extensive steady-state fluorescence spectral measurements on two commercially available phenazines, phenazine-1-carboxylic acid and pyocyanine, were performed independently at JILA and Oak Ridge National Laboratory (ORNL). Stable fluorescence emission necessary for both two-photon spectral and imaging acquisition could not be established under various reduced conditions using different concentrations of sodium dithionite and incubation times. The findings suggest that use of bacterial phenazines as intrinsic biomarkers for live cell imaging of *Pseudomonas* strains isolated from the rhizosphere is unlikely. The team is now focused on determining the feasibility of using green fluorescent protein for these studies. In parallel, researchers have built and tested a stable squeezed light source capable of producing up to ~3 milliwatts of twin beam power at JILA. The team used an electro-optic modulator and a temperature-stabilized etalon to red-shift the seed beam and filter the undesired frequencies respectively. This simplification resulted in a stable seed beam power (~1% standard deviation) comparable to that of the commonly used acousto-optic modulator and could allow for an easier implementation into instrumentation. A gain of up to ~8.5 has been observed in the amplified probe beam that is also comparable to past four-wave mixing squeezed light experiments. With the range of powers and gains available and recently assembled fluorescence detection system, researchers can now move onto the next stage of determining and optimizing the quantum-enhanced two-photon excitation rates of common fluorescent dyes such as fluorescein and rhodamine b. This same design will be used for building a second light source at ORNL. In the meantime, a two-photon fluorescence spectral system based on signal photon detection was built at ORNL and will be used to validate the linear intensity dependence using a recently developed squeezed light source in the Materials Science and Technology Division at ORNL while the dedicated light source is being built for this project. These instruments will be used to evaluate photoinduced

stress and toxicity in microbial communities resulting from the squeezed light source versus classical light irradiation to assess whether the predicted quantum advantage is realized.

Funding Statement: This research was supported by the DOE Office of Science, BER program, grant DE-SC0023538. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for DOE under contract DE-AC05-00OR22725. This program is supported by the DOE Office of Science through the Genomic Science program, BER program, under FWP ERKPA62.

Fluorescence Lifetime-Based Imaging of *Bacillus subtilis* Membrane Potential

Debjit Roy[1]* (deb9804@g.ucla.edu), Xavier Michalet[2], Kiran Bharadwaj[1], Evan W. Miller[3], **Shimon Weiss[1,4]** (PI)

[1]UCLA-DOE Institute for Genomics and Proteomics; [2]University of California–Los Angeles; [3]University of California–Berkeley; [4]Bar-Ilan University

Membrane potential (MP) changes can provide a simple readout of bacterial functional and metabolic state or stress levels. While several optical methods exist for measuring fast changes in MP in excitable cells, there is a dearth of such methods for precise (and calibrated) measurements of steady-state MPs in bacterial cells. Conventional electrode-based methods for the measurement of MP are not suitable for small bacterial cells. Existing optical electrophysiological techniques based on fluorescent Nernstian probes have been successfully used in many studies, but they do not provide precision or absolute quantification of MP or their changes. This team presents a novel, calibrated MP recording approach to address this gap. This group's method uses (1) a unique optical transducer (a chromophorewire-donor construct), that utilizes intrinsic photoinduced electron transfer (PeT) mechanism to measure MP via its fluorescence lifetime and (2) a quantitative fluorescence lifetime imaging microscopy (FLIM) data analysis based on phasor analysis. In order to visualize individual bacterial cells' MPs under different extracellular conditions, amplitude-averaged lifetime maps were computed from pixel-wise phasor fractions. This allows group members to accurately measure even small MP changes in single bacterial cells. Calibration of membrane potential estimation via phasor-FLIM measurements has been achieved by modulating MP artificially through changing ionic (potassium +) concentration gradients across the membrane utilizing ionophores. Applying this technique to *Bacillus subtilis*, researchers estimated their normal MP at -86 millivolts and a chemically modulated depolarized state at +1 mV. This breakthrough work paves the way for deeper insights into bacterial electrophysiology and bioelectricity research.

Mid-Infrared Single Photon Counting Photodetectors for Quantum Biosensing

Leon Shterengas* (leon.shterengas@stonybrook.edu, PI), Egor Portyankin, Gela Kipshidze, Jinze Zhao, Dmitri Donetski

State University of New York at Stony Brook

The overall project goal is to develop single photon counting avalanche photodiodes (SPAD) operating in mid-infrared spectral range above 3 micrometers. Application of the novel devices as a bucket detector in mid-infrared quantum ghost imaging of biological tissue is envisioned. The team designed and fabricated GaSb-based separate absorption, charge, multiplication (SACM) heterostructures optimized for hole-initiated impact ionization. The devices were grown onto tellurium-doped GaSb substrates and contained 1 μm–thick nominally undoped $InAs_{0.91}Sb_{0.09}$ absorber, ~100 nm–thick tellurium-doped $Al_{0.9}Ga_{0.1}As_{0.07}Sb_{0.93}$ and 300 nm–thick nominally undoped $Al_{0.9}Ga_{0.1}As_{0.07}Sb_{0.93}$ multiplier layer terminated with ~300 nm–thick phosphorous-doped contact layers. The epitaxial wafers were processed into 40 and 80 diameter shallow etched mesa devices with, correspondingly, 20 and 60 μm diameter windows in top contact metallization. The SACM APD were indium-soldered epi-side-up with onto gold-plated carriers and characterized in wide temperature range.

The punch-through voltage was about 10 volts at liquid nitrogen temperature; dark and photocurrent current increase due to avalanche breakdown was observed at voltages above 13 V, which was taken as a unity gain voltage reference. The linear regime multiplication gains exceeding 200 were observed at voltages near 17.5 V at liquid nitrogen temperature. The dark current values of several nanoamperes have been recorded for all devices before the breakdown. The analysis of the temperature dependence of the dark current above punch-through confirms diffusion limited absorber operation at temperatures above 150°K (activation energy ~370 megaelectron volts). At temperatures below ~150°K, the dark current became ~nA and its dependence on temperature was characterized by activation energy ~10 millivolts indicating other current controlling mechanisms. The dark current values below

punch-through voltage at temperatures below 200°K remained under 100 picoamperes and were virtually temperature independent for all devices. Independently characterized responsivities values above 5 amperes per watt at bias voltages corresponding to linear gain values below 10. The device cutoff wavelength at liquid nitrogen temperature was ~3.9 μm as determined at the half maximum level. Experiment confirmed efficiency of the proposed device architecture, which does not require etching through absorber section. The observed mismatch of about several volts between punch-through and start of the avalanche breakdown indicates that thickness of the multiplier section and doping level of the charge control section will need to be optimized to further reduce dark current values.

Funding Statement: This research was supported by the DOE Office of Science, BER program, grant DE-SC0023165.

Quantum-Entangled Hyperpolarized Spin States for Non-Invasive Imaging of Nitrogen Assimilation

Thomas Theis* (ttheis@ncsu.edu, PI)

North Carolina State University

See abstract, p. 13.

Quantum Ghost Imaging of Water Content and Plant Health with Entangled Photon Pairs

James Werner[1]* (jwerner@lanl.gov, PI), Duncan Ryan[1], Kristina Meier[1], Kati Seitz[1], David Thompson[1], Rebecca Holmes[1], Demosthenes Morales[1], Buck Hanson[1], David Hanson[2], Peter Goodwin[1], Raymon Newell[1], David Thompson[1]

[1]Los Alamos National Laboratory; [2]University of New Mexico

See abstract, p. 15.

Probing Product Redistribution During Photosynthesis Dark Conditions Using Quantum Imaging with Undetected Photons

Joshua Yuan[1]* (joshua.yuan@wustl.edu, PI), Yining Zeng[2] (yining.zeng@nrel.gov)

[1]Washington University in St. Louis; [2]National Renewable Energy Laboratory

Mitigation of sample photodamage allows for a longer observation window to trace biological processes. In particular, the natural photosynthesis systems that use light to drive chemical reactions are very sensitive to optical imaging, as the probing light itself inevitably perturbs photosynthetic reactions. This challenge has prevented the effective real-time monitoring of photosynthesis reactions in dark conditions, where the chemicals synthesized by light redistribute in the plant. Understanding the mechanism of this redistribution is important to close the circle of knowledge on plant photosynthesis. Direct monitoring of the distribution in dark conditions would provide straightforward evidence but has never been achieved ever due to perturbation from the probing light. To address the above challenge, the team applied quantum imaging with undetected photons (QIUP) to probe the photosynthesis processes using a very low dose of photons. QIUP uses two beams consisting of entangled photons separated into infrared (IR) and visible wavelengths. The IR photons probe the sample and obtain sample information, but they are not detected by the detector due to the very low sensitivity in the IR range. The visible photons do not touch the sample but carry the sample information through entanglement, and they are detected at high sensitivity and produce sample images. This study demonstrated that QIUP can produce sample images by using only picowatts of illumination power on the sample. The power density is many orders of magnitude lower than that of classic light microscopy, for example, confocal fluorescence microscopy. Researchers applied QIUP to image squalene in tobacco leaves by using squalene's IR absorption bands. The illumination power density used in QIUP to achieve similar chemical imaging capability is even more drastically lower than that of stimulated Raman scattering microscopy, which uses a picosecond pulsed laser to induce a high illumination field. The imaging data has informed the metabolic engineering of more efficient squalene production in tobacco.

CHAPTER 14 | POSTERS

Structural Biology

DOE Synchrotron Light Sources, Neutron Scattering and Cryo-Electron Microscopy User Capabilities for Biological Systems Science (Structural Biology and Imaging User Resources)

berstructuralbioportal.org

BER's Biological Systems Science research seeks to understand the fundamental genome-encoded properties of plants and microbes that can be harnessed or redesigned for beneficial purposes. Deciphering the complex and essential interactions among plants, microbial communities, and soils, particularly under stressful or changing environmental conditions, is essential for optimizing the production of biomass for a range of bioenergy and bioproduct applications in a sustainable bioeconomy.

Research efforts in bioenergy, environmental microbiome and biosystems design require innovative methods and technologies to elucidate the foundational principles that drive biological systems of interest to DOE's energy and environmental missions. Characterizing biological systems involves analytical approaches that illuminate cellular components and their form, structure, size, function, spatial location, dynamics, and interactions with the environment.

A wide array of synchrotron- and neutron- and electron- based techniques supported by BER are available for characterizing structure, function, and interrelationships among complex chemical, macromolecular and cellular components that are relevant to biological systems science. The spatial and temporal resolutions available from neutron, photon and electron beams enable characterization and imaging of system components and interactions among plants, microbes, water, and soil constituents. Accessible scales range from subnanometer to centimeter length and over time dimensions from femtoseconds to seconds and longer.

Capabilities and resources at the DOE synchrotron and neutron facilities, especially suited for biological systems science, provide expert guidance, user support and training for designing experiments and measuring data. Cryo-electron microscopy and tomography are also available techniques. Free access is available for competitive proposals. Please visit the poster and peruse the informational brochures.

A DOE BER User Facility for Structural and Chemical Insights on Plant-Soil-Microbial Systems

Aina Cohen* (acohen@slac.stanford.edu), Thomas Weiss (weiss@slac.stanford.edu), Jocelyn Richards, Ritimukta Sarangi, Britt Hedman, **Keith Hodgson** (PI)

SLAC National Accelerator Laboratory

www-ssrl.slac.stanford.edu/smb

The Structural Molecular Biology (SMB) resource at the Stanford Synchrotron Radiation Lightsource (SSRL) develops, operates, and supports state-of-the-art synchrotron radiation capabilities for enabling biological and environmental research using macromolecular crystallography (MC), small angle X-ray scattering (SAXS), X-ray absorption spectroscopy (XAS) and X-ray fluorescence (XRF) imaging techniques. As a user facility, the SMB resource provides the national scientific community access to these advanced capabilities primarily through general user proposals as well as a strong user support program, which includes outreach, training, and dissemination. The suite of instruments operated by the resource encompasses seven full time synchrotron beamlines at SSRL, four for MC, one for SAXS, two for XAS, and ~1.4 beamline equivalents spread out over six beamlines for μXRF imaging and advanced spectroscopy. Two state-of-the-art undulator microfocus beamlines are optimized for challenging micro-crystal and time-resolved MC measurements. The SAXS beamline is equipped with highly automated solution scattering robotics and features a state-of-the-art chromatography coupled SAXS setup as well as an exchangeable high flux multilayer monochromator for time-resolved experiments in the millisecond timescale. Three dedicated XRF imaging beamlines cover a range of spatial scales (micrometer to centimeter) and elements of biological importance (phosphorus, sulfur, potassium, calcium, and metals).

A powerful aspect of the XRF imaging beamlines is that they can perform μ-XAS to characterize the oxidation state, or chemical species, at a single point within a sample. Combining XRF with XAS is a tool for generating spatial distribution images of individual chemical species of an element within a sample. The synchrotron resource is managed and operated in a fully integrated and centrally coordinated manner, across all beamlines and techniques, facilitating cross technique structural investigations in biological and environmental research, covering length scales from Angstrom to centimeters. Recent results enabled by the SMB resource will be presented, highlighting the scientific potential of interlaboratory collaborations for a multi-technique approach to BER science made possible by a dedicated BER-supported outreach program at SSRL.

Advances in Small-Angle X-Ray Scattering for Structural Biology at the SIBYLS Beamline

Joshua T. Del Mundo* (jdelmundo@lbl.gov), Daniel T. Murray, Lee Joon Kim, Daniel J. Rosenberg, Brandon Russell, George Meigs, Blake Wienker, Scott Classen, David Shin, Kathryn Burnett, Susan E. Tsutakawa, Michal Hammel, **Gregory L. Hura** (PI)

Lawrence Berkeley National Laboratory

The SIBYLS beamline conducts small-angle X-ray scattering (SAXS) to reveal the structure of biological macromolecules in solution (proteins complexes, RNA, lipid nanoparticles) and is supported by BER Integrated Diffraction Analysis Technologies (IDAT). The beamline primarily operates in high throughput (HT-SAXS) and size exclusion chromatography (SEC-SAXS) modes. The mail-in SAXS program offers users timely, high-quality data. Between September 2022 and September 2023, this program has resulted in 29 publications supported by IDAT. This poster summarizes some high impact studies that were made possible using the SAXS tools developed at this beamline.

First, SEC-SAXS was used in a comparative structural study to investigate the diversity in assembly of the forms of ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco), an enzyme that catalyzes the first step in carbon fixation in plants. Characterization of the oligomeric states of deep-branching form Iα and I" rubiscos revealed structural origins of form I, shedding light onto the evolutionary path of rubisco and its transition from a homo-oligomer to a hetero-oligomer (Liu et al. 2023). Another high impact plant study investigated synthesis of pectin, an essential polysaccharide required for plant cell wall expansion. SEC-SAXS of the enzyme galactan synthase (GalS1), which catalyzes side chains in the pectin rhamnogalacturonan, was found to take on an antiparallel dimer orientation in solution, which differed from an alternative crystal structure (Prabhakar et al. 2023), giving insight into the mechanics of pectin synthesis. SAXS is also essential in understanding microbe metabolism. SEC-SAXS was used to elucidate the conformational changes within each step of electron bifurcation in EtfABCX, a membrane-bound superdimer of supermonomers (Murray et al. 2024). Electron bifurcation is a process that generates a high energy electron at the expense of the loss of energy in a second electron and is key in modelling the metabolisms of microbes for the design of novel bioenergetic systems. Finally, the team presents the utilization of SAXS as a tool in protein engineering. HT-SAXS was used to verify the architectures of various helical heterotrimer assemblies designed for use in complex protein nanostructures (Bermeo et al. 2022), as well as tunable protein crystals in solution and dried states (Li et al. 2023) for potential applications in biological sensing, catalysis, separations, and drug delivery.

After the upcoming Advanced Light Source upgrade period, the team will further improve data quality, efficiency of data collection, and implement new sample environments. Additionally, analysis pipelines are in development that will provide precise protein structural conformations by submitting sequences along with samples. The further insights that these improved SAXS studies will provide on biological macromolecules is the foundation for their applications in biomass utilization, microbe engineering, and advanced biomaterials.

Funding Statement: DOE BER

References

Bermeo, S., et al. 2022. "*De Novo* Design of Obligate ABC-type Heterotrimeric Proteins," *Nature Structural & Molecular Biology* **29**(12),1266–76. DOI: 10.1038/s41594-022-00879-4.

Li, Z., et al. 2023. "Accurate Computational Design of Three-Dimensional Protein Crystals," *Nature Materials* **22**, 1556–63. DOI:10.1038/s41563-023-01683-1.

Liu, A. K., et al. 2023. "Deep-Branching Evolutionary Intermediates Reveal Structural Origins of Form I Rubisco," *Current Biology* **33**(24), 5316–25. DOI:10.1016/j.cub.2023.10.053.

Murray, D. T., et al. 2024. "Correlating Conformational Equilibria with Catalysis in the Electron Bifurcating EtfABCX of *Thermotoga maritima*," *Biochemistry* **63**(1), 128–40. DOI:10.1021/asc.biochem/3c00472.

Prabhakar, P. K., et al. 2023. "Structural and Biochemical Insight into a Modular β-1,4-Galactan Synthase in Plants," *Nature Plants* **9**(3), 486–500. DOI:10.1038/s41477-023-01358-4.

Cryo-Electron Microscopy at Environmental Molecular Sciences Laboratory: New Advances, User Science and How to Access

James E. Evans* (james.evans@pnnl.gov, PI), Trevor Moser, Irina Novikova, Amar Parvate, Samantha Powell

Pacific Northwest National Laboratory

This project is focused on the operation a Krios cryogenic transmission electron microscope (Krios G3i) at the Environmental Molecular Sciences Laboratory (EMSL) to advance DOE BER user research in protein and 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-electron microscopy (cryo-EM) with other EMSL capabilities such as mass spectrometry or super-resolution fluorescence microscopy. Access is offered free of charge for BER users, with Pacific Northwest National Laboratory staff time funded by this current project. The BER access mechanism allows for an expedited submission and review process for "cryo-EM only" projects.

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 subnanometer 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 start from a provided gene of interest and employ the cell-free expression system to produce enough protein for structural characterization. The team 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. The team will also present an overview of EMSL's 1,000 Fungal Proteins project, which will use the Krios as a core capability and is accepting user proposals for structural and functional characterization of conserved fungal proteins.

Funding Statement: Pacific Northwest National Laboratory is operated by Battelle for DOE under contract DE-AC05-76RL01830. This program is supported by the DOE Office of Science, through the Genomic Science program of BER program, under FWPs 74194 and 74195. The work was performed at Environmental Molecular Sciences Laboratory (grid.436923.9), a DOE Office of Science user facility sponsored by the BER program.

Biological Soft X-Ray Tomography at the Advanced Light Source

Carolyn A. Larabell[1,2]* (carolyn.larabell@ucsf.edu, PI), Valentina Loconte[1,2], Jian-Hua Chen[1,2], Mark Le Gros[1,2]

[1]University of California–San Francisco; [2]Lawrence Berkeley National Laboratory

ncxt.org

Soft X-ray tomography (SXT) visualizes and quantifies the structural organization of biological organisms up to 20 micrometers in diameter. Specimens are imaged in the near-native state—rapidly frozen in their normal growth conditions—at a resolution up to 35 nanometers. Large numbers of cells can be imaged since it takes only five to 10 minutes to go from the frozen specimen to a reconstructed tomogram. Imaging is based primarily on the absorption of carbon, a common element of all known life. At the same time, water (ice) is virtually invisible so that high-contrast images are obtained based solely on the inherent properties of the structures examined. This is accomplished by imaging with X-ray photons in the 'water window' (between 284 to 543 electron volts), where X-ray photons are absorbed an order of magnitude more strongly by carbon- and nitrogen-containing organic material than by water. The absorption of soft X-rays adheres to the Beer-Lambert Law and is, therefore, a function of the chemical

composition and concentration of organic material, yielding unique quantitative Linear Absorption Coefficient measurements for specimen components. The team has used this label-free imaging technology to image and quantify a wide variety of structures, including bacteria, yeast, spores, algae, larger mammalian cells and isolated organic particles. This poster will present examples of SXT data that enabled biological findings that couldn't be obtained with other technologies, including the simultaneous visualization and quantification of carbon, effects of altered environments on cell structures, and novel findings about N_2 cycling in an endosymbiotic organism.

Funding Statement: This research was supported by the DOE Office of Science, BER program, under contract DE-AC025CH11231.

eBERlight—A User Program for Biological and Environmental Research at the Advanced Photon Source

Karolina Michalska* (kmichalska@anl.gov, PI), Zou Finfrock, Alex Lavens, Gyorgy Babnigg, Youngchang Kim, Changsoo Chang, Kemin Tan, Jerzy Osipiuk, Michael Endres, Hannah Parraga, Kenneth Kemner

Argonne National Laboratory

The eBERlight program at Argonne National Laboratory's Advanced Photon Source (APS) offers comprehensive support to research communities specializing in biological and environmental science. Utilizing synchrotron radiation, the program aids in the understanding of complex Earth systems and offers user support for project development, proposal design, workflow creation, and data analysis. Operating across multiple APS beamlines, eBERlight provides access to the state-of- the-art instrumentation for macromolecular crystallography, X-ray full field imaging (computed tomography), X-ray fluorescence microscopy, X-ray absorption spectroscopy, scattering and coherent diffractive imaging (including ptychography). To ensure an optimal infrastructure for demanding experiments, eBERlight leverages additional campus resources for sample preparation and computational data analysis. The ongoing upgrade of the APS facility will greatly improve X-ray capabilities, enabling imaging of larger samples at high resolution, enhancing spatial resolution and addressing dynamic processes. Through high-throughput, multidimensional data collection, unprecedented statistical analysis of complex, heterogenous systems will become attainable. Additionally, simultaneous recording of structural and kinetic data will enable tracking correlations between molecular motions and chemistry, providing valuable insights through serial and time-resolved macromolecular crystallography. These developments will enable researchers to address complex questions relevant to the biological, geological, geochemical, biogeochemical, and environmental sciences.

Funding Statement: Argonne National Laboratory is operated by UChicago Argonne, LLC for DOE under contract number DE-AC02-06CH11357.

Recent Developments at the Center for Structural Molecular Biology at Oak Ridge National Laboratory

Hugh O'Neill* (oneillhm@ornl.gov, PI), Sai Venkatesh Pingali, Wellington Leite, Kevin Weiss, Hong-Hai Zhang, Qiu Zhang, Felicia Gilliland, Volker Urban

Oak Ridge National Laboratory

ornl.gov/facility/csmb

Project Goals: The Center for Structural Molecular Biology (CSMB) at Oak Ridge National Laboratory (ORNL) is funded to support and develop the user access and science research program of the Biological Small-Angle Neutron Scattering (Bio-SANS) instrument at the High Flux Isotope Reactor (HFIR). Bio-SANS is dedicated to the analysis of the structure, function, and dynamics of complex biological systems. The CSMB also operates a Bio-Deuteration Laboratory (BDL) for expression and purification of deuterium labeled biomacromolecules and for synthesis of small molecules and ligands in support of the biology neutron scattering program. The CSMB supports a vibrant biological research community from academia, industry, and government laboratories.

The Bio-SANS instrument is ideally suited for studies of biomacromolecules including proteins, DNA/RNA, lipid membranes and other hierarchical complexes. The Bio-SANS detector system is designed to allow simultaneous access to a wide spatial range that enables utilization of the full potential of the high neutron flux from the ORNL HFIR cold source. This team has recently completed the next development stage of the detector system by installation and commissioning of a mid-range detector to complement the existing main and wing detectors. This development will improve data quality for hierarchical systems, decrease Q-

resolution mismatch, increase angular coverage, and enable sub-minute time resolution.

Sample environment (SE) capabilities that can accommodate sample types ranging from biomacromolecules in solution to biomass are critically important to realize the full potential of Bio-SANS. One recent SE development effort was to upgrade the robotic sample changer originally installed in 2019 with a Universal Robot (UR5), which has an expanded a temperature-controlled holding area for up to 66 sample cells. A Peltier heating block at the sample position allows rapid temperature change between 10 to 100°C. New science opportunities include *in situ* kinetic processes of complex biological systems using time-resolved SANS with simultaneous access to multiple length scales. Further development is underway to expand this capability to allow liquid handling at Bio-SANS for mixing samples directly before measurement. Another example is chromatography—SANS for in beam fractionation of biomacromolecules that can operate in continuous flow mode as well as fractionation of complex mixtures of biomacromolecules. The flow cell design accommodates four cells to minimize down time during sequential purifications of multiple proteins.

To broaden the impact of the CSMB and catalyze the synergy between BER program–funded structural biology resources, the team established collaborative programs with the National Synchrotron Light Source II for joint access to SANS and SAXS and with the BER Facilities Integrating Collaborations for User Science (FICUS) program between the DOE Joint Genome Institute at Lawrence Berkeley National Laboratory and the Environmental Molecular Sciences Laboratory at Pacific Northwest National Laboratory.

Development and Deployment of New Structure Prediction and Determination Capabilities at the UCLA-DOE Institute

Jose A. Rodriguez* (jrodriguez@mbi.ucla.edu), Samantha Zink, Jeff Qu, Niko Vlahakis, Theodosia Bartashevitch, Lukasz Salwinski, Roger Castells-Graells, Todd O. Yeates, David Eisenberg, **Matteo Pellegrini** (PI)

UCLA-DOE Institute for Genomics and Proteomics, University of California–Los Angeles

doe-mbi.ucla.edu/ucla-doe-institute-missions

Project Goals: Research in the DOE-University of California–Los Angeles (UCLA) Institute for Genomics and Proteomics (IGP) includes major efforts in the area of imaging science, proteomics, structure prediction and atomic structure determination. These new capabilities help scientists better understand microbial biosystems, their genomics and molecular biology. This team is pioneering new enabling capabilities that facilitate the discovery of molecular structural features affecting protein function and specificity, to better understanding of bioenergy crops and microbes. These capabilities span the broad areas of X-ray diffraction, electron microscopy, and micro-electron diffraction (MicroED), along with computational structure and function prediction methods. This team is also enabling rapid access to robust public-facing tools for use by the BER community.

This group's efforts in imaging science and protein characterization bridge a number of technological areas to address pressing problems in protein structure and function.

Breakthroughs in cryo-electron microscopy (cryo-EM): Numerous technical advances have made cryo-EM an attractive method for atomic structure determination. Cryo-EM is ideally suited for very large structures; symmetrical structures like viruses are especially amenable. However, problems of low-signal-to-noise in imaging small proteins makes it practically impossible to determine structures smaller than about 50 kilodaltons, leaving a great many cellular proteins and enzymes (and nucleic acid molecules) outside the reach of this important structural technique. The DOE-UCLA IGP team has broken through this barrier by engineering novel scaffolds with sufficient rigidity and modularity to achieve resolution useful for interpreting atomic structure. This team has applied this system to image a 19 kDa protein, obtaining multiple structures of its sequence variants unbound and bound to a small molecule. The findings highlight the promise of these novel scaffolds for advancing the design of drug molecules against small therapeutic protein targets in cancer and other human diseases as well as other important targets. Recent efforts have been aimed at imaging microbial and plant protein targets.

Enabling microcrystal electron diffraction (MicroED) methods: A broad array of atomic structures has now been determined by MicroED; they include naturally occurring peptides, synthetic protein fragments and peptide-based natural products. This team is further enhancing the capabilities of electron diffraction (ED) by improving understanding of electron counting detectors and their application to diffraction measurements. In addition, the group is broadening comprehension of electron beam-induced

radiation damage and its consequences for molecular systems and their characterization at atomic resolution. Collectively, these efforts have yielded new insights into how ED data are impacted by electron beam-induced lattice reorientation and the impact of radiation damage on the ability to determine the chiral nature of handed molecules.

Tools for analysis of condensate or aggregate-forming proteins: The recent revolution in artificial intelligence (AI) and machine learning methods has dramatically improved scientists' ability to predict protein structure and sequence characteristics. This team has exploited the growing capacity of AI models to train a fully connected neural network to emulate the predictive abilities of computationally time-consuming 3D profiling approaches. This method relies on the network to calculate the propensity of segments in a sequence to form amyloid-like contacts or structures. Whereas the previous approach required weeks or months of compute time to evaluate an entire proteome, the new approach can evaluate the entire yeast proteome in 15 minutes and is available as an online server for public use.

The institute's enabling capabilities will broadly facilitate the determination and prediction of unknown macromolecular structures with importance for bioenergy.

Funding Statement: This work is supported by DOE grant DE-FC02-02ER63421.

References

Castells-Graells, R., et al. 2023. "Cryo-EM Structure Determination of Small Therapeutic Protein Targets at 3 Å-Resolution Using a Rigid Imaging Scaffold," *Proceedings of the National Academy of Sciences of the United States of America* **120**(37), e2305494120. DOI:10.1073/pnas.2305494120.

Thompson, M. C., et al. 2020. "Advances in Methods for Atomic Resolution Macromolecular Structure Determination," *F1000Research* 9. DOI:10.12688/f1000research.25097.1.

Berkeley Synchrotron Infrared Structural Biology (BSISB) Imaging Program

Neslihan Taş*, Peter Zwart*, Nowzesh Hasan, **Hoi-Ying N. Holman** (hyholman@lbl.gov, PI)

Lawrence Berkeley National Laboratory

bsisb.lbl.gov

BSISB Missions:

1. Increase synchrotron radiation-based Fourier-transform infrared (SR-FTIR) capabilities accessible to BER users through Berkeley Synchrotron Infrared Structural Biology (BSISB) for tracking the locations and concentrations of chemical events in biological samples.
2. Enhance chemical identification capabilities available to users through subsequent mass spectrometry characterization of SR-FTIR informed regions of interest.
3. Enable autonomous experimentation for improved temporal resolution and faster, more efficient experiments for scientific discoveries at infrared beamlines at Lawrence Berkeley National Laboratory's Advanced Light Source (ALS).
4. Support the diverse and evolving research interests and adapt to the changing needs of the user community through user-stimulated technology development and refinement.

The Berkeley Synchrotron Infrared Structural Biology (BSISB) imaging program is a BER-funded national user resource at three infrared (IR) beamlines at the Advanced Light Source (ALS) in Berkeley, Calif. Synchrotron IR (SIR) radiation spans the far-, mid-, and near-IR. It is 100 to 1,000 times brighter than a conventional thermal source, enabling broadband spectroscopic imaging with high signal-to-noise ratios. Spatial resolution is diffraction-limited for SIR spectromicroscopy or microspectroscopy, and well beyond the diffraction limit for SIR nanospectroscopy. By probing molecular and lattice vibrations, low-energy electronic excitations, and related collective plasmon and phonon resonances, SIR spectroscopy enables high spatial resolution measurements of heterogeneity in biological, chemical, and physical properties. Current capabilities enable the imaging of engineered, natural/living samples at the micro- and nanoscale. This poster presentation offers an overview of SIR as a unique method for chemical imaging. The team then provides an

overview of the following capabilities available through BSISB as well as example applications: SR-FTIR spectromicroscopy, SR-FTIR nanospectroscopy, Autonomous Adaptive Data Acquisition, Integrated SR-FTIR with ambient atmospheric infrared ablation mass spectrometry, time-resolved imaging of chemical events, and membrane microfluidics to circumvent water interference.

Funding Statement: The Berkeley Synchrotron Infrared Structural Biology (BSISB) is supported by DOE Office of Science, BER program. Advanced Light Source is supported by DOE Office of Science user facility operating. Both BSISB and ALS are under contract DE-AC02-05CH11231.

Visualizing Biological Systems at the Molecular and Cellular Level at the Laboratory for BioMolecular Structure

Liguo Wang* (lwang1@bnl.gov, PI), Guobin Hu, Jake Kaminsky, Qun Liu, Sean McSweeney

Brookhaven National Laboratory

Cryo-electron microscopy (cryo-EM) is a powerful imaging technique used to visualize biological specimens; it has experienced exponential growth in the past decade, marking a 'resolution revolution'. Currently, there are more than 32,000 entries of EM maps and other results in the Electron Microscopy Data Bank. In addition to the atomic structures of biological macromolecules, cryo-EM has been employed to study protein-protein and protein-cell interactions and offers insights into cellular and tissue organizations at a resolution unsurpassed by other imaging techniques. This technique plays a pivotal role in advancing scientists' insight into biological processes at the molecular and cellular level.

With the establishment of the Laboratory for BioMolecular Structure (LBMS), Brookhaven National Laboratory provides peer-reviewed research access, support, and training for the use of cryo-EM. By allowing science-driven use of these instruments, LBMS meets the urgent need to advance the molecular understanding of biological processes, enabling deeper insight and opening the possibility to engineer biological functions in a predictable fashion. Last year, LBMS supported more than 100 sessions, and collected more than one million cryo-EM images, which resulted in 116 high-resolution (better than 4 angstroms) structures and 15 publications.

LBMS also offers three-tiered trainings to current and potential users: (1) annual four-day cryo-EM course to the public; (2) quarterly cryo-EM workshops for current and potential LBMS users; (3) on-demand five-day training for LBMS users, either in person on screening EMs or remote training on the high-end EM, as needed. The average rating of the workshops is 4.4 out of 5.0, with 91% of participants indicating they would recommend the workshop to others.

In recent years, cryo-electron tomography (cryo-ET) has garnered increasing attention due to its unique capability for direct visualization of interactions between complexes in their cellular environment. It offers unparalleled insights into molecular organization, cellular structure, and cell physiology, making it a powerful tool for probing intricate details at the nanoscale within a cellular context. To expand the cryo-ET capability at LBMS, the team will establish and operate a cryo-ET user program to support a broad range of projects funded by DOE. Three distinct routes will be offered based on the nature of the sample and the specific regions of interest. With the development of the cryo-ET program, researchers can study cells/organelles and tissues. This bridges a critical imaging gap in the biomedical size spectrum, connecting studies of molecules at atomic resolution to cellular and tissue investigations.

The Availability of Inorganic Nitrogen and Organic Carbon Manipulates Ectomycorrhizal Fungi-Mediated Iron Acquisition in the Forest Ecosystem

Haihua Wang[1]*, Hui-Ling Liao[1] (sunny.liao@ufl.edu), Kaile Zhang[1], Tiffany Victor[2], Jennifer Bhatnagar[3], Rytas Vilgalys[4], Kerrie Barry[5], John Cliff[6], Jeremy Bougoure[6], Dehong Hu[6], Sarah Leichty[6], **Ryan Tappero[2]** (PI)

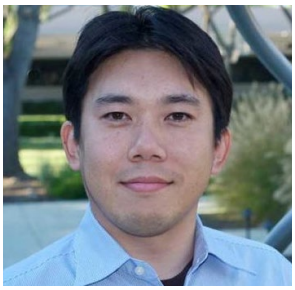
[1]University of Florida; [3]Brookhaven National Laboratory; [3]Boston University; [4]Duke University; [5]DOE Joint Genome Institute; [6]Pacific Northwest National Laboratory

Ectomycorrhizal fungi (EMF) play a crucial role in aiding plant nutrition, specifically by extracting nitrogen (N) from organic compounds in soil organic matter (SOM)—a process known as N-mining. In iron deficient soils, EMF can strengthen iron (Fe) acquisition at both hypha-minerals and fungal-plant cell interfaces. However, the effect of EMF induced N-mining and SOM decomposition on Fe processing in mycorrhizal plants remains unclear. Furthermore, researchers aim to explore the interactive effects of inorganic N

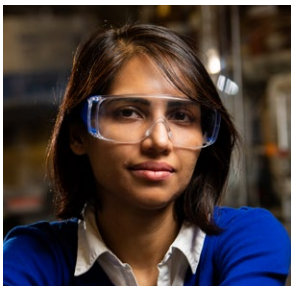
fertilization and SOM on shaping EMF-mediated Fe processes and plant Fe uptake, a topic that remains largely unexplored. To address these questions, the team performed a mesocosm study using the *Pinus-Suillus* model system. Specifically, researchers inoculated *Pinus taeda* with *Suillus cothurnatus* and grew them in conditions treated with +/- Fe-coated sand, +/- SOM, and a gradient of ammonium nitrate concentrations. Using the synchrotron pink beam X-ray microfluorescence imaging (PB-XRF) on cross-sections of ectomycorrhizal roots two months post-fungal inoculation, the team found that the effect of inorganic N availability on Fe acquisition in ectomycorrhiza largely depended on SOM supply. Among the combinations of SOM and inorganic N treatments, mycorrhization demonstrated a greatest preference for +SOM/-inorganic N conditions, while mostly exhibiting negative responses to +SOM/high inorganic N conditions. With the addition of SOM, the Fe concentration in mycorrhizae was significantly decreased with the rising levels of treated inorganic N. Conversely, in the absence of SOM, an opposite trend was observed. Spatial analysis of Fe across ectomycorrhizal compartments showed that Fe was primarily accumulated in the fungal mantle underlying the Fe-enriched condition, while Fe was transferred more to the inner compartments, specifically the cortex and vascular tissues, when less Fe was acquired. These findings imply that in EMF-predominant forests, EMF may possess the capacity to facilitate Fe-associated SOM processing and mycorrhizal N/Fe uptake, enhancing the formation of mycorrhization. However, this ability may be compromised under elevated inorganic N conditions. Further studies on the molecular and biochemical aspects of plant-EMF interactions are necessary to precisely evaluate this implication. The team's ongoing studies are dedicated to conducting these aspects of the research. For example, group members are utilizing the NanoSIMS tool to visualize and quantify the flux of N transformed from N-labeled SOM to EMF hyphae. The team also employs metatranscriptomics and fluorescence *in situ* hybridization imaging to visualize the activity of fungal genes responsible for cellulose metabolism in mycorrhizae, influenced by the combinations of SOM and inorganic N addition.

2024 Oral Presentation Principal Investigators and Presenting Authors

Multidisciplinary Science Enabled by Facility Collaboration



Yasuo Yoshikuni
U.S. DOE Joint Genome Institute
yyoshikuni@lbl.gov



Arunima Bhattacharjee
Pacific Northwest National Laboratory
arunimab@pnnl.gov

Malak Tfaily
University of Arizona
tfaily@arizona.edu

Jennifer Bhatnagar
Boston University
jmbhat@bu.edu

Hui-Ling (Sunny) Liao*
University of Florida
sunny.liao@ufl.edu

Victoria Orphan
California Institute of Technology
vorphan@caltech.edu

Quantum Imaging and Sensing

Theodore Goodson
University of Michigan
tgoodson@umich.edu

Shiva Abbaszadeh
University of California–Santa Cruz
sabbasza@ucsc.edu

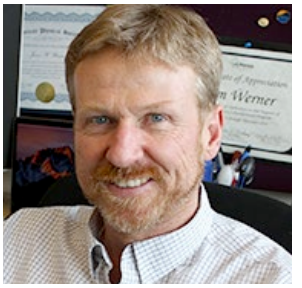
Sean McSweeney
Brookhaven National Laboratory
smcsweeney@bnl.gov

Justin C. Goodrich*
Brookhaven National Laboratory
jgoodrich@bnl.gov

**Presenting for the PI*

Continued from previous page – Quantum Imaging and Sensing

Thomas Theis
North Carolina State University
ttheis@ncsu.edu



James Werner
Los Alamos National Laboratory
jwerner@lanl.gov

Ted Laurence
Lawrence Livermore National Laboratory
laurence2@llnl.gov

Audrey Eshun*
Lawrence Livermore National Laboratory
eshun1@llnl.gov

James Evans
Pacific Northwest National Laboratory
james.evans@pnnl.gov

Phenomics and Function

James Neal
Broad Institute of Massachusetts Institute of Technology
jtneal@broadinstitute.org

Thomas O. Metz
Pacific Northwest National Laboratory
thomas.metz@pnnl.gov

Benjamin Cole
U.S. DOE Joint Genome Institute
bjcole@lbl.gov

Gerald Tuskan
Oak Ridge National Laboratory
tuskanga@ornl.gov

*Presenting for the PI

Continued from previous page – Phenomics and Function

John Lagergren*
Oak Ridge National Laboratory
lagergrenjh@ornl.gov

BER Data Science and Infrastructure

Kerstin Kleese van Dam
Brookhaven National Laboratory
kleese@bnl.gov

Emiley Eloe-Fadrosh
Lawrence Berkeley National Laboratory
eaeloefadrosh@lbl.gov

Lee Ann McCue*
Pacific Northwest National Laboratory
leeann.mccue@pnnl.gov

Kjiersten Fagnan
U.S. DOE Joint Genome Institute
kmfagnan@lbl.gov

Adam P. Arkin
Lawrence Berkeley National Laboratory
aparkin@lbl.gov

Nikos Kyrpides
U.S. DOE Joint Genome Institute
nckyrpides@lbl.gov

Shantenu Jha
Princeton Plasma Physics Laboratory
sjha@pppl.gov

**Presenting for the PI*

KBase Science: Data Integration to Support (or Refute) Predictions

Mitchel J. Doktycz
Oak Ridge National Laboratory
doktyczmj@ornl.gov

Dale A. Pelletier*
Oak Ridge National Laboratory
pelletierda@ornl.gov

Paramvir S. Dehal
Lawrence Berkeley National Laboratory
psdehal@lbl.gov

Shinjae Yoo*
Brookhaven National Laboratory
sjyoo@bnl.gov

Christopher S. Henry
Argonne National Laboratory
chenry@anl.gov

Robert Egbert
Pacific Northwest National Laboratory
robert.egbert@pnnl.gov

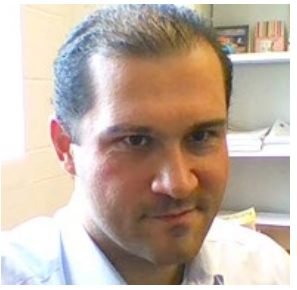
William C. Nelson*
Pacific Northwest National Laboratory
william.nelson@pnnl.gov

Adam P. Arkin
Lawrence Berkeley National Laboratory
aparkin@lbl.gov

Elisha Wood-Charlson*
Lawrence Berkeley National Laboratory
elishawc@lbl.gov

*Presenting for the PI

Bioimaging

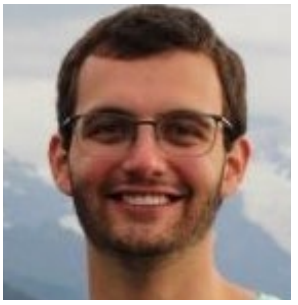


Andreas E. Vasdekis
University of Idaho
andreasv@uidaho.edu

Na Ji
University of California–Berkeley
jina@berkeley.edu

Marcus T. Cicerone
Georgia Institute of Technology
cicerone@gatech.edu

Mark Kasevich
Stanford University
kasevich@stanford.edu

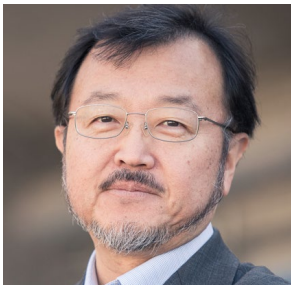


Adam Bowman*
Stanford University
abowman2@stanford.edu

Leslie Sieburth
University of Utah
sieburth@biology.utah.edu

Andrew Groover*
U.S. Forest Service
andrew.t.groover@usda.gov

Scott Lea
Pacific Northwest National Laboratory
scott.lea@pnnl.gov



Soichi Wakatsuki
SLAC National Accelerator Laboratory
soichi@slac.stanford.edu

Shimon Weiss
University of California–Los Angeles
sweiss@chem.ucla.edu

Debjit Roy*
University of California–Los Angeles
deb9804@g.ucla.edu

**Presenting for the PI*

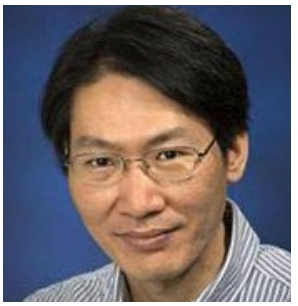
BER Biopreparedness Research Virtual Environment (BRaVE)

Martin Schoonen
Brookhaven National Laboratory
mschoonen@bnl.gov

Shawn Starkenburg
Los Alamos National Laboratory
shawns@lanl.gov

Margaret Cheung
Pacific Northwest National Laboratory
margaret.cheung@pnnl.gov

Vivek K. Mutalik
Lawrence Berkeley National Laboratory
vkmutalik@lbl.gov



Qun Liu
Brookhaven National Laboratory
qunliu@bnl.gov