Pacific Northwest
NATIONAL LABORATORY

Persistence Control
@PNNL

DOE BSSD Performance Management Metrics Report Q1: 1/17/2023

SFA Laboratory Research Manager: Jon Magnuson[1], Jon.Magnuson@pnnl.gov
SFA Principal Investigator: Robert G. Egbert[1], Robert.Egbert@pnnl.gov
[1]**Pacific Northwest National Laboratory**, Richland, WA 99354

Investigators: Robert G. Egbert[1], Jayde A. Aufrecht[1], Adam M. Deutschbauer[2], Joshua R. Elmore[1], Adam M. Guss[3], Pubudu P. Handakumbura[1], Caroline S. Harwood[6], Vivian S. Lin[1], Ryan S. McClure[1], Jason E. McDermott[1], Ernesto S. Nakayasu[1], William C. Nelson[1], Devin Coleman-Derr[4], Enoch Yeung[5]

Participating Institutions: [1]**Pacific Northwest National Laboratory**, Richland, WA 99354; [2]Lawrence Berkeley National Laboratory, 94720; [3]Oak Ridge National Laboratory, Oak Ridge, TN 37831; [4]University of California, Berkeley, CA 94704; [5]University of California, Santa Barbara, CA 93106; [6]University of Washington, Seattle, WA 98195.

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

Q1 Target: Describe overall approaches to securely designing new functions into organisms for bioenergy and bioeconomy applications.

Executive Summary

As part of the Secure Biosystems Design program, the PNNL Persistence Control (PerCon) Science Focus Area (SFA) aims to use high-throughput genetic and bioinformatic tools to understand, predict, and control plant-microbe interactions in the rhizosphere of the bioenergy crop sorghum. Specifically, we investigate approaches to reshape the environmental niche of native sorghum rhizosphere bacterial isolates by creating an engineered metabolic addiction to root exudate compounds. Pursuing the secure biocontainment aims of our research will position the U.S. to safely harness the full potential of the rhizosphere microbiome to support ecosystems that fulfill national goals for a resilient bioeconomy through sustainable bioenergy cropping systems that durably deposit atmospheric carbon in soils.

In the inaugural funding cycle of the SFA, we have established techniques for interrogating microbial function and plant-microbe interactions in seven areas:

1. We developed a strain-agnostic method for the high-throughput genetic manipulation of bacteria, including rhizosphere isolates[1].

2. We established a methodology to perform genome-wide mutant fitness assays in soil environments.

3. We cultivated synthetic microbial communities derived from competitive enrichment and defined isolates by formulating a data-driven synthetic sorghum growth medium.

4. We created a scalable software pipeline to cluster and search protein libraries for shared functions using amino acid recoding schemes [2].

5. We investigated spatially resolved root-microbiome interactions using RhizoGrid, a plant cultivation platform we developed to collect multi-omics data on root structure, taxonomic profiles, and metabolomics [3].

6. We synthesized an activity-based chemical probe to identify proteins that interact with the sorghum-specific root exudate compound sorgoleone.

7. We developed a multi-omics platform to discover genes responsible for bacterial metabolism of sorghum root exudate compounds.

Background

The potential to employ emerging genetic and computational tools to design and securely deploy synthetic rhizosphere microbiomes, defined communities of microbes that colonize plant roots and benefit plant growth, offers tremendous opportunity to realize highly productive and stress-tolerant biomass cropping systems. A critical obstacle to realizing this vision is understanding fundamental principles of microbiome persistence in complex environments, notably amid the dynamics of root exudation and microbial colonization. Further, we need predictive tools to assess the risks associated with the deployment or unintended release of engineered microbes in plant and soil ecosystems and to mitigate those risks through effective strategies to control persistence. To meet these challenges, high-throughput genetic manipulation and bioinformatic tools provide a platform to elucidate the genetic elements underpinning rhizosphere community function, to create beneficial communities, and to control their environmental persistence.

The Persistence Control SFA investigates approaches to reshape the environmental niche of native sorghum rhizosphere bacterial isolates by creating an engineered metabolic addiction to root exudate compounds. We aim to understand the genes and networks that underpin environmental persistence of microbiomes to create dependencies of engineered microbes on plants such that the microbes can no longer persist in the environment in the absence of the target plant (Figure 1). Our strategy to control the environmental persistence of engineered microbes is to establish metabolic addiction to plant root exudate compounds by installing compound-specific catabolic pathways while removing pathways used for scavenging nutrients in soil microbiome environments. This strategy requires exquisite control of genome content to, for instance, eliminate metabolic niches that allow survival cross-feeding of nutrients by other microbiome members as well as to introduce complete pathways to thrive in the plant rhizosphere, but not in the rhizosphere of other plants. It also demands a deep understanding of plant or microbe interactions that could overcome control through complementation.

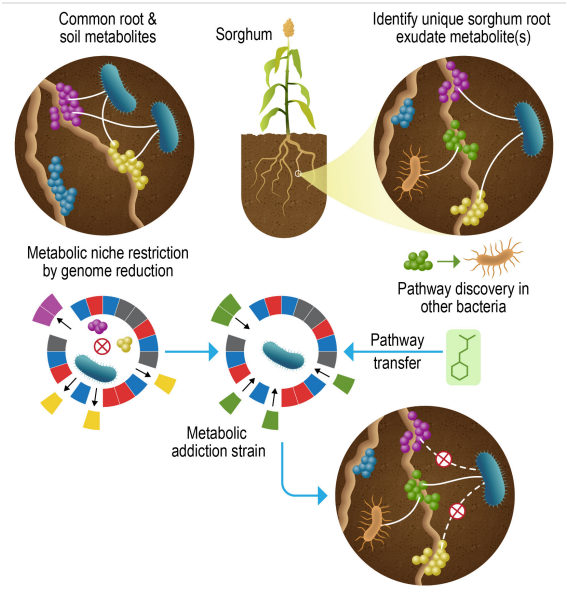

Figure 1. The Persistence Control SFA aims to control the proliferation of engineered microbes by establishing metabolic addiction of microbe to a bioenergy crop supported by the engineered microbiome.

To help accomplish the goals of the Secure Biosystems Design program, the Persistence Control SFA team draws expertise from across three DOE Office of Science National Laboratories – Pacific Northwest National Laboratory, Lawrence Berkeley National Laboratory, and Oak Ridge National Laboratory – and three Universities: the University of California Berkeley, the University of California Santa Barbara, and the University of Washington. To understand the genes and networks that control environmental fitness we have developed and integrated research strategies that draw from our team's expertise in functional genomics, synthetic biology, microbial ecology, chemical biology, bioinformatics, machine learning, and plant-microbe interactions. In this report, we share a selection of our approaches for persistence control to securely design new functions into organisms for bioenergy and bioeconomy applications.

Strain-agnostic high-throughput genome integration for rhizosphere isolates

Efficient genome engineering is critical to understand and utilize microbial functions. In the Persistence Control SFA we seek to add non-native functions, such as the ability to consume sorghum-specific root exudate compounds for metabolic addiction, into non-model bacteria that have been isolated from the sorghum rhizosphere. Most standard approaches to add non-native functions include the use of unstable, autonomously replicating plasmid DNA and cannot be used in complex environments such as soil. Thus, to engineer strains for applications in these environments, DNA encoding non-native functions must be stably integrated into the bacterial chromosome. Despite recent development of tools such as CRISPR/Cas gene editing, the ability to efficiently integrate multiple exogenous DNA sequences into bacterial chromosomes remains limited to model bacteria.

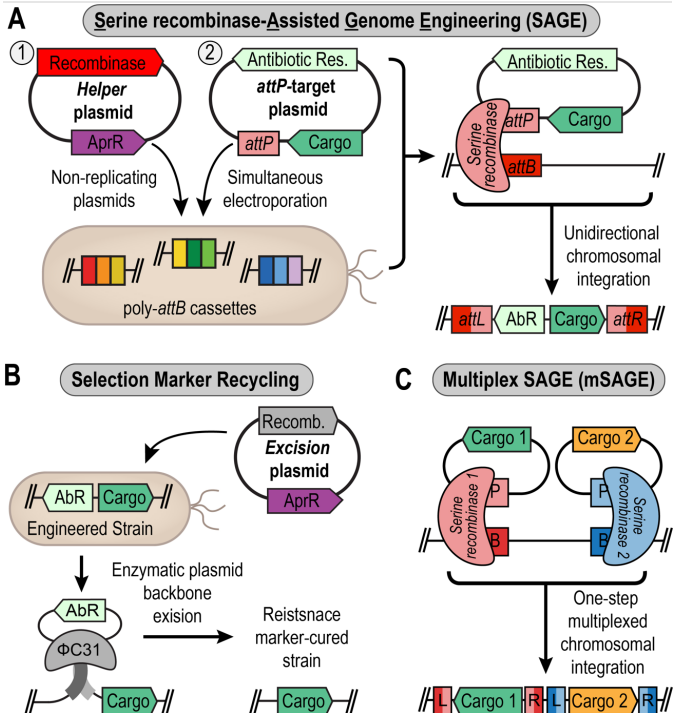

Figure 2. Iterative genome integration of genetic constructs using SAGE. The SAGE cycle enables rapid, efficient, and site-specific integration of multiple DNA fragments into bacterial chromosomes. (A) A base strain with *attB* sequences is created using standard methods. The *attB* sequences are landing pads for the integration of non-replicating plasmids through transformation and expression of a matching recombinase. Unnecessary components of the integrated plasmid are excised with a specific recombinase (B), enabling iterative SAGE cycles (C).

We developed a bacterial genome engineering tool called Serine recombinase-Assisted Genome Engineering, or SAGE, with which we sought to achieve parity for engineering model, non-model, and undomesticated bacterial strains [1]. SAGE utilizes up to 10 site-specific serine recombinases that each catalyze DNA recombination between two distinct attachment sequences (*attP* and *attB* sites). This process is unidirectional, which unlike many similar technologies generates a stably modified chromosome with no further steps (**Figure 2A**).

An innovative aspect of SAGE is that we transiently express the recombinases from non-replicating DNA molecules, which both expands the range of bacteria we can engineer and abolishes the need for extra steps required with other technologies to remove genome integration machinery. By excising the selection marker and other extraneous DNA with a second recombinase (**Figure 2B**), we can sequentially integrate multiple DNA fragments – a requirement for installing and optimizing complex engineered functions. More recently, we developed multiplex SAGE (mSAGE), which allow installation of complex functions in a single step – thus saving time, effort, and resources. For this, we simultaneously introduce multiple serine recombinases to incorporate up to 3 DNA cargos into the chromosome at once. All three selection markers can be excised in a single transformation step. This effectively doubles or triples the speed of chromosomal modifications versus standard SAGE – which was already significantly faster and enables more modifications than comparable genome engineering technologies. The exceptional efficiency of SAGE enables construction of strain libraries with millions of expression variants, enabling experiments such as high-throughput analysis of genetic control elements

(e.g. transcriptional promoters) and genome-scale genetic perturbations (e.g., CRISPR interference) for evaluating gene function.

Barcoded transposon libraries to assess environmental niches in soil

Understanding the genes and associated functions that contribute to environmental fitness phenotypes in microbes is critical to developing persistence control strategies for engineered microbial functions in rhizosphere environments. We have developed an approach to apply high-throughput functional genomics methods [4] to understand which genes contribute to fitness in sterile and live soil environments. This understanding will drive genome reduction efforts to establish metabolic addiction phenotypes for our engineered hosts wherein they colonize the sorghum rhizosphere but are unfit in soil without sorghum.

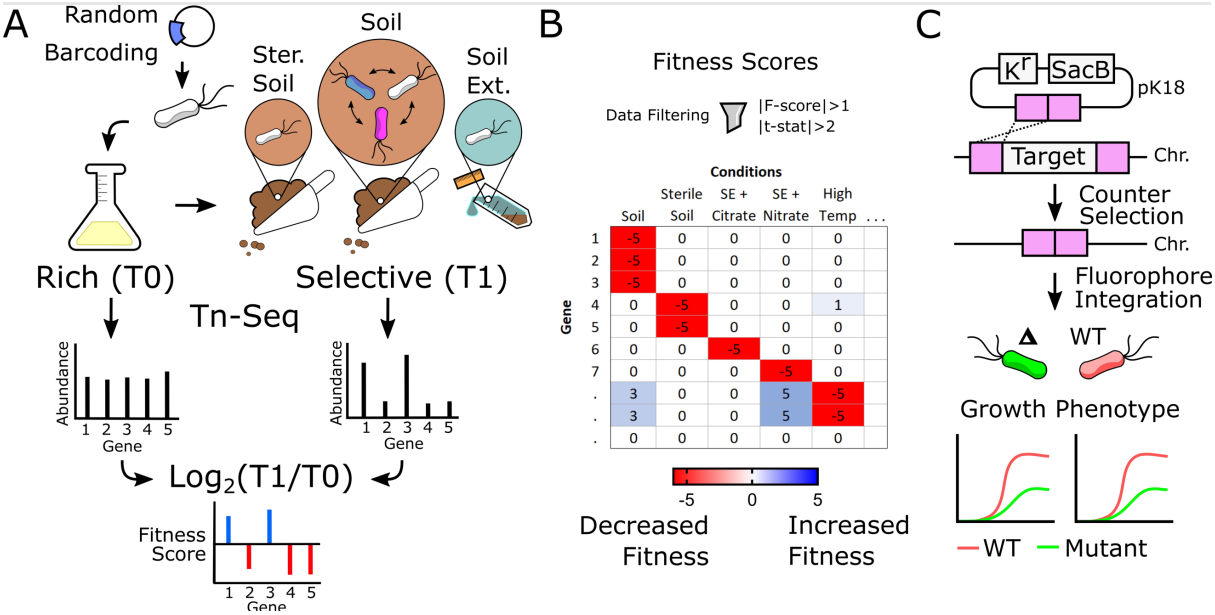

Figure 3. Overview of randomly barcoded transposon mutagenesis sequencing (RB-TnSeq) assays to identify genes required for fitness in soil. (A) Mutagenized strain libraries of plant growth-promoting rhizobacteria, e.g. *P. fluorescens* SBW25 and *P. putida* KT2440, are grown in a selective media (T1) for comparison of barcode abundances against the starting population (T0). Selective media include sterile soil, living soil, and many soil-extract conditions. Amplicon sequencing quantifies the relative abundance of each barcoded mutant strain. The abundance ratio for each mutant provides a statistical fitness score for all barcoded genes. (B) Genes with significant contributions to soil fitness are determined using quality control, statistical, and fitness-score thresholds with hierarchical clustering analysis. (C) RB-TnSeq observations are validated using flow cytometry assays of co-inoculated wild-type and mutant strains with distinct fluorescent protein tags.

We have constructed randomly barcoded transposon sequencing (RB-TnSeq) libraries in the plant growth promoting rhizobacterium *Pseudomonas fluorescens* SBW25, a model host for persistence control engineering efforts, and in novel sorghum rhizosphere isolates. We conducted genome-wide gene fitness surveys for SBW25 and the model bioremediation and bioproduction host *Pseudomonas putida* KT2440 in soil and in soil-extract media environments. These surveys identified genes that contribute to fitness under multiple environmental stresses, including pH, temperature extremes and abrupt shifts, osmotic stress, and for the utilization of many soil-relevant carbon and nitrogen sources under defined minimal media and soil extract conditions. Notably, these surveys identify metabolic and regulatory genes that are conditionally essential to environmental persistence under either sterile or live soil conditions. Linking these functional genomics datasets with our growing collection of environmental isolates and computational tools for

defining and predicting complementation dynamics is likely to lead to powerful insights and innovations regarding the genetic determinants of environmental persistence and strategies to control engineered microbes in complex environments.

Testbeds for rhizosphere colonization dynamics of engineered microbes

Development of a Synthetic Sorghum Rhizosphere Community

To demonstrate control over the persistence of engineered microbial functions in rhizosphere environments, we developed a synthetic rhizosphere sorghum community (RSC1) testbed. Members of RSC1 were selected from cultured isolates based on co-abundance network analysis from sorghum rhizosphere microbiome taxonomic data. Using 16S amplicon data and selecting only abundant taxa we inferred a network of species and cross-referenced the amplicon sequence variant (ASV) list against our isolate collection. Starting with rhizosphere amplicon data we selected ASVs that were present in at least 50% of samples. We then down selected for ASVs with average abundance in the top 50th percentile. These abundant taxa were then used to infer a network from Pearson correlation co-efficients. Our network contained 440 co-abundant ASVs, which we hypothesized could support a sorghum rhizosphere-representative synthetic community. With isolate collection matches of 293 out of 440 ASVs, we selected 66 isolates from genera represented multiple times in the network. suggesting that isolates of this genera may be particularly important. In order to test the stability of the RSC1 testbed, we inoculated and experimentally assessed the RSC1 synthetic community in three environments of increasing complexity: defined agar media, sterile plant boxes, and a sorghum field site.

Cultivation of a defined synthetic microbial community on a data-driven synthetic growth media

We formulated a defined synthetic sorghum growth media to simulate the carbon sources available in the sorghum rhizosphere. Our formula is based on the peak area of known compounds from metabolomics mass spectrometry data of soil collected from the Kearney Agricultural Research and Extension Center field site in California. Using the buffered MME media base[5] that contains no carbon or nitrogen compounds, we introduced 20 carbon and/or nitrogen sources, including amino acids, sugars and other compounds. To assess community stability over time, we prepared agar plates with synthetic sorghum media and cultivated RSC1 over twelve consecutive week-long passages. We collected and extracted genomic DNA and cold storage stocks from each passage to allow longitudinal taxonomic analysis and revival of the evolved RSC1 community at multiple stages.

Lab-based in planta cultivation of RSC1

To assess RSC1 stability *in planta*, sorghum seeds were sterilized and germinated on sterilized wet paper filter at 30 C without light. Each strain in RSC1 was grown at 30 C from one to three days according to individual growth rates, and six inoculating loopfuls of growth for each strain were suspended in 1.2 mL of sterile PBS 1x buffer. Cell suspensions were pooled and inoculated into plant microboxes containing 1 kg of calcined clay per box (Figure 4). Four germinated seeds were transplanted individually in 5 cm beneath soil surface in a microbox. Plants were watered three times a week, and after two weeks the drought

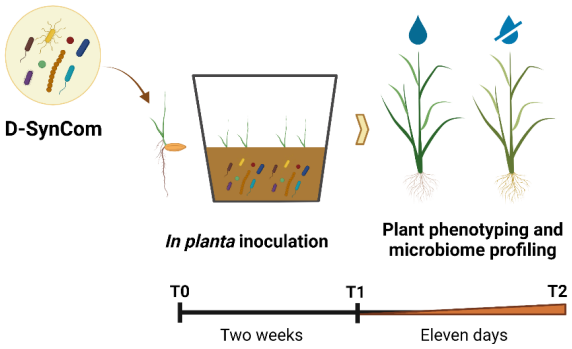

Figure 4. Lab-based D-SynCom colonization analysis in sorghum plants under normal irrigation (T0-T2) and drought stress (T1-T2).

treatment was applied for 11 days (T2), with additional 26 days before harvesting (T0-T2). Shoot phenotypes were recorded and the roots were used to profile 16S rRNA taxonomic diversity at T2.

Field trail with RSC1 inoculation

To investigate the persistence of the RSC1 synthetic microbial community in field grown plants, sorghum seeds inoculated RSC1 prepared as described above or with a mock treatment were planted under two types of irrigation, drought-stress (complete lack of irrigation) and normal watering. Five weeks after planting, a second inoculation was applied to a subset of plants per row of each block (Figure 5). At the 16th week, 10 plants from each treatment/inoculation combination were bagged and harvested for shoot phenotyping and microbiome analysis using 16S rRNA community profiling. Additional yield quantification was performed on additional plants at the 25th week after planting.

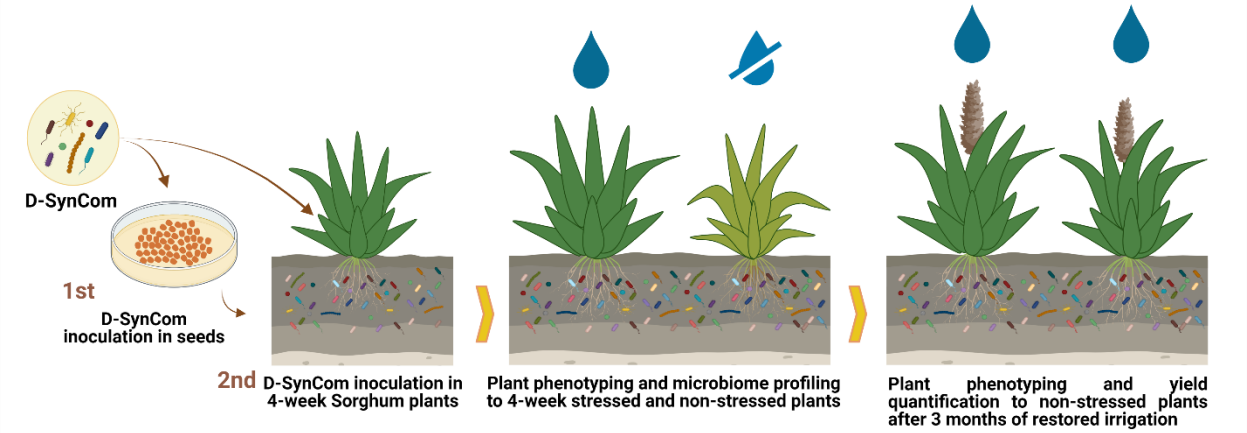

Figure 5. Field trial with sorghum plants inoculated once (seeds) or twice (seeds and 4-week plants) with mock or the D-SynCom. Drought stress was applied after the second inoculation to one plot and irrigation was restored after 4 weeks. Microbiome profiling was analyzed at the end of drought stress and plant phenotyping was recorded at the same time point and after 3 months of restored irrigation.

Computational assessment of the complementation landscape

Fundamental goals of the Persistence Control SFA are to understand the ecological and biochemical mechanisms that underpin environmental persistence in microbiomes to enable colonization, persistence, and proliferation control of engineered organisms when introduced into rhizosphere microbiomes. To accomplish this goal, we must gain a comprehensive understanding of the genetic determinants and metabolic routes that comprise the dynamic functional profile of a microbiome, a concept we term the *complementation landscape*. We have dedicated a concentrated effort to establish and evaluate analytical tools that will allow an improved description of the complementation landscape in the microbiomes we are currently experimentally characterizing.

Analysis of protein functional annotation assignment has revealed that, on average, only ~50-60% of genes are assigned any functional annotation using current standard annotation techniques, and only about half of those receive specific functional assignment [6, 7]. To address this knowledge gap, we have developed a scalable pipeline for building models of protein functional families, searching those families against novel sets of protein sequences (e.g., from metagenomes), and clustering novel sets of protein sequences for the purpose of classification into functional families. The framework, called Snekmer, is based on protein fingerprinting, whereby protein sequences are treated as a profile comprising overlapping subsequences of a fixed length, or kmer. Snekmer models are constructed for each protein family using machine learning to

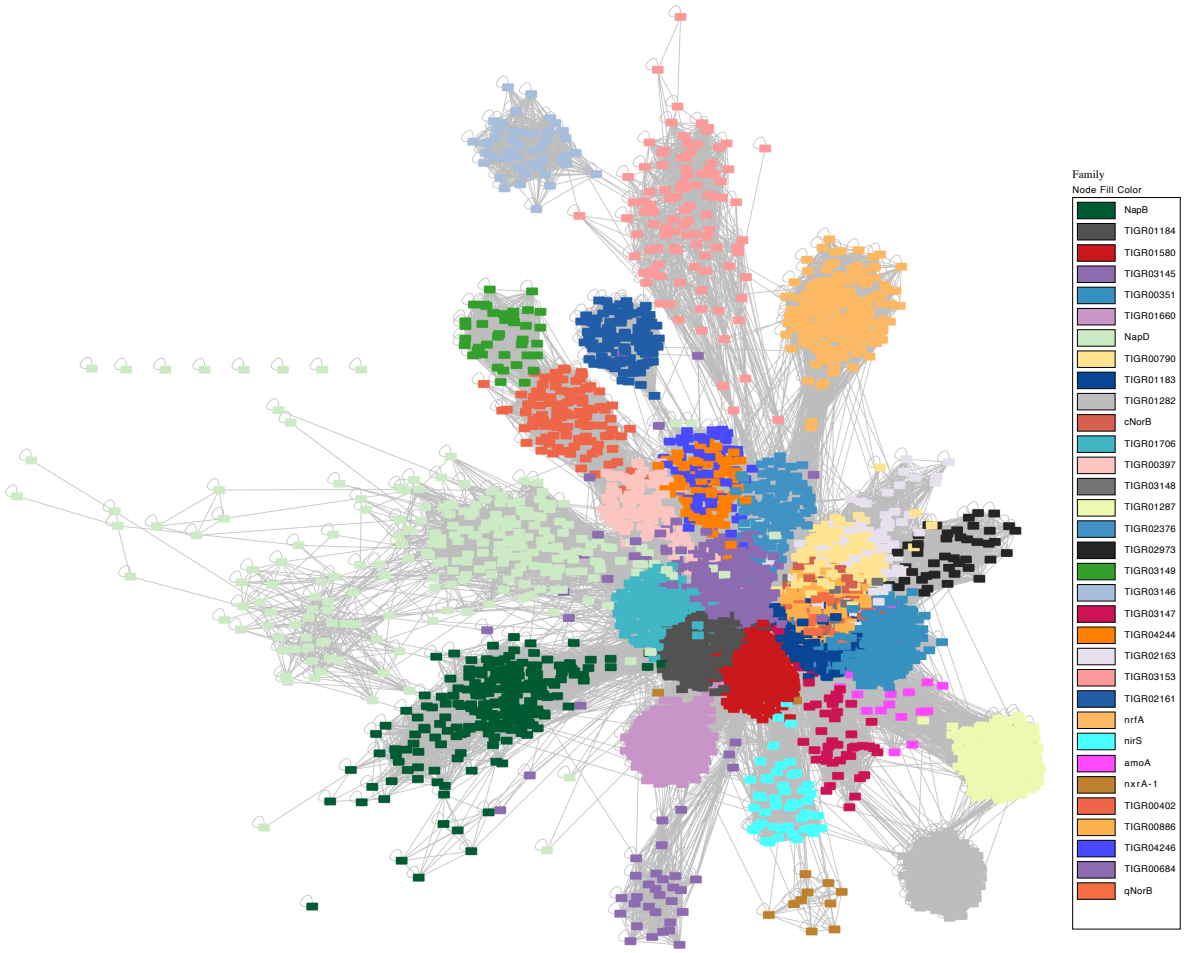

Figure 6. **Network representation of de novo similarity relationships determined by Snekmer** (kmer length of four, no amino acid recoding) showing nitrogen cycling families as different colors (legend). Nodes represent proteins and edges represent similarity relationships with a correlation of 0.1 or greater as determined by comparing kmer vectors for both proteins.

identify patterns that uniquely and specifically identify the family of interest (Figure 6). Our innovation on this idea was to reduce the complexity of the underlying protein sequences using physicochemical similarities and evolutionary relationships. This re-coding of the protein sequence 1) provides a computational advantage by significantly reducing the amount of memory and computation required to perform searches, and 2) improves the ability of these models to recognize divergent sequences. We have published the method [2], and the code is available to the public through Github and as a new application in KBase. We are currently using Snekmer to construct models for a large number of well-described protein families derived from one thousand taxonomically diverse genomes to allow faster, more comprehensive characterization of the complementation landscape for complex plant rhizosphere communities.

Spatially resolved multi-omics of plant-microbe interactions

Visualizing rhizosphere plant-microbe interactions in a field-relevant environment is a grand challenge in science. The rhizosphere, one of the most dynamically regulated soil environments, is influenced by plant roots and root-exudated metabolites. Root exudates favor the recruitment of beneficial microorganisms to

the rhizosphere to build resilience and face the constantly changing surrounding environment. Determining the chemical signals involved in recruiting and maintaining specific root microbiomes under changing climates is a major challenge in rhizosphere biology. Improving targeted microbial functions within a rhizosphere, for example through synthetic biology tools, requires a fundamental understanding of the spatiotemporal assembly of microbiomes along the root system. Current approaches that homogenize and analyze the whole root overlook the microscale spatial organization of the rhizosphere, and thus fail to capture the unique microhabitats occupied by specific microbial communities. The lack of spatiotemporal resolution hampers our ability to attain a more holistic view of the plant-microbe interactions at a resolution meaningful for synthetic biology-enabled system optimizations.

To understand the colonization dynamics of our engineered hosts and synthetic communities, we developed an platform to investigate spatially resolved root-microbiome interactions using Rhizogrids (**Figure 7**). A Rhizogrid is an innovative plant cultivation platform to integrate taxonomic and metabolomic data with a three-dimensional root cartography workflow to enable unprecedented spatially mapping of root exudates and microbes using plants grown in soil. We designed Rhizogrids to anchor a pot-grown root system and to provide 3D coordinates for root and rhizosphere samples harvested from the pot. With our approach, we can link microbial taxa and metabolites corresponding to specific locations in the root system. By linking molecular-level measurements to spatial information for the root system we can begin to understand microenvironment effects on rhizosphere colonization and interrogate colonization dynamics using engineered rhizosphere isolates and synthetic microbial communities. The resulting information is critical to shed light into persistence control principles in field-relevant laboratory environments and, more broadly, interactions between plants and microbes in the rhizosphere which can play significant roles in the resilience of plants to climate change.

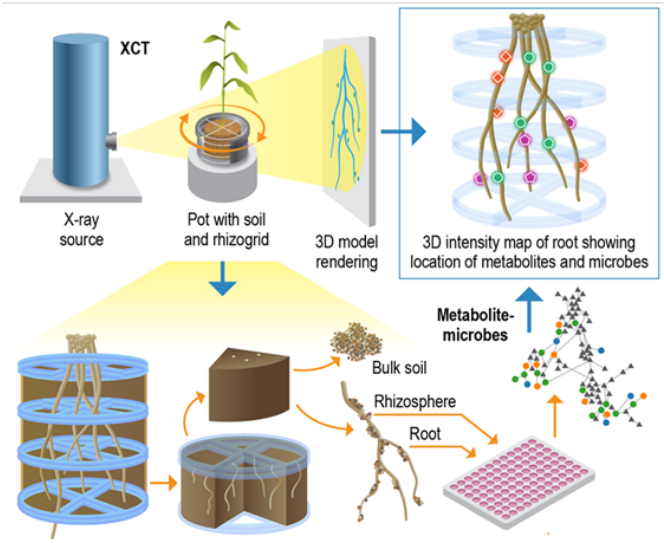

Figure 7. Three-dimensional root-rhizosphere cartography. A plant is grown in soil using pots equipped with Rhizogrids. A 3D image of the root is generated by X-ray computed tomography (XCT). Following XCT, the root and rhizosphere soil are extracted from the pot. The root with its rhizosphere is segmented into 1-cm vertical quadrants for metabolite and taxonomy profiling using liquid chromatography-mass spectrometry and 16S rRNA sequencing, respectively. Metabolite and taxonomy data are mapped to the XCT image for spatial 'omics analysis.

Synthesis of chemical probes for root exudate compounds

Activity-based protein profiling (ABPP) is a chemical biology technique that uses chemical probes designed to selectively and covalently react with protein targets of interest in a complex sample. Probe-labeled proteins can be isolated from the sample and identified using bottom-up proteomics. Chemical probes can be readily applied to diverse samples without requiring genetic manipulation or prior knowledge of protein target sequence or structure, making ABPP ideally suited for rapid discovery of protein-ligand interactions in complex systems. *By designing chemical probes that structurally mimic root exudate*

compounds, we can identify proteins that recognize and bind to these metabolites in lysates from microbial isolates and microbiome samples. This approach has previously been applied for discovery of proteins that interact with specific metabolites or drug molecules in various setting, although no studies have used this approach to study plant-microbe interactions specific to sorghum.

Due to the selective enrichment of probe-labeled proteins, ABPP provides more sensitive detection of protein targets compared to bulk proteomics techniques, particularly for low abundance proteins with high affinity for the metabolite, such as transporters or regulators, which may otherwise not be detected through global proteomics analysis. Proteins identified through ABPP can then be structurally modeled and empirically validated through biochemical experiments to investigate the nature of the protein-ligand interaction, such as potential ligand binding sites and/or the functional impact of the interaction. This workflow provides a distinct protein profile of protein-metabolite interactions by targeting the physical interaction between the plant metabolite and active proteins present in a complex sample and provides complementary data to other omics approaches such as global proteomics and transcriptomics.

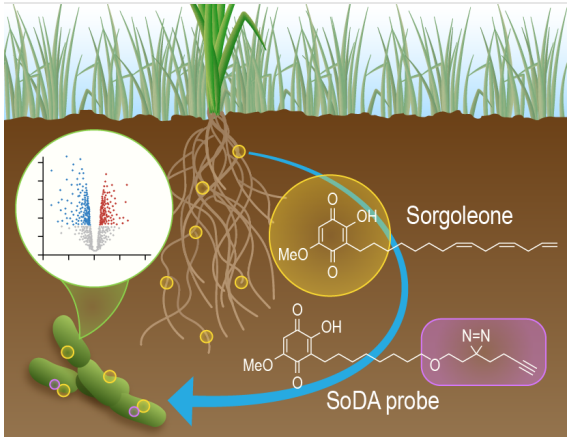

Figure 8. A sorgoleone diazirine alkyne (SoDA) probe has been synthesized and applied to SO1 cultured on sorgoleone as a sole carbon source to identify proteins that interact with this abundant hydrophobic secondary metabolite produced in sorghum root exudates.

To complement our ongoing work involving isolation and characterization of microbes capable of using sorgoleone as a substrate for metabolic addiction, we designed and synthesized a novel chemical probe based on the structure of sorgoleone to enable ABPP of proteins that interact with sorgoleone (**Figure 8**). This probe features the benzoquinone head group of sorgoleone and a lipid-like tail which incorporates a diazirine photocrosslinking group and alkyne "click" tag for further attachment of reporter groups such as a fluorophore or biotin to allow for fluorescence detection or streptavidin enrichment, respectively, of probe-labeled proteins. We have applied the sorgoleone diazirine alkyne (SoDA) probe to *Acinetobacter pittii* SO1, a novel sorgoleone catabolizing isolate we isolated from sorghum field soil, to profile sorgoleone-interacting proteins. We identified proteins that are known to bind benzoquinones, as well as lipid metabolizing proteins, transporters, and transcriptional regulators, which are now being explored further using structural prediction and biochemical approaches.

Discovery of pathways for metabolic addiction to root exudate compounds

A novel approach to biocontainment the Persistence Control SFA is investigating is an engineered *metabolic addiction* to sorghum-specific root exudate compounds. The concept is to first reduce the ability of the engineered microbe to compete with native microbes in non-sorghum rhizosphere environments by removing functions that allow it to use the most abundant carbon sources found in bulk soil (e.g., chitin-derived metabolites), plant necromass (e.g., lignocellulose metabolites), and root exudates from other plants (e.g. sucrose). In parallel, we will introduce new functions that enable it to establish a defined niche in the sorghum rhizosphere. For this, we aim to introduce metabolic pathways that allow our engineered host to rhizosphere colonization by consuming abundant sorghum-specific root exudate compounds (e.g., sorgoleone, MHPP). In principle, genome reduction coupled to exudate metabolism should restrict growth

of the engineered microbe to the sorghum rhizosphere, where the nutrients that support its growth are present.

Despite significant research on the production and characteristics of many sorghum-specific root exudate compounds, the metabolic pathways through which microbes consume these compounds are unknown and must be discovered to create metabolic addiction modules to transfer into new hosts. Broadly speaking, this process involves: (1) isolating microbes that grow using the exudate compound as a sole carbon source, (2) sequencing their genomes, (3) performing 'omics analysis to identify genes upregulated during or required for growth with the carbon source, and (4) expressing these genes in our engineered hosts (**Figure 9)**. Bacteria capable of using sorghum-specific exudates are largely unknown. To isolate these organisms we first enrich for them by incubating soils where sorghum has been cultivated in growth media containing the exudate as the sole carbon source. Only microbes able to consume the compound should grow, rapidly enriching for bacteria harboring the target metabolic pathways. Novel environmental isolates are distinct from even closely related bacteria, and thus it is critical to sequence their genomes before we can perform downstream steps for pathway discovery. Genes that are specifically activated during consumption of a compound are often involved in metabolic pathways for its degradation, and thus as a first step we perform transcriptomic analysis (RNAseq) to identify these genes. However, some enzymes in a pathway are not specifically activated by the compound and not all genes activated by the compound are directly involved in its metabolic pathway. Notably, we have found that combining transcriptomic data with functional genomics data available using RB-TnSeq methods[4] is a powerful method to filter out uninvolved genes and identify critical genes with no differential expression profile. The characterization of candidate genes is laborious and combining these 'omics methods has allowed us to reduce a candidate gene set by an order-of-magnitude. We also employ computational methods to measure evolutionary conservation of genes across taxonomically diverse hosts shown to catabolize our target compound, which can add further confidence to our selection of gene targets. For isolates amenable to genetic modification, targeted gene deletions allow us to confirm involvement of encoded proteins in metabolic pathways. We determine enzymatic activities by either performing metabolite analysis with gene deletion mutants or activity assays with recombinantly expressed proteins. In parallel with genetic and metabolic analyses, we express candidate pathway genes in our engineered hosts and evaluate the resulting strains for the ability to consume the exudate compound. Serine-recombinase Assisted Genome Engineering (SAGE, above) allows us to rapidly characterize subsets of candidate genes.

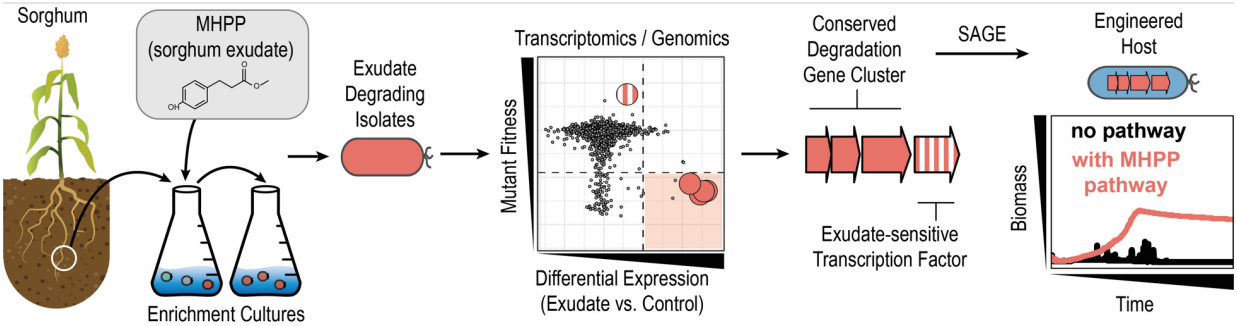

Figure 9. An experimental workflow to eludicate root exudate metabolic pathways. Microbes from bulk soil are enriched and isolated in defined culture media with purified root exudate compounds. Transcriptomic and functional genomic profiles are assessed under conditions that allow discovery of genes that contribute to the catabolic phenotype. Candidate genes and pathways are transferred to a naïve rhizosphere host using Serine-recombinase Assisted Genome Engineering to establish a necessary and sufficient set of genes to comprise a metabolic addiction module.

References

1. Elmore, J.R., et al., *High-throughput genetic engineering of non-model and undomesticated bacteria via iterative site-specific genome integration* Science Advances (in press), 2023.
2. Chang, C.H., et al., *Snekmer: A scalable pipeline for protein sequence fingerprinting based on amino acid recoding.* Bioinformatics Advances (in press), 2023.
3. Handakumbura, P.P., A. Rivas Ubach, and A.K. Battu, *Visualizing the Hidden Half: Plant-Microbe Interactions in the Rhizosphere.* mSystems, 2021. **6**(5): p. e0076521.
4. Price, M.N., et al., *Mutant phenotypes for thousands of bacterial genes of unknown function.* Nature, 2018. **557**(7706): p. 503-509.
5. Elmore, J.R., et al., *Engineered Pseudomonas putida simultaneously catabolizes five major components of corn stover lignocellulose: Glucose, xylose, arabinose, p-coumaric acid, and acetic acid.* Metab Eng, 2020. **62**: p. 62-71.
6. Salzberg, S.L., *Next-generation genome annotation: we still struggle to get it right.* Genome Biol, 2019. **20**(1): p. 92.
7. Lobb, B., et al., *An assessment of genome annotation coverage across the bacterial tree of life.* Microb Genom, 2020. **6**(3).

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov